



# Glycoenzonto is Ontology for Glycoenzyme Pathways and Molecular Functions

Bella Edward\*

Department of Biochemistry, Molecular and Cellular Biology, USA

\*Corresponding Author's E-mail: [edwardbella@rediff.com](mailto:edwardbella@rediff.com)

**Received:** 02-Dec-2022, Manuscript No. IRJBB-22-84550; **Editor assigned:** 05-Dec-2022, PreQC No. IRJBB-22-84550 (PQ); **Reviewed:** 19-Dec-2022, QC No. IRJBB-22-84550; **Revised:** 24-Dec-2022, Manuscript No. IRJBB-22-84550 (R); **Published:** 31-Dec-2022, DOI: 10.14303/2250-9941.2022.39

## Abstract

The development of glycoproteomics techniques and technologies has made significant strides in recent years, which has led to a steady growth in the reported proteins, their associated glycans, and their glycosylation sites. However, relatively few of these reports end up in databases or other data storage systems. One of the main causes is the lack of a digital standard for representing glycoproteins and the difficult glycan annotations. Such a standard must be able to store not only single glycans but also glycoforms on a specific glycosylation site, deal with partially missing site information if no site mapping was done, and store abundances or ratios of glycans within a glycoform of a specific site, depending on the experimental method (Tsankova 2006). We created the GlycoConjugate Ontology (GlyCoCoO) as a common semantic framework to characterise and represent glycoproteomics data in order to enable the aforementioned. GlyCoCoO may be used as the foundation for data sharing formats and to encode glycoproteomics data in triplestores.

**Keywords:** Glycoconjugate, Glycolipid, Glycoprotein, Ontology

## INTRODUCTION

Glycobiology is the study of saccharides, which are found abundantly throughout nature and are also known as carbohydrates, sugar chains, or glycans. Given that glycobiology includes some of the most significant posttranslational modifications of proteins, it is important to comprehend how the relatively few genes in a typical genome can result in the enormous biological complexity necessary for the development, growth, and function of a variety of organisms. In the building of complex multicellular organs and animals, which needs interactions between cells and the surrounding matrix, the biological functions of carbohydrates are particularly apparent (Lubin 2008). Every cell and countless macromolecules in nature have a variety of covalently linked glycans; there is no known exception (albeit glycans can also be freestanding entities). In order to regulate or mediate a variety of events in cell-cell, cell-matrix, and cell-molecule interactions crucial to the development and function of a complex multicellular organism, such as cellular activation, embryonic development, differentiation, and malignancy,

glycoproteins are frequently found on the cell membrane or secreted (Lubin et al., 2008). As an example, they can moderate interactions between a host and a parasite, disease, or symbiont. To increase our understanding of biological systems, it is crucial to comprehend the functions of glycans, variations in glycoforms/abundance of glycans, and site-occupancy. Glycobiologists now have a greater understanding of the roles that glycoproteins play because to advancements in bioinformatics tools and databases, including data standards and interoperability (Jakobsson 2008).

Several projects over the past few decades have catalogued and arranged data on glycans in databases. These initiatives had their start with CarbBank, a glycan structure database project that was begun in 1987 but stopped operating in 1997 owing to a lack of financing. The database's final edition included 50,000 entries with over 23,000 glycan sequences and related biological context, experimental approach, and publishing data (Denning et al., 2012). Following database projects, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) Glycan, the US Consortium

for Functional Glycomics (CFG), GlycoSuiteDB, UniCarbKB, Carbohydrate Structure Database (CSDB), GlycomeDB, and EUROCarbDB, have used this data set to seed new databases with a fundamental set of glycan structure records (Johnsen et al., 2016).

