GLYCOBIOINFORMATICS IN DECIPHERING THE MAMMALIAN GLYCOCODE: RECENT ADVANCES

Arun K. Datta, Ph.D.[1*] and **Nitin Sukhija, Ph.D.**[2]

[1]Department of Engineering and Computing, National University, San Diego, CA, USA;
[2]Department of Computer Science, Slippery Rock University of Pennsylvania, Slippery Rock, PA, USA

ABSTRACT

Experimental evidence supports the notion that glycans exert a significant role on the physiological and pathological processes in a mammalian cell. The importance of identification of the specific structural motif of the glycan or the *glycocode* is realized for various purposes including identifying the biomarkers and therapeutic purposes. Incidentally, glycan synthesis is neither template-driven nor it is a linear molecule like DNA, RNA or protein. It is synthesized by a set of glycosyltransferases under a tightly controlled cellular regulation forming a branched structure often with high complexity. Although analytical techniques have significantly improved to obtain glycomics data, needed improvement for bioinformatics analysis is yet to be materialized. A number of glycobioinformatics tools have been developed over the years. However, a major effort is needed to make those accessible to the larger Glyco-community. This article focuses on the glycobioinformatics tools that have been developed earlier and discusses the grid-based technology that can be used for developing a web-based Glycomics Workbench for glycomes analysis.

Keywords: glycocode, glycobiology, bioinformatics, glycome, glycomics, glycoinformatics, glycobioinformatics, grid technology.

Abbreviations: CMAHP, CMP-NeuAc hydroxylase-like protein; IUPAC, International Union of Pure and Applied Chemistry; KCF, KEGG Chemical Function; LINUCS, LInear Notation for Unique description of Carbohydrate Sequences; NIGMS, National Institute of General Medical Sciences; RDBMS, Relational Database Management System; SQL, Structured Query Language; URL, Universal Resource Locator.

[*] Corresponding Author: Arun K. Datta
Email: arun.datta@natuniv.edu
Author's draft version

INTRODUCTION

Glycomics focuses on the study of defining the structures and functions of complex carbohydrates, as found in the glycoproteins, glycolipids and proteoglycans [1]. These complex carbohydrates, often referred to as glycans, are synthesized by glycosyltransferses that are tightly regulated at multiple levels ([2]; [3]). These are one of the four major classes of biomolecules, next to nucleic acids, proteins, and lipids. Among these, carbohydrates are the most abundant and most complex ([4]; [5]). These molecules can exist in the biological systems either as free or conjugated with proteins or lipids forming glycoproteins or glycolipids respectively. These glycoconjugates also cover most of the mammalian cell surface forming glycocalyx ([6] and the references therein). While the estimated number of translated proteins in human is ~ 20,000 ([7]; [8]), post-translational modification with glycans exceeds this number by multiple folds, the total number of which is yet to be determined [9]. Recent experiment has shown the evidence of more unknown glycoproteins than were previously estimated [10]. These authors developed a novel algorithm for quantitative approach to identify intact glycopeptides from comparative proteomic datasets. Using their approach, this group uncovered a significant number of novel glycoproteins in the nucleus and cytosol of human and murine embryonic stem cells (also see, [5]).

These glycans exert a tight control on the biological pathways and regulation. These have two vital roles: i. Glycans can contribute to protein conformation, folding, oligomerization, stability, and turnover, and ii. Glycans can be directly recognized by proteins that accounts for various physiological and pathological events [1]; [5]). In most of these cases these glycans are covalently joined to a protein or a lipid to form a glycoconjugate, which is the biologically active molecule that accounts for various physiological functions and are covered in detail in the other chapters in this book (readers can also see, [11]). Mammalian cells use specific glycan motifs to encode important information, termed as *glycocode*, for intracellular targeting of proteins, cell-cell interactions, extracellular signals, and cell differentiation and tissue development [6].

Glycoproteomics specifically involves the study of glycosylation events on protein sequences. Formation of the glycan-amino acid linkage that participates in various biological events is a crucial event in the tightly regulated biosynthesis of the saccharide units of glycoproteins. With the exception of the Asn-linked carbohydrate and the GPI anchor, which are transferred to the polypeptide *en bloc* [12], the glycan-amino acid linkages are formed by the enzymatic transfer of an activated monosaccharide directly to the protein (Chapter 51 in [11]). Experimental evidence indicated that 13 different monosaccharides and 8 amino acids are involved in glycoprotein linkages leading to a total of at least 41 bonds, if the anomeric configurations, the phosphoglycosyl linkages, as well as the GPI (glycophosphatidylinositol) phosphoethanolamine bridges are also considered [13]. These bonds represent the products of N- and O-glycosylation, C-mannosylation, phosphoglycation, and glypiation (for more recent, readers can see the relevant Chapters in [11]). Formation of these glycan-peptide linkages are tightly regulated at multiple levels, dysregulation of which often leads to the development of

pathological conditions. For example, in human, dysregulation of O-GlcNAcylation of protein has been observed to occur in a wide range of diseases, including cancer, diabetes, neurodegenerative diseases, and cardiovascular diseases [14]. Bennun, *et. al*. [15] summarized some of the recent advances in both experimental profiling and analytical methods involving N- and O-linked glycosylation processing for biotechnological and medically relevant cells. At least 16 enzymes involved in their formation were identified and in many cases cloned [16].

On the other hand, proteoglycans (PG), which are complex heterogeneous molecules consisting of one or more glycosaminoglycan (GAG) chains, are attached like bristles to a core protein molecule that make up a major part of the extracellular matrix [17]. These GAGs-are characterized by repeating disaccharide units and variable sulfation patterns along the chain. GAG length and sulfation patterns impact disease etiology [18]; [19]; [20]), cellular signaling ([21]; [22]), and structural support for cells [23]. Because of the high degree of glycosylation by glycosaminoglycan (GAG), N-glycan and mucin-type O-glycan classes, the peptide sequence coverage of complex proteoglycans is revealed poorly by standard mass spectrometry-based proteomics methods. As a result, not much information could be deciphered concerning how proteoglycan site specific glycosylation changes during normal and pathological processes. Earlier, Ly *et. al*. [24] published a review on proteoglycomics and Frey [25] on the bioinformatics tools for the analysis of proteoglycans. Since then, a significant progress has been made in the analytical techniques that led to a better understanding on the structure and function of *glycocodes* associated with the proteoglycans and glycosaminoglycans. Several methods have been developed for proteoglycan analysis since then. Klein, *et. al*. [26], as for example, developed a workflow utilizing bioinformatics approach and LC-MS to improve the analytical processes and improved the identification of glycosylated peptides in proteoglycans. Improved analytical techniques also identified various proteoglycans as potential biomarkers [27]. Following data mining in multiple databases, Cui, et. al. [28] showed a positive correlation between heparan sulfate proteoglycan 2 (HSPG2) and Maternally Expressed Gene 3 (MEG3) expression in breast cancer tissues. Thus, this combination can be used as a biomarker to predict the prognosis of breast cancer.

Alternations in glycosylation are often associated with vast numbers of pathological conditions. Experimental evidence clearly suggests that the *glycocodes* are responsible for various functions ranging from physiological phenomena to pathological events. Therefore, the need for identifying specific structural motif in the glycans and glyconjugates is critical to correlate with its functions [29]. Such information will not only be helpful in identifying biomarkers, such information is also essential for therapeutic purpose. Because of such biological importance, detailed analysis were performed by various groups [30-32]) to identify the codes associated with the carbohydrate motifs, peptides and the glycans involved.

Incidentally, glycans can be branched at multiple points and have multiple options in the sequence of monosaccharides, the reducing-carbon linkage, and anomeric type of linkage, causing a given set of monosaccharides composing a glycan with many possible isomers [33]. Until recently, such complexity posed at least three challenges: i). Selecting appropriate experimental analytical tool to identify the structure of the glycocode. Simple mass spectrometry analysis does not provide information about the branching, sequence, linkages,

or anomeric state [33]; ii). To obtain enough glycan material from a clinical specimen for structural analysis. The analytical tools are not sensitive enough to identify the glycan structure from the minute amount of clinical specimen, and iii). Lack of appropriate database and software tools available for bioinformatics analysis of glycomes that can provide needed knowledge. Even when appropriate tools have been developed in some specific labs, technology platform used for such developments restrict its use by the majority in the Glyco-community. Most of those software tools and databases were custom made and used as standalone tools by the research groups that developed those tools. Recent efforts are underway to improve that situation. A significant number of publications are now available including those in the other chapters in this book that addresses the first two challenges. This article focuses on the recent advances made to address the bioinformatics challenges for the mammalian glycomics to decifer the *glycocode* and proposes a need for a robust workbench, termed here as *Glycomics Workbench*[1]. Readers can check Toukach and Egorova [34] for bacterial glycomics tools including bacterial Carbohydrate Structure Database. Throughout this article, the term glycobio-informatics or simply 'glycobioinformatics' has been used to describe the bioinformatics of glycobiology (**Figure 1**), a term that was also used earlier by other authors [35].

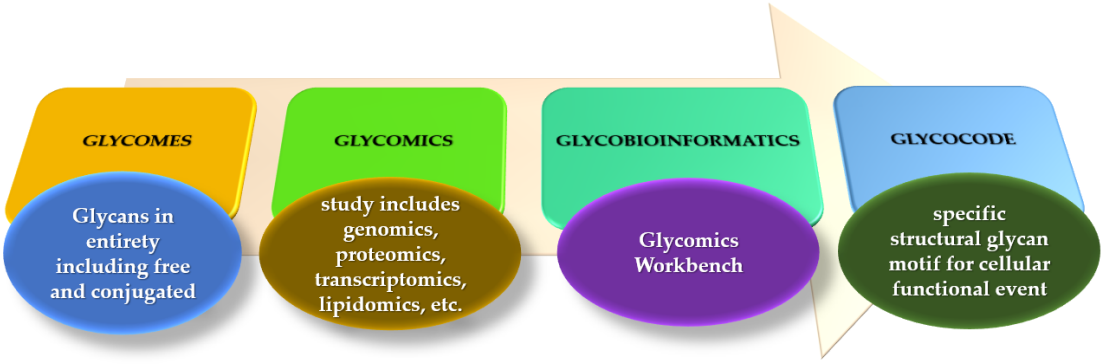

Figure 1: Glycocode in relation to Glycans. For understanding and identification of glycocode, glycobioinformatics analysis is an integral part for the data analysis, the data that are generated by various analytical techniques.

Bioinformatics analysis provides convenient and efficient means of searching, visualizing, comparing, and often predicting, interactions in numerous and diverse molecules and molecular biology applications related to the -omics fields [36]. However, compared to Bioinformatics analysis for proteomics, transcriptomics and genomics [37], glycobioinformatics adds additional and much more difficult challenges [38, 39]. It is because

[1]Presented by these authors at the NIH & FDA *Virtual* Glycoscience Research Day 2020 meeting.

glycans are not template driven linear molecules, rather exhibit a complex branched structure with high diversity. Significant improvements of sensitivity in experimental analytical methods over recent years have led to a tremendous increase in the amount of glycan structural data. Consequently, the availability of robust databases and hardware-software tools to store, retrieve and analyze these data in an efficient way is of fundamental importance to progress in glycobiology. Incidentally, data processing software for glycomics is relatively underdeveloped compared to software for other '–omics' fields [40].

COMPUTATIONAL TOOLS & RESOURCES

.....Blank pages.....

