



FGDB: Database of follicle stimulating hormone glycans

Sushil K Shakyawar^a, Sanjit Pandey^a, David J Harvey^b, George Bousfield^c, Chittibabu Guda^{a,*}

^aDepartment of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, NE 68198, United States

^bTarget Discovery Institute, Nuffield Department of Medicine, Oxford University, Oxford OX3 7FZ, United Kingdom

^cFairmount College of Liberal Arts and Sciences, Wichita State University, KS 67260, United States



ARTICLE INFO

Article history:

Received 19 November 2020

Received in revised form 4 March 2021

Accepted 6 March 2021

Available online 22 March 2021

Keywords:

Follicle stimulating hormones (FSH)

Glycan structures

Glycoprotein database

Glycomics

Glycan abundance analysis

ABSTRACT

Glycomics, the study of the entire complement of sugars of an organism has received significant attention in the recent past due to the advances made in high throughput mass spectrometry technologies. These analytical advancements have facilitated the characterization of glycans associated with the follicle-stimulating hormones (FSH), which play a central role in the human reproductive system both in males and females utilizing regulating gonadal (testicular and ovarian) functions. The irregularities in FSH activity are also directly linked with osteoporosis. The glycoanalytical studies have been tremendously helpful in understanding the biological roles of FSH. Subsequently, the increasing number of characterized FSH glycan structures and related glycoform data has thrown a challenge to the glycoinformatics community in terms of data organization, storage and access. Also, a user-friendly platform is needed for providing easy access to the database and performing integrated analysis using a high volume of experimental data to accelerate FSH-focused research.

FSH Glycans DataBase (FGDB) serves as a comprehensive and unique repository of structures, features, and related information of glycans associated with FSH. Apart from providing multiple search options, the database also facilitates an integrated user-friendly interface to perform the glycan abundance and comparative analyses using experimental data. The automated integrated pipelines present the possible structures of glycans and variants of FSH based on the input data, and allow the user to perform various analyses. The potential application of FGDB will significantly help both glycoinformaticians as well as wet-lab researchers to stimulate the research in this area.

FGDB web access: <https://fgdb.unmc.edu/>

© 2021 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Advances made in the development and application of high-throughput technologies including Mass Spectrometry (MS), and Nuclear Magnetic Resonance (NMR) have revolutionized glycomic studies focusing on structures, glycosylation, and biological roles of glycans in the cellular systems. Previous studies have widely explored the involvement of glycans in a variety of cellular functions such as protein folding to provide specialized functions in eukaryotes [1,2], occurrence of nucleocytoplasmic glycosylation to regulate cellular metabolism [3,4], disease progression [5,6], cell proliferation and differentiation [7], cell-to-cell interactions [8,9], immune evasion [10] and many more. Our understanding of the

involvement of glycans in human reproductive systems is still in its infancy; however, owing to the access to advanced technologies, in-depth characterization of FSH associated glycans has been carried out in the recent past [11–13]. FSH is a heterodimeric glycoprotein with a common α and the hormone-specific β -subunits. α -Subunit is N-glycosylated at positions Asn52 and Asn78, while, β -subunit is N-glycosylated at Asn7 and Asn24 in their amino acid sequences. Previous researches have demonstrated the regulatory role of FSH in reproduction [14,15] and osteoporosis in humans [16].

Recent glycoanalytical innovations have provided a large amount of experimental data for structural analysis of complex glycan molecules in various organisms, which warranted the need to develop bioinformatic and computational solutions for data storage and organization, and build analytical platforms for easy access, analysis and visualization [17]. In this context, GlycomeDB [18] (now part of GlyTouCan 1.0 [19], CFG (Consortium for Functional Glycomics), and GlycoWorkbench [20] are among the

* Corresponding author at: Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, NE 68198-5805, United States.