Briefly stated, KEGG Glycan is an integrated knowledge base of protein networks with genomic and chemical information that gives access to glycan structures through hand constructed pathway maps that reflect the current understanding of glycan production and metabolism for different species. The technical specifications for creating centralised and uniform database architecture for analytical data from liquid chromatography, mass spectrometry, and nuclear magnetic resonance (NMR) tests as well as data relating to carbohydrate structure were created by EUROCarbDB (Lauzon et al., 2000). Numerous tools, such as MonosaccharideDB and the separation-focused database GlycoBase that was eventually relocated to GlycoStore, were created under EUROCarbDB. GLYCOSCIENCES.de, which just received an upgrade from Glycosciences, focuses on the three-dimensional conformations of carbohydrates as retrieved from PDB and imported the whole CarbBank dataset. DB. Data from glycan microarray binding and glycan mass spectrometry profiling on human and mouse tissue and cell lines are combined in the CFG database by the consortium's members. The NCFG recently replaced the CFG, which had been concentrating on developing glycan microarray technologies with accompanying informatics (Lauzon et al., 1993).

Complex carbohydrates or glycans are biosynthesized through glycosylation and are found on lipids, nuclear proteins, and cell surfaces. The metabolic processing of monosaccharides generated from either dietary sugars (typically glucose and fructose) or through salvage pathways that break down recycled glycoconjugates results in the formation of nucleotide-sugar donors (for example, GDP-fucose). Once in the endoplasmic reticulum (ER) and Golgi, these nucleotide-sugars serve as donors for enzymatic activities carried out by a group of enzymes known as glycosyltransferases (GTs). These GTs work in a coordinated manner to transfer monosaccharides from nucleotide-sugar donors to target substrates expressed on protein/lipid scaffolds (Kim et al., 2015). Humans are endowed with a wide variety of glycans as a result of these biosynthetic processes, with four main subclasses being (i) N-linked glycans, (ii) O-linked glycans (O-GalNAc, O-GlcNAc, etc.), (iii) glycolipids, and (iv) glycosaminoglycans (heparan sulphates). All glycans collectively make up the cellular "glycome."

Glycosylation studies using systems-based bioinformatics demand structured information, including ontologies. GlycoRDF and Glycoconjugate Ontology (GlyCo) are two of these that aid in the synchronisation of glycan structural data from diverse glycoscience sources (Pálka et al., 2015). Although specific annotations of glycoEnzymes, pathways, and molecular activities are not a part of this project,

GlycoRDF outlines the framework to represent instances of glycoGenes and reactions here. Glycan inquiries at the GlyGen repository are made easier by the Glycan Naming and Subsumption Ontology (GNOme), which explains the topological connectedness of monosaccharides and substructures inside complex carbohydrates. The genetic Glyco-Disease Ontology (GGDonto), in particular the congenital diseases of glycosylation, collects data on glycoEnzyme malfunction and associated pathways in disease situations. There is currently no site that offers an exhaustive glycoEnzyme functional hierarchy.

## DISCUSSION

As demonstrated by the SPARQL queries described in the Results, it is possible to query various databases with a single query and get integrated data on a single glycoprotein. Numerous papers contained a wide range of information, including connections between diseases, tissues, and cell lines. Since all of these databases are constantly being updated, the information that is currently available only represents the situation as of this writing (James et al., 2003).

It is clear from the glycan data that, despite the GlyTouCan IDs not overlapping, the allocated IDs might still be mapped to different glycans because of variations in fragmentation annotations and confusing connections. For such ambiguity, GlyTouCan gives connection information; additional investigation of these glycan linkages is left for future effort. Examples of the RDF data for glycoproteins created by GlycoNAVI, GlyConnect, and UniCarbKB are presented in this work. GlyGen is a resource that offers information about glycoproteins in RDF format and is implementing GlyCoO ideas to promote data interoperability. In order to explore how ideas may be integrated or mapped with one another, we also intend to get in touch with database and lipid ontology developers. The GlySpace Alliance's members will eventually make all of this integrated data accessible.

We will also look at ways to link the GlyCoO ontology with other relevant ontologies after demonstrating its efficacy. For instance, the Protein Ontology (PRO) offers a reliable and expandable framework for ontological study on proteins. It acts as a standard representation of proteoforms utilising PSI-MOD as a posttranslational modification reference and UniProtKB as a sequence reference to precisely and fully model protein entities and their connections in biological systems. By harmonising with the GlyCoO principles mentioned, the PRO project will be broadened to capture the complexity of glycoproteoforms, particularly the variability of site-specific protein glycosylation.