Table I: **List of portals that are commonly used for glycobioinformatics analysis**. However, not all the portals have been described here. The selection of these portals is partly based on their popularity, usefulness and importance.

Name	Reference	Domain address[2]	Use	Comment
Glycomics@ExPaSy	[86,78]	expasy.org/glycomics	Gateway to multiple resources	A part of SIB resources; also includes GlyConnect
Glycosciences.de	[36]	glycosciences.de	Gateway to multiple resources	Description is provided above
GlyCosmos	This article	glycosmos.org	Glycome database	Supported by JSCR
Glyco3D	[51]	glyco3d.cermav.cnrs.fr/	Identification of 3D structure	A database with glycan and lectin structures
GlyGen	This article	glygen.org	A data integration and dissemination project	Supported by UGA (CCRC) and GWU; While this article is in press, York et.

[2] 'http' is now the standard protocol for hyperlinking and 'www' is the standard address for any Uniform Resource Locator (URL) in the Internet for any browser; so, these are omitted in the domain address unless required.

				al. reported on GlyGen [210].
KEGG Glycan	[93]	genome.jp/kegg/glycan/	Gateway to multiple resources including genes and diseases	One of the most well-known projects starting in 1995 by Kanehisa Lab
NCFG	This article	ncfg.hms.harvard.edu	Gateway to multiple resources including CFG	A rich source of data from multiple experimental analysis
RINGS	[83]	rings.t.soka.ac.jp	Gateway to multiple resources	Provides multiple algorithmic and data mining tools

Databases: **Table II** shows the list of glycan databases that are commonly used for

.....Blank pages....

Table II: List of some selected databases that are often used for glycobioinformatics analysis. Some of these databases are for special purpose and has been indicated in the comment section.

Name	Reference	Domain address	Use	Comment
CAZy	[82]	cazy.org	Carbohydrate-Active EnZyme database	More information is available in the Web site
EuroCarbDB	[104]	code.google.com/archive/p/eurocarb/	For identifying carbohydrate structures	Published Web site is no longer available
GlycoGeneDB (GGDB)	[105]	acgg.asia/ggdb2/	Glyco enzyme database	More information is available in the Web site
GlycoEpitope	[106]	glycoepitope.jp	On Glycan Binding Proteins	Also available at: glycosmos.org

O-GlycBase	[108]	cbs.dtu.dk/databases/oglycbase/[3]	O- and C-glycosylated proteins	Part of CBS Prediction Server; version 6.00 has 242 glycoprotein entries.
Glycosciences.DB	[36]	glycosciences.de	On glycan structure, MS, NMR, etc.	Primary database of glycosciences.de
GlycoStore	[111, 112]	glycostore.org	curated chromato-graphic, electro-phoretic and mass-spectrometry composition database	An integrated platform of Glycan Retention Properties; accessible via Web.
GlyTouCan	[98]	glytoucan.org	International glycan structure repository	Data from GlycomeDB [120] is now integrated into this portal.
JCGGDB	[113]	jcggdb.jp	Querying across multiple Japanese databases	The English version is no longer available[4].
PolySac3DB	[114]	polysac3db.cermav.cnrs.fr	Glycan 3D structure	Also, provides information on structure determination techniques
SugarBindDB	[115]	sugarbind.expasy.org	on Glycan Binding Proteins and its ligands	Plan to include synthetic glycan data
UniCarb-DB	[117, 118]	unicarb-db.expasy.org	MS-glycomic data repository	Now, a part of ExPaSy
UniLectin3D	[75]	unilectin.eu	On Lectin 3D structures	SIB ExPASy external link
UniProt	[121]	uniprot.org	On glycosylation sites	A protein database with notation on glycosylation site/s

Glycoscientists often need to search other relevant databases (a short list is provided in **Table III**). Among these, EMBL-EBI (ebi.ac.uk/) and neXtProt (nextprot.org) provides tools for

[3] Under migration to https://services.healthtech.dtu.dk

[4] JCGGDB is proceeded by ACGG-DB and now at https://acgg.asia/db/

proteomics [122]. *UniProt* of EMBL-EBI is a major protein database that also includes glycosylation-site annotations [121]. These annotations are marked as predicted from computer simulations or experimentally verified. *ViralZone* [123,124]), available at ExPASy, is a major resource on viruses that includes DNA viruses (both ds- and ss-), RNA viruses (ds-, circular-,and both ss+ and ss-) and Reverse-Transcribing Viruses. Glycans often serve as their receptors and glycocodes are known for some of those interactions [125]. Additionally, PDB and Cambridge Structural databases provide experimentally determined 3D carbohydrate structures.

Table III: **Relevant Portals, databases, etc. that are useful for glycobioinformatics analysis.**

Name	References	Domain address	Use	Comment
BRENDA	[126]	brenda-enzymes.org/	A Comprehensive Enzyme Information system	Also provides structure-based search function
EMBL-EBI	[127]	ebi.ac.uk	Gateway to multiple resources	Provides advanced bioinformatics training
neXtProt	[122]	nextprot.org	Protein knowledge database	A SIB product
wwPDB	[128]	wwpdb.org	Protein structural database	Includes glycosylation-site annotations
UniProt	[121]	uniprot.org/	Protein database	Includes glycosylation-site annotations
ViralZone	[124]	viralzone.expasy.org	Resource on viruses	A SIB resource

Software Tools: A list of software tools for Glycome analysis is provided in **Table IV**. This

...Blank pages...

Other Relevant Computational tools and Resources

...Blank pages....

TABLE IV: A list of specialized software tools that are used for analysis of glycomes. This list also specifies the usage, such as, for (a) mass spectral analysis. (b) NMR analysis, (c) modelling and simulation, and (d) other miscellaneous usage.

(a) List of commonly used specialized software tools for Mass Spectral (MS/MS/LC-MS) Data Analysis:

Name	Reference	Domain address	Use	Comment
Byonic	[142]		Glycopeptide identification	Proprietary software
Cartoonist	[143]		For MS analysis of glycan structure	A standalone tool
GAGfinder	[71]		MS analysis specifically for GAGs	A standalone software
GAG-ID	[64]		MS analysis of heparin/heparan sulfate	A standalone software
Glycoforest	[144]		LC-MS analysis of glycan	A standalone tool
GlycoFragment	[36]	Glycosciences.de/tools/	For MS analysis of glycan	A part of glycosciences.de portal
Glycolyzer	[145]		Automated glycan annotation software from MS analysis data	A standalone software
GlycoMiner	[146]	szki.ttk.mta.hu/ms/glycominer/	Glycan structure determination by MS analysis	Published Web site is no longer available
GlycoMod	[47]	expasy.org/tools/glycomod/	To determine the glycan structure from MS analysis data	Part of ExPaSy portal
GlycoPAT	[43]	virtualglycome.org/glycopat	Designed to analyze shotgun glycoproteomics data with parallel computing facilities	Fully documented source code is provided
GlycoSearchMS	[147]	glycosciences.de/database/start.php?action=form_ms_search	For MS analysis of glycans	Web-based tool; part of glycosciences.de portal
GlycanSolver	[33]		To obtain information on the isomeric variants of glycans	A standalone software

Glyco-peakfinder	[148]	glyco-peakfinder.org/	MS Glyco-Analytic tool	Forwarded to glycome-db.org/; a part of GlyTouCan
GlycoWorkbench	[60, 149]	code.google.com/archive/p/glycoworkbench/	MS Glyco-Analytic tool	Available for download from Google site
GP Finder	[150]		N- and O-site specific glycosylation	A standalone software
GRITS Toolbox	[151]	grits-toolbox.org	MS data analysis for glycans	A standalone software; source code freely available
MotifFinder	[73]	haablab.vai.org	For identifying glycan motifs present in a clinical sample	Source code available for download at the Web site
MultiGlycan	[139]		LC-MS analysis	A standalone software; can be downloaded from sourceforge.net
pGlyco 2.0	[67]		MS identification of intact glycopeptide	A standalone software
PepSweetener	[72]	expasy.org/glycomics	Annotation of intact glycopeptides in MS spectra	Available at ExPaSY
Pinnacle	[152]		O-Glycopeptide analysis	Proprietary software
Protein Prospector	[152, 153]	prospector.ucsf.edu/prospector/mshome.htm	Glycopeptide analysis	Web-based
SimGlycan	[140, 141]	premierbiosoft.com/glycan	Glycans and glycopeptides characterization	Proprietary software
Skyline	[152]	skyline.ms/project/home/begin.view?	O-glycopeptide analysis	Open source
SweetNET	[65]		MS data evaluation in glycomics	A standalone software
SysBioWare	[154]; also		MS data evaluation in glycomics	Was a part of GlycoWorkbench

	see: [63]			

(b) List of software tools for NMR data Analysis:

Name	Reference	Domain address	Use	Comment
CARP	[53]	glycosciences.de/tools/carp/	For NMR analysis	Now a part of glycosciences.de
CASPER	[61]	organ.su.se/gw/doku.php	For NMR analysis	Supported by a database on NMR
CCPN	[129, 133]	ccpn.ac.uk	For NMR analysis	A major initiative dedicated to NMR users' group; accessible via Web
GODESS	[155]	csdb.glycoscience.ru/database/	NMR analysis of glycan	A web-based software
GlyNest	[156]	glycosciences.de/database/nmr/	NMR analysis	A part of glycosciences.de

(c) List of software tools for 3D Modelling and Simulation:

Name	Reference	Domain address	Use	Comment
CarbBuilder	[157]	people.cs.uct.ac.za/~mkuttel/Downloads.html	Building 3D molecular model of glycan	A standalone software; however, source code is available
Glycam-Web	[135]	dev.glycam.org	Modeling of oligosaccharides and glycoproteins	Web-based tool
Glycan Reader	[138]	glycanstructure.org/glycanreader/	Glycosylation sites determination in	Web-based tool

			the PDB structure	
GlyProt	[158]	glycosciences.de/tools/	For 3D model of glycoproteins	Part of glycoscience.de portal
Shape	[56]	sourceforge.net/projects/shapega/	fully automated conformation prediction software	Source code can be downloaded
SweetII	[159]	glycosciences.de/modeling/sweet2/doc/	For constructing 3D models of glycans	Available at glycoscience.de

(d) Other tools for Glycome analysis:

Name	**Reference**	**Domain address**	**Use**	**Comment**
GlycanBuilde2	[102]	rings.t.soka.ac.jp/downloads.html	Glycan drawing	Part of RINGS portal; also, can be downloaded from code.google.com/archive/p/glycanbuilder/
GlycoBase, autoGU	[110]		HPLC data analysis	A standalone software
GlycomeAtlas	[100]	rings.t.soka.ac.jp/	To visualize and perform queries of glycome data	A part of RINGS portal
GlycoPattern	[96]	glycopattern.emory.edu/	For analysis of glycan array data at CFG	Part of NCFG/CFG portal
GlycoRDF	[57]	github.com/ReneRanzinger/GlycoRDF/wiki	GlycoRDF is a standard representation for storing Glycomcis data	Source code is available for download from the Web site
GlycoViewer	[160]	glycoviewer.babs.unsw.edu.au	To visualize glycan structure	Codes are publicly available