E-mail addresses: sushil.shakyawar@unmc.edu (S.K Shakyawar), sanjit.pandey@unmc.edu (S. Pandey), david.harvey@ndm.ox.ac.uk (D.J Harvey), george.bousfield@wichita.edu (G. Bousfield), babu.guda@unmc.edu (C. Guda).

most used databases and tools that provide structural data of glycans accompanied with analytical tools [21]. Notably, GlyTouCan 1.0 [19] is one of the largest repositories of glycan structures. Some databases such as GLYCOSCIENCES.de harbor more specific data on NMR-based glycan 3D structures [22] and facilitate structure-based computational analyses including molecular modeling and drug designing [23,24] for investigating biological roles of glycans [25].

The glycomic research for characterizing glycans associated with FSH in human reproductive systems is advancing at a rapid pace and a large amount of structural information on FSH glycans is available in the literature. However, these data are archived in the literature by independent groups and are not easily accessible in a user-friendly manner, thus limiting their use by the research community. There are several hundreds of characterized FSH glycans, of which 91 are core-fucosylated while 139 lacked fucose residue [13]. To the best of our knowledge, there is no specific database which provides structural information of FSH glycans. Existing public databases poorly store metadata on FSH glycans and are inconsistent in the data formats used to represent structural information. They also lack a user-friendly interface to perform analysis on FSH-specific glycan structures or their relative abundance. To address these issues, glycobiologists and glycoinformaticians are encouraged to develop bioinformatics-based solutions to support large scale data-driven analyses focusing on FSH glycans.

We address this issue by developing an open-source webserver that provides a platform to store curated FSH glycan structures, and supports searching and retrieval of pertinent information using various features. In addition, we provide an integrated interface with analytical tools for abundance calculation and comparison of data between experiments. This web server is expected to significantly promote research in the FSH glycomics domain. With this objective, we developed an FSH Glycans

DataBase (FGDB) using the Python framework, which provides access not only to the glycan structural data but also facilitates analytics using the raw experimental data. The FGDB will uniquely serve as a central hub for accessing FSH glycans data, depositing new glycan structures, and performing analyses using mass spectrometry data using its user-friendly features.

2. Architecture of FGDB and web interface

The FGDB primarily stores information of glycan structures and their features in flat files, and python scripts are used to process queries to the database, as illustrated in Fig. 1. The web interface was built using Flask (<https://flask.palletsprojects.com/en/1.1.x/>) in a python environment, where most information related to glycan structures and their features are generated using python package glyppy [26]. All glycan abundance calculations and graph plotting tasks are carried out by scripts developed using python libraries as they provide better capabilities for integration with existing open-source algorithms and tools designed for biological data analysis.

3. Data formats

3.1. Glycan structure representation

3.1.1. Graphical representation

FGDB uses two graphical formats for representing glycan structures: 1) the symbol nomenclature for glycans (SNFG) [27], and 2) the OGI format, which is recommended by Oxford Glycobiology Institute [28,29]. SNFG structures were generated programmatically using python package glyppy, whereas, OGI structures were drawn manually and curated by our experts following Oxford Glycobiology Institute system’s conventions, which display

