



Review

Databases and Bioinformatic Tools for Glycobiology and Glycoproteomics

Xing Li, Zhijue Xu, Xiaokun Hong, Yan Zhang * and Xia Zou *

Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China; xingli87@126.com (X.L.); xuzhijue@gmail.com (Z.X.); hongxk@sjtu.edu.cn (X.H.)

* Correspondence: yanzhang2006@sjtu.edu.cn (Y.Z.); zouxia0206@sjtu.edu.cn (X.Z.);
Tel.: +86-21-34206778 (Y.Z.); +86-21-34206966 (X.Z.)

Received: 31 July 2020; Accepted: 11 September 2020; Published: 14 September 2020



Abstract: Glycosylation plays critical roles in various biological processes and is closely related to diseases. Deciphering the glycode in diverse cells and tissues offers opportunities to develop new disease biomarkers and more effective recombinant therapeutics. In the past few decades, with the development of glycobiology, glycomics, and glycoproteomics technologies, a large amount of glycoscience data has been generated. Subsequently, a number of glycobiology databases covering glycan structure, the glycosylation sites, the protein scaffolds, and related glycomics have been developed to store, analyze, and integrate these data. However, these databases and tools are not well known or widely used by the public, including clinicians and other researchers who are not in the field of glycobiology, but are interested in glycoproteins. In this study, the representative databases of glycan structure, glycoprotein, glycan–protein interactions, glycomics, and the newly developed bioinformatic tools and integrated portal for glycoproteomics are reviewed. We hope this overview could assist readers in searching for information on glycoproteins of interest, and promote further clinical application of glycobiology.

Keywords: glycosylation; glycoinformatics; glycan; glycoprotein; glycomics; lectin

1. Introduction

Glycosylation is known as one of the most ubiquitous and important post-translational modifications (PTMs) in nature. It is estimated that more than 50% of mammal proteins and 90% of secreted and membrane proteins are glycosylated [1]. Glycosylation controls and defines a wide range of biological events, including cellular signaling, ligand–receptor interactions, cell–cell communication, pathogen–host recognition, immunological responses, and, consequently, it is involved in many human diseases, including inflammation and cancer [2]. Furthermore, the importance of glycosylation is highlighted by the fact that it can influence the therapeutic properties of many therapeutic proteins [3]. Therefore, deciphering the glycode in diverse cells and tissues offers opportunities to develop new disease biomarkers and more effective recombinant therapeutics.

However, studies of the glycome are more complicated compared with those of the genome and proteome due to the diversity and branched nature of glycans. The complexity is further increased by the fact that glycan expression on a single protein is subject to both macroheterogeneity (variable occupancy of different glycosylation sites) and microheterogeneity (variable distribution of different glycans attached to a single glycosylation site) [4]. In addition, due to the different expression of glycomics such as glycosyltransferase and glucosidase in different cells and tissues, the glycans attached to the same protein are cell-, tissue-, organism-, and physiological state-dependent. Furthermore, the same glycan structure on different proteins may have different functions. Therefore, the interpretation of

the glycosylation sites, the attached glycan structure, the protein scaffolds, and related glycogenes are crucial for understanding the biological function of glycosylation.

In past decades, with improvements in analytical methods including high performance separation techniques and mass spectrometry (MS)-based analysis techniques [5,6], the qualitative and quantitative data of glycans, glycosites, glycopeptides, and glycoproteins have increased tremendously. In addition, extensive application of lectin chips, glycan microarrays, and glycogene chips greatly increased the data of glycan–protein interactions and the glycogenome [7–9]. Consequently, the development of databases and informatics tools to store, retrieve, integrate, and interpret these data is one of the most active fields in glycobiology.

The first publicly available glycan structural database, the Complex Carbohydrate Structural Database (CCSD)—also known as the CarbBank—was developed in the late 1980s [10,11]. Inspired by this work, more and more databases have been developed, such as the GLYCOSCIENCE.de database [12] and the KEGG (Kyoto Encyclopedia of Genes and Genomes) GLYCAN database [13]. In the early 2000s, the Consortium for Functional Glycomics (CFG) project began with a comprehensive database containing information on glycan structures, glycan-binding proteins (GBPs), glycosyltransferases, and glycan-related gene knockout mice [14], while the Japan Consortium for Glycobiology and Glycotechnology DataBase (JCGGDB) collected information on mass spectrometry analysis, glycogenes, and the affinity constants between lectins and pyridylaminated glycans [15–17]. In recent years, various glycoinformatic databases and tools have been developed. Of note, significant progress has been made lately through a new generation of centralized and integrated resources, including GlyGen [18], Glycomics@ExPASy [19], and GlyCosmos [20]. In addition to these comprehensive resources, a variety of databases on glycan structure, glycogenes, glycoproteins, and a series of bioinformatic tools based on these databases have been developed all over the world, which also promote the development of glycobiology. However, generally speaking, these databases and tools have not been well known and widely used by the public, including clinicians and other researchers who are not in the field of glycobiology, but are interested in glycoproteins.

As glycoproteins have become the promising candidates of disease markers and therapeutic targets, we mainly focus on providing an overview of main features and functionalities of the representative databases of glycan structure, glycoproteins, glycan–protein interactions, and glycogenes in humans and mammals (Figure 1). In addition, we summarized some newly developed bioinformatic tools and integrated resources for glycoproteomics with the goal of making these bioinformatics resources more widely known to the public, especially to researchers in other disciplines. We hope this overview assists readers in searching for information on glycoproteins of interest, and promotes further clinical application of glycobiology.

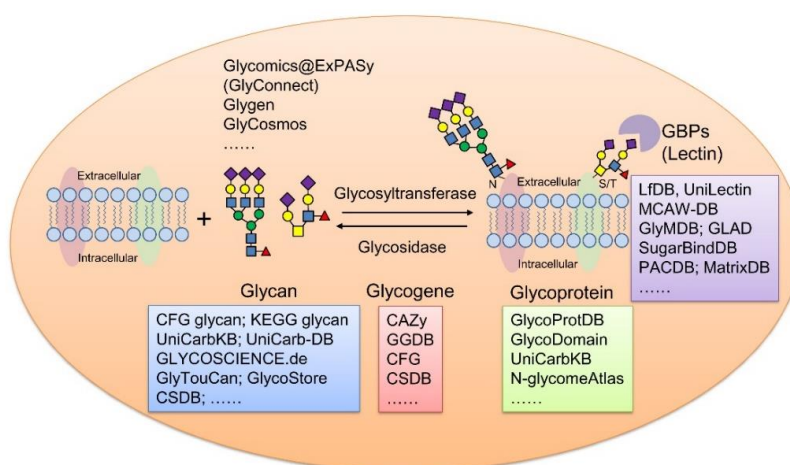


Figure 1. Summary of representative glycoinformatic databases using membrane glycoproteins as example. GBPs, glycan binding proteins.

2. Glycan Structure Databases

As different glycan structures have different regulatory effects on protein function, the analysis of glycan structure is one of the most important parts of glycobiology. According to the attached amino acid residues, glycans are divided into *N*-glycan, *O*-glycan, and other minor modifications. *N*-glycans are attached to the amino group of the Asn residue on the protein and can be released by endoglycosidases such as PNGaseF, while *O*-glycans are attached to the hydroxyl group of Ser or Thr and rarely on Tyr, and can be released by β -elimination. To determine the structure of a glycan, identification and quantification of monosaccharides, as well as identification of glycosidic bonds between monosaccharides, are both required. Useful information can be obtained by comparing the chromatography and mass spectra of unidentified glycans and known structures, which will accelerate the analysis [21–24]. For this purpose, storage and management of glycan structure data are very important. The first database in glycobiology, CarbBank, was constructed for collecting the glycan structures and the original mass spectra. There were more than 40,000 glycan structure entries in CarbBank until it was discontinued in 1997. These data are currently stored in multiple databases, including the CFG glycan databases, KEGG glycan, GLYCOSCIENCE.de, and CSDB (Carbohydrate Structure DataBase). Benefiting from the accumulating data of glycan structure, as well as the development of mass spectrometry technology, glycan structure identification is becoming more and more rapid, accurate, and automatic [23].

2.1. CFG Glycan Structure Database

The Consortium for Functional Glycomics (CFG, <http://www.functionalglycomics.org/>) is one of the largest resource platforms in glycomics [14]. It contains databases on glycan structure, glycosyltransferases, GBPs, and glycogene knockout mice, and also provides services such as glycogene microarrays and glycan arrays. The CFG Glycan Structure Database (<http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp>) offers detailed structural and chemical information for thousands of *N*- and *O*-glycans, including both synthetic glycans and glycans derived from human and mouse tissues and cell lines. Glycan structures stored in CFG can be searched by glycan name, composition, molecular weight, motifs, cell lines, or tissue samples. The search results show the glycan structure, molecular weight, composition, biological source and references, and provide links to relevant entries in CFG and external databases including a 3D modeling feature. Of note, CFG also provides carbohydrate standards for mass spectrometry and glycosyltransferases for synthesis of glycans, which has greatly improved the analysis of glycan structure. CFG has previously provided glycans profiling service using a variety of analytical methods such as MALDI–MS and ESI–MS. However, this service was closed after 2011 and was transferred to the Dell/Haslam biopolymer mass spectrometry laboratory at Imperial College, London.

2.2. Glycan Mass Spectral DataBase

JCGGDB is a meta-database involving 15 original databases including glycan profile data from mass spectral, lectin array data, glycoprotein data, glycogene data, glycoepitope data, and experiment support. Glycan Mass Spectral DataBase (GMDB, https://jcgddb.jp/rcmg/glycodb/Ms_ResultSearch) in JCGGDB is a multi-stage tandem mass spectral database, which stores MS^2 , MS^3 , and MS^4 spectra of structurally-defined *N*- and *O*-linked glycans, and glycolipid glycans, as well as the partial structures of these glycans using a MALDI–QIT–TOF mass spectrometer [16]. *N*-glycans and glycolipid glycans are mostly tagged with 2-aminopyridine (PA), which can be used for fluorescence detection in HPLC. More importantly, this database also provides their protocol for MS^n spectra acquisition, which greatly helps users to compare and analyze their results correctly through spectral matching. In this database, the MS data of glycans can be easily searched by composition or the m/z value of the precursor ion. Although this database has not been updated, it will still be useful for some users' data analysis.

2.3. UniCarbKB

UniCarbKB (<http://unicarbkb.org/>), composed of GlycoSuiteDB, GlycoBase, and EUROCarbDB, aims to construct an information storage and search platform [25–27]. GlycoSuiteDB collects all the glycan structures and glycosylation site information published from 1990 to 2005. In this database, glycan structures can be searched by taxonomy, tissue, protein, and disease. Entries in GlycoSuiteDB contain the glycan structure, the glycosylation site on the protein, the biological source and literature references, and information on diseases [28,29]. In addition, UniCarbKB includes more than 350 glycan structures analyzed by HPLC previously available from GlycoBase [30], and MS, NMR, and HPLC experimental data for glycans from EUROCarbDB [31]. UniCarbKB also provides a series of bioinformatic tools for glycan analysis. For example, GlycoDigest can simulate the digestion of oligosaccharides by exoglycosydases [32], and UniCorn is a theoretical *N*-glycan structure database [21], which could be used to improve the efficiency of glycan analysis and validate glycan structures. UniCarbKB provides useful information on site-specific *N*- and *O*-glycans of glycoprotein, and connects glycan structure with its attached protein(s) which can be annotated in UniProtKB. Unfortunately, UniCarbKB cannot be accessed currently, possibly due to technical problems. As UniCarbKB stores valuable data for glycobiology studies, we hope this database can be functional again.

2.4. KEGG Glycan

KEGG is an integrated database containing the knowledge of molecular networks such as signaling and metabolic pathways. KEGG glycan (<http://www.genome.jp/kegg/glycan/>) collects the experimentally determined glycan structures, along with the glycan biosynthesis and metabolism pathways [13]. The glycan structures stored in KEGG glycan could be searched through the G number as well as the DBGET tool. Each entry includes the molecular weight, composition, detailed structure of the glycan, and the references. KEGG glycan also provides external links to other glycan databases for cross reference. In addition, glycan-related pathways and diseases are also important information provided by KEGG. The glycan-related pathways include biosynthesis and degradation of glycans, glycan-involved signaling molecules and interaction pathways, and cancer-associated carbohydrates, which could help researchers understand the biological role of glycans.

2.5. GLYCOSCIENCE.de

GLYCOSCIENCE.de (<http://www.glycosciences.de/>) is an integrated portal containing a combination of databases and tools for glycome analysis. Its main focus is on glycan 3D structures, which are extracted from CarbBank, Protein Data Bank (PDB), and literature research [12,33]. In addition to standard molecular information, GLYCOSCIENCES.de provides many useful tools for performing various quality checks and structural analyses of 3D structure modeling, detection, and validation of carbohydrates in PDB files, and supports MS/NMR analyses of glycans. For example, Sweet-II is a powerful tool in the GLYCOSCIENCE.de database, which can rapidly convert the sequence of a complex glycan into a reliable 3D molecular model for molecular dynamic simulations and other further analyses [34].