GLYDE-II	[161]		For glycan data exchange format	Information for the GUI used are in the publication
GlySEQ	[53]	glycosciences.de/tools/glyseq/	Statistical Analytical tool	Part of glycosciences.de portal
GlyVicinity	[162]	glycosciences.de/tools/	Statistical analytical tool for carbohydrate notation in PDB database	Part of glycosciences.de portal
GPI-SOM	[163]	gpi.unibe.ch/	To predict GPI-anchored proteins	Source code available from the Web site
GUcal	[70]		For calculating the Glucose Unit values	A standalone software
NETD	[164]		For GAG analysis	A standalone software
Pdb-care	[52]	glycosciences.de/tools/pdb-care/	To check carbohydrate structure in PDB database	Now a part of glycoscience.de portal
PredGPI	[165]	gpcr2.biocomp.unio.it/predgpi	Glycosylation sites prediction tool for GPI	Web-based tool
ProfilePSTMM	[101]	rings.t.soka.ac.jp/	Data mining tool	Part of RINGS portal
Qrator	[166]	code.google.com/archive/p/qrator/	Curation tool for glycan structures	Source code is available
SugarSketcher	[79]	expasy.org/glycomics	Glycan drawing tool	Part of ExPaSY
UniCarbKB	[118]; also see: [41]	unicarbkb.org/	To support online data storage and search on glycans	Now, a part of GlyGen (glygen.org)

CHALLENGES AND OPPORTUNITIES

Presently, there are multiple challenges for a researcher to perform glycobioinformatics analysis. Until recently, there was no standard notation for representing glycan structures. Therefore, most of the glycan databases available had have lack of standards representing

glycan structures thereby causing a challenge to obtain the desired output by searching. The format that is returned by query of such a database often needs to be converted to another format for subsequent meaningful glycobioinformatics analysis. Various tools, such as, GlycanBuilder2 [102], SugarSketcher [79], REStLESS [167], WURCS [168], and relevant tools available at multiple portals, such as in RINGS [83], were developed to resolve this issue: converting output format to the desired format for subsequent analysis. Such lack of standard notation also possesses a challenge while searching glycan substructure [169]. Recently, a consensus has been reached for notation and a universal **S**ymbol **N**omenclature **F**or the **G**raphical representation of glycan structures (SNFG) has been formulated [116,170]). Creating any new glycan database or updating the already developed glycan databases will now be possible following SNFG that should solve this notation related problems. Moreover, supporting software tools (available at https://www.ncbi.nlm.nih.gov/glycans/snfg.html), namely, 3D-*SNFG* [171], *DrawGlycan* [172] and *GlycanBuilder*2 (the updated version of GlycanBuilder; [102]) will greatly benefit to advance glycan structure analysis and identifying glycocodes. Adoption of this notation should also simplify the representation of glycans for visualization. Presently, researchers are using various visualization tools for various purposes. For example, *GlycoViewer* is to visualize glycan structures and summarizes their variations in 2D [160], *GlycomeAtlas* is to visualize the expression of glycans in a specific tissue [100], *GlycoDomainViewer* is to visualize the glycosylation sites in the glycoproteins [173]. 3D modeling tools, such as, in *Glyco3D* [51], and *LiteMol* have also generated interactive images of glycans and glycoconjugates [174]. Recently, a plug-in, 3D-SNFG, has also been published [175] for LiteMol in order to visualize a 3D depiction of glycan ligands in the glycoprotein. More recently, *GlyConnect* has also been developed for resolving this issue [78]. Additionally, the consensus for assigning an unique ID for every known glycan structure repository in the *GlyTouCan* [98] will also help a glycobioinformatics analyst to obtain a desired output by searching the glycan database.

Another major challenge is the accessibility to useful glycobioinformatics tools that have been developed over the years and used for some specific analysis. It is because most of those tools are standalone. Appropriate Web services via which these tools can be used are lacking. Even when it is Web-based, the web site address either no longer exists or changed to a different address that is harder to find. This challenge is also augmented by the lack of information on the underlying software technology that was used for developing those software tools, making the integration of these tools in a Web-based system difficult, if not impossible. Except a few (e.g., [48] for CFG; [78] for GlyConnect, etc.), most often the publications describing the development of a software tool do not specify what sort of architecture (e.g., Model-View-Controller) and/or technology were used for such development. Moreover, use of multiple technology platforms (e.g., asp.net vs. php) for application development also makes it difficult to serve from the same server. Same is true for the databases as well. Often, the information on how the database was developed even if that is relational, what type of Database Management System (MSSQL vs. MySQL, etc.) was used, often remains missing in the description. Moreover, lack of information on database schema makes it difficult for a

developer to understand whether the said database has a single table or multiple tables and whether those satisfy the normalization (e.g., 3rd Normalization Form or 3NF). Such information is needed for database integration and transferring data from one database to another for bringing uniformity and eliminating the redundancy. All of these challenges become more obvious when the glycan analysis generates multiple data types with multiple formats.

Such complexity often demands accessing multiple resources appropriate for glycobioinformatics analysis. Mercier *et. al*. [176], for example, had to access multiple databases (Table 1 in their article) and software resources (Table 2 in their artcle) for analysis of glycan-virus interaction using bioinformatics approach. This effort needed mapping of the resources (Figure 1 in their article) to begin with. The resources used for this analysis by these authors include ViralZone [123], UniProtKB [121], neXtProt [122], GlyConnect [78], CAZy [82], GlyTouCan [98], SugarBindDB [115], MatrixDB [177], UniLectin3D [75], IMGT [178], IEDB [179], DAGR [180], PDB [128], PubMed [181], etc. One can easily understand the amount of efforts needed for bioinformatics analysis of a single event (a single instance on glycan-virus interactions in this case). When a System Biology approach is needed to identify biomarkers ([2, 182, 183]) or for developing a strategy for personalized medicine [184], such approach poses more challenges. This creates a demand for more efficient system where integration of databases and tools are flawless and the workflow is semi-automated, if not fully automated. Nevertheless, presently cross-referencing to multiple databases and tools, which often lacks in a portal, would certainly benefit an analyst.

NEED FOR A ROBUST WORKBENCH

An workbench, termed *GlycoWorkbench*, was developed earlier by Anne Dell and her group as a tool for the computer-assisted annotation of mass spectral data of glycans [60,149]. Its *GlycanBuilder* was designed to display glycan structures [46,61]. Software like *Glycomod*, *GlycoQuest*, *Glyco-Peakfinder,* and *SysBioWare* were also employed in GlycoWorkbench for glycomic data processing. This might have been an attempt to provide necessary software tools under an informatics framework. Incidentally, that project was discontinued although the codes are now available at https://code.google.com/archive/p/glycoworkbench/. More recently, Barnett, et. al. [35] described Glycome Analytics Platform to address this issue in which a set of tools were integrated into the analytic functionality of the Galaxy bioinformatics platform. Nevertheless, considering the amount of data generated now from various experiments, a need for a more robust workbench is needed where necessary resources and tools can be integrated under a common framework that can provide seemless access for glycomes analysis.

Since last 15-20 years, particularly under the Consortium of Functional Glycomics (CFG) [48]), the effort on glycobiology research has generated a lot of useful data that can be exploited for various purposes including developing new drugs. Data generated from various experimental methods ranging from gene microarray screening, glycan array screening, glycan profiling, mouse phenotyping experiments, Glycan–protein interaction screening to glycome related gene sequencing, has become so voluminous that the need for a better management of

data is realized for meaningful glycobioinformatics analysis especially if these are to be used for developing a therapeutic strategy.

Since the Internet was made available to the public ([185, 186]), a number of tools and technologies have been developed to utize the resources through the Web. Such approach can also be followed for the development of glycobioinformatics tools. The developers, however, need to address the challenges posed by the incompatibilities in interfaces and data types. Such an approach was earlier initiated by Katayama *et al*. [187]. These authors developed *TogoWS* (accesible at http://togows.dbcls.jp) to integrate various databases and software tools. They developed REST-based Web service for accessing the existing databases and client-side SOAP-based Web service to address the incompatibilities in the interfaces (it may be noted that while this article is in press, Vos *et al*. [211] reported on Semantic Web Technology for resolving these issues). A more recent effort to resolve such issues has been elaborated by Alocci, *et al*. [78]. These authors provided detailed description of the development of *GlyConnect* and integrated multiple databases and tools that are now made available through the Web (available at https://glyconnect.expasy.org). With the advancement of these various technologies, it is now possible to make the standalone but useful software tools (shown in **Table IV**) available to the Glyco-community. In a collaborative environment, such as in GitHub (github.com), programmers can participate in developing Web services for making those standalone tools available to any researchers from anywhere with appropriate authentication.

The above examples clearly indicate that despite the availability of multiple databases and multiple software tools, accessibility to those are often restricted because of lack of a uniform developmental framework. While which framework should be adopted as standard can be debatable, a framework based on grid infrastructure for portal development offers multiple advantages including high scalability ([188,189]). In this context, grid infrastructure includes grid services, grid computing, and data grid. Grid computing provides accessibility to High Performance Computing (HPC) similar to a power grid; a user simply accesses to the power/electricity via a power point in the building; does not need to generate power in the building. For accessing to the high-performance computing, a user simply can use a laptop or desktop to the grid computing infrastructure, such as, the one provided by the NSF supported Extreme Science and Engineering Discovery Environment (XSEDE; more information available at https://www.xsede.org). Earlier, this author was involved in developing neoGrid for the analysis of sialylmotifs [190], a cardinal feature of mammalian sialyltransferases [191]. Similar to our earlier effort [192], this system was developed using Open Source Portal Technology with three-tired Open Grid Services Architecture (OGSA; e.g., Figure 1 in [192]), an accepted standard for accessing Grid and other services (ogce.org) under Open Grid Collaborating Environments (OGCE; collab-ogce.org/ogce). OGSA, based on several Web service technologies, is a distributed interaction and computing architecture based around services, assuring interoperability on heterogenous systems so that different types of resources can communicate and share information [193]. OGCE Software system has a bundled set of Java Portlet Specifications JSR 168/286 compatible portlets and services for building Grid Portals and related tools for Web access to Grid and Cloud computing resources. A number of

Science Gateways have been built utilizing this technology [194]. As a part of XSEDE program [195], this author (AKD) was involved in developing neoGrid in Quarry, a virtual hosting environment, for working with Taverna-based workflow utilizing grid computing. Taverna (taverna.org.uk/) is a graphical workbench often used for biomedical informatics ([196] and the references therein). neoGRID was designed to offer a HPC-supported collaborative environment for the researchers from multidisciplinary scientific fields to gather data, integrate and analyze using XSEDE resources. Moreover, a significant number of bioinformatics resources including InterPro [197] are available in XSEDE. Besides, the myGrid team (mygrid.org.uk/about-us/) produced a suite of tools that can be used for analysis of glycosyltransferases. In addition, their myExperiment site makes it easy to find, use and share scientific workflows and building scientific communities with common interests. This Cyberinfrastructure (CI)-supported neoGRID was utilized for glycomotifs analysis ([198]; [190]). neoGrid was made available in HUBzero (hubzero.org) for drug discovery research using members of glycosyltransferase enzyme family as target protein(s) [199]. Currently, XSEDE resources include more than a petaflop of computing capability and more than 30 petabytes of online and archival data storage. Moreover, researchers can access more than 100 discipline specific databases through XSEDE. Through a set of Web Application Programming Interface (API) and built on authenticated services, neoGrid was designed for a user to access such resources remotely from anywhere anytime. Its interface to access XSEDE supported bioinformatics tools allowed a clinician/researcher for comparative genomics/proteomics analysis for personalized medicine. Such analysis demanding high-performance computing power was made available in neoGRID. Incidentally, this project has been discontinued. Nevertheless, similar approach can be undertaken for developing a Grid technology-based portal, termed here as *Glycomics Workbench* (**Figure 2**), in a collaborative environment that can serve the Glyco-community.