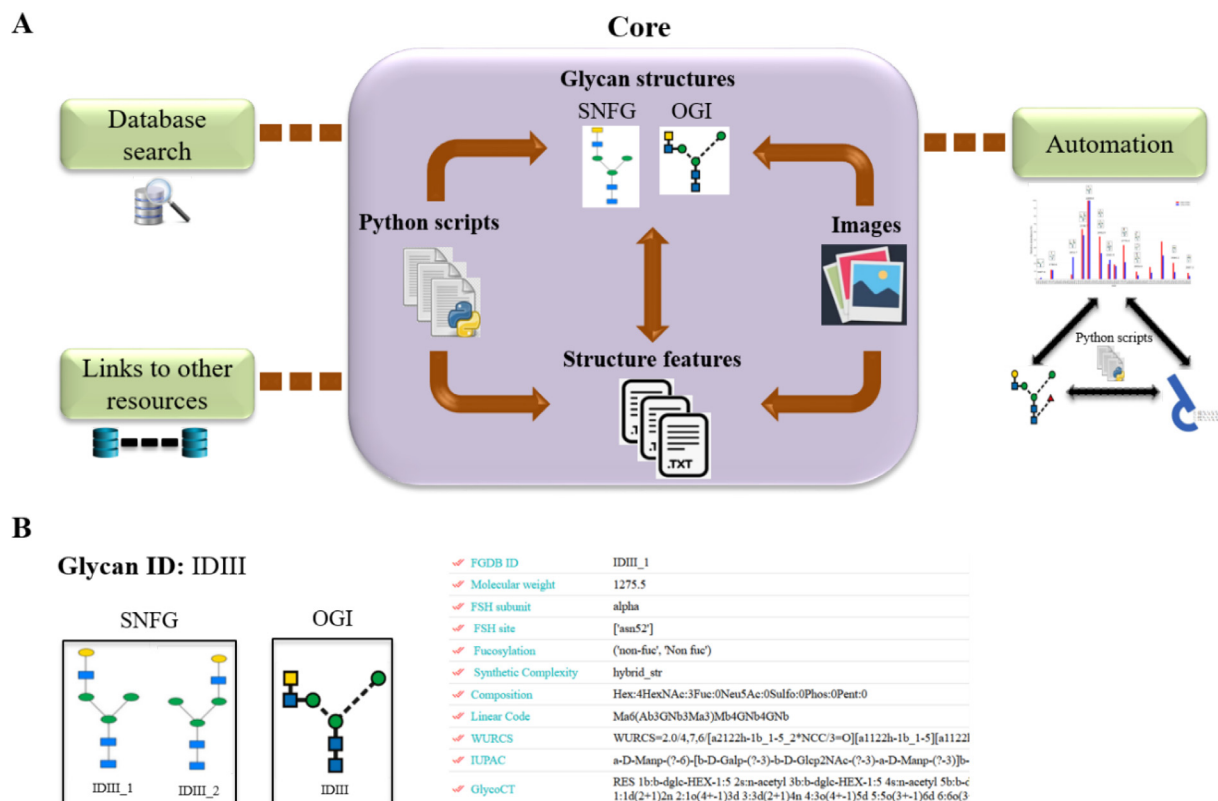
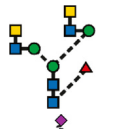
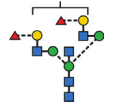
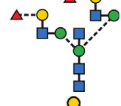
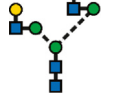
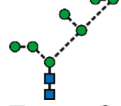
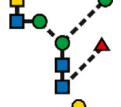
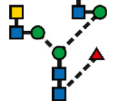
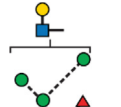
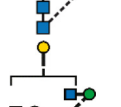
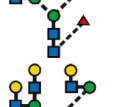
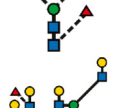
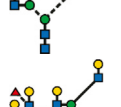
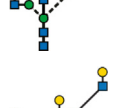


Fig. 1. A) Architecture of FGDB and associated functionalities. B) An example showing the core information that FGDB stores on a glycan structure.

Table 1
Structural features of glycans that are incorporated into the FGDB database. A brief description of each feature is provided below.

Features	Sub-features	Description	Examples
Fucosylation	Core fucosylation	Fucose residue attached to the reducing terminal GlcNAc residue attached to Asn	
	Branch fucosylation	Fucose residue attached to GlcNAc or Gal residues in one or more branches	
	Terminal fucosylation	Fucose residue terminating a branch	
	No fucosylation	No fucose residue attached to oligosaccharide	
Synthetic complexity	High mannose	Glycans with two N-acetylglucosamines and 4–9 mannose residues	
	Hybrid	Glycans contain 1 to 6 mannose residues on the α1-6 mannose branch while one or more complex branches are present on the α1-3 mannose branch	
	Complex	Glycans possessing two or more antenna composed of GlcNAc, Gal, GalNAc, Fucose, or sialic acid residues	
Branching complexity	Mono-antennary	Single, complex branches initiated with GlcNAc residues to one of the core mannose residues, either the α1-3 or α1-6 mannose	
	Bi-antennary	Two GlcNAc-initiated complex branches linked to the pentasaccharide core	
	Tri-antennary	Three GlcNAc-initiated branches linked to the pentasaccharide core	
	Tetra-antennary	Four GlcNAc-initiated branches linked to the pentasaccharide core	
	Penta-antennary	Five GlcNAc-initiated branches linked to the core	
Sialylation	Neutral	No charged moieties, such as sulfate, phosphate or sialic acid in glycan	