Glycosciences.DB (<http://www.glycosciences.de/database/>) is the main glycan structure database of GLYCOSCIENCE.de, collecting various kinds of data on glycan structures linking glycomics and proteomics data [35]. Glycosciences.DB provides three entries, including a glycan structure entry, literature entry, and worldwide Protein Data Bank (wwPDB) entry, all of which are linked to each other. Glycan structure can be searched by glycan (sub-)structure, monosaccharide composition, molecular formula, structure classification and motifs, as well as NMR, MS, PDB query, or bibliography queries. The graphical interface GlycanBuilder is also supported for database search. When users search specific glycan, the detailed information including glycan composition, glycan structure, NMR data, PDB link (if available), and references are provided. As the PDB database has not provided a way to search 3D structures of glycans, this database provides the only way to search for specific carbohydrate structures

in PDB. In addition, more and more literature has been added to Glycosciences.DB, which increases the number of new glycan structure entries in this database. At the time of this writing, Glycosciences.DB contains more than 26,000 glycan structure entries with 13,500 3D structure models, which makes it one of the largest databases for glycan 3D structure analysis.

2.6. UniCarb-DB

UniCarb-DB (<https://unicarb-db.expasy.org/>) is an emerging public database providing access to a collection of LC–MS/MS glycan fragments released from glycoproteins for glycomic discovery [36]. UniCarb-DB provides a user-friendly search interface. Users can search glycan data by taxonomy, tissue, reference, mass, composition, or precursor mass using basic search or advanced search. The result shows detailed information, such as glycan types, linkages, structures, and the corresponding MS data obtained from the literature and experimental evidence. The links to PubMed and Uniprot entries are also presented. By comparing with the experimental spectra in the database, UniCarb-DB can be used as reference to aid manual annotation of glycan structure. UniCarb-DB contains over 1500 spectra of both *N*- and *O*-glycans of glycoproteins derived from 13 taxonomies and 20 tissues, and has grown to be one of the largest experimental glycomic MS databases.

2.7. GlyTouCan

GlyTouCan (<https://glytoucan.org/>) is an international glycan structure repository, which was developed on the basis of GlycomeDB [37]. GlycomeDB is established for integrating glycan structure information from different sources (CFG, BCSDB, GLYCOSCIENCES.de, KEGG glycan, EUROCarbDB, and CarbBank) and excluding the redundant parts [38,39]. Glycan structures in GlyTouCan can be searched by text input, motif, or drawing glycan structures in GlycanBuilder. Each entry shows the glycan structure, molecular weight, biological source, references and external links to other databases. The most important feature of GlyTouCan is that each glycan structure is assigned a globally unique accession number. This unique accession number, like the mRNA and protein accessions, can be cited when referring to a specific glycan, and will simplify and unify the description of glycans in scientific literature, making it possible to search between different databases. In addition, the developers of GlyTouCan encourage users to register identified glycans and obtain unique accessions for publication, which is similar to GeneBank [40].

2.8. GlycoStore

GlycoStore (<https://www.glycostore.org/>), developed based on the publicly available experimental datasets GlycoBase, is an annotated database of retention properties of *N*-, *O*-, glycosphingolipid (GSL) glycans and free oligosaccharides of glycoproteins, glycolipids, and biotherapeutics [41]. GlycoStore focuses on ultra-high performance liquid chromatography (U/HPLC), reversed phase (RP)-U/HPLC, porous graphitized carbon (PGC) chromatography, and capillary electrophoresis (CE) elution positions for approximately 850 unique glycan structures with links to taxonomy, glycoprotein, and supporting literature. Data stored in this database can be searched by experimental values (GU, AU, or time), monosaccharide composition or metadata labels (taxonomy, sample name, and the Oxford linear notation). Each glycan structure entry page lists all methods for determining this structure and substances in which this glycan is present. Interestingly, GlycoStore provides a comparison function. Users can choose two to three proteins of interest, and the results will show common structures of these proteins.

2.9. CSDB

Carbohydrate Structure DataBase (CSDB, <http://csdb.glycoscience.ru/database/>) is the largest database focusing on the structures of glycans and glycoconjugates in prokaryotes, plants, and fungi. The current version of CSDB is a merger of the Bacterial (BCSDB) and Plant and Fungal (PFCSD) databases [42]. CSDB contains manually curated structural, taxonomic, and NMR data of carbohydrates.

CSDB has a high coverage of natural carbohydrates in bacteria, archaea, fungi, and plants, and is still being updated [43]. Users can search the database by IDs, bibliographic data and keywords, biological source, structural fragments, and NMR data, and also browse all the identified carbohydrates in a species. The results show the glycan structural data with an extended bibliography, assigned NMR spectra, taxon identification, and other information. Of note, CSDB provides high data quality by manual curation of original publications. In addition, CSDB also provides bioinformatic tools to interpret the carbohydrate structures from NMR spectra and predict it according to carbohydrate structures [44], as well as a number of computational services, such as NMR simulation and taxon clustering.

3. Glycoprotein Databases

N-and *O*-glycosylation are the most common glycosylation modifications on proteins. *N*-glycans are attached to the amino group of Asn residue on protein, with a consensus motif as Asn-Xaa-Ser/Thr, where Xaa could be any amino acid except Pro. *O*-glycosylation includes *O*-GalNAc, *O*-GlcNAc, *O*-Man, *O*-Fuc, and other types, in which *O*-GalNAc glycans initiated by *N*-acetyl- α -D-galactosamine (GalNAc) attached to the hydroxyl group of Ser or Thr residues is the most abundant type [45]. In addition, *O*-GlcNAcylation is a monosaccharide *N*-acetyl- β -D-glucosamine (GlcNAc) modification on nuclear, cytoplasmic, and mitochondrial proteins [46], which is different from other glycosylation types with complex glycans mainly attached to secreted and membrane proteins.

In order to identify glycoproteins, the glycosylated proteins should be enriched with analytical, affinity, or chemical techniques. A variety of technologies such as hydrophilic chromatography, hydrazide chemistry, lectin chromatography, or metabolic labeling have been widely used [17,47–51]. Sequentially, the complex glycans are removed or shortened by glycosidase treatment or chemical approaches and usually a unique tag is left or labeled at the glycosylation site. The peptides are then applied to mass spectrometry for identification. To identify the glycopeptides and glycosylation sites, different strategies are applied for different types of glycosylation. As the release of *N*-glycans by endoglycosidase facilitates high-throughput and automated identification for *N*-glycoproteins, a variety of *N*-glycosylation sites have been annotated. Based on these data, the Technical University of Denmark developed the NetNGlyc server to predict the *N*-glycosylation sites according to the well-known Asn-Xaa-Ser/Thr sequon. However, the identification of *O*-glycoproteins such as *O*-GalNAc and *O*-Man has lagged because of a lack of efficient glycosidases to release *O*-glycans. Recently, genetic technology named SimpleCell was developed to simplify the structure of *O*-glycans to a single GalNAc for lectin affinity enrichment and tag for glycosite identification by MS [52–54]. As a result, a large number of *O*-glycoproteins were identified, which greatly expanded the *O*-glycoprotein data pool. Here we describe the represented databases for *N*-glycoprotein and *O*-glycoprotein.

3.1. GlycoProtDB (GPDB)

GlycoProtDB (<https://acgg.asia/gpdb2>) is a database containing curated experimental data of *N*-glycoproteins. The *N*-glycosylation sites were identified through lectin affinity column combined with isotope-coded glycosylation site-specific tagging (IGOT) method. In detail, the proteins were digested and the corresponding *N*-glycosylated peptides were enriched by ConA or WGA lectin column. PNGaseF treatments were performed in a solvent using stable isotope-labeled water, H₂¹⁸O, to remove the *N*-glycan and label the glycosylated Asn residue with ¹⁸O. The *N*-glycosylation sites were identified by LC-MS/MS [55]. This method was first applied to the model organism *Caenorhabditis elegans* and identified 250 glycoproteins carrying 400 unique *N*-glycosylation sites [17]. The current version of GlycoProtDB contains *N*-glycoproteins from *C. elegans*, different mouse tissues (C57BL/6, male) [56,57], and human cell lines [58]. Proteins in GlycoProtDB can be searched by gene symbol, gene name, or protein name. The results show a map of *N*-glycosylation sites on the protein, the amino acid sequence annotated with potential and identified *N*-glycosylation sites. Of note, protein sequences with common glycopeptide sequence(s) are linked each other, which may be convenient for users to analyze the glycosylation of homologous proteins.

3.2. UniPep and N-GlycositeAtlas

Unlike GlycoProtDB, UniPep (<http://www.unipep.org/>) is a database that focuses only on the human *N*-glycoprotein for glycoprotein biomarker discovery [59]. *N*-glycoproteins from plasma, bladder, breast cancer cells, liver, lymphocytes, cerebrospinal fluid, prostate tissue, and prostate cancer cells were captured using the hydrazide chemistry method, released by specific glycosidase PNGaseF, and identified by MS/MS [49]. The 1552 unique *N*-linked glycosylation sites were identified and mapped on the associated proteins, and then imported into the UniPep database [59]. The *N*-glycoproteins can be searched by gene symbol, gene name, Swiss-Prot ID, IPI ID, amino acid sequence, or peptide mass. In addition, users can also browse the identified list and the metabolic and signaling pathways for *N*-glycoproteins. Each *N*-glycoprotein is provided with detailed information including tables listing predicted and identified *N*-linked glycopeptides, protein sequence annotated with *N*-glycopeptides, and a map of *N*-glycosylation sites on the protein.

Recently, a larger database called *N*-GlycositeAtlas (<http://nglycositeatlas.biomarkercenter.org>) containing more than 30,000 glycosite-containing peptides with >14,000 *N*-glycosylation sites from over 7200 *N*-glycoproteins was developed by the same lab [22]. These human glycosite-containing peptides were collected from over 100 publications and unpublished datasets, and then mapped to UniProt database. Users can perform basic search or advanced search by gene/protein name, accession number, glycosylation site location, glycosite containing peptide, tissue/liquid/cell line, or publication. Overall, the most important feature of these two databases is that lots of *N*-glycoproteins were derived from clinical samples, including plasma, human-derived tissues, body fluids, and cell lines, which may be of interest to clinical researchers.

3.3. O-GalNAc Protein Databases

The GlycoDomain Viewer (<https://glycodomain.glycomics.ku.dk/>) established by the Copenhagen Center for Glycomics to organize and share the *O*-GalNAc proteome identified by SimpleCell technology [53], has been considered as one of the largest databases of *O*-GalNAc glycoprotein. Presently, this database includes 629 experimentally identified *O*-GalNAc glycoproteins and 2942 *O*-glycosylation sites from human and animal cell lines. It allows users to search by the NCBI gene name and the UniProt ID, or browse all the identified list. Each *O*-GalNAc glycoprotein in this dataset is shown with its sequence and domain topology. Importantly, the verified and predicted glycosylated sites of *N*-glycan, *O*-GalNAc, *O*-Mannose and *O*-Xylose are mapped on the protein sequence. By showing data produced experimentally as well as retrieved from other databases, GlycoDomain Viewer presents the interplay between relevant protein and post-translational modification information to explore the possible effects of glycosylation on a protein. In addition, based on this dataset, NetOGlyc was upgraded to version 4.0 (www.cbs.dtu.dk/services/NetOGlyc/), which could provide more accurate predictions for *O*-GalNAc glycosylation sites in mammalian proteins [53].

3.4. O-GlcNAc Protein Database

YinOYang 1.2 (<http://www.cbs.dtu.dk/services/YinOYang/>) is a prediction database for identifying potential *O*-GlcNAcylation sites for any submitted protein [60]. This server has incorporated results from NetPhos (<http://www.cbs.dtu.dk/services/NetPhos/>) [61]. The Ser/Thr residues which are predicted to be *O*-GlcNAcylation, as well as phosphorylated, are marked. Such sites may be reversibly and dynamically modified by *O*-GlcNAc or phosphate groups at different times in the cell. However, in comparison with *O*-GalNAc glycosylation, there is still a lack of a comprehensive available *O*-GlcNAc protein database.

4. Glycogene Databases

Different from nucleic acids and proteins, the synthesis of glycans is not template-driven. Instead, the glycosyltransferases control the structures of glycans and their attached sites on the carrier

proteins, and the glycosidases further modify the glycan structures. These glycan-related genes are named glycogenes. There are about 200 glycogenes in humans, which cover about 1% of the human genome [15]. These glycogenes are expressed in a tissue- and time-specific pattern. The glycome of a specific tissue at a specific time point is determined by the combination of all present glycogenes. To understand the regulation of glycans, information about glycogenes should be collected. In this section, we describe several representative databases of glycogenes.

4.1. CAZy

The Carbohydrate-Active Enzymes database (CAZy, <http://www.cazy.org/>), the largest database of glycan-related genes, has collected enzymes that degrade, modify, or form glycosidic bonds and proteins containing carbohydrate-binding modules since 1998 [62]. Enzymes in CAZy are divided into five categories according to sequence and structure similarity including glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and auxiliary activities (AAs). Users can search enzymes by protein name, organism name, GeneBank or UniProt accession, and EC number. The results contain the genomic, structural, and biochemical information on glyco-enzymes, as well as external links to GenBank, UniProt, CFG, and PDB. Application of this information can significantly facilitate the synthesis of biologically active glycan products. However, it is noteworthy that although CAZy stores information on several hundred thousands of enzymes from multiple taxa, but less than 5% of them have experimentally established activities [23].

To create the most reliable encyclopedia of carbohydrate-active enzymes possible, CAZy provides a community-driven resource named CAZypedia (<http://www.cazypedia.org>) [63]. Currently, the database contains a series of curator-approved content of glyco-enzymes, which can help researchers who are not in this field understand the basic knowledge of glycobiology. In addition, just like a wiki, the search and display pages of CAZypedia are very simple and very friendly for the beginners.