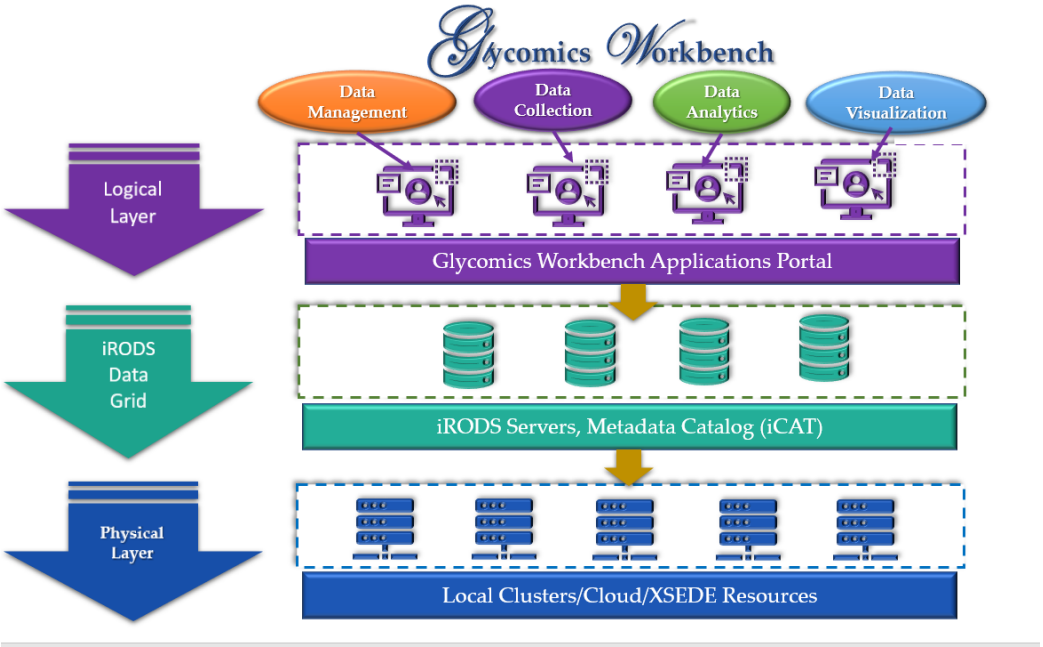

Figure 2: Proposed Grid technology-based portal, Glycomics Workbench. Supported by XSEDE resources for High-Performance computing and other available resources, this portal will enable data analytics and data visualization in addition and will integrate C-Grid for creating, storing, and managing virtual data collection ([200, 201]).

Any '-omics' analysis needs a large digital space for storing the data that includes genomics, proteomics, glycomics, transcriptomics, interactomics, metabolomics, to name a few. Eficient storage, retrieval, and analysis of such large volume multi-omics data, in the order of petabytes to hexabytes, is a huge obstacle. As a solution, we worked on data grid technology to develop a community grid, termed as *C-Grid*, to store and manage large amount of user collaborative data objects ([200, 201]). It has been developed to serve as a distributed computing environment and data management system for sharing data with the collaborators. Data grid technology enables a user to store large amount of data or files into a huge distributed structure that is fabricated using many geographically dispersed heterogeneous storage instruments. Data grid technology provides two fundamental services: allowing access to data, and its metadata (data about data). Data access services provide mechanism to access, manage, and initiate third-party data transfer in distributed storage systems. Metadata access services provide mechanisms to access and manage information about the data stored in storage systems [202]. A number of large scale data grid projects, such as, the Biomedical Informatics Research Network (BIRN) [203], have been developed that uses data grid to discover, transfer, store and manage distributed heterogeneous data ([204] and the references therein). Remote management of this data grid is performed using a middleware. Among the various middleware

available, these authors [200] utilized Integrated Rule-Oriented Data System (iRODS; [205]) because of multiple advantages as elaborated by Chiang and colleagues [206]. iRODS allows metadata to be applied to all components of its system, collections, users, files and entire zones. The metadata can be used for data discovery and workflow automation within the system. iRODS is equipped with a rule engine system that when certain conditions are met, background processes are automatically triggered [207]. This allows automatic data discovery and provides methods to search through the metadata stored on the system. Moreover, since iRODS installation metadata is stored in a SQL database (either MySQL, PostgreSQL, or Oracle) [207], there is opportunity of analyzing that data independently of iRODS for data analytics. We also developed a front-end-interface (GUI), termed ez-iRODS, to interact with this middleware. Similar to ezSRB that was developed earlier using PHP [200], we have recently developed a Python-based ez-iRODS to access and interact with iRODS server of C-Grid with 10 TB size and RAID 5 configuration [201]. The ez-iRODS was developed as a direct wrapper of the iRODS Client API, Python-iRODS Client (PRC; [207]). ez-iRODS provides input forms for users to interact with iRODS-server. This data grid, supported by a PostgreSQL database application, has been designed to help a user for creating and managing 'virtual data collection' that could be stored in heterogenous data resources across the distributed network in a collaborating environment (**Figure 3**). The server for C-Grid web portal and the iRODS, is located at the Slippery Rock University (SRU)'s Obsidian server that runs with Linux operating system (Red Hat Enterprise). This server provides functions for storing user accounts, verifying user accounts, communicating with the iRODS-server, and transmitting data objects. This Web-based system, supported by the LAVA supercomputing system at SRU, has been developed with an objective of long-term data preservation, unified data access and sharing domain specific data amongst the scientific research collaborators [201].

Another major hurdle in the glycobioinformatics analysis is curating data from the published literature, which is now mostly done manually. Such hurdle can be overcome by taking a similar approach that Wikipedia has undertaken - involving volunteers from global community to develop web pages. It may be envisioned that Glyco-community members would be involved in developing Web-based glyco-related 'Molecule page' based on their research interest. A Web-based form can be made available to insert data for each molecule, submission of which will populate corresponding tables (or tuples, a database terminology) in the appropriate glyco-related database, a task that can be accomplished with an Object-Oriented approach. Earlier, this author developed a molecule page on ST6Gal I (**Appendix I**[5]) as a participating member of James Shaw's effort on PROW (Protein Reviews On the Web) at NCBI. This Web-based 'Molecule Page', peer-reviewed by two experts from the glycoscience community, was then made available at NCBI for viewing from anywhere via Web. Incidentally, Shah's effort was discontinued and the earlier web site address at NCBI no longer exists. Anyway, later, Raman *et. al*. [48] used a similar approach for developing 'Molecule Page' as a part of CFG initiative. More recently, Saunders *et. al*. [208] published a detailed description of 'Molecule Page' design for signaling molecules. Similar approach can be

[5]Also, made available at http://www.arundatta.info/PROWGuideST6GalI.html for view.

undertaken for developing glyco-related 'Molecule Page'. While XML-based approach can develop proposed database/s, anyone from anywhere can view such a page on-the-fly by clicking on the molecule of interest. Experts from the Global Glyco-Community can also be provided 'edit' feature after authenticated access.

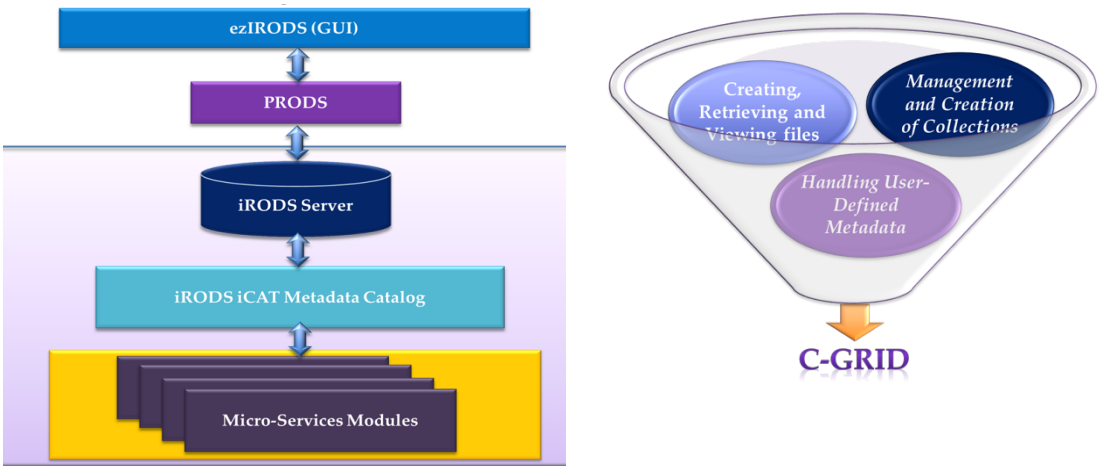

Figure 3. C-Grid or Community Grid Portal framework. It also shows core functionality offered by C-grid web portal based on Integrated Rule-Oriented Data System (iRODS) (adapted from [201]).

CONCLUSION

Developing a fully automated technology infrastructure for glycobioinformatics analysis is a monumental job. This can be accomplished by involving 'volunteers' from the Glyco-community *globally*. A researcher can be involved in developing peer-reviewed 'Molecule Page' that can relatively quickly populate appropriate glyco-related database/s. As has been demonstrated by the XSEDE supporting team, more tools can be integrated in a grid infrastructure-based portal (Glycomics Workbench) whenever such need arises for glycomes analysis. Considering that a large number of databases and software tools are already available, a team of scientists can determine their usefulness, while the Application developers can utilize collaborating environment, such as GitHub (github.com/), for developing Web-services for making those available to the Glyco-community. Some of those tools, developed outside of the Glyco-Community can also be considered because of the relevancy, such as, InterPro [197]. Any new software tool can be developed based on Model-View-Controller (MVC), which is a standard architecture for software development, by using open source yet portable (e.g., Java) technology that can easily be made available though a portal. Regarding the choice of database management system, a search engine works well if a data is stored in a relational database (although one may argue for NoSQL, [209]). A relational database has less redundancy, especially in a 3NF (third normal form) where every table represents an entity. Stadleman *et. al*. [10] showed the evidence that a significant number of glycoproteins are still unknown.

Therefore, it may be argued that such a database should be developed with an Object-Oriented approach where any new 'glycoprotein' be considered as an 'object'. Authenticated access to a collaborative environment can be provided by the grid infrastructure. In these authors' opinion, portal development utilizing grid infrastructure may provide a solution to the informatics challenges encountered by a glycobiologist while performing bioinformatics analysis for deciphering the glycocodes.

ACKNOWLEDGEMENT

This article is the culmination of this author's (AKD) fruitful discussion with several glycobiologists and computer scientists over the years. Particularly notable is the discussion with James C. Paulson of TSRI while this author wrote the bioinformatics part of the Glue Grant. These authors also acknowledge the expert opinion on data grid technology and help for C-Grid development by Reagan Moore and his group while at the San Diego Supercomputing Center (SDSC). Expert opinion and help of Nancy Wilkins-Diehr, and Amit Mazumdar of the SDSC for XSEDE resources is also acknowledged. For bioinformatics analysis, this author (AKD) acknowledges the collaborative support of Eric Jakobson of the National Center for Scientific Applications (NCSA) at the University of Illinois- Urbana Champaign.