(continued on next page)

Table 1 (continued)

Features	Sub-features	Description	Examples
	Full	One sialic acid residue on mono-antennary, 2 on bi-antennary, 3 on tri-antennary, or 4 on tetra-antennary glycan	
	Partial	One sialic acid residue on bi-antennary, 1–2 on tri-antennary, or 1–3 on tetra-antennary glycan	
GlcNAc bisection	Yes	Attachment of beta1-4 GlcNAc to the branching, beta-1–4 mannose residue	
	No	No GlcNAc residue attached to the branching, beta-1–4 mannose residue	

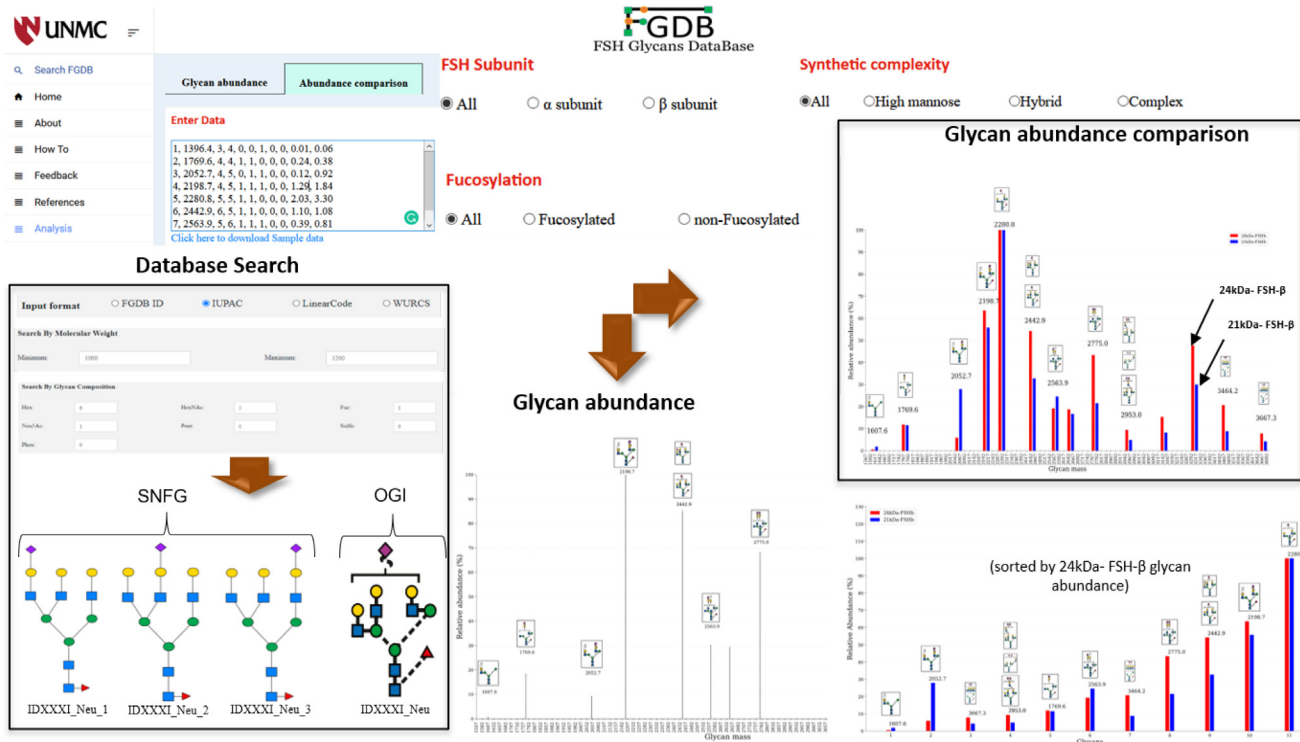


Fig. 2. Demonstration of FGDB search and glycan abundance analysis with user input data.

embedded specificity and anomericity. More information about monosaccharide linkage and orientations defined in OGI structures are provided on the database webpage.