4.2. GGDB

The GlycoGene DataBase (GGDB, <http://acgg.asia/ggdb2/>), established by JCGGDB, includes identified genes related to glycan synthesis, such as glycosyltransferases, sugar nucleotide synthases, and sugar-nucleotide transporters [15]. As some proteoglycans such as heparan sulfate and chondroitin sulfate contain sulfo groups, a category of 34 sulfotransferases is also included in GGDB. Users can search by gene symbol or designation. Each gene provides mRNA and protein sequences, chromosome location, EC number, and gene ontology, as well as links to GeneBank, CAZy, and OMIM. References with brief annotations are also provided. In comparison with CAZy, GGDB provides more detailed information, such as donor and acceptor substrates, as well as the expression pattern for each gene. Moreover, researchers can order the plasmid conveniently according to the biological resource of each gene. GGDB also provides information on the tissue expression and a link to the glycogene knock-out mice resource if available, which can be very helpful for the functional study of glycogenes.

4.3. CFG Glycosyltransferases Database

The glycan structures, glycosyltransferases, and GBPs are the three main contents of the CFG databases. The CFG Glycosyltransferases database (<http://www.functionalglycomics.org/glycomics/molecule/jsp/glycoEnzyme/geMolecule.jsp>) provides a user-friendly graphical interface showing the structure of different glycans. By clicking a monosaccharide, users are directed to the information of the glycosyltransferase which forms this structure. General information such as enzyme name, EC number, organism, and detailed information such as nucleotide accession in GeneBank, expression profile, Swiss-Prot ID, amino acid sequence, and biochemical reaction are provided. There are also external links to PubMed, KEGG pathway, CAZy, and Swiss-Prot for cross reference. CFG also offers several types of resources, including glycogene microarray for determining glycogene expression profiles, glycotransferases, and reagents for studying the catalytic activity of glycotransferases, antibodies for histochemical staining and purification of glycoconjugates, synthetic glycans for studying the

glycan–protein interactions, and glycogene knock-out mice for characterizing the biological role of glycogenes.

4.4. CSDB_GT Subdatabase

CSDB is a database of the glycan and glycoconjugate structures in prokaryotes, plants, and fungi. In 2017, CSDB established a curated database on carbohydrate-active enzymes called CSDB_GT (<http://csdb.glycoscience.ru/gt.html>). Currently, CSDB_GT contains glycosyltransferases found in *Arabidopsis thaliana*, *Escherichia coli* [64] and recently was expanded to *Saccharomyces cerevisiae*. Users can search for glycosyltransferases using the name or protein/gene database ID, type of glycan the enzyme takes part in, glycosidic bond the enzyme synthesizes, and the enzyme's donor or acceptor. The results provide the name and links to NCBI and UniProt. Moreover, CSDB_GT provides information on carbohydrate structures and enzyme activity, which are supported by different levels of evidence that can be traced to original publications.

4.5. GlyMAP

Mutations in most glycogenes could cause a global defect of glycosylation. While some glycosyltransferase families are composed of homologous isoenzymes, mutations on one member may not affect glycosylation globally. Large-scale whole exome sequencing (WES) could provide information on mutations in glycosyltransferase genes in populations, and could be useful to analyze and predict the functional relationship between the glycogenes and diseases. From WES of 2000 Danes, Hansen et al. constructed a database of Functional Mutational Map of glycogenes (GlyMAP, <http://glymap.glycomics.ku.dk/>) [65] to provide the global map of glycogenome genetic stability. All missense mutations were collected in this database and deleterious mutation maps were drawn by prediction algorithms, manual inspection, and additional experimental analysis in CAZy family GT27. From these data, mutations with unknown functions could be related to specific disorders and may help in the discovery of novel congenital disorders of glycosylation (CDG).

5. Glycan-Protein Interaction Databases

The specific interactions between glycans and glycan binding proteins is an important part of the biological function of glycans. In recent years, because of the development of high-throughput technologies such as lectin arrays and glycan arrays, glycan–protein interaction data is growing rapidly. Several databases have been established to store and share the glycan–protein interaction data. In this section, we introduce some representative databases closely related to mammals and human health.

5.1. LfDB

Lectins is one of the major categories of GBPs, which can selectively interact with glycans. The Lectin Frontier Database (LfDB, <http://acgg.asia/lfdb2/>), belonging to JCGGDB, is a database established to describe the quantitative interaction data between a number of lectins and various glycans. Users can simply search by keywords or choose categories among lectin family, monosaccharide specificity, or 3D-fold. The results are provided on two pages: the lectin information page including monosaccharide specificity, source of lectin and 3D structure if available; the interaction page displaying the affinity constants (K_a) of lectins toward a panel of glycans obtained by an automated frontal affinity chromatography system [66]. In the future, other GBPs such as anti-glycan antibodies will also be included in LfDB [67]. Information from LfDB could help researchers understand the structural basis of lectin–glycan interactions.

5.2. UniLectin

The recently released UniLectin (<https://www.unilectin.eu/>) is a new platform designed to cover the knowledge of lectins, their classification, and their biological role [68]. Currently, the platform

consists of two modules UniLectin3D and PropLec. UniLectin3D, the main module of UniLectin, is a curated database of lectin 3D structures and interacting ligands. User can search lectins by keyword, kingdom order, historical classification, monosaccharide, associate IUPAC sequence, fold of the binding site, or multiple criteria. For each lectin, a detailed page with 3D visualization, interactions, and links to external databases is displayed. Another module, PropLec, is focused on β -propeller lectin prediction in all species. A quick search can be performed by keyword, accession number, species name, or protein name. Currently, UniLectin has exceeded 2000 lectin structures, which will make it an important tool in glycobiology research.

5.3. PACDB

The interaction between pathogen and host is mediated by cell surface molecules, such as proteins and glycolipids. The Pathogen Adherence to Carbohydrate Database (PACDB, <https://acgg.asia/db/diseases/pacdb>) is a database that was established to collect the information on pathogens (e.g., bacteria, fungus, toxin and virus) adhering to glycan expressed on the cell surface of host animals or plants [69]. Currently, there are more than 1800 interactions related to over 180 diseases in PACDB. Users can browse the database by pathogen names or disease names. Interaction information extracted from references are listed on the page of each pathogen or disease, and annotated with [binding] or [not binding]. The glycans can be linked to the JCGGDB database and the references are linked to PubMed for further interest. Although some interactions require confirmation, it provides a lead for further investigation of correlations between pathogen and host cells. Currently, the data of PACDB has been summarized in the GlyCosmos.

5.4. SugarBindDB

The SugarBind Database (SugarBindDB, <https://sugarbind.expasy.org/>) is another curated resource, covering knowledge of glycan-mediated host–pathogen interactions based on glycan–protein binding pairs [70]. A set of five inseparable components including the pathogenic agent, lectin adhesin, glycan ligand, disease, and references constitutes the core information. Users can search the database by several terms, such as pathogenic agents (such as influenza virus), ligands (such as A Lewis b), recognizing lectins or adhesins (BabA), affected area in the pathology (e.g., intestine), references, diseases, and multi-criteria. The database can also be queried by a glycan composition or a glycan structure drawn by GlycanBuilder. Each entry lists all related information with as much precision as possible in the form of graphs and text. Similar to PACDB, SugarBindDB provide literature for each glycan–protein binding pair. Moreover, this database also provides 3D structure from PDB and offers external links to protein and glycan-related resources such as UniProtKB, UniCarbKB, and CFG. As accumulating evidences show that glycans play important roles in the recognition between pathogen and host which is crucial to the entry and release of pathogen, SugarBindDB is therefore a valuable tool for mechanism study of pathogen infection and the toxicity of glycan-binding toxins.

5.5. GLAD: Glycan Array Dashboard

The glycan array is a high-throughput tool for profiling protein–glycan interactions. So far, thousands of glycan array experiments have been performed and a huge amount of data have been collected. However, the glycan array results were mostly stored as excel files (for example the CFG Core H glycan array data), which is not convenient to present the protein–glycan interactions visually. To address this limitation, GLAD (GLycan Array Dashboard, <https://glycokit.com/Tools/GLAD/>) is a web-based tool developed to visualize, analyze, and present glycan array data [71]. Users input data as tab-delimited text files in the correct format, and then can visualize and select data for display using various types of charts, including grouped bar charts, heatmaps, and interaction networks. GLAD also allows users to filter, sort and normalize data to accentuate key data and binding relationships. Overall, GLAD is a useful tool to uncover hidden relationships between glycan array datasets.

5.6. MCAW-DB

Glycan recognition patterns are often obtained by glycan microarrays. Multiple Carbohydrate Alignment with Weights Database (MCAW-DB, <https://mcawdb.glycoinfo.org/>) is a glycan profiling database containing the multiple alignment analysis results of 1081 glycan microarray samples collected from the CFG to find the glycan substructures having higher binding affinity to GBPs [72]. On the search page, users can filter taxa, protein family, investigator and array version. Text search is also permitted. The results show detailed sample information, multiple alignment analysis results and data set details, as well being as linked to the CFG database.

5.7. GlyMDB

Unlike MCAW-DB only focuses on glycan sequence alignment, the Glycan Microarray DataBase (GlyMDB, <http://www.glycanstructure.org/glymdb>) is a comprehensive glycan microarray database and analysis tool for data visualization, binder/non-binder classification, glycan-binding motif discovery, and glycan array sample comparison [73]. There are two options on the main pages. On one hand, users can upload microarray spreadsheet files. On the other hand, a query can be made by protein name, protein sequence or PDB ID within the 5203 glycan microarray samples collected from the CFG in the database. The GlyMDB results show glycan ligand information, fluorescence intensity, and common motifs. In addition, this database also provides some commonly used tools, such as binder/non-binder classification, glycan-binding motif discovery, glycan array sample comparison, and cross-linking of the glycan microarray to PDB. In addition, GlyMDB also provides bioinformatic tools such as Glycan Reader and Glycan Modeler for visualization and simulation of glycan 3D structures.

5.8. MatrixDB

Glycoproteins, proteoglycans and glycosaminoglycans (GAGs) are highly enriched in the extracellular matrix (ECM). MatrixDB (<http://matrixdb.univ-lyon1.fr/>) is a freely available database focused on these extracellular interactions [74–77]. The interaction data in the MatrixDB is collected from experiments and literatures with curation, as well as adopted from the International Molecular Exchange consortium (IMEx) databases. As many extracellular proteins usually assemble into complex multimers, MatrixDB provides not only the interactions with individual polypeptides, but also the interactions with multimers, which provide a more comprehensive perspective to understand the extracellular interaction network. In addition, the interacting proteins are linked to their expression and localization data in multiple databases including UniGene, the Human Protein Atlas, and Expression Atlas, which enable the users to create tissue- and cell-specific interaction networks.

6. Software Tools for Glycan and Intact Glycopeptide Analysis

Mass spectrometry (MS) has long been considered as one of the most powerful techniques for glycomic and glycoproteomic study. In general, there are two main strategies for elucidating glycosylation information using MS techniques [24]. The first strategy is for global analysis of glycan structure released from glycoproteins using endoglycosidases or chemical methods, which is useful for a rapid glycan profiling analysis, but lost the information of carrier protein. Another strategy is the analysis of intact glycopeptides after proteolytic digestion. As it provides information on both glycan composition and the attached protein, it is attracting attention in recent years. However, due to the complexity of glycans, interpreting MS output, in terms of glycan structures, attachment sites and glycopeptide is still challenging. Fortunately, recent advances in various searching softwares and tools greatly increased high throughput glycomic and glycoproteomic analysis. In addition, many notable reviews have summarized the informatics softwares and tools for glycan analysis [47,48] and intact glycopeptide analysis [49]. In this section, we introduce some representative software tools for glycan and intact *N*-glycopeptide analysis, as well as newly developed tools for intact *O*-glycopeptide.

6.1. Software Tools for Glycan Analysis

In order to determine the possible glycan structure based on experimentally determined masses, firstly we must know the theoretic MS fragments. GlycoWorkbench (Division of Molecular Biosciences, School of Life Sciences, Imperial College London, London, UK. Download link: <https://glycoworkbench.software.informer.com/2.1/>) is such a tool designed to facilitate manual annotation of mass spectrometry data by matching experimental MS/MS peak lists against theoretical fragments [78]. A particular feature of this tool is that it provides a user-friendly interface for drawing various glycans structure. After selection of fragmentation type and user-defined annotation options, the MS data can be automatically interpreted in several minutes or hours, which greatly accelerates the glycomics study. Currently, Glycoworkbench has become the most widely used software tool in MS-based glycomic studies.

Another software tool for interpretation of glycan profiling from LC/MS data, GlycReSoft (Program for Bioinformatics, Boston University, Boston MA, USA. Download link: <http://www.bumc.bu.edu/msr/glycresoft/>), has been updated recently. GlycReSoft is a software package implementing supervised and unsupervised scoring methods to enable assignment of peaks to both known and unknown glycan compositions [79]. In the updated version, they developed an optimized algorithm by using network Laplacian regularization to smooth LC-MS assignments of glycan compositions across multiple experimental protocols and thus improve the sensitivity and specificity of glycan composition assignment for LC-MS based experiments [50].