REFERENCES

1. Cummings, R.D., and Pierce, J.M. (2014). The challenge and promise of glycomics. Chem Biol **21**, 1-15.
2. Agrawal P, Fontanals-Cirera B, Sokolova E, Jacob S, Vaiana CA, Argibay D, Davalos V, McDermott M, Nayak S, Darvishian F, Castillo M, Ueberheide B, Osman I, Fenyö D, Mahal LK, and Hernando E. (2017). A Systems Biology Approach Identifies FUT8 as a Driver of Melanoma Metastasis. Cancer Cell **31**, 804-819.e7.
3. Neelamegham, S., and Mahal, L.K. (2016). Multi-level regulation of cellular glycosylation: from genes to transcript to enzyme to structure. Curr Opin Struct Biol **40**, 145–52.
4. Hart, G.W., and Copeland, R.J. (2010). Glycomics hits the big time. Cell **143**, 672-76.
5. Cummings, R.D. (2017) The Haystack Is Full of Needles: Technology Rescues Sugars! Mol Cell **68**, 827-2.
6. Varki, A. (2017). Biological roles of glycans. Glycobiol **27,** 3-49 .
7. ENCODE Project Consortium. Dunham, I., Kundaje, A., Aldred, S.F. *et al*. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature **489**, 57-74.
8. Kim, M. S. *et al*. (2014). A draft map of the human proteome. Nature **509**, 575-81.
9. Ponomarenko, E A. *et al*. (2016). The Size of the Human Proteome: The Width and

Depth. Intl J Analyt Chem 7436849.
10. Stadlmann, J., Taubenschmid, J., Wenzel, D., Gattinger, A., Dürnberger, G., Dusberger, F., Elling, U., Mach, L., Mechtler, K. and Penninger, J.M. (2017). Comparative glycoproteomics of stem cells identifies new players in ricin toxicity. Nature **549**, 538-42.
11. Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, .G., Kinoshita, T., Packer, N.H., Prestegard, J.H., Schnaar, R.L., and Seeberger, P. (eds; 2017). Essentials of glycobiology, third edition. Cold Spring Harbor Laboratory Press. ISBN 9781621821328.
12. Kinoshita, T., and Fujita, M. (2016). Biosynthesis of GPI-anchored proteins: Special emphasis on GPI lipid remodeling. J Lipid Res **57**, 6-24.
13. Spiro, R.G. (2002). Protein glycosylation: Nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiol **12**, 43R-56R.
14. Nie, H., and Yi, W. (2019). O-GlcNAcylation, a sweet link to the pathology of diseases. Journal of Zhejiang University: Science B **20**, 437-48.
15. Bennun, S.V., Hizal, D.B., Heffner, K, Can, O, Zhang, H, B. M. (2016). Systems Glycobiology: Integrating Glycogenomics, Glycoproteomics, Glycomics, and Other 'Omics Data Sets to Characterize Cellular Glycosylation Processes. J Mol Biol **428**, 3337–52.
16. Rini, J.M., and Esko, J.D. (2017). Chapter 6: Glycosyltransferases and Glycan-Processing Enzymes. In: Essentials of Glycobiology. 3rd edition (Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., Schnaar, R.L., and Seeberger, P., eds). Cold Spring Harbor Laboratory Press; 2015-17.
17. Iozzo, R.V., and Schaefer, L. (2015). Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. Matrix Biol **42**, 11-55.
18. Vu, T.T., Marquez, J., Le, L.T., Nguyen, A.T.T., Kim, H.K., Han, J. (2018). The role of decorin in cardiovascular diseases: more than just a decoration. Free Radical Research **52**, 1210-19.
19. Zou, W., Wan, J., Li, M., Xing, J., Chen, Q., Zhang, Z., Gong, Y. (2019). Small leucine rich proteoglycans in host immunity and renal diseases. J Cell Commun and Signaling **13**, 463-71.
20. Uchimido, R., Schmidt, E.P. and Shapiro, N.I. (2019). The glycocalyx: A novel diagnostic and therapeutic target in sepsis. Critical Care **23**, 16.
21. Wigén, J., Elowsson-Rendin, L., Karlsson, L., Tykesson, E., and Westergren-Thorsson, G. (2019). Glycosaminoglycans: A Link between Development and Regeneration in the Lung. Stem Cells and Dev **28**, 823-32.
22. Xie, M. and Li, J. P. (2019). Heparan sulfate proteoglycan – A common receptor for diverse cytokines. Cell Signal **54**,115-21.
23. Ughy, B., Schmidthoffer, I., and Szilak, L. (2019). Heparan sulfate proteoglycan (HSPG) can take part in cell division: inside and outside. Cell Molec Life Sci **76**, 865-71.
24. Ly, M., Laremore, T.N. and Linhardt, R.J. (2010). Proteoglycomics: Recent progress

and future challenges. OMICS A Journal of Integrative Biology 1**4**, 389-99.
25. Frey, L.J. (2015). Informatics tools to advance the biology of glycosaminoglycans and proteoglycans. Methods Mol Biol **1229**, 271-87.
26. Klein, J.A., Meng, L. and Zaia, J. (2018). Deep sequencing of complex proteoglycans: A novel strategy for high coverage and sitespecific identification of glycosaminoglycanlinked peptides. Mol Cell Proteomics **17**, 1578-90.
27. Tanaka, Y., Tateishi, R. and Koike, K. (2018). Proteoglycans are attractive biomarkers and therapeutic targets in hepatocellular carcinoma. Intl J Molec Sci **19**, 3070.
28. Cui, X., Yi, Q., Jing, X., Huang, Y., Tian, J., Long, C., Xiang, Z., Liu, J., Zhang, C., Tan, B., Li, Y., and Zhu, J. (2018). Mining Prognostic Significance of MEG3 in Human Breast Cancer Using Bioinformatics Analysis. Cell Physiol Biochem **50** , 41-51.
29. Gabius, H.J. (2018). The sugar code: Why glycans are so important. BioSystems **164,** 102-11.
30. Palaniappan, K.K. and Bertozzi, C.R. (2016). Chemical Glycoproteomics. Chem Rev **116**, 14277-306.
31. Jung, H.J. (2017). Chemical proteomic approaches targeting cancer stem cells: A review of current literature. Cancer Genomics and Proteomics **14**, 315-27.
32. Zhu, J., Warner, E., Parikh, N.D. and Lubman, D.M. (2019). Glycoproteomic markers of hepatocellular carcinoma-mass spectrometry based approaches. Mass Spectrometry Rev **38**, 265-90.
33. Klamer, Z., Hsueh, P., Ayala-Talavera, D. and Haab, B. (2019). Deciphering protein glycosylation by computational integration of on-chip profiling, glycan-array data, and mass spectrometry. Mol Cell Proteomics **18**, 28-40.
34. Toukach, P.V. and Egorova, K.S. (2016). Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts. Nucleic Acids Res **44** (D1), D1229-36.
35. Barnett, C.B., Aoki-Kinoshita, K.F. and Naidoo, K.J. (2016). The Glycome Analytics Platform: An integrative framework for glycobioinformatics. Bioinformatics **32**, 3005-11.
36. Böhm, M., Bohne-Lang, A., Frank, M., Loss, A., Rojas-Macias, M.A., Lütteke, T. (2019). An annotated data collection linking glycomics and proteomics data (2018 update). Nucleic Acids Res **47**, D1195–D1201.
37. Chen, C., Huang, H. and Wu, C.H. (2017). Protein bioinformatics databases and resources. in Methods in Molecular Biology **1558**, 3-39.
38. Campbell, M.P., Aoki-Kinoshita, K.F., Lisacek, F., York, W.S., and Packer, N. (2017). Chapter 52. Glycoinformatics. in Essentials of Glycobiology. 3rd edition. (Ajit Varki, Executive Editor, Richard D Cummings, Jeffrey D Esko, Pamela Stanley, Gerald W Hart, Markus Aebi, Alan G Darvill, Taroh Kinoshita, Nicolle H Packer, James H Prestegard, Ronald L Schnaar, eds) Cold Spring Harbor Laboratory Press, 2015-2017.
39. Egorova, K.S. and Toukach, P.V. (2018). Glycoinformatics: Bridging Isolated Islands in the Sea of Data. Angew Chemie - Int Ed **57**, 14986-90.
40. Veillon, L., Zhou, S. and Mechref, Y. (2017). Quantitative Glycomics: A Combined Analytical and Bioinformatics Approach. Methods in Enzymol **585**, 431-77.
41. Aoki-Kinoshita, K.F. (2013). Using databases and web resources for glycomics

research. Mol Cell Proteomics **12**, 1036–45.
42. Baycin, H.D., Wolozny, D., Colao, J., Jacobson, E., Tian, Y., Krag, S.S., Betenbaugh, M.J., Zhang, H. (2014). Glycoproteomic and glycomic databases. Clin Proteomics **11**, 1–10.
43. Liu, G., Cheng, K., Lo, C.Y., Li, J., Qu, J., Neelamegham, S. (2017). A comprehensive, open-source platform for mass spectrometry-based glycoproteomics data analysis. Mol Cell Proteomics **16**, 2032–47.
44. Doubet, S., and Albersheim, P. (1992). Letter to the glyco-forum: Carbbank. Glycobiol **2**, 505.
45. Doubet, S., Bock, K., Smith, D., Darvill, A., Albersheim, P. (1989). The complex carbohydrate structure database. Trends Biochem Sci **14**, 475-77.
46. Damerell, D., Ceroni, A., Maass, K., Ranzinger, R., Dell, A., Haslam, S.M. (2012). The GlycanBuilder and GlycoWorkbench glycoinformatics tools: Updates and new developments. Biol Chem **393**, 1357-62.
47. Cooper, C.A., Gasteiger, E. and Packer, N.H. (2001). GlycoMod - A software tool for determining glycosylation compositions from mass spectrometric data. Proteomics **1**, 340-49.
48. Raman, R., Venkataraman, M., Ramakrishnan, S., Lang, W., Raguram, S., and Sasisekharan, R. (2006). Advancing glycomics: Implementation strategies at the consortium for functional glycomics. Glycobiol **16**, 82R-90R.
49. Yuriev, E., and Ramsland, P.A. (2015). Carbohydrates in cyberspace. Front Immunol **6**, 4–7.
50. Emsley, P., Brunger, A.T. and Lütteke, T. (2014). Tools to assist determination and validation of carbohydrate 3d structure data. Methods Mol Biol **1273**, 229-40.
51. Pérez, S., Sarkar, A., Rivet, A., Breton, C. & Imberty, A. (2015). Glyco3d: A portal for structural glycosciences. Methods Mol Biol **1273**, 241-58.
52. Lütteke, T. and von der Lieth, C.W. (2004). pdb-care (PDB CArbohydrate REsidue check): A program to support annotation of complex carbohydrate structures in PDB files. BMC Bioinformatics **5**, 69.
53. Lütteke, T., Frank, M. and von der Lieth, C.W. (2005). Carbohydrate Structure Suite (CSS): Analysis of carbohydrate 3D structures derived from the PDB. Nucleic Acids Res **33** (Database issue), D242-6.
54. Theillet, F., Simenel, C., Guerreiro, C., Phalipon, A., Mulard, L.A., Delepierre, M. (2011). Effects of backbone substitutions on the conformational behavior of Shigella flexneri O-antigens: Implications for vaccine strategy. Glycobiol **21**, 109-21.
55. Frank, M., Lütteke, T. and von der Lieth, C.W. (2007). GlycoMapsDB: A database of the accessible conformational space of glycosidic linkages. Nucleic Acids Res **35** (Database issue), 287-90.
56. Rosen, J., Miguet, L. and Pérez, S. (2009). Shape: Automatic conformation prediction of carbohydrates using a genetic algorithm. J Cheminform **1**, 1–7.
57. Ranzinger, R., Aoki-Kinoshita, K.F., Matthew P Campbell, M.F. *et al*. (2015). GlycoRDF: An ontology to standardize glycomics data in RDF. Bioinformatics **31**, 919–25.