3.1.2. Text-based representation

Similar to drawing SNFG structures, glypy package was also used to parse glycan structures in the widely used text-based formats, such as IUPAC, WURCS [30], LinearCode [31], and GlycoCT [32].

3.2. Glycan structures: sources and features

FGDB stores information of glycans that are associated with FSH-α subunit (attached at Asn52 and Asn78 positions), and FSH-

β subunit (attached at Asn7 and Asn24 positions). A variety of features based on the interlinkage of monosaccharides in the glycan structures are also incorporated. The current version of the database accommodates the following features of the glycan structures: “Fucosylation”, “Synthetic complexity”, “Branching complexity”, “Sulfation”, “Phosphorylation”, “Sialylation”, and “GlcNAc bisection”, each of which is briefly described in Table 1.

4. Database accessibility and current status

FGDB database can be accessed on the web at <https://fgdb.unmc.edu/> to perform search queries and glycan data analyses. Both operations require input data in a specific format, which has been

shown in the example text file on corresponding web pages. FGDB output tables are downloadable to local desktop along with appropriate data labels. The database is available as an open-source resource under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial reuse, redistribution, and reproduction in any medium provided the original work is properly cited. For commercial reuse, permission in writing should be taken from the developers. The full list of FGDB entries can be provided to users up on request. We also request the user community to submit glycan structures to FGDB, curate, and update annotations by using the correspondence form provided on the FGDB web page.

FGDB on its first release (FGDB 1.0) includes 230 N-glycans (represented in OGI format) from FSH alpha and beta subunits, and our group will continue to update the database with recombinant human FSH glycans as well as FSH from horse and other species. From these 230 glycans, we generated the images of over 850 possible glycan variants in SNFG format. For each glycan structure, features as mentioned above can be accessed from the 'glycan details page'.

5. Database usage: search and analysis of glycans abundance

FGDB facilitates web-based queries using a variety of user input data such as molecular weight range, monosaccharide composition, and text-based IDs such as IUPAC, LinearCode, WURCS, or database-specific FGDB ID (Fig. 2). Every search can be coupled with different filtering criteria such as FSH subunit (α or β or both) as glycan source, glycosylation site location on the protein chain, glycan fucosylation, complexity of glycan structures from synthetic perspectives, and other features, which allow users to interactively narrow down the results. The results page simply lists glycans along with SNFG and OGI structures and other information. Each glycan entry in the output table is hyperlinked to the corresponding 'glycan details page' which contains detailed information of the glycan structure such as glycan source, structural features, molecular weight, and names in IUPAC, LinearCode, WURCS, and GlycoCT formats, and monosaccharide composition, as shown in Fig. 1B.

An integrated interface in FGDB was developed to facilitate analysis and plotting of output data such as the relative abundance of glycans and comparison between the datasets. The monosaccharide composition along with the abundance of information in a specific format can be entered or uploaded in text files to perform simple glycan abundance analysis (as shown for 24 kDa-FSH β glycans in Fig. 2). Moreover, advanced analysis can also be performed to compare glycan abundance from either different experimental settings or different sources in a whole glycan population. For instance, Fig. 2 displays an example from FGDB on the relative abundance of 24 kDa-FSH β and 21 kDa-FSH β glycans. Along with configuring the output plots, we also facilitate additional options on the input form to sort the order of glycans on the bar chart based on their abundance levels (Fig. 2).

6. Future work

In the future versions, we plan to update FGDB to include more glycan structures and variants from human, horse and other species along with their experimental sources and biosynthetic pathway information. We will continue to add more functionalities, especially to perform abundance focused analysis that includes structural features as mentioned above. Also, we will emphasize on interlinking FGDB with other glycomics and glycoproteomic databases such as GlyTouCan and KEGG GLYCAN Database [33] to facilitate more robust and interactive analyses.

CRediT authorship contribution statement

Sushil K Shakyawar: Conceptualization, Methodology, Data curation, Writing - original draft, Writing - review & editing. **Sanjit Pandey:** Resources, Software. **David J Harvey:** Data curation. **George Bousfield:** Conceptualization, Data curation, Funding acquisition. **Chittibabu Guda:** Conceptualization, Project Administration, Resources, Supervision, Validation, Review & Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by NIH Grant # 2P01AG029531 to GB (subcontract to CG). The authors would like to thank the Bioinformatics and Systems Biology Core (BSBC) facility at UNMC for providing the computational infrastructure and support. BSBC is partly supported by the Nebraska Research Initiative and NIH awards [5P20GM103427, 5P30CA036727] to CG.