6.2. Software Tools for Intact N-Glycopeptide Analysis

In the last decade, the development of a number of bioinformatic tools for glycopeptide identification have facilitated the increase in the *N*-glycoproteome coverage. GPQuest (Center for Biomarker Discovery & Translation, the Johns Hopkins School of Medicine, Baltimore, MA, USA. Download link: <https://www.biomarkercenter.org/gpquest>) is one of the representative tools for large-scale identification of *N*-glycopeptides. In this algorithm, a spectral library of glycosite-containing peptides in the sample was built. By comparing with the relevant precursor ion of the intact glycopeptides, GPQuest assigns each intact glycopeptide MS/MS spectrum to a specific glycosite-containing peptide [80]. Combined this algorithm with other analytical methods, thousands of *N*-glycopeptides have been identified in different cell lines [51,81].

pGlyco (Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. Download link: <http://pfind.ict.ac.cn/software/pGlyco/index.html>) is another representative tool for global characterization of intact *N*-glycopeptides developed in recent years. pGlyco [82] and pGlyco 2.0 [83] use the optimized stepped-energy higher-energy collision dissociation (HCD) to select the mass spectra of glycopeptides and analyze the peptide sequence, and use collision-induced dissociation (CID) to analyze the glycan structure. It is noteworthy that although quality control has been considered in some other bioinformatic tools, pGlyco includes comprehensive quality control when matching mass spectra to glycans, peptides, and glycopeptides. Therefore, pGlyco and pGlyco 2.0 provide accurate identification about glycopeptides with a relatively lower false discovery rate compared to other bioinformatic tools such as Byonic™, the most commonly used commercial software. Using this search engine pGlyco 2.0, more than ten thousand intact *N*-glycopeptides have been identified in five mouse tissues [83], which achieved the deepest and largest scale ever reported [49].

6.3. Software Tools for Intact O-Glycopeptide Analysis

Compared to the mature and generic software tools for *N*-glycopeptide, the development of software tools for intact *O*-glycopeptide lags behind. In recent two years, several strategies and interpretation tools for *O*-glycopeptides are emerging [52–54,84]. *O*-*O*-Search is a new search scheme for the interpretation of *O*-glycopeptide HCD spectra [52]. By setting variable mass tags on the peptide level instead of Ser/Thr residues level, the search space is significantly reduced and many heterogeneous glycan structures could be considered. Currently, this method has been proven to

significantly outperform the conventional searching scheme for identification of *O*-glycopeptide in serum.

AOGP is a newly developed bioinformatic tool for intact *O*-glycopeptide analysis on single proteins [84]. By utilizing de novo sequencing for *O*-glycans, a database search strategy for peptide backbones and a false discovery rate (FDR) validation, AOGP automatically interprets intact *O*-glycopeptide mass spectra and provide information on both the glycan and possible glycosylated sites. Currently, AOGP has been found to well exhibit superior performance in identifying intact *O*-glycopeptides of model *O*-glycoproteins. Further improvement and optimization are needed for complex samples.

Overall, these intact glycopeptide identification methods could elucidate the glycan structures and the glycosylation sites precisely. However, there are still several disadvantages. First, the throughput of intact glycoprotein identification is relatively lower than released glycans. Second, the mapping of mass spectrums to glycan structures largely relies on the existing glycan structures in databases. It is still difficult to identify novel glycan structures at large scale. Until recently, a GlycoNovoDB tool provides a de novo sequencing approach to identify glycopeptides carrying novel *N*-glycans [85]. Third, these methods are widely used to identify *N*-glycopeptides but less often used for *O*-glycopeptides. Therefore, more high-throughput methods which can identify novel glycans and *O*-glycopeptides still need to be developed.

7. The Latest Integrated Glycoscience Portal

To understand the biological role of glycosylation, we need a comprehensive view of multiple pieces of glycobiology information including glycan structures, glycoprotein sequences and glycosylation sites, interactions between glycan and glycan binding proteins, and relative pathways. Although many resources are technically available, it still takes glycobiologists and researchers a lot of efforts to extract their needed information from different resources. Therefore, integrated portals where all glycoscience-related resources can be accessed from a single website are required. In this section, we introduce three representative integrated glycoscience portals which will benefit researchers in the glycoscience field.

7.1. Glycomics@ExPASy

Glycomics@ExPASy (<https://www.expasy.org/glycomics>) is the glycomics tab of ExPASy, centralizing glycoinformatic resources developed by glycoscientists. The aim of this database is to promote bioinformatics research in glycoscience [19]. It contains comprehensive databases and tools developed and maintained by SIB (such as GlyConnect, SugarBind and UniCarb-DB databases) and external resources (such as CAZy, CSDB, EPS-DB and Glyco3D databases).

GlyConnect (<https://glyconnect.expasy.org/>) is an integrated glycodata platform of Glycomics@ExPASy, which helps characterize the molecular components of protein glycosylation [86]. User can easily search the database by protein names (Figure 2b), monosaccharide compositions, structure properties of glycans (Figure 2c), and free glycans or browse by protein, structure, tissue, disease or cell lines (Figure 2d). A particular feature of this platform is that the results are presented in the form of octopus, which is very easy to understand the relation between protein, glycan, tissue and disease (Figure 2).

7.2. Glygen

GlyGen (<https://glygen.org/>) is another data integration and dissemination project for carbohydrate and glycoconjugate related data [18]. In the integration process, data are firstly retrieved and extracted from multiple international data sources including the National Center for Biotechnology Information (NCBI), UniProt, the Protein Data Bank (PDB), UniCarbKB, and the GlyTouCan glycan structure repository, and then standardized and harmonized [87]. The current version of the GlyGen Portal provides relevant data about *N*- and *O*-glycans from human, mice, rats, hepatitis C virus, SARS-CoV-1, and SARS-CoV-2. Users can simply search by protein accession, sequences, glycan

structure or monosaccharide composition, and then diverse information relevant to glycosylation would be displayed due to the powerful integrated ability of this portal. When searching for a glycan, the information including motifs, associated protein, biosynthetic enzymes and reference will be displayed (Figure 3b). And the information such as function, GO annotation, glycosylation including the structure and glycosite, pathway, disease, mutation, expression, and reference will be present for a protein (Figure 3c). More importantly, the data integrated in the GlyGen project are publicly available in standard formats supported by NCBI and EMBL-EBI, which greatly promotes standardization and sharing of data within the glycomics community.

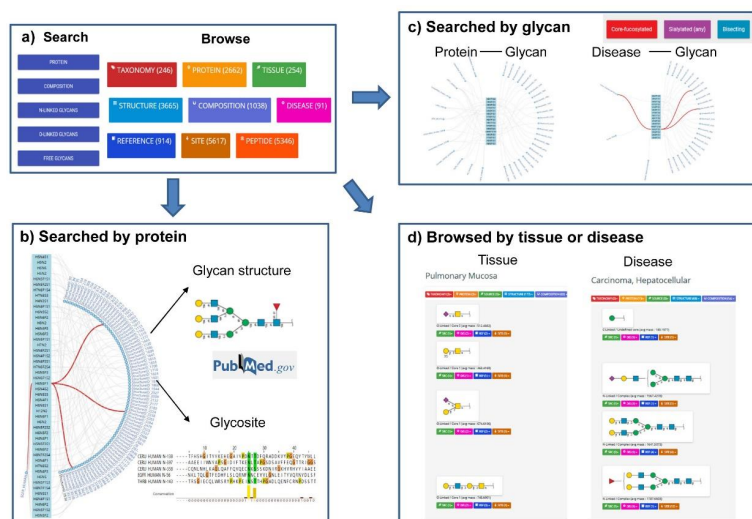


Figure 2. Search interfaces and integrated results of GlyConnect. (a) The two interfaces available in GlyConnect: Search and Browse. (b) The octopus result of glycan structures and relevant glycosites generated by querying protein UniProt accession (e.g., P00533). (c) The octopus results generated by querying glycan structures (e.g., “core-fucosylated”, “sialylated (any)”, and “bisecting”). The associated proteins with these glycans, and disease related to these glycans were shown. (d) The Browse interface displays the results of glycan structures related to a certain tissue (e.g., “pulmonary mucosa”) or disease (e.g., “carcinoma, hepatocellular”) in lists of items and proposes a global view of the data.

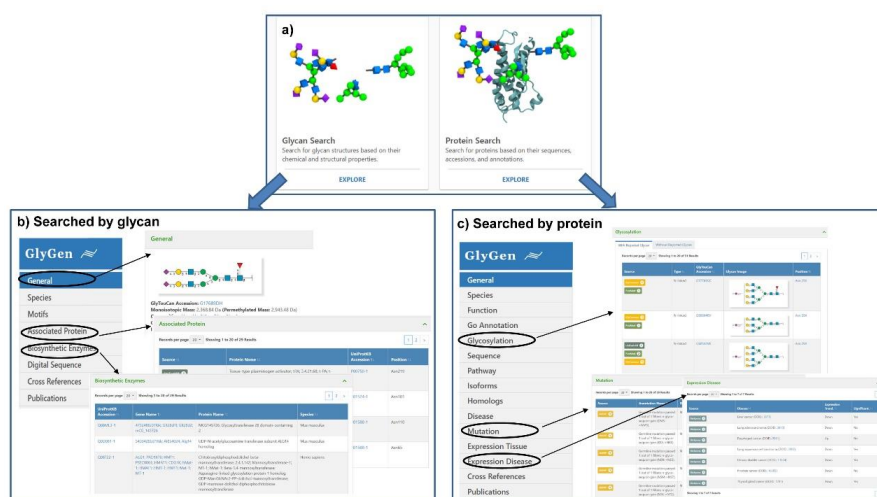


Figure 3. Search interfaces and integrated results of GlyGen. (a) The two interfaces available in GlyGen: Glycan Search and Protein Search. (b) The integrated result interface when searching with glycan accession (e.g., G17689DH) was shown. The detailed results of general information, associated protein, and biosynthetic enzyme were displayed. (c) The integrated result interface when searching with protein accession (e.g., P14210) was shown. The detailed results of glycosylation, mutation, and expression disease were displayed.

7.3. GlyCosmos

The GlyCosmos Glycoscience Portal (<https://glycosmos.org>) is the latest integrated web resource for accessing various kinds of glycoscience data resources including glycan-related genes, proteins, lipids, glycomes, pathways and diseases [20]. GlyCosmos integrates multiple resources including the databases developed by JCGGDB and updates with a four-month release cycle. Currently, GlyCosmos Glycogenes has integrated the glycan-related genes from GDGDB, CAZy, FlyGlycoDB, Lipid Maps and KEGG Orthology, while Glycoproteins has integrated glycoproteins from GlycoProtDB and UniProt with relevant links to Reactome pathways, Protein Data Bank, and the Human Proteome Atlas for further functional annotation of each glycoprotein. These data resources are listed on the main page of GlyCosmos (Figure 4a); users can simply click on the icon for a dataset of interest to access it. Also, user can search interest protein by name or accession, then the information on sequence, feature, related pathway, disease and PDB images if available will be displayed (Figure 4b). Compared to GlyConnect and Glygen, the glycoprotein data including their glycosylation sites and the binding lectins which identified by IGOT-MS and lectin microarray experiments are unique to GlyCosmos. Another feature of GlyCosmos is that it provides visualization of glycome profiling data on human, mouse and zebrafish tissue samples (GlycomeAtlas, Figure 4c), and also a lectin microarray-based glycome analysis of mouse tissue (LM-GlycomeAtlas, Figure 4d).

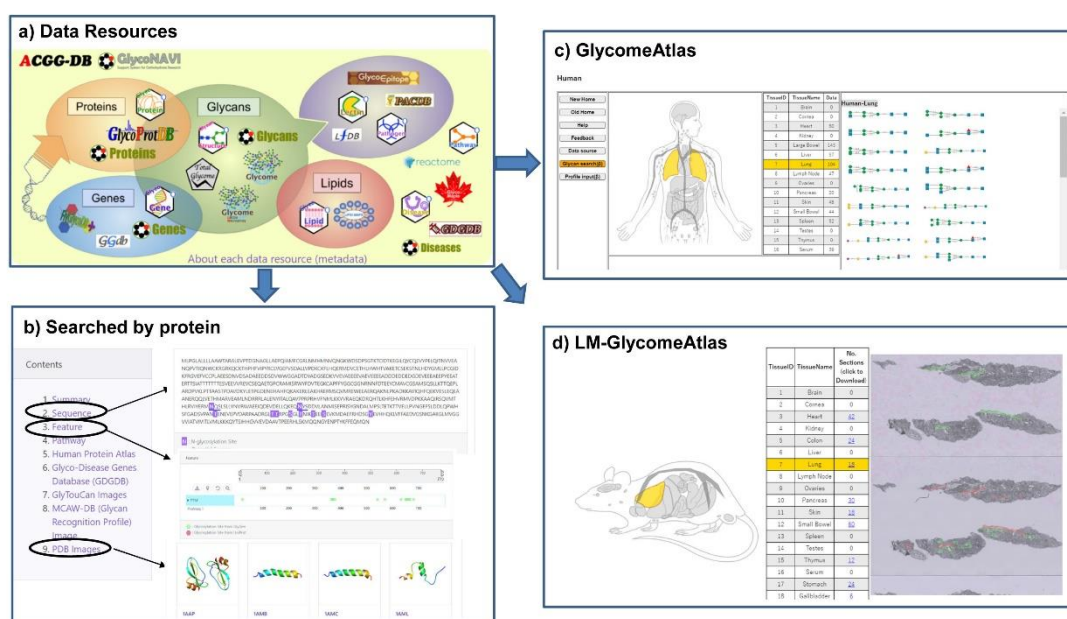


Figure 4. Data resources and integrated results in GlyCosmos. (a) An overview of the data resources available in GlyCosmos. Users simply click on the icon for a dataset of interest to access it. (b) The integrated result interface when searching with protein accession (e.g., P05067) was shown. The detailed results of sequence and feature with glycosites, and PDB images were displayed. The result interface of GlycomeAtlas (c) and LM-GlycomeAtlas (d) were shown.