58. Agravat, S.B., Saltz, J.H., Cummings, R.D., and Smith, D.F. (2014). GlycoPattern: A Web platform for glycan array mining. Bioinformatics **30**, 3417–18.
59. Lisacek, F., Mariethoz, J., Alocci, D. *et al*. (2017). Databases and associated tools for glycomics and glycoproteomics. Methods in Molec Biol **1503**, 235-64.
60. Damerell, D., Ceroni, A., Maass, K., Ranzinger, R., and Dell, A. (2015). Annotation of glycomics MS and MS/MS spectra using the GlycoWorkbench software tool. Methods Mol Biol **1273**, 3–15.
61. Ceroni, A., Dell, A.and Haslam, S.M. (2007). The GlycanBuilder: A fast, intuitive and flexible software tool for building and displaying glycan structures. Source Code Biol Med **2**, 1–13.
62. Souady, J., Dadimov, D., Kirsch, S., Bindila, L., Peter-Katalinić, J., and Vakhrushev, S.Y. (2010). Software utilities for the interpretation of mass spectrometric data of glycoconjugates: Application to glycosphingolipids of human serum. Rapid Commun. Mass Spectrom **24**, 1039-48.
63. Tsai, P.-L., and Chen, S.-F. (2017). A Brief Review of Bioinformatics Tools for Glycosylation Analysis by Mass Spectrometry. Mass Spectrom **6**, S0064–S0064.
64. Chiu, Y., Huang, R., Orlando, R., and Sharp, J. S. (2015). GAG-ID: Heparan sulfate (HS) and heparin glycosaminoglycan high-throughput identification software. Mol Cell Proteomics **14**, 1720–30.
65. Nasir, W., Toledo, A.G., Noborn, F., Nilsson, J., Wang, M., Bandeira, N., and Larson, G. (2016). SweetNET: A bioinformatics workflow for glycopeptide MS/MS spectral analysis. J Proteome Res **15**, 2826-40.
66. Hu, H., Khatri, K., and Zaia, J. (2017). Algorithms and design strategies towards automated glycoproteomics analysis. Mass Spectrometry Rev **36**, 475-98.
67. Liu, M., Zeng, W., Fang, P. *et al*. (2017). PGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. Nat Commun **8**, 438.
68. Khatri, K., Klein, J. A., and Zaia, J. (2017). Use of an informed search space maximizes confidence of site-specific assignment of glycoprotein glycosylation. Anal Bioanal Chem **409**, 607–18.
69. Cox, N.J., Luo, P.M., Smith, T.J., Bisnett, B.J., Soderblom, E.J., and Boyce, M. (2018). A novel glycoproteomics workflow reveals dynamic O-GlcNAcylation of COPγ1 as a candidate regulator of protein trafficking. Front Endocrinol (Lausanne) **9**, 606.
70. Jarvas, G., Szigeti, M., and Guttman, A. (2018). Structural identification of N-linked carbohydrates using the GUcal application: A tutorial. J Proteomics **171**, 107-15.
71. Hogan, J.D., Klein, J.A., Wu, J., Chopra, P., Boons, G.J., Carvalho, L., Lin, C., and Zaia, J. (2018). Software for peak finding and elemental composition assignment for glycosaminoglycan tandem mass spectra. Mol Cell Proteomics **17**, 1448-56.
72. Domagalski, M.J., Alocci, D., Almeida, A., Kolarich, D., and Lisacek, F. (2018). PepSweetener: A Web-Based Tool to Support Manual Annotation of Intact Glycopeptide MS Spectra. Proteomics - Clin Appl **12**, e1700069.
73. Klamer, Z., Staal, B., Prudden, A.R., Liu, L., Smith, D.F., Boons, G.J., and Haab, B. (2017). Mining High-Complexity Motifs in Glycans: A New Language to Uncover the

Fine Specificities of Lectins and Glycosidases. Anal Chem **89**, 12342–50.
74. Jiang, H., Aoki-Kinoshita, K.F. and Ching, W.-K. (2011). Extracting glycan motifs using a biochemically-weighted kernel. Bioinformation **7**, 405–12.
75. Bonnardel, F., Mariethoz, J., Salentin, S., Robin, X., Schroeder, M., Perez, S., Lisacek, F., and Imberty, A. (2019). Unilectin3d, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. Nucleic Acids Res **47**, D1236–D1244.
76. Klein, J., and Zaia, J. (2019). Glypy: An Open Source Glycoinformatics Library. J Proteome Res **18**, 3532-37.
77. Lemmin, T., and Soto, C. (2019). Glycosylator: A Python framework for the rapid modeling of glycans. BMC Bioinformatics **20**, 513.
78. Alocci, D., Mariethoz, J., Gastaldello, A., Gasteiger, E., Karlsson, N.G., Kolarich, D., Packer, N.H., and Lisacek, F. (2019). GlyConnect: Glycoproteomics Goes Visual, Interactive, and Analytical. J. Proteome Res **18**, 664–77.
79. Alocci, D., Suchánková, P., Costa, R., Hory, N., Mariethoz, J., Vařeková, R.S., Toukach, P., and Lisacek, F. (2018). Sugarsketcher: Quick and intuitive online glycan drawing. Molecules **23**, 3206.
80. Shajahan, A., Heiss, C., Ishihara, M., and Azadi, P. (2017). Glycomic and glycoproteomic analysis of glycoproteins—a tutorial. Anal. Bioanal Chem **409**, 4483-505.
81. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res **45** (D1), D353-D361.
82. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res **42**, 490–95.
83. Akune, Y., Hosoda, M., Kaiya, S., Shinmachi, D., and Aoki-Kinoshita, K.F. (2010). The RINGS resource for glycome informatics analysis and data mining on the web. Omi A J Integr Biol **14**, 475-86.
84. Struwe, W.B., Pagel, K., Benesch, J.L.P., Harvey, D.J., and Campbell, M.P. (2016). GlycoMob: an ion mobility-mass spectrometry collision cross section database for glycomics. Glycoconj J **33**, 399-404.
85. Gotz, L., Abrahams, J.L., Mariethoz, J., Rudd, P.M. *et al*. (2014). GlycoDigest: A tool for the targeted use of exoglycosidase digestions in glycan structure determination. Bioinformatics **30**, 3131–33.
86. Mariethoz, J., Alocci, D., Gastaldello, A. *et al*. (2018). Glycomics@ExPASy: Bridging the gap. Mol Cell Proteomics **17**, 2164–76.
87. Kumar, S., Lütteke, T., and Schwartz-albiez, R. (2012). GlycoCD: A repository for carbohydrate-related CD antigens. Bioinformatics **28**, 2553-55.
88. Frank, M. (2015). Conformational analysis of oligosaccharides and polysaccharides using molecular dynamics simulations. Methods Mol Biol **1273**, 359-77.
89. Rigden, D.J., and Fernández, X.M. (2019). The 26th annual Nucleic Acids Research database issue and Molecular Biology Database Collection. Nucleic Acids Res **47** (D1),

D1-D7.
90. Kolarich, D., Rapp, E., Struwe, W.B., Haslam, S.M., Zaia, J., McBride, R., Agravat, S., Campbell, M.P., Kato, M., Ranzinger, R., Kettner, C., and York, W.S. (2013). The minimum information required for a glycomics experiment (MIRAGE) project: Improving the standards for reporting mass-spectrometry-based glycoanalytic data. Mol Cell Proteomics **12**, 991–95.
91. Liu, Y., McBride, R., Stoll, M. *et al*. (2017). The minimum information required for a glycomics experiment (MIRAGE) project: Improving the standards for reporting glycan microarray-based data. Glycobiol **27**, 280–84.
92. Campbell, M.P., Abrahams, J.L., Rapp, E., Struwe, W.B., Costello, C.E., Novotny, M., Ranzinger, R., York, W.S., Kolarich, D., Rudd, P.M., and Kettner, C. (2019). The minimum information required for a glycomics experiment (MIRAGE) project: LC guidelines. Glycobiol **29**, 349-54.
93. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K, and Tanabe, M. (2019). New approach for understanding genome variations in KEGG. Nucleic Acids Res **47** (D1), D590-D595.
94. Aoki-Kinoshita, K.F. (2013). Mining frequent subtrees in glycan data using the rings Glycan Miner Tool. Methods Mol Biol **939**, 87-95.
95. Brodeur, G.M. (2003). Neuroblastoma: Biological insights into a clinical enigma. Nature Reviews Cancer **3**, 203-16.
96. Agravat, S.B., Saltz, J.H., Cummings, R.D., and Smith, D.F. (2014). GlycoPattern: a web platform for glycan array mining. Bioinformatics **30**, 3417–18.
97. Mehta, A.Y., Cummings, R.D., and Wren, J. (2019). GLAD: GLycan Array Dashboard, a visual analytics tool for glycan microarrays. Bioinformatics **35**, 3536-37.
98. Tiemeyer, M., Aoki, K., Paulson, J. *et al*. (2017). GlyTouCan: An accessible glycan structure repository. Glycobiol **27**, 915-19.
99. Krambeck, F.J., Bennun, S.V., Narang, S., Choi, S., Yarema, K.J., and Betenbaugh, M.J. (2009). A mathematical model to derive N-glycan structures and cellular enzyme activities from mass spectrometric data. Glycobiol **19**, 1163-75.
100. Konishi, Y., and Aoki-Kinoshita, K.F. (2012). The GlycomeAtlas tool for visualizing and querying glycome data. Bioinformatics **28**, 2849-50.
101. Aoki-Kinoshita, K.F. (2015). Analyzing glycan-binding patterns with the profilePSTMM tool. Methods Mol Biol **1273**, 193-202.
102. Tsuchiya, S., Aoki, N.P., Shinmachi, D., Matsubara, M., Yamada, I., Aoki-Kinoshita, K.F., and Narimatsu, H. (2017). Implementation of GlycanBuilder to draw a wide variety of ambiguous glycans. Carbohyd Res **445**, 104-16.
103. von der Lieth, C., Freire, A.A., Blank, D., Campbell, M.P. *et al*. (2011). EUROCarbDB: An open-access platform for glycoinformatics. Glycobiol **21**, 493–502.
104. Jadda, K.A., Porterfield, M.P., Bridger, R., Heiss, C., Tiemeyer, M., Wells, L., Miller, J.A., York, W.S., and Ranzinger, R. (2015). EUROCarbDB(CCRC): A EUROCarbDB node for storing glycomics standard data. Bioinformatics **31**, 242–45.
105. Narimatsu, H. (2004). Construction of a human glycogene library and comprehensive functional analysis. in Glycoconj J **21**, 17-24.