Author contributions

SKS implemented all the programming scripts for designing database architecture, developing web interface, performing different analysis, and wrote the manuscript. GB and DJP provided the data on glycan structures. SP provided technical help for deploying the database on the UNMC server. CG guided and supervised the project from conception to completion, and significantly edited to improve the manuscript. All authors have read and approved the final manuscript.

References

- [1] Vasudevan D, Haltiwanger RS. Novel roles for O-linked glycans in protein folding. *Glycoconj J* 2014;31:417–26. <https://doi.org/10.1007/s10719-014-9556-4>.
- [2] Xu C, Ng DTW. Glycosylation-directed quality control of protein folding. *Nat Rev Mol Cell Biol* 2015;16:742–52. <https://doi.org/10.1038/nrm4073>.
- [3] West CM, van der Wel H, Gaucher EA. Complex glycosylation of Skp1 in Dictyostelium: Implications for the modification of other eukaryotic cytoplasmic and nuclear proteins. *Glycobiology* 2002;12:17R–27R. <https://doi.org/10.1093/glycob/12.2.17R>.
- [4] West CM. Nucleocytoplasmic glycosylation. *Biochim Biophys Acta - Gen Subj* 2010;1800:47–8. <https://doi.org/10.1016/j.bbagen.2009.12.008>.
- [5] Gomes PS, Feijó DF, Morrot A, Freire-de-Lima CG. Decoding the role of glycans in malaria. *Front Microbiol* 2017;8. <https://doi.org/10.3389/fmicb.2017.01071>.
- [6] Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. *Nat Rev Nephrol* 2019;15:346–66. <https://doi.org/10.1038/s41581-019-0129-4>.
- [7] Lau KS, Partridge EA, Grigorian A, Silvescu CI, Reinhold VN, Demetriou M, et al. Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. *Cell* 2007;129:123–34. <https://doi.org/10.1016/j.cell.2007.01.049>.
- [8] Forestier C-L, Gao Q, Boons G-J. Leishmania lipophosphoglycan: how to establish structure-activity relationships for this highly complex and multifunctional glycoconjugate?. *Front Cell Infect Microbiol* 2015;4. <https://doi.org/10.3389/fcimb.2014.00193>.
- [9] Hall MK, Weidner DA, Dayal S, Schwalbe RA. Cell surface N-glycans influence the level of functional E-cadherin at the cell-cell border. *FEBS Open Bio* 2014. <https://doi.org/10.1016/j.fob.2014.10.006>.
- [10] Clark GF. The role of glycans in immune evasion: the human foetoembryonic defence system hypothesis revisited. *Mol Hum Reprod* 2014. <https://doi.org/10.1093/molehr/gat064>.
- [11] Bousfield GR. Comparison of follicle-stimulating hormone glycosylation microheterogeneity by quantitative negative mode nano-electrospray mass spectrometry of peptide-N-glycanase-released oligosaccharides. *J Glycomics Lipidomics* 2015. <https://doi.org/10.4172/2153-0637.1000129>.
- [12] T Rajendra Kumar JSD. Naturally occurring follicle-stimulating hormone glycosylation variants. *J Glycomics Lipidomics* 2014. <https://doi.org/10.4172/2153-0637.1000e117>.