In addition, GlyCosmos also provides three available repositories for data submission. GlyYouCan assigns unique accession codes to all unique glycans, while GlycoPOST and UniCarb-DR archives raw data generated from mass MS experiments on glycans and glycoproteins.

8. Discussion and Conclusions

With the development of glycobiology, the association between glycans, glycoproteins, and diseases has been disclosed gradually. For instance, in infectious diseases, glycosylation regulates host–virus interactions, viral immune evasion, and viral release for a range of pathogens such as SARS-CoV, influenza, HIV, and EBOV [88]. In the field of cancer, tumor cells usually bear a specific glycan expression

pattern, including truncated *O*-glycans, branched *N*-glycans, and diverse fucosylated and sialylated terminal structures [89]. Based on these facts, glycoproteins have become the promising candidates of disease markers and therapeutic targets, such as in liver fibrosis [90,91], cholangiocarcinoma [92], and lung disease [93]. Thus, fully integrated databases collecting the data on glycan-related molecules, glycogenes and their biological functions are needed. In past decades, with the development of various analytical techniques in glycomics and glycoproteomics, many databases with huge information are emerging. Recently, many notable reviews have summarized these databases and informatics tools [24,48,49,94]; however, most of them are from the perspective of analytical methods and data interpretation, which requires readers to have a certain background in glycomics. In this study, we reviewed the related databases of glycan, glycogenes, and interacting proteins in the process of glycoprotein formation (Figure 5, and Table 1), and gave a brief introduction of representative software tools for glycopeptide (Tables 2 and 3). We provided an overview of the main features, functionalities, and how to use these representative databases. We believe that this may make it easier for glycobiologists without glycomics or glycoproteomics background or researchers in other fields to use these databases for functional study or interdisciplinary study, ultimately promoting the clinical application of glycobiology.

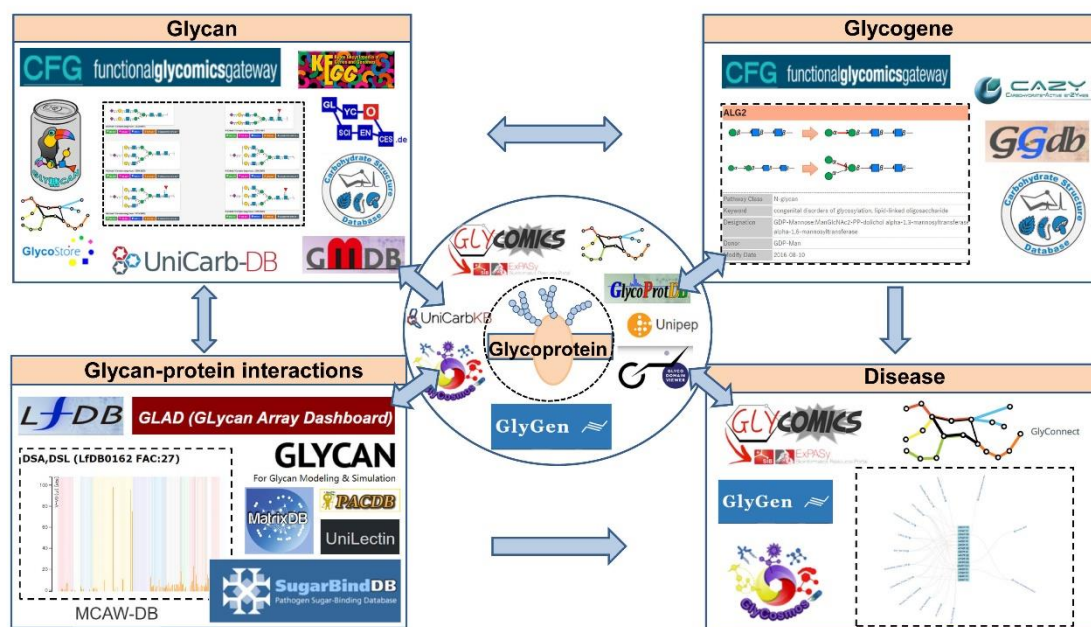


Figure 5. Illustration of integration of various databases to promote functional glycobiology study.

Table 1. Summary of representative glycoinformatic databases.

Name	Description	Retrievable by	URL
Glycan structure databases			
CFG glycan structure database [14]	Database providing structural and chemical information on thousands of glycans, including both synthetic glycans and glycans for mammalian species.	Searched by glycan names, composition, molecular weight, motifs, cell lines or tissue samples.	http://www.functionalglycomics.org/glycomics/molecule/jsp/carbohydrate/carbMoleculeHome.jsp
JCGGDB Glycan Mass spectral DataBase [16]	Database containing multi-stage tandem mass spectral of structurally defined <i>N</i> - and <i>O</i> -linked glycans, and glycolipid glycans.	Searched by glycan composition or m/z value of precursor ion.	https://jcgddb.jp/rcmg/glycodb/Ms_ResultSearch
UniCarbKB [25–27]	A curated database of information on glycan structures of glycoproteins, with descriptions of its biological source, supporting reference and experimental methods.	Searched by monosaccharide composition, attached protein, taxonomy or tissue by using an auto completion feature.	http://unicarbkb.org/
KEGG glycan [13]	Database providing information on experimentally determined glycan structures and their metabolic pathways.	Searched by the G number for each glycan structure.	http://www.genome.jp/kegg/glycan/
GLYCOSCIENCE.de [12,33]	An integrated portal containing databases and tools mainly glycan 3D structure analysis.	Searched by monosaccharide composition, molecular formula, structure classification and motifs, as well as NMR atoms or peaks.	http://www.glycosciences.de/
Glycosciences.DB [35]	The main glycan structure database of GLYCOSCIENCE.de, providing published data on glycan structures, their taxonomy, MS and NMR-experimental data, 3D structure models as well as references to PDB entries.	Searched by glycan (sub-)structure, monosaccharide composition, molecular formula, structure classification and motifs, as well as NMR, MS, PDB query, or bibliography queries.	http://www.glycosciences.de/database/
UniCarb-DB [36]	Database providing LC-MS/MS data of glycan structures.	Searched by taxonomy, tissue, reference, mass, composition or precursor mass.	https://unicarb-db.expasy.org/
GlyTouCan [37]	A international glycan sequence repository with a globally unique accession number assigned to each structure.	Searched by text input, motif, or drawing glycan structures in GlycanBuilder. Registered users can additionally register new glycan structures to obtain unique IDs for each structure.	https://glytoucan.org/

Table 1. Cont.

Name	Description	Retrievable by	URL
Glycan structure databases			
GlycoStore [41]	A curated database of information on glycan retention properties with chromatographic, electrophoretic and mass-spectrometry composition data.	Searched by experimental values (GU, AU or time), monosaccharide composition or metadata labels (taxonomy, sample name and the Oxford linear notation).	https://www.glycostore.org
CSDB [42]	Database on the structures of glycans and glycoconjugates in prokaryotes, plants and fungi.	Searched by CSDB ID, glycan substructure, composition, taxonomy, bibliography, NMR signals, conformation ID or GT name.	http://csdb.glycoscience.ru/database/
Glycoprotein databases			
GlycoProtDB [56]	Database providing information on <i>N</i> -glycoproteins and their glycosylated site(s) identified from <i>C. elegans</i> , mouse tissues and human.	Searched by gene ID, gene name, and its description (protein name).	https://acgg.asia/db/gpdb2/
UniPep [59]	Database providing information on <i>N</i> -glycopeptides identified from human plasma and tissues including bladder, breast, liver, lymphocytes, cerebrospinal fluid, and prostate.	Searched by gene name, gene symbol, Swiss Prot ID, IPI ID, protein sequence or peptide mass.	http://www.unipep.org/
<i>N</i> -GlycositeAtlas [22]	Database providing information on <i>N</i> -glycopeptides identified from over 100 publications and unpublished datasets	Searched by gene/protein name, accession number, glycosylation site location, glycosite containing peptide, tissue/liquid/cell line, or publication	http://nglycositeatlas.biomarkercenter.org
GlycoDomain Viewer [95]	Database of <i>O</i> -GalNAc proteins identified by SimpleCell technology from human and animal cell lines, associated with the verified and predicted glycosylated sites of <i>N</i> -glycan, <i>O</i> -GalNAc, <i>O</i> -Mannose and <i>O</i> -Xylose mapping on the protein sequence	Searched by the NCBI gene name or the Uniprot ID	https://glycodomain.glycomics.ku.dk/

Table 1. Cont.

Name	Description	Retrievable by	URL
Glycogene databases			
CAZy [62]	The largest database for display and analysis of genomic, structural and biochemical information on glyco-enzymes	Searched by enzyme family, protein name, organism name, GeneBank or UniProt accession, or EC number.	http://www.cazy.org/
CAZypedia [63]	A comprehensive encyclopedic of detailed structural, and biochemical information on glyco-enzymes, and relevant reference.	Searched by enzyme name or enzyme ID	http://www.cazypedia.org
GlycoGene DataBase [15]	Database providing information of glycogenes on gene sequences, substrate specificities, homologous genes, EC numbers, tissue distribution, KO mouse as well as external links to various databases.	Searched by gene symbols or designations or selected from the list of glycogenes	https://acgg.asia/ggdb2/
CFG glycosyltransferases database	Database providing information of glycosyltransferase on enzyme name, EC number, organism, relevant CFG data, and other data from public databases (PubMed, KEGG, CAZy, SwissProt, and others).	Providing a graphical interface of different glycans. By clicking a monosaccharide, users are directed to the information of the glycosyltransferase which forms this structure.	http://www.functionalglycomics.org/glycomics/molecule/jsp/glycoEnzyme/geMolecule.jsp
CSDB_GT [96]	A curated database of glycosyltransferases in <i>Arabidopsis thaliana</i> , <i>Escherichia coli</i> and <i>Saccharomyces cerevisiae</i> .	Searched by CSDB ID, glycan structure, composition, taxonomy, bibliography, NMR signals, conformation, or GT activity.	http://csdb.glycoscience.ru/gt.html
Glycan-protein interaction database			
Lectin Frontier Database [67]	Database providing quantitative interaction data between various glycan and lectins, as well as basic information such as kingdom, monosaccharide specificity on lectins	Searched by keyword or choose categories among Lectin family, Monosaccharide Specificity, or 3D-fold.	http://acgg.asia/lfdb2/

Table 1. Cont.

Name	Description	Retrievable by	URL
Glycan-protein interaction database			
UniLectin [68]	A interactive database for the classification and curation of lectins (with UniLectin3D module), and the prediction of β -propeller lectins (with PropLec module).	Searched by keywords, kingdom order, historical classification, monosaccharide, associate IUPAC sequence, fold of the binding site, or multiple criteria	https://www.unilectin.eu/
PACDB [69]	Database providing information on the interaction of microbial glycan-binding proteins and glycans with host glycan ligands	Selected from the list of disease, pathogen names, monosaccharides, or glycoepitopes	https://acgg.asia/db/diseases/pacdb
SugarBindDB [70]	A curated database providing information on known glycan structure interacted with pathogenic organisms (bacteria, toxins and viruses) in various disease	Searched by pathogenic agents, ligands, recognizing lectins, affected area, references, diseases or multi-criteria.	https://sugarbind.expasy.org/
GLAD [71]	A web-based tool to visualize, analyze, present, and mine glycan array data.	Input data as tab-delimited text files in the correct format	https://glycotookit.com/Tools/GLAD/
MCAW-DB [72]	Database providing information on binding affinity of glycan binding proteins to glycan substructures by multiple alignment analysis of glycan array data	Searched by filtering taxa, protein family, investigator and array version.	https://mcawdb.glycoinfo.org/
GlyMDB [73]	Database enabling users to upload their own microarray data, query binder/non-binder classification, discover glycan-binding motif, compare glycan array sample, and cross-link microarray samples to PDB structures	Searched by protein name, protein sequence or PDB ID, or upload microarray spreadsheet file.	http://www.glycanstructure.org/
MatrixDB [74]	A curated database providing information on interactions between extracellular matrix proteins, proteoglycans and polysaccharides.	Searched by a biomolecule, keyword, author, publication or IMEX identifier.	http://matrixdb.univ-lyon1.fr/

Table 1. Cont.

Name	Description	Retrievable by	URL
Latest integrated glycoscience portal			
Glycomics@ExpASY [19]	The glycomics tab of ExpASY, centralizing web-based glycoinformatics databases and tools resources developed by SIB (such as GlyConnect, SugarBind and UniCarb-DB databases) and other external resources to (such as CAZy, CSDB, GlyTouCan and UniCarbKB) to bridge the glycobiology and protein-oriented bioinformatics resources	Click on the link of interest	https://www.expasy.org/glycomics
GlyConnect [86]	The central platform of the Glycomics@ExpASY, providing interactive diagrams that help the user understand relations between glycans, proteins, tissues, diseases, and taxonomy	Either browsed or searched by protein name, ID, or monosaccharide composition, linkage type	https://glyconnect.expasy.org/
Glygen [18]	A web portal data integration, harmonization and dissemination web portal for integrate data and knowledge from diverse disciplines relevant to glycobiology, carbohydrate and glycoconjugate-related data retrieved from multiple international data sources including UniProtKB, GlyTouCan, UniCarbKB and other key resources.	Searched by protein accession, sequences, glycan structure or monosaccharide composition.	https://glygen.org/
GlyCosmos [20]	An integrated web resource including the database of JCGGDB and providing information on glycan-related genes, proteins, lipids, glycomes, pathways and diseases to integrate the glycosciences with the life sciences	Searched by protein name, protein accession, species, or various glycan search tools, such as by mass, composition, graphical glycan structure or monosaccharide composition.	https://glycosmos.org

Table 2. Bioinformatic tools for glycopeptide identification.

