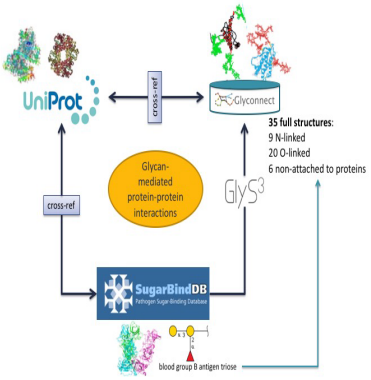
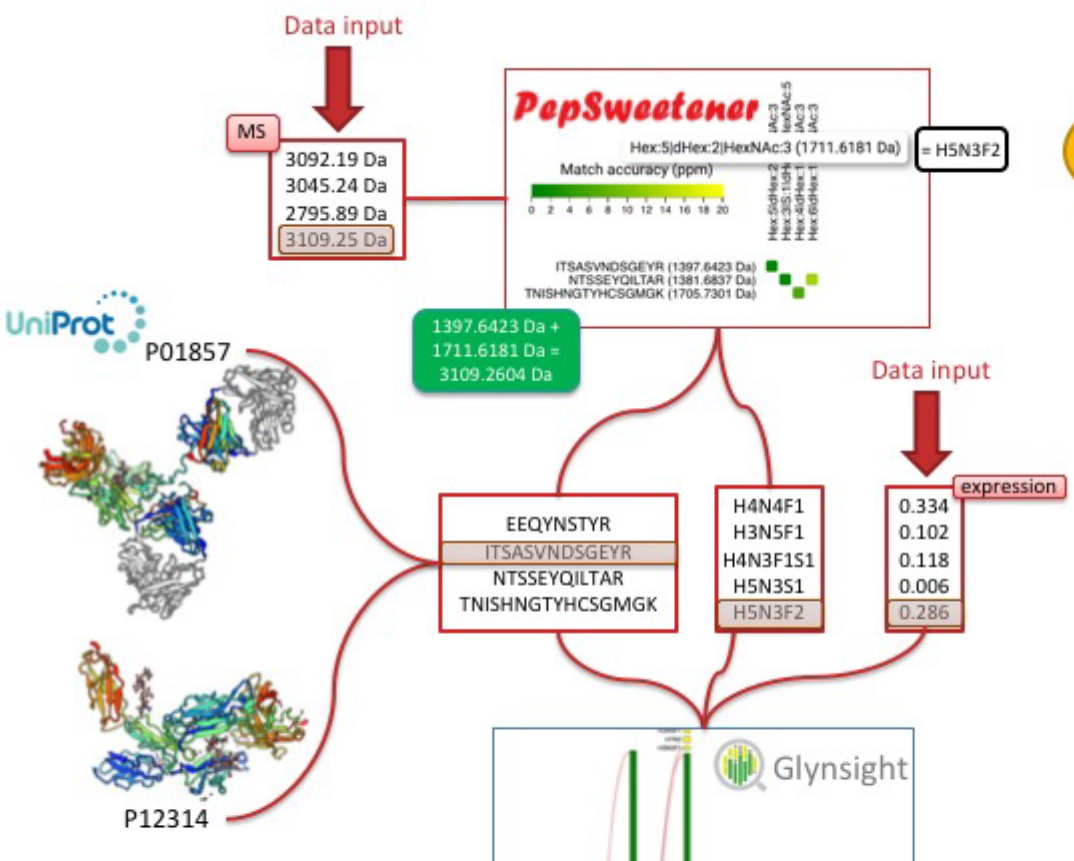


Integrative Tools in Practice

Description

The collection of tools and databases described in detail in the previous sections can be grouped into two categories. They can either be dedicated to solving a specific question or be used in an integrative way in several applications. In the context of exploring and understanding the biological functions where glycans are involved, toolboxes are required to navigate, investigate and correlate data. One such a facility is offered by the crosslinks of databases such as GlyS3, SugarBindDB, GlyConnect and the respective crosslinks to UniProt. A user can seek to establish the consistency of interactions taking place at the cell surface. In the following, three possible use cases are brought to the practitioner.