106. Kawasaki, T., Nakao, H., and Tominaga, T. (2009). GlycoEpitope: A Database of Carbohydrate Epitopes and Antibodies, pp 429-431. In: Taniguchi N., Suzuki A., Ito Y., Narimatsu H., Kawasaki T., Hase S. (eds) Experimental Glycoscience. Springer, Tokyo. ISBN: 978-4-431-77921-6.
107. Okuda, S., Nakao, H., and Kawasaki, T. (2015). Glycoepitope: Database for Carbohydrate Antigen and Antibody, pp 267-273. In: Taniguchi N., Endo T., Hart G., Seeberger P., Wong CH. (eds) Glycoscience: Biology and Medicine. Springer, Tokyo. ISBN: 978-4-431-54840-9.
108. Gupta, R., Birch, H., Rapacki, K., Brunak, S., and Hansen, J. E. (1999). O-GLYCBASE version 4.0: A revised database of O-glycosylated proteins. Nucleic Acids Res **27**, 370-72.
109. Lobeta, A. (2002). SWEET-DB: an attempt to create annotated data collections for carbohydrates. Nucleic Acids Res **30**, 405-08.
110. Campbell, M.P., Royle, L., and Rudd, P.M. (2015). Glycobase and autoGU: Resources for interpreting HPLC-glycan data. Methods Mol Biol **1273**, 17-28.
111. Zhao, S. *et al*. (2018). GlycoStore: a database of retention properties for glycan analysis. Bioinformatics **34**, 3231-32.
112. Abrahams, J.L., and Campbell, M.P. (2018). Building a PGC-LC-MS N-glycan retention library and elution mapping resource. Glycoconj J **35**, 15–29.
113. Maeda, M., Fujita, N., Suzuki, Y., Sawaki, H., Shikanai, T., and Narimatsu, H. (2015). JCGGDB: Japan consortium for glycobiology and glycotechnology database. Methods Mol Biol **1273**, 161-79.
114. Sarkar, A., and Pérez, S. (2012). PolySac3DB: An annotated database of 3 dimensional structures of polysaccharides. BMC Bioinformatics, **13**, 302.
115. Mariethoz, J., Khatib, K., Alocci, D., Campbell, M.P., Karlsson, N.G., Packer, N.H., Mullen, E.H., and Lisacek, F. (2016). SugarBindDB, a resource of glycan-mediated host-pathogen interactions. Nucleic Acids Res **44**, D1243–D1250.
116. Varki, A., Cummings, R.D., Aebi, M. *et al*. (2015). Symbol nomenclature for graphical representations of glycans. Glycobiol **25**, 1323-24.
117. Hayes, C.A., Karlsson, N.G., Struwe, W.B., Lisacek, F., Rudd, P.M., Packer, N.H., and Campbell, M.P. (2011). UniCarb-DB: A database resource for glycomic discovery. Bioinformatics **27**,1343-44.
118. Campbell, M.P., Peterson, R., Mariethoz, J., Gasteiger, E., Akune, Y., Aoki-Kinoshita, K.F., Lisacek, F., and Packer, N.H. (2014). UniCarbKB: Building a knowledge platform for glycoproteomics. Nucleic Acids Res **42**, 215–21.
119. Aoki-Kinoshita, K.F., Sawaki, H., An, H.J., Cho, J.W. *et al*. (2013). The Fifth ACGG-DB Meeting Report: Towards an International Glycan Structure Repository. in Glycobiol **23**, 144-46.
120. Ranzinger, R., Herget, S., Von Der Lieth, C.W., and Frank, M. (2011). GlycomeDB-A unified database for carbohydrate structures. Nucleic Acids Res **39** (Database issue), D373-6.
121. Bateman, A. (2019). UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res **47** (Database issue), D506-D515.

122. Gaudet, P., Michel, P., Zahn-Zabal, M. *et al*. (2017). The neXtProt knowledgebase on human proteins: 2017 update. Nucleic Acids Res **45** (Database issue), D177-D182.
123. Masson, P., Hulo, C., Castro, E.D., Bitter, H., Gruenbaum, L., Essioux, L., Bougueleret, L., Xenarios, I., and Mercier, P.L. (2013). ViralZone: Recent updates to the virus knowledge resource. Nucleic Acids Res **41** (Database issue), D579-83.
124. Hulo, C., Masson, P., de Castro, E. *et al*. (2017). The ins and outs of eukaryotic viruses: Knowledge base and ontology of a viral Infection. PLoS One **12**, e0171746.
125. Shanker, S., Hu, L., Ramani, S., *et al*. (2017). Structural features of glycan recognition among viral pathogens. Current Opin Struct Biol **44**, 211-18.
126. Jeske, L., Placzek, S., Schomburg, I., Chang, A., and Schomburg, D. (2019). BRENDA in 2019: A European ELIXIR core data resource. Nucleic Acids Res **47** (Database issue), D542-D549.
127. Cook, C.E., Lopez, R., Stroe, O., Cochrane, G., Brooksbank, C., Birney, E., and Apweiler, R. (2019). The European Bioinformatics Institute in 2018: Tools, infrastructure and training. Nucleic Acids Res **47** (Database issue), D15-D22.
128. Burley, S.K., Berman, H.M., Bhikadiya, C. *et al*. (2019). Protein Data Bank: The single global archive for 3D macromolecular structure data. Nucleic Acids Res **47** (Database issue), D520-D528.
129. Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins Struct Funct Genet **59**, 687-96.
130. Stevens, T.J., Fogh, R.H., Boucher, W. *et al*. (2011). A software framework for analysing solid-state MAS NMR data. J. Biomol NMR **51**, 437-47.
131. Chignola, F., Mari, S., Stevens, T.J., Fogh, R.H., Mannella, V., Boucher, W., and Musco, G. (2011). The CCPN metabolomics project: A fast protocol for metabolite identification by 2D-NMR. Bioinformatics **27**, 885-86.
132. Skinner, S.P., Fogh, R.H., Boucher, W., Ragan, T.J., Mureddu, L.G., and Vuister, G.W. (2016). CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. J Biomol NMR **66**, 111-24.
133. Mureddu, L., and Vuister, G.W. (2019). Simple high-resolution NMR spectroscopy as a tool in molecular biology. FEBS J **286**, 2035-42.
134. Singh, A., Montgomery, D., Xue, X., Foley, B. L., and Woods, R.J. (2019). GAG builder: A web-tool for modeling 3D structures of glycosaminoglycans. Glycobiol **29**, 515-18.
135. Kirschner, K.N., Yongye, A.B., Tschampel, S.M. *et al*. (2008). GLYCAM06: A generalizable biomolecular force field. carbohydrates. J Comput Chem **29**, 622-55.
136. Park, S., Lee, J., Qi, Y. *et al*. (2019). CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. Glycobiol **29**, 320-31.
137. Lee, J., Patel, D.S., Ståhle, J. *et al*. (2019). CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. J Chem Theory Comput **15**, 775-86.
138. Park, S., Lee, J., Patel, D.S., Ma, H., Lee, H.S., Jo, S, and Im, W. (2017). Glycan Reader

is improved to recognize most sugar types and chemical modifications in the Protein Data Bank. Bioinformatics **33**, 3051-57.

139. Hu, Y., Zhou, S., Yu, C.-Y., Tang, H., and Mechref, Y. (2015). Automated annotation and quantitation of glycans by liquid chromatography/electrospray ionization mass spectrometric analysis using the MultiGlycan-ESI computational tool. Rapid Commun Mass Spectrom **29**, 135-42.
140. Apte, A., and Meitei, N.S. (2010). Bioinformatics in glycomics: glycan characterization with mass spectrometric data using SimGlycan. Methods Mol Biol **600**, 269-81.
141. Meitei, N.S., Apte, A., Snovida, S.I., Rogers, J.C., and Saba, J. (2015). Automating mass spectrometry-based quantitative glycomics using aminoxy tandem mass tag reagents with SimGlycan. J Proteomics **127** (Pt A), 211-22.
142. Lee, L.Y., Moh, E.S.X., Parker, B.L., Bern, M., Packer, N.H., and Thaysen-Andersen, M. (2016). Toward Automated N-Glycopeptide Identification in Glycoproteomics. J Proteome Res **15**, 3904-15.
143. Goldberg, D., Sutton-Smith, M., Paulson, J., and Dell, A. (2005). Automatic annotation of matrix-assisted laser desorption/ionization N-glycan spectra. Proteomics **5**, 865-75.
144. Horlacher, O., Jin, C., Alocci, D., Mariethoz, J., Müller, M., Karlsson, N.G., and Lisacek, F. (2017). Glycoforest 1.0. Anal Chem **89**, 10932-40.
145. Kronewitter, S.R., A De Leoz, M.L., Strum, J.S. *et al*. (2012). The glycolyzer: Automated glycan annotation software for high performance mass spectrometry and its application to ovarian cancer glycan biomarker discovery. Proteomics **12**, 2523–38.
146. Ozohanics, O., Turiák, L., Puerta, A., Vékey, K., and Drahos, L. (2012). High-performance liquid chromatography coupled to mass spectrometry methodology for analyzing site-specific N-glycosylation patterns. J Chromatogr A **1259**, 200-12.
147. Lohmann, K.K., and von der Lieth, C.W. (2004). GlycoFragment and GlycoSearchMS: Web tools to support the interpretation of mass spectra of complex carbohydrates. Nucleic Acids Res **32** (Web Server issue), W261-6.
148. Maass, K., Ranzinger, R., Geyer, H., Von Der Lieth, C.W., and Geyer, R. (2007). 'Glyco-peakfinder' - De novo composition analysis of glycoconjugates. Proteomics **7**, 4435-44.
149. Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A., and Haslam, S.H. (2008). GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of glycans. J Proteome Res **7**, 1650-59.
150. Strum, J.S., Nwosu, C.C., Hua, S. *et al*. (2013). Automated assignments of N- and O-site specific glycosylation with extensive glycan heterogeneity of glycoprotein mixtures. Anal Chem **85**, 5666-75.
151. Weatherly, D.B., Arpinar, F.S., Porterfield, M., Tiemeyer, M., York, W.S., and Ranzinger, R. (2019). GRITS Toolbox-a freely available software for processing, annotating and archiving glycomics mass spectrometry data. Glycobiol **29**, 452-60.
152. Pap, A., Prakash, A., Medzihradszky, K., and Darula, Z. (2018). Assessing the reproducibility of an O-glycopeptide enrichment method with a novel software, Pinnacle. Electrophoresis **39**, 3142-47.
153. Chalkley, R.J., and Baker, P.R. (2017). Use of a glycosylation site database to improve