Bioinformatic Tools	Glycan Identification Method	Peptide Identification Method
GlycoWorkbench [78]	MS, MS/MS	MS
GlycReSoft [50,79]	LC-MS	LC-MS/MS
Byonic [97]	match by glycan mass	ETD, HCD
Protein Prospector [98]	match by glycan mass	ETD
pGlyco [82,83]	CID	HCD
GlycoNovoDB [85]	HCD	HCD
GPQuest [80]	HCD	HCD
GlycoPeptide Finder (GPFinder) [4]	QTOF-CID	QTOF-CID
MAGIC [99]	QTOF-CID	QTOF-CID
GlycoMaster DB [100]	HCD	ETD
GlycoFinder [101]	low energy HCD	HCD
Sweet-Heart [102]	CID	MS ³
Sweet-Heart for HCD [103]	CID	HCD
GlycoFragWork [104]	CID	ETD
pMatchGlyco [105]	match by MS/MS spectra	HCD
GlycoPeptideSearch [106,107]	CID	ETD
ArMone [108]	CID	HCD
GlypID 2.0 [109]	CID	HCD
O-O-Search [52]	HCD	HCD
AOGP [84]	HCD	HCD

Table 3. Bioinformatic tools for glycosylation site prediction.

Bioinformatic Tools	Description	URL
NetNGlyc	N-glycosylation site prediction	http://www.cbs.dtu.dk/services/NetNGlyc/
NetOGlyc [110]	O-GalNAc site prediction	http://www.cbs.dtu.dk/services/NetOGlyc/
YinOYang [60]	O-(beta)-GlcNAc and phosphorylation site prediction	http://www.cbs.dtu.dk/services/YinOYang/
DictyOGlyc [111]	O-(alpha)-GlcNAc site prediction	http://www.cbs.dtu.dk/services/DictyOGlyc/
NetCGlyc [112]	C-mannose site prediction	http://www.cbs.dtu.dk/services/NetCGlyc/
Big-PI Predictor [113]	GPI-anchor prediction	http://mendel.imp.ac.at/sat/gpi/gpi_server.html
GPI-SOM [114]	GPI-anchor prediction	http://genomics.unibe.ch/cgi-bin/gpi.cgi
PredGPI [115]	GPI-anchor prediction	http://gpcr.biocomp.unibo.it/predgpi/pred.htm
FragAnchor [116]	GPI-anchor prediction	http://navet.ics.hawaii.edu/~fraganchor/NNHMM/NNHMM.html

The integration is an important direction in the development of glycobiology database. The first level of integration should facilitate the cross-reference between databases. Presently, glycan structures are described in different formats including International Union of Pure and Applied Chemistry (IUPAC), LinearCode (r), KEGG Chemical Function (KCF), GlycoCT, GLYDE-II, Oxford, and LINUCS. Therefore, a universal nomenclature for glycan structures is needed for standardization and cross-referencing between databases. For example, the GlycanBuilder, which is developed for drawing glycan structures, also supports the conversion of various glycan structure formats. This tool is included in UniCarbKB and GlyTouCan, and facilitates the collection of all identified glycans in one database.

The second level of integration should link the data of glycan structure and glycoprotein. Due to the limitations of technology, glycan structures and glycosylation sites were identified separately in earlier research. However, only protein-specific glycosylation reveals the exact biological function. As the intact glycopeptide identification technologies are developing rapidly, the glycan information is integrated with glycoprotein data and leads to the establishment of comprehensive glycoprotein databases. In a future glycobiology database, the glycosylated protein will be annotated with the type of glycosylation with all the glycosites mapped. On one hand, all the known glycan structures linked to

each glycosylation site will be listed with detailed information. On the other hand, each known glycan structure will be assigned a unique ID, and all the proteins carrying this glycan could be provided.

The third level of integration would collect the information of glycosylated molecules, GBPs, glycogenes, their biological functions and relationship to diseases, and finally establish a comprehensive knowledge database. In the early days, some databases have made efforts in this integration. For example, JCGGDB has integrated MS data of glycans and glycoproteins, lectin affinity data, and glycogene information. KEGG glycan has summarized the glycan structure and their biosynthesis pathways in maps linked to metabolic and signaling pathways, while UniCarbKB has connected the glycan data to protein data in UniProtKB [117]. In recent years, with the development of several international cooperative initiatives, some new diseases with stronger integration abilities have been established, such as Glycomics@ExPASy, GlyGen, and GlyCosmos. These databases establish the standard format of glycans, and attempt to integrate glycoscience data with proteomics and genomics. On this basis, future databases would further strengthen the cross-linking of glycobiology resources with other omics including genomic, proteomic, and metabolomic data, and also include the biological function annotations of glycans and glycoproteins in diseases. Such a comprehensive glycomics database still needs plenty of additional experimental data and bioinformatic analysis. As the glycoscience community has prompted several international cooperative initiatives, this vision will become a reality in the near future.

Funding: This research was funded by the National Science and Technology Major Project of China (2018ZX10302205), the National Natural Science Foundation of China (31570796, 31770850 and 81802100), Shanghai Sailing Program (18YF1410500), Shanghai Jiao Tong University Interdiscipline with Medicine Program (YG2017MS63).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Apweiler, R.; Hermjakob, H.; Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the swiss-prot database. *Biochim. Biophys. Acta* **1999**, *1473*, 4–8. [[CrossRef](#)]
2. Mereiter, S.; Balmana, M.; Campos, D.; Gomes, J.; Reis, C.A. Glycosylation in the era of cancer-targeted therapy: Where are we heading? *Cancer Cell* **2019**, *36*, 6–16. [[CrossRef](#)] [[PubMed](#)]
3. Dalziel, M.; Crispin, M.; Scanlan, C.N.; Zitzmann, N.; Dwek, R.A. Emerging principles for the therapeutic exploitation of glycosylation. *Science* **2014**, *343*, 1235681. [[CrossRef](#)] [[PubMed](#)]
4. Strum, J.S.; Nwosu, C.C.; Hua, S.; Kronewitter, S.R.; Seipert, R.R.; Bachelor, R.J.; An, H.J.; Lebrilla, C.B. Automated assignments of n- and o-site specific glycosylation with extensive glycan heterogeneity of glycoprotein mixtures. *Anal. Chem.* **2013**, *85*, 5666–5675. [[CrossRef](#)] [[PubMed](#)]
5. Kailemia, M.J.; Xu, G.; Wong, M.; Li, Q.; Goonatilake, E.; Leon, F.; Lebrilla, C.B. Recent advances in the mass spectrometry methods for glycomics and cancer. *Anal. Chem.* **2018**, *90*, 208–224. [[CrossRef](#)]
6. Xiao, K.; Han, Y.; Yang, H.; Lu, H.; Tian, Z. Mass spectrometry-based qualitative and quantitative n-glycomics: An update of 2017–2018. *Anal. Chim. Acta* **2019**, *1091*, 1–22. [[CrossRef](#)]
7. Hirabayashi, J.; Kuno, A.; Tateno, H. Development and applications of the lectin microarray. *Top. Curr. Chem.* **2015**, *367*, 105–124.
8. Hyun, J.Y.; Pai, J.; Shin, I. The glycan microarray story from construction to applications. *Acc. Chem. Res.* **2017**, *50*, 1069–1078. [[CrossRef](#)]
9. Saravanan, C.; Cao, Z.; Head, S.R.; Panjwani, N. Analysis of differential expression of glycosyltransferases in healing corneas by glycogene microarrays. *Glycobiology* **2019**, *29*, 188–189. [[CrossRef](#)]
10. Doubet, S.; Bock, K.; Smith, D.; Darvill, A.; Albersheim, P. The complex carbohydrate structure database. *Trends Biochem. Sci.* **1989**, *14*, 475–477. [[CrossRef](#)]
11. Doubet, S.; Albersheim, P. Carbbank. *Glycobiology* **1992**, *2*, 505. [[CrossRef](#)] [[PubMed](#)]
12. Lutteke, T.; Bohne-Lang, A.; Loss, A.; Goetz, T.; Frank, M.; von der Lieth, C.W. Glycosciences. De: An internet portal to support glycomics and glycobiology research. *Glycobiology* **2006**, *16*, 71–81. [[CrossRef](#)] [[PubMed](#)]
13. Hashimoto, K.; Goto, S.; Kawano, S.; Aoki-Kinoshita, K.F.; Ueda, N.; Hamajima, M.; Kawasaki, T.; Kanehisa, M. Kegg as a glycome informatics resource. *Glycobiology* **2006**, *16*, 63–70. [[CrossRef](#)]

14. Raman, R.; Venkataraman, M.; Ramakrishnan, S.; Lang, W.; Raguram, S.; Sasisekharan, R. Advancing glycomics: Implementation strategies at the consortium for functional glycomics. *Glycobiology* **2006**, *16*, 82–90. [[CrossRef](#)]
15. Narimatsu, H. Construction of a human glycogene library and comprehensive functional analysis. *Glycoconj. J.* **2004**, *21*, 17–24. [[CrossRef](#)] [[PubMed](#)]
16. Kameyama, A.; Kikuchi, N.; Nakaya, S.; Ito, H.; Sato, T.; Shikanai, T.; Takahashi, Y.; Takahashi, K.; Narimatsu, H. A strategy for identification of oligosaccharide structures using observational multistage mass spectral library. *Anal. Chem.* **2005**, *77*, 4719–4725. [[CrossRef](#)]
17. Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.; Takahashi, N.; Isobe, T. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify n-linked glycoproteins. *Nat. Biotechnol.* **2003**, *21*, 667–672. [[CrossRef](#)]
18. York, W.S.; Mazumder, R.; Ranzinger, R.; Edwards, N.; Kahsay, R.; Aoki-Kinoshita, K.F.; Campbell, M.P.; Cummings, R.D.; Feizi, T.; Martin, M.; et al. Glygen: Computational and informatics resources for glycoscience. *Glycobiology* **2020**, *30*, 72–73. [[CrossRef](#)] [[PubMed](#)]
19. Mariethoz, J.; Alocci, D.; Gastaldello, A.; Horlacher, O.; Gasteiger, E.; Rojas-Macias, M.; Karlsson, N.G.; Packer, N.H.; Lisacek, F. Glycomics@expasy: Bridging the gap. *Mol. Cell. Proteom.* **2018**, *17*, 2164–2176. [[CrossRef](#)] [[PubMed](#)]
20. Yamada, I.; Shiota, M.; Shinmachi, D.; Ono, T.; Tsuchiya, S.; Hosoda, M.; Fujita, A.; Aoki, N.P.; Watanabe, Y.; Fujita, N.; et al. The glycosmos portal: A unified and comprehensive web resource for the glycosciences. *Nat. Methods* **2020**, *17*, 649–650. [[CrossRef](#)]
21. Akune, Y.; Lin, C.H.; Abrahams, J.L.; Zhang, J.; Packer, N.H.; Aoki-Kinoshita, K.F.; Campbell, M.P. Comprehensive analysis of the n-glycan biosynthetic pathway using bioinformatics to generate unicorn: A theoretical n-glycan structure database. *Carbohydr. Res.* **2016**, *431*, 56–63. [[CrossRef](#)] [[PubMed](#)]
22. Sun, S.; Hu, Y.; Ao, M.; Shah, P.; Chen, J.; Yang, W.; Jia, X.; Tian, Y.; Thomas, S.; Zhang, H. N-glycositeatlas: A database resource for mass spectrometry-based human n-linked glycoprotein and glycosylation site mapping. *Clin. Proteom.* **2019**, *16*, 35. [[CrossRef](#)] [[PubMed](#)]
23. Egorova, K.S.; Toukach, P.V. Glycoinformatics: Bridging isolated islands in the sea of data. *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 14986–14990. [[CrossRef](#)] [[PubMed](#)]
24. Abrahams, J.L.; Taherzadeh, G.; Jarvas, G.; Guttman, A.; Zhou, Y.; Campbell, M.P. Recent advances in glycoinformatic platforms for glycomics and glycoproteomics. *Curr. Opin. Struct. Biol.* **2020**, *62*, 56–69. [[CrossRef](#)] [[PubMed](#)]
25. Campbell, M.P.; Hayes, C.A.; Struwe, W.B.; Wilkins, M.R.; Aoki-Kinoshita, K.F.; Harvey, D.J.; Rudd, P.M.; Kolarich, D.; Lisacek, F.; Karlsson, N.G.; et al. Unicarbkb: Putting the pieces together for glycomics research. *Proteomics* **2011**, *11*, 4117–4121. [[CrossRef](#)]
26. Campbell, M.P.; Packer, N.H. Unicarbkb: New database features for integrating glycan structure abundance, compositional glycoproteomics data, and disease associations. *Biochim. Biophys. Acta* **2016**, *1860*, 1669–1675. [[CrossRef](#)]
27. Campbell, M.P.; Peterson, R.; Mariethoz, J.; Gasteiger, E.; Akune, Y.; Aoki-Kinoshita, K.F.; Lisacek, F.; Packer, N.H. Unicarbkb: Building a knowledge platform for glycoproteomics. *Nucleic Acids Res.* **2014**, *42*, D215–D221. [[CrossRef](#)]
28. Cooper, C.A.; Harrison, M.J.; Wilkins, M.R.; Packer, N.H. Glycositedb: A new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res.* **2001**, *29*, 332–335. [[CrossRef](#)]
29. Cooper, C.A.; Joshi, H.J.; Harrison, M.J.; Wilkins, M.R.; Packer, N.H. Glycositedb: A curated relational database of glycoprotein glycan structures and their biological sources. 2003 update. *Nucleic Acids Res.* **2003**, *31*, 511–513. [[CrossRef](#)]
30. Royle, L.; Radcliffe, C.M.; Dwek, R.A.; Rudd, P.M. Detailed structural analysis of n-glycans released from glycoproteins in sds-page gel bands using hplc combined with exoglycosidase array digestions. *Methods Mol. Biol.* **2006**, *347*, 125–143.
31. von der Lieth, C.W.; Freire, A.A.; Blank, D.; Campbell, M.P.; Ceroni, A.; Damerell, D.R.; Dell, A.; Dwek, R.A.; Ernst, B.; Fogh, R.; et al. Eurocarbdb: An open-access platform for glycoinformatics. *Glycobiology* **2011**, *21*, 493–502. [[CrossRef](#)] [[PubMed](#)]