Several sources, such as Reactome, KEGG (Kanehisa and Goto, 2000), GO, and UniProt, curate protein activities and pathways associated with glycosylation (Uniprot Consortium, 2018). Additional databases, such as Rhea, curate glycoenzyme reactions employing ChEBI molecular identities as glycan substrates by using ontology words.

Despite being important, these sites do not yet include enough information about glycosylation biosynthesis. The hierarchical aspect of glycosylation, which comprises the initiation, elongation/branching, and capping/termination stages, is especially poorly captured by them. GlycoEnzOnto fills this hole. Even though the focus is on human biology, other animals can benefit from much of the information and the data analysis approach. After include a few enzymes that are absent in humans, such as CMAH (Cytidine monophospho-N-acetylneuraminic acid hydroxylase) and A3GALT2, this same ontology, for instance, might be used to mice (Alpha-1, 3-galactosyltransferase 2).

Using linear strings obtained from IUPAC-condensed nomenclature, GlycoEnzOnto curates reaction rules for 403 glycoEnzymes. This schema has three key elements that describe: (i) substrate ambiguity linked to substrate ambiguity and enzyme-specific substrates; (ii) five categories of glycobiochemically relevant biochemical reactions; and (iii) reaction restrictions that restrict the type of substrate transformation. For the example of N-linked glycosylation, it was shown that the rules and restrictions may be effectively parsed to construct a glycosylation reaction network. This project's scope might be expanded to include different kinds of glycoconjugates. Automating the creation of glycosylation reaction networks enables modelling of biological reaction networks and fitting with experimental data on glycan structure. It enables the thorough integration of metabolic processes and biochemical pathways. Such networks may also be overlaid with transcriptomic data to establish glycoenzyme-glycan expression correlations.

## REFERENCES

1. Tsankova NM (2006). Sustained hippocampal chromatin regulation in a mouse model of depression and antidepressant action. *Nat Neurosci.* 9: 519-532.
2. Lubin FD, Roth TL, Sweatt (2008). Epigenetic regulation of *bdnf* gene transcription in the consolidation of fear memory. *J Neurosci.* 28: 10576-10586.
3. Jakobsson J (2008). KAP1-mediated epigenetic repression in the forebrain modulates behavioral vulnerability to stress. *Neuron. Behav neur.* 60: 818-831.
4. Denning DP, Hatch V, Horvitz HR (2012). Programmed elimination of cells by caspase-independent cell extrusion in *C. elegans*. *Nature.* 488: 226-230.
5. Johnsen HL, Horvitz HR (2016). Both the apoptotic suicide pathway and phagocytosis are required for a programmed cell death in *Caenorhabditis elegans*. *BMC Biol.* 14: 39-56.
6. Lauzon RJ, Rinkevich B, Patton CW, Weissman IL (2000). A morphological study of nonrandom senescence in a colonial urochordate. *Biol Bull.* 198: 367-378.
7. Lauzon RJ, Patton CW, Weissman IL (1993). A morphological and immunohistochemical study of programmed cell death in *Botryllus schlosseri* (Tunicata, Ascidiacea). *Cell Tissue Res.* 272: 115-127.
8. Kim MY, Kim HS, Choi N, Yang JH, Yoo YB, et al (2015). Screening mammography-detected ductal carcinoma in situ: mammographic features based on breast cancer subtypes. *Clinical Imag.* 39: 983-986.
9. Pálka I, Ormándi K, Gaál S, Boda K, Kahán Z (2015). Casting-type calcifications on the mammogram suggest a higher probability of early relapse and death among high-risk breast cancer patients. *Acta Onco.* 46: 1178-1183.
10. James JJ, Evans AJ, Pinder SE, Macmillan RD, Wilson ARM, et al. (2003). Is the presence of mammographic comedo calcification really a prognostic factor for small screen-detected invasive breast cancers. *Clinical Radiology.* 58: 54-62.