- [13] Bousfield GR, Harvey DJ. Follicle-stimulating hormone glycobiochemistry. *Endocrinology* 2019. <https://doi.org/10.1210/en.2019-00001>.
- [14] Orłowski M, Sarao MS. Physiology, Follicle Stimulating Hormone. 2019.
- [15] Bousfield GR, May JV, Davis JS, Dias JA, Kumar TR. In vivo and in vitro impact of carbohydrate variation on human follicle-stimulating hormone function. *Front Endocrinol (Lausanne)* 2018. <https://doi.org/10.3389/fendo.2018.00216>.
- [16] Agrawal M, Zhu G, Sun L, Zaidi M, Iqbal J. The role of FSH and TSH in bone loss and its clinical relevance. *Curr Osteoporos Rep* 2010;8:205–11. <https://doi.org/10.1007/s11914-010-0028-x>.
- [17] Liu G, Neelamegham S. Integration of systems glycobiochemistry with bioinformatics toolboxes, glycoinformatics resources, and glycoproteomics data. *Wiley Interdiscip Rev Syst Biol Med* 2015;7:163–81. <https://doi.org/10.1002/wsbm.1296>.
- [18] Ranzinger R, Herget S, Von Der Lieth CW, Frank M. GlycomeDB-A unified database for carbohydrate structures. *Nucleic Acids Res* 2011;39. <https://doi.org/10.1093/nar/gkq1014>.
- [19] Aoki-Kinoshita K, Agravat S, Aoki NP, Arpinar S, Cummings RD, Fujita A, et al. GlyTouCan 1.0 - The international glycan structure repository. *Nucleic Acids Res* 2016;44:D1237–42. <https://doi.org/10.1093/nar/gkv1041>.
- [20] Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM. GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of glycans. *J Proteome Res* 2008;7(4):1650–9.
- [21] Hizal DB, Wolozny D, Colao J, Jacobson E, Tian Y, Krag SS, et al. Glycoproteomic and glycomic databases. *Clin Proteomics* 2014;11:15. <https://doi.org/10.1186/1559-0275-11-15>.
- [22] Lütteke T, Bohne-Lang A, Loss A, Goetz T, Frank M, von der Lieth CW. GLYCOSCIENCES.de: An internet portal to support glycomics and glycobiochemistry research. *Glycobiology* 2006;16:71R–81R.
- [23] de Ruyck J, Brysbaert G, Blossey R, Lensink MF. Molecular docking as a popular tool in drug design, an in silico travel. *Adv Appl Bioinform Chem* 2016;9:1–11. <https://doi.org/10.2147/AABC.S105289>.
- [24] Meng X-Y, Zhang H-X, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des* 2011;7:146–57. <https://doi.org/10.2174/157340911795677602>.
- [25] Varki A. Biological roles of glycans. *Glycobiology* 2017;27:3–49. <https://doi.org/10.1093/glycob/cww086>.
- [26] Klein J, Zaia J. Glypy: an open source glycoinformatics library. *J Proteome Res* 2019;18:3532–7. <https://doi.org/10.1021/acs.jproteome.9b00367>.
- [27] Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, et al. Symbol nomenclature for graphical representations of glycans. *Glycobiology* 2015;25:1323–4.
- [28] Harvey DJ, Merry AH, Royle L, P. Campbell M, Dwek RA, Rudd PM. Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics* 2009;9:3796–801. <https://doi.org/10.1002/pmic.200900096>.
- [29] Harvey DJ, Merry AH, Royle L, Campbell MP, Rudd PM. Symbol nomenclature for representing glycan structures: extension to cover different carbohydrate types. *Proteomics* 2011;11:4291–5. <https://doi.org/10.1002/pmic.201100300>.
- [30] Tanaka K, Aoki-Kinoshita KF, Kotera M, Sawaki H, Tsuchiya S, Fujita N, et al. WURCS: The Web3 unique representation of carbohydrate structures. *J Chem Inf Model* 2014;54:1558–66.
- [31] Banin E, Neuberger Y, Altshuler Y, Halevi A, Inbar O, Nir D, et al. A novel Linear Code((R)) nomenclature for complex carbohydrates. *TRENDS Glycosci Glycotechnol* 2002;14:127–37.
- [32] Herget S, Ranzinger R, Maass K, Lieth CWvd. GlycoCT-a unifying sequence format for carbohydrates. *Carbohydr Res* 2008;343:2162–71.
- [33] Hashimoto K, Goto S, Kawano S, Aoki-Kinoshita KF, Ueda N, Hamajima M, et al. KEGG as a glycome informatics resource. *Glycobiology* 2006;16. <https://doi.org/10.1093/glycob/cwj010>.