32. Gotz, L.; Abrahams, J.L.; Mariethoz, J.; Rudd, P.M.; Karlsson, N.G.; Packer, N.H.; Campbell, M.P.; Lisacek, F. Glycodigest: A tool for the targeted use of exoglycosidase digestions in glycan structure determination. *Bioinformatics* **2014**, *30*, 3131–3133. [[CrossRef](#)] [[PubMed](#)]
33. Lutteke, T.; von der Lieth, C.W. Data mining the pdb for glyco-related data. *Methods Mol. Biol.* **2009**, *534*, 293–310. [[PubMed](#)]
34. Bohne, A.; Lang, E.; von der Lieth, C.W. Sweet-www-based rapid 3d construction of oligo- and polysaccharides. *Bioinformatics* **1999**, *15*, 767–768. [[CrossRef](#)] [[PubMed](#)]
35. Bohm, M.; Bohne-Lang, A.; Frank, M.; Loss, A.; Rojas-Macias, M.A.; Lutteke, T. Glycosciences.Db: An annotated data collection linking glycomics and proteomics data (2018 update). *Nucleic Acids Res.* **2019**, *47*, D1195–D1201. [[CrossRef](#)]
36. Campbell, M.P.; Nguyen-Khuong, T.; Hayes, C.A.; Flowers, S.A.; Alagesan, K.; Kolarich, D.; Packer, N.H.; Karlsson, N.G. Validation of the curation pipeline of unicarb-db: Building a global glycan reference ms/ms repository. *Biochim. Biophys. Acta* **2014**, *1844*, 108–116. [[CrossRef](#)]
37. Aoki-Kinoshita, K.; Agravat, S.; Aoki, N.P.; Arpinar, S.; Cummings, R.D.; Fujita, A.; Fujita, N.; Hart, G.M.; Haslam, S.M.; Kawasaki, T.; et al. Glytoucan 1.0—The international glycan structure repository. *Nucleic Acids Res.* **2016**, *44*, D1237–D1242. [[CrossRef](#)]
38. Ranzinger, R.; Herget, S.; Wetter, T.; von der Lieth, C.W. Glycomedb-integration of open-access carbohydrate structure databases. *BMC Bioinform.* **2008**, *9*, 384. [[CrossRef](#)]
39. Ranzinger, R.; Herget, S.; von der Lieth, C.W.; Frank, M. Glycomedb—A unified database for carbohydrate structures. *Nucleic Acids Res.* **2011**, *39*, D373–D376. [[CrossRef](#)]
40. Tiemeyer, M.; Aoki, K.; Paulson, J.; Cummings, R.D.; York, W.S.; Karlsson, N.G.; Lisacek, F.; Packer, N.H.; Campbell, M.P.; Aoki, N.P.; et al. Glytoucan: An accessible glycan structure repository. *Glycobiology* **2017**, *27*, 915–919. [[CrossRef](#)]
41. Zhao, S.; Walsh, I.; Abrahams, J.L.; Royle, L.; Nguyen-Khuong, T.; Spencer, D.; Fernandes, D.L.; Packer, N.H.; Rudd, P.M.; Campbell, M.P. Glycostore: A database of retention properties for glycan analysis. *Bioinformatics* **2018**, *34*, 3231–3232. [[CrossRef](#)] [[PubMed](#)]
42. Toukach, P.V.; Egorova, K.S. Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts. *Nucleic Acids Res.* **2016**, *44*, D1229–D1236. [[CrossRef](#)]
43. Toukach, P.V.; Egorova, K.S. Bacterial, plant, and fungal carbohydrate structure databases: Daily usage. *Methods Mol. Biol.* **2015**, *1273*, 55–85. [[PubMed](#)]
44. Toukach, P.; Egorova, K. Carbohydrate structure database (csdb): Examples of usage. In *A Practical Guide to Using Glycomics Databases*; Aoki-Kinoshita, K.F., Ed.; Springer: Tokyo, Japan, 2017; pp. 75–113.
45. Bennett, E.P.; Mandel, U.; Clausen, H.; Gerken, T.A.; Fritz, T.A.; Tabak, L.A. Control of mucin-type o-glycosylation: A classification of the polypeptide galnac-transferase gene family. *Glycobiology* **2012**, *22*, 736–756. [[CrossRef](#)] [[PubMed](#)]
46. Holt, G.D.; Hart, G.W. The subcellular distribution of terminal n-acetylglucosamine moieties. Localization of a novel protein-saccharide linkage, o-linked glcnac. *J. Biol. Chem.* **1986**, *261*, 8049–8057.
47. Walsh, I.; Zhao, S.; Campbell, M.; Taron, C.H.; Rudd, P.M. Quantitative profiling of glycans and glycopeptides: An informatics' perspective. *Curr. Opin. Struct. Biol.* **2016**, *40*, 70–80. [[CrossRef](#)] [[PubMed](#)]
48. Lisacek, F.; Mariethoz, J.; Alocci, D.; Rudd, P.M.; Abrahams, J.L.; Campbell, M.P.; Packer, N.H.; Stahle, J.; Widmalm, G.; Mullen, E.; et al. Databases and associated tools for glycomics and glycoproteomics. *Methods Mol. Biol.* **2017**, *1503*, 235–264.
49. Cao, W.; Liu, M.; Kong, S.; Wu, M.; Zhang, Y.; Yang, P. Recent advances in software tools for more generic and precise intact glycopeptide analysis. *Mol. Cell Proteom.* **2020**. [[CrossRef](#)]
50. Klein, J.; Carvalho, L.; Zaia, J. Application of network smoothing to glycan lc-ms profiling. *Bioinformatics* **2018**, *34*, 3511–3518. [[CrossRef](#)]
51. Sun, S.; Shah, P.; Eshghi, S.T.; Yang, W.; Trikanad, N.; Yang, S.; Chen, L.; Aiyetan, P.; Hoti, N.; Zhang, Z.; et al. Comprehensive analysis of protein glycosylation by solid-phase extraction of n-linked glycans and glycosite-containing peptides. *Nat. Biotechnol.* **2016**, *34*, 84–88. [[CrossRef](#)]
52. Mao, J.; You, X.; Qin, H.; Wang, C.; Wang, L.; Ye, M. A new searching strategy for the identification of o-linked glycopeptides. *Anal. Chem.* **2019**, *91*, 3852–3859. [[CrossRef](#)] [[PubMed](#)]
53. Ye, Z.; Mao, Y.; Clausen, H.; Vakhrushev, S.Y. Glyco-dia: A method for quantitative o-glycoproteomics with in silico-boosted glycopeptide libraries. *Nat. Methods* **2019**, *16*, 902–910. [[CrossRef](#)] [[PubMed](#)]

54. Yang, W.; Ao, M.; Hu, Y.; Li, Q.K.; Zhang, H. Mapping the o-glycoproteome using site-specific extraction of o-linked glycopeptides (exoo). *Mol. Syst. Biol.* **2018**, *14*, e8486. [[CrossRef](#)]
55. Kaji, H.; Yamauchi, Y.; Takahashi, N.; Isobe, T. Mass spectrometric identification of n-linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable isotope tagging. *Nat. Protoc.* **2006**, *1*, 3019–3027. [[CrossRef](#)]
56. Kaji, H.; Shikanai, T.; Sasaki-Sawa, A.; Wen, H.; Fujita, M.; Suzuki, Y.; Sugahara, D.; Sawaki, H.; Yamauchi, Y.; Shinkawa, T.; et al. Large-scale identification of n-glycosylated proteins of mouse tissues and construction of a glycoprotein database, glycoprotodb. *J. Proteome Res.* **2012**, *11*, 4553–4566. [[CrossRef](#)]
57. Sugahara, D.; Kaji, H.; Sugihara, K.; Asano, M.; Narimatsu, H. Large-scale identification of target proteins of a glycosyltransferase isozyme by lectin-igot-ic/ms, an ic/ms-based glycoproteomic approach. *Sci. Rep.* **2012**, *2*, 680. [[CrossRef](#)]
58. Kaji, H.; Ocho, M.; Togayachi, A.; Kuno, A.; Sogabe, M.; Ohkura, T.; Nozaki, H.; Angata, T.; Chiba, Y.; Ozaki, H.; et al. Glycoproteomic discovery of serological biomarker candidates for hcv/hbv infection-associated liver fibrosis and hepatocellular carcinoma. *J. Proteome Res.* **2013**, *12*, 2630–2640. [[CrossRef](#)]
59. Zhang, H.; Loriaux, P.; Eng, J.; Campbell, D.; Keller, A.; Moss, P.; Bonneau, R.; Zhang, N.; Zhou, Y.; Wollscheid, B.; et al. Unipep—A database for human n-linked glycosites: A resource for biomarker discovery. *Genome Biol.* **2006**, *7*, R73. [[CrossRef](#)]
60. Gupta, R.; Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* **2002**, 310–322.
61. Blom, N.; Sicheritz-Ponten, T.; Gupta, R.; Gammeltoft, S.; Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **2004**, *4*, 1633–1649. [[CrossRef](#)] [[PubMed](#)]
62. Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The carbohydrate-active enzymes database (cazy): An expert resource for glycogenomics. *Nucleic Acids Res.* **2009**, *37*, D233–D238. [[CrossRef](#)] [[PubMed](#)]
63. Consortium, C.A. Ten years of cazypedia: A living encyclopedia of carbohydrate-active enzymes. *Glycobiology* **2018**, *28*, 3–8. [[CrossRef](#)] [[PubMed](#)]
64. Egorova, K.S.; Knirel, Y.A.; Toukach, P.V. Expanding csdb_gt glycosyltransferase database with escherichia coli. *Glycobiology* **2019**, *29*, 285–287. [[CrossRef](#)] [[PubMed](#)]
65. Hansen, L.; Lind-Thomsen, A.; Joshi, H.J.; Pedersen, N.B.; Have, C.T.; Kong, Y.; Wang, S.; Sparso, T.; Grarup, N.; Vester-Christensen, M.B.; et al. A glycogene mutation map for discovery of diseases of glycosylation. *Glycobiology* **2015**, *25*, 211–224. [[CrossRef](#)]
66. Narimatsu, H.; Sawaki, H.; Kuno, A.; Kaji, H.; Ito, H.; Ikehara, Y. A strategy for discovery of cancer glyco-biomarkers in serum using newly developed technologies for glycoproteomics. *FEBS J.* **2010**, *277*, 95–105. [[CrossRef](#)]
67. Hirabayashi, J.; Tateno, H.; Shikanai, T.; Aoki-Kinoshita, K.F.; Narimatsu, H. The lectin frontier database (lfd), and data generation based on frontal affinity chromatography. *Molecules* **2015**, *20*, 951–973. [[CrossRef](#)]
68. Bonnardel, F.; Perez, S.; Lisacek, F.; Imberty, A. Structural database for lectins and the unilectin web platform. *Methods Mol. Biol.* **2020**, *2132*, 1–14.
69. Maeda, M.; Fujita, N.; Suzuki, Y.; Sawaki, H.; Shikanai, T.; Narimatsu, H. Jcggdb: Japan consortium for glycobiology and glycotecchnology database. *Methods Mol. Biol.* **2015**, *1273*, 161–179.
70. Mariethoz, J.; Khatib, K.; Alocci, D.; Campbell, M.P.; Karlsson, N.G.; Packer, N.H.; Mullen, E.H.; Lisacek, F. Sugarbinddb, a resource of glycan-mediated host-pathogen interactions. *Nucleic Acids Res.* **2016**, *44*, D1243–D1250. [[CrossRef](#)]
71. Mehta, A.Y.; Cummings, R.D. Glad: Glycan array dashboard, a visual analytics tool for glycan microarrays. *Bioinformatics* **2019**, *35*, 3536–3537. [[CrossRef](#)]
72. Hosoda, M.; Takahashi, Y.; Shiota, M.; Shinmachi, D.; Inomoto, R.; Higashimoto, S.; Aoki-Kinoshita, K.F. Mcaw-db: A glycan profile database capturing the ambiguity of glycan recognition patterns. *Carbohydr. Res.* **2018**, *464*, 44–56. [[CrossRef](#)] [[PubMed](#)]
73. Cao, Y.; Park, S.J.; Mehta, A.Y.; Cummings, R.D.; Im, W. Glymdb: Glycan microarray database and analysis toolset. *Bioinformatics* **2020**, *36*, 2438–2442. [[CrossRef](#)]
74. Chautard, E.; Ballut, L.; Thierry-Mieg, N.; Ricard-Blum, S. Matrixdb, a database focused on extracellular protein-protein and protein-carbohydrate interactions. *Bioinformatics* **2009**, *25*, 690–691. [[CrossRef](#)]