glycopeptide identification from complex mixtures. Anal Bioanal Chem **409**, 571-77.
154. Vakhrushev, S.Y., Dadimov, D., and Peter-Katalinić, J. (2009). Software platform for high-throughput glycomics. Anal Chem **81**, 3252-60.
155. Kapaev, R.R., and Toukach, P.V. (2016). Simulation of 2D NMR Spectra of Carbohydrates Using GODESS Software. J Chem Inf Model **56**, 1100-04.
156. Loß, A., Stenutz, R., Schwarzer, E., and von der Lieth, C.W. (2006). GlyNest and CASPER: Two independent approaches to estimate 1H and 13C NMR shifts of glycans available through a common web-interface. Nucleic Acids Res **34** (Web Server issue), W733-7.
157. Kuttel, M.M., Ståhle, J., and Widmalm, G. (2016). CarbBuilder: Software for building molecular models of complex oligo- and polysaccharide structures. J Comput Chem **37**, 2098-105.
158. Bohne-Lang, A., and Von der Lieth, C.W. (2005). GlyProt: In silico glycosylation of proteins. Nucleic Acids Res **33**, 214–19.
159. Bohne, A., Lang, E., and Von Der Lieth, C.W. (1999). SWEET - WWW-based rapid 3D construction of oligo- and polysaccharides. Bioinformatics **15**, 767-68.
160. Joshi, H.J., von der Lieth, C.W., Packer, N.H., and Wilkins, M. R. (2010). GlycoViewer: A tool for visual summary and comparative analysis of the glycome. Nucleic Acids Res **38** (Web Server issue), W667-70.
161. Ranzinger, R., Kochut, K.J., Miller, J.A., Eavenson, M., Lütteke, T., and York, W.S. (2017). GLYDE-II: The GLYcan data exchange format. Perspect Sci **11**, 24-30.
162. Rojas-Macias, M.A., and Lütteke, T. (2015). Statistical analysis of amino acids in the vicinity of carbohydrate residues performed by glyvicinity. Methods Mol Biol **1273**, 215-26.
163. Fankhauser, N., and Mäser, P. (2005). Identification of GPI anchor attachment signals by a Kohonen self-organizing map. Bioinformatics **21**, 1846-52.
164. Wu, J., Wei, J., Hogan, J.D. *et al*. (2018). Negative Electron Transfer Dissociation Sequencing of 3-O-Sulfation-Containing Heparan Sulfate Oligosaccharides. J Am Soc Mass Spectrom **29**, 1262-72.
165. Pierleoni, A., Martelli, P., and Casadio, R. (2008). PredGPI: A GPI-anchor predictor. BMC Bioinformatics **9**, 392.
166. Eavenson, M., Kochut, K.J., Miller, J.A., Ranzinger, R., Tiemeyer, M., Aoki, K., and York, W.S. (2015). Qrator: A web-based curation tool for glycan structures. Glycobiol **25**, 66–73.
167. Chernyshov, I.Y., and Toukach, P.V. (2018). REStLESS: Automated translation of glycan sequences from residue-based notation to SMILES and atomic coordinates. Bioinformatics **34**, 2679-81.
168. Tanaka, K., Aoki-Kinoshita, K.F., Kotera, M. *et al*. (2014). WURCS: The Web3 unique representation of carbohydrate structures. J Chem Inf Model **54**, 1558-66.
169. Alocci, D., Mariethoz, J., Horlacher, O., Bolleman, J.T., Campbell, M.P., Lisacek, F. (2015). Property Graph vs RDF triple store: A comparison on glycan substructure search. PLoS One **10**, 1–17.
170. Neelamegham, S., Aoki-Kinoshita, K., Bolton, E. *et al*. (2019). Updates to the Symbol

Nomenclature for Glycans guidelines. Glycobiol **29**, 620-24.
171. Thieker, D.F., Hadden, J.A., Schulten, K. and Woods, R.J. (2016). 3D implementation of the symbol nomenclature for graphical representation of glycans. Glycobiol **26**, 786-87.
172. Cheng, K., Zhou, Y. and Neelamegham, S. (2017). DrawGlycan-SNFG: A robust tool to render glycans and glycopeptides with fragmentation information. Glycobiol **27**, 200-05.
173. Joshi, H.J., Jørgensen, A., Schjoldager, K.T. *et al*. (2018). GlycoDomainViewer: A bioinformatics tool for contextual exploration of glycoproteomes. Glycobiol **28**, 131-36.
174. Woods, R.J. (2018). Predicting the Structures of Glycans, Glycoproteins, and Their Complexes. Chemical Rev **118**, 8005-24.
175. Sehnal, D. and Grant, O.C. (2019). Rapidly Display Glycan Symbols in 3D Structures: 3D-SNFG in LiteMol J Proteome Res **18**, 770-74.
176. Mercier, P.L., Mariethoz, J., Lascano-Maillard, J., Bonnardel, F., Imberty, A., Ricard-Blum, S., Lisacek, F. (2019). A bioinformatics view of glycan–virus interactions. Viruses **11**, 1–14.
177. Clerc, O., Deniaud, M., Vallet, S.D., Naba, A., Rivet, A., Perez, S., Thierry-Mieg, N., Ricard-Blum, S. (2019). MatrixDB: Integration of new data with a focus on glycosaminoglycan interactions. Nucleic Acids Res **47** (Database issue), D376-D381.
178. Lefranc, M., Giudicelli, V., Duroux, P. *et al*. (2015). IMGT, the international ImMunoGeneTics information system R 25 years on. Nucleic Acids Res 43, (Database issue), D413–D422.
179. Vita, R., Mahajan, S., Overton, J.A. *et al*. (2019). The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res **47** (Database issue), D339-D343.
180. Sterner, E., Flanagan, N. and Gildersleeve, J. C. (2016). Perspectives on Anti-Glycan Antibodies Gleaned from Development of a Community Resource Database. ACS Chem Biol **11**, 1773-83.
181. Sayers, E.W., Barrett, T., Benson, D.A., *et al*. (2011). Database resources of the national center for biotechnology information. Nucleic Acids Res **39** (Database issue), D38-51.
182. Krambeck, F.J. and Betenbaugh, M.J. (2005). A mathematical model of N-linked glycosylation. Biotechnol Bioeng **92**, 711-28.
183. Liu, G., Marathe, D.D., Matta, K.L. and Neelamegham, S. (2008). Systems-level modeling of cellular glycosylation reaction networks: O-linked glycan formation on natural selectin ligands. Bioinformatics **24**, 2740-47.
184. Siest, G., Auffray, C., Taniguchi, N. *et al*. (2015). Systems medicine, personalized health and therapy. in Pharmacogenomics **16**, 1527-39.
185. Freire, M. and Pereira, M. (2007). Encyclopedia of Internet technologies and applications. Encyclopedia of Internet Technologies and Applications. Information Science Reference - Imprint of: IGI Publishing, Hershey PA (pages 716). ISBN:978-1-59140-993-9.
186. Juneau, J. and Juneau, J. (2018). RESTful Web Services. in Java EE 8 Recipes - A Problem-Solution Approach (2nd Edition; Apress, New York, NY). ISBN-13: 978-1-

4842-3593-5
187. Katayama, T., Nakao, M. and Takagi, T. (2010). TogoWS: Integrated SOAP and REST APIs for interoperable bioinformatics web services. Nucleic Acids Res **38**, 706–11.
188. Foster, I., Kesselman, C. and Tuecke, S. (2001). The anatomy of the grid: Enabling scalable virtual organizations. Int J High Perform Comput Appl August 1.
189. Foster, I. Kesselman, C. Tuecke, S. (2003). Chapter 6, The Anatomy of the Grid: Enabling Scalable Virtual Organizations, Pages: 169-197. In: Grid Computing: Making the Global Infrastructure a Reality (Fran Berman, Geoffrey Fox, Tony Hey, eds.), John Wiley & Sons, Ltd., March, 2003. ISBN:9780470853191.
190. Datta, A.K., Eller, B., Sinai, A. (2011). neoGRID: Enabling access to teragrid resources - Application for sialylmotif analysis in the protozoan Toxoplasma gondii. in TeraGrid 11: Extreme Digital Discovery. July 18-21.
191. Datta, A.K. and Paulson, J.C. (1997). Sialylmotifs of sialyltransferases. Indian J Biochem Biophys **34**, 157-65.
192. Datta, A.K., Jackson, V., Nandkumar, R., Sproat, J., Zhu, W. , Krahling, H. (2010). CHOIS: Enabling grid technologies for obesity surveillance and control. In: 'Healthgrid Applications and Core Technologies (T. Solomonides, I. Blanquer, V. Breton, T. Glatard, and Y. Legre, eds.), **159**, 191-202. IOS Press, Washington D.C. ISBN 978-1-60750-582-2.
193. Foster, I. and Kesselman, C. (2011). The history of the grid. Advances in Parallel Computing **20**, 3-30.
194. Barker, M., Delgado labarriaga, S., Wilkins-Diehr, N., Gesing, S., Katz, DS., Shahand, S., Henwoodg, S., Glatard, T., Jeffery, K., Corrie, B., Treloar, A., Glaves, H., Wyborn, L., Hong, NPC., Costa, A. (2019). The global impact of science gateways, virtual research environments and virtual laboratories. Futur Gener Comput Syst **95**, 240–48.
195. Hunt, K., Datta, A., Moye, R., von Oehsen, B., Lathrop, S. (2009). Campus Champions – Bringing High Performance Computing to your Campus. in Proceedings in EDUCAUSE 2009 Annual Conference (Panel Discussion).
196. Tan, W., Madduri, R., Nenadic, A., Soiland-Reyes, S., Sulakhe, D., Foster, I., Goble, C.A. (2010). CaGrid workflow toolkit: A taverna based workflow tool for cancer grid. BMC Bioinformatics **11**, 542.
197. Hunter, S., Jones, P., Mitchell, A. *et al*. (2012). InterPro in 2011: New developments in the family and domain prediction database. Nucleic Acids Res **40**, 306–12.
198. Datta, A. (2009). Cyberinfrastructure for Glycome Research. Glycoconj J **26**, 849–50.
199. Datta, A. (2010). High-throughput motif analysis for drug discovery research. In: Proceedings of the BIT 8th Annual Congress of International Drug Discovery Sciences and Technology. p137, 23 - 26 October 2010, Beijing, China.
200. Sukhija, N. and Datta, A.K. (2013). C-Grid: Enabling iRODS-based grid technology for community health research. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), **8060**, 17-31.
201. Sukhija, N., Sevin, S., Datta, A. K. & Coulter, E. (2018). Grid technology for supporting health education and measuring the health outcome. In: ACM International Conference

Proceeding Series. ACM ISBN 978-1-4503-6446-1/18/07.
202. Baru, C., Moore, R., Rajasekar, A. and Wan, M. (2010). The SDSC storage resource broker. In: Proceedings of CASCON'10 - CASCON First Decade High Impact Papers, November 2010.
203. Olson, J.S., Ellisman, M., James, M., Grethe, J.S. and Puetz, M. (2013). The Biomedical Informatics Research Network. in Scientific Collaboration on the Internet (Gary M. Olson, Ann Zimmerman, and Nathan Boss, eds). Pages 221-232. The MIT Press, Cambridge, MA. ISBN: 978-0-262-15120-7.
204. Billah, M.M, Goodall, J.L., Narayan, U., Essawy, B.T., Lakshmi, V., Rajasekar, A., Moore, R.W. (2016). Using a data grid to automate data preparation pipelines required for regional-scale hydrologic modeling. Environ Model Softw **78**, 31-39.
205. Rajasekar, A., Moore, R., Wan, M., Schroeder, W. and Hasan, A. (2010). Applying rules as policies for large-scale data sharing. In: ISMS 2010 - UKSim/AMSS 1st International Conference on Intelligent Systems, Modelling and Simulation, January 2010, Pages 322–327.
206. Chiang, G.T., Clapham, P., Qi, G., Sale, K. and Coates, G. (2011). Implementing a genomic data management system using iRODS in the Wellcome Trust Sanger Institute. BMC Bioinformatics **12**, 361.
207. Venugopal, S., Buyya, R. and Ramamohanarao, K. (2006). A Taxonomy of Data Grids for Distributed Data Sharing, Management, and Processing. ACM Comput Surv **38**, June 2006.
208. Saunders, B., Lyon, S., Day, M., Riley, B., Chenette, E., Subramaniam, S., Vadivelu, I. (2008). The molecule pages database. Nucleic Acids Res **36** (Database issue), D700-6.
209. Obay, M.A.M., Altrafi, G. and Ismail, M.O. (2014). Relational vs . NoSQL Databases : A Survey. Int J Comput Inf Technol (ISSN: 2279 – 0764), **3** (3), May 2014.
210. York, W.S., Mazumder, R., Ranzinger, R. *et al*. (2020). GlyGen: Computational and Informatics Resources for Glycoscience. Glycobiol **30**, 72-73.
211. Vos, R.A., Katayama, T., Mishima, H. *et al*. (2020). BioHackathon 2015: Semantics of data for life sciences and reproducible research. F1000Research 9, 136.