75. Chautard, E.; Fatoux-Ardore, M.; Ballut, L.; Thierry-Mieg, N.; Ricard-Blum, S. Matrixdb, the extracellular matrix interaction database. *Nucleic Acids Res.* **2011**, *39*, D235–D240. [[CrossRef](#)]
76. Launay, G.; Salza, R.; Multedo, D.; Thierry-Mieg, N.; Ricard-Blum, S. Matrixdb, the extracellular matrix interaction database: Updated content, a new navigator and expanded functionalities. *Nucleic Acids Res.* **2015**, *43*, D321–D327. [[CrossRef](#)]
77. Clerc, O.; Deniaud, M.; Vallet, S.D.; Naba, A.; Rivet, A.; Perez, S.; Thierry-Mieg, N.; Ricard-Blum, S. Matrixdb: Integration of new data with a focus on glycosaminoglycan interactions. *Nucleic Acids Res.* **2019**, *47*, D376–D381. [[CrossRef](#)]
78. Damerell, D.; Ceroni, A.; Maass, K.; Ranzinger, R.; Dell, A.; Haslam, S.M. The glycanbuilder and glycoworkbench glycoinformatics tools: Updates and new developments. *Biol. Chem.* **2012**, *393*, 1357–1362. [[CrossRef](#)]
79. Maxwell, E.; Tan, Y.; Tan, Y.; Hu, H.; Benson, G.; Aizikov, K.; Conley, S.; Staples, G.O.; Slysz, G.W.; Smith, R.D.; et al. Glycosoft: A software package for automated recognition of glycans from lc/ms data. *PLoS ONE* **2012**, *7*, e45474. [[CrossRef](#)] [[PubMed](#)]
80. Toghi Eshghi, S.; Shah, P.; Yang, W.; Li, X.; Zhang, H. Gpquest: A spectral library matching algorithm for site-specific assignment of tandem mass spectra to intact n-glycopeptides. *Anal. Chem.* **2015**, *87*, 5181–5188. [[CrossRef](#)] [[PubMed](#)]
81. Yang, G.; Hu, Y.; Sun, S.; Ouyang, C.; Yang, W.; Wang, Q.; Betenbaugh, M.; Zhang, H. Comprehensive glycoproteomic analysis of chinese hamster ovary cells. *Anal. Chem.* **2018**, *90*, 14294–14302. [[CrossRef](#)]
82. Zeng, W.F.; Liu, M.Q.; Zhang, Y.; Wu, J.Q.; Fang, P.; Peng, C.; Nie, A.; Yan, G.; Cao, W.; Liu, C.; et al. Pglyco: A pipeline for the identification of intact n-glycopeptides by using hcd- and cid-ms/ms and ms3. *Sci. Rep.* **2016**, *6*, 25102. [[CrossRef](#)] [[PubMed](#)]
83. Liu, M.Q.; Zeng, W.F.; Fang, P.; Cao, W.Q.; Liu, C.; Yan, G.Q.; Zhang, Y.; Peng, C.; Wu, J.Q.; Zhang, X.J.; et al. Pglyco 2.0 enables precision n-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. *Nat. Commun.* **2017**, *8*, 438. [[CrossRef](#)] [[PubMed](#)]
84. Huang, J.; Jiang, B.; Zhao, H.; Wu, M.; Kong, S.; Liu, M.; Yang, P.; Cao, W. Development of a computational tool for automated interpretation of intact o-glycopeptide tandem mass spectra from single proteins. *Anal. Chem.* **2020**, *92*, 6777–6784. [[CrossRef](#)] [[PubMed](#)]
85. Sun, W.; Liu, Y.; Lajoie, G.A.; Ma, B.; Zhang, K. An improved approach for n-linked glycan structure identification from hcd ms/ms spectra. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2019**, *16*, 388–395. [[CrossRef](#)] [[PubMed](#)]
86. Alocchi, D.; Mariethoz, J.; Gastaldello, A.; Gasteiger, E.; Karlsson, N.G.; Kolarich, D.; Packer, N.H.; Lisacek, F. Glyconnect: Glycoproteomics goes visual, interactive, and analytical. *J. Proteome Res.* **2019**, *18*, 664–677. [[CrossRef](#)]
87. Khasay, R.; Vora, J.; Navelkar, R.; Mousavi, R.; Fochtman, B.C.; Holmes, X.; Pattabiraman, N.; Ranzinger, R.; Mahadik, R.; Williamson, T.; et al. Glygen data model and processing workflow. *Bioinformatics* **2020**, *36*, 3941–3943. [[CrossRef](#)]
88. Watanabe, Y.; Bowden, T.A.; Wilson, I.A.; Crispin, M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochim. Biophys. Acta Gen. Subj.* **2019**, *1863*, 1480–1497. [[CrossRef](#)]
89. Pinho, S.S.; Reis, C.A. Glycosylation in cancer: Mechanisms and clinical implications. *Nat. Rev. Cancer* **2015**, *15*, 540–555. [[CrossRef](#)]
90. Kuno, A.; Ikehara, Y.; Tanaka, Y.; Ito, K.; Matsuda, A.; Sekiya, S.; Hige, S.; Sakamoto, M.; Kage, M.; Mizokami, M.; et al. A serum “sweet-doughnut” protein facilitates fibrosis evaluation and therapy assessment in patients with viral hepatitis. *Sci. Rep.* **2013**, *3*, 1065. [[CrossRef](#)] [[PubMed](#)]
91. Zou, X.; Zhu, M.Y.; Yu, D.M.; Li, W.; Zhang, D.H.; Lu, F.J.; Gong, Q.M.; Liu, F.; Jiang, J.H.; Zheng, M.H.; et al. Serum wfa(+)-m2bp levels for evaluation of early stages of liver fibrosis in patients with chronic hepatitis b virus infection. *Liver Int.* **2017**, *37*, 35–44. [[CrossRef](#)]
92. Matsuda, A.; Kuno, A.; Matsuzaki, H.; Kawamoto, T.; Shikanai, T.; Nakanuma, Y.; Yamamoto, M.; Ohkohchi, N.; Ikehara, Y.; Shoda, J.; et al. Glycoproteomics-based cancer marker discovery adopting dual enrichment with wisteria floribunda agglutinin for high specific glyco-diagnosis of cholangiocarcinoma. *J. Proteom.* **2013**, *85*, 1–11. [[CrossRef](#)] [[PubMed](#)]

93. Zou, X.; Yao, F.; Yang, F.; Zhang, F.; Xu, Z.; Shi, J.; Kuno, A.; Zhao, H.; Zhang, Y. Glycomic signatures of plasma igg improve preoperative prediction of the invasiveness of small lung nodules. *Molecules* **2019**, *25*, 28. [[CrossRef](#)] [[PubMed](#)]
94. Rojas-Macias, M.A.; Mariethoz, J.; Andersson, P.; Jin, C.; Venkatakrisnan, V.; Aoki, N.P.; Shinmachi, D.; Ashwood, C.; Madunic, K.; Zhang, T.; et al. Towards a standardized bioinformatics infrastructure for n- and o-glycomics. *Nat. Commun.* **2019**, *10*, 3275. [[CrossRef](#)] [[PubMed](#)]
95. Steentoft, C.; Vakhrushev, S.Y.; Joshi, H.J.; Kong, Y.; Vester-Christensen, M.B.; Schjoldager, K.T.; Lavrsen, K.; Dabelsteen, S.; Pedersen, N.B.; Marcos-Silva, L.; et al. Precision mapping of the human o-galnac glycoproteome through simplecell technology. *EMBO J.* **2013**, *32*, 1478–1488. [[CrossRef](#)]
96. Egorova, K.S.; Toukach, P.V. Csdg: A new curated database on glycosyltransferases. *Glycobiology* **2017**, *27*, 285–290. [[CrossRef](#)]
97. Bern, M.; Kil, Y.J.; Becker, C. Byonic: Advanced peptide and protein identification software. *Curr. Protoc. Bioinform.* **2012**, *40*, 13–20. [[CrossRef](#)]
98. Medzihradzky, K.F.; Kaasik, K.; Chalkley, R.J. Tissue-specific glycosylation at the glycopeptide level. *Mol. Cell Proteom.* **2015**, *14*, 2103–2110. [[CrossRef](#)]
99. Lynn, K.S.; Chen, C.C.; Lih, T.M.; Cheng, C.W.; Su, W.C.; Chang, C.H.; Cheng, C.Y.; Hsu, W.L.; Chen, Y.J.; Sung, T.Y. Magic: An automated n-linked glycoprotein identification tool using a y1-ion pattern matching algorithm and in silico ms(2) approach. *Anal. Chem.* **2015**, *87*, 2466–2473. [[CrossRef](#)]
100. He, L.; Xin, L.; Shan, B.; Lajoie, G.A.; Ma, B. Glycomaster db: Software to assist the automated identification of n-linked glycopeptides by tandem mass spectrometry. *J. Proteome Res.* **2014**, *13*, 3881–3895. [[CrossRef](#)]
101. Cao, L.; Tolic, N.; Qu, Y.; Meng, D.; Zhao, R.; Zhang, Q.; Moore, R.J.; Zink, E.M.; Lipton, M.S.; Pasa-Tolic, L.; et al. Characterization of intact n- and o-linked glycopeptides using higher energy collisional dissociation. *Anal. Biochem.* **2014**, *452*, 96–102. [[CrossRef](#)]
102. Wu, S.W.; Liang, S.Y.; Pu, T.H.; Chang, F.Y.; Khoo, K.H. Sweet-heart-an integrated suite of enabling computational tools for automated ms2/ms3 sequencing and identification of glycopeptides. *J. Proteom.* **2013**, *84*, 1–16. [[CrossRef](#)] [[PubMed](#)]
103. Wu, S.W.; Pu, T.H.; Viner, R.; Khoo, K.H. Novel lc-ms(2) product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. *Anal. Chem.* **2014**, *86*, 5478–5486. [[CrossRef](#)] [[PubMed](#)]
104. Mayampurath, A.; Yu, C.Y.; Song, E.; Balan, J.; Mechref, Y.; Tang, H. Computational framework for identification of intact glycopeptides in complex samples. *Anal. Chem.* **2014**, *86*, 453–463. [[CrossRef](#)] [[PubMed](#)]
105. Pagan, J.D.; Kitaoka, M.; Anthony, R.M. Engineered sialylation of pathogenic antibodies in vivo attenuates autoimmune disease. *Cell* **2018**, *172*, 564–577e513. [[CrossRef](#)]
106. Pompach, P.; Chandler, K.B.; Lan, R.; Edwards, N.; Goldman, R. Semi-automated identification of n-glycopeptides by hydrophilic interaction chromatography, nano-reverse-phase lc-ms/ms, and glycan database search. *J. Proteome Res.* **2012**, *11*, 1728–1740. [[CrossRef](#)]
107. Chandler, K.B.; Pompach, P.; Goldman, R.; Edwards, N. Exploring site-specific n-glycosylation microheterogeneity of haptoglobin using glycopeptide cid tandem mass spectra and glycan database search. *J. Proteome Res.* **2013**, *12*, 3652–3666. [[CrossRef](#)]
108. Cheng, K.; Chen, R.; Seebun, D.; Ye, M.; Figeys, D.; Zou, H. Large-scale characterization of intact n-glycopeptides using an automated glycoproteomic method. *J. Proteom.* **2014**, *110*, 145–154. [[CrossRef](#)]
109. Mayampurath, A.M.; Wu, Y.; Segu, Z.M.; Mechref, Y.; Tang, H. Improving confidence in detection and characterization of protein n-glycosylation sites and microheterogeneity. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 2007–2019. [[CrossRef](#)]
110. Hansen, J.E.; Lund, O.; Tolstrup, N.; Gooley, A.A.; Williams, K.L.; Brunak, S. Netoglyc: Prediction of mucin type o-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.* **1998**, *15*, 115–130. [[CrossRef](#)]
111. Gupta, R.; Jung, E.; Gooley, A.A.; Williams, K.L.; Brunak, S.; Hansen, J. Scanning the available dictyostelium discoideum proteome for o-linked glcnac glycosylation sites using neural networks. *Glycobiology* **1999**, *9*, 1009–1022. [[CrossRef](#)]
112. Julenius, K. Netglyc 1.0: Prediction of mammalian c-mannosylation sites. *Glycobiology* **2007**, *17*, 868–876. [[CrossRef](#)] [[PubMed](#)]

113. Eisenhaber, B.; Bork, P.; Eisenhaber, F. Sequence properties of gpi-anchored proteins near the omega-site: Constraints for the polypeptide binding site of the putative transamidase. *Protein Eng.* **1998**, *11*, 1155–1161. [[CrossRef](#)] [[PubMed](#)]
114. Fankhauser, N.; Maser, P. Identification of gpi anchor attachment signals by a kohonen self-organizing map. *Bioinformatics* **2005**, *21*, 1846–1852. [[CrossRef](#)]
115. Pierleoni, A.; Martelli, P.L.; Casadio, R. Predgpi: A gpi-anchor predictor. *BMC Bioinform.* **2008**, *9*, 392. [[CrossRef](#)] [[PubMed](#)]
116. Poisson, G.; Chauve, C.; Chen, X.; Bergeron, A. Fraganchor: A large-scale predictor of glycosylphosphatidylinositol anchors in eukaryote protein sequences by qualitative scoring. *Genom. Proteom. Bioinform.* **2007**, *5*, 121–130. [[CrossRef](#)]
117. Campbell, M.P.; Aoki-Kinoshita, K.F.; Lisacek, F.; York, W.S.; Packer, N.H. Glycoinformatics. In *Essentials of Glycobiology*, 3rd ed.; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2015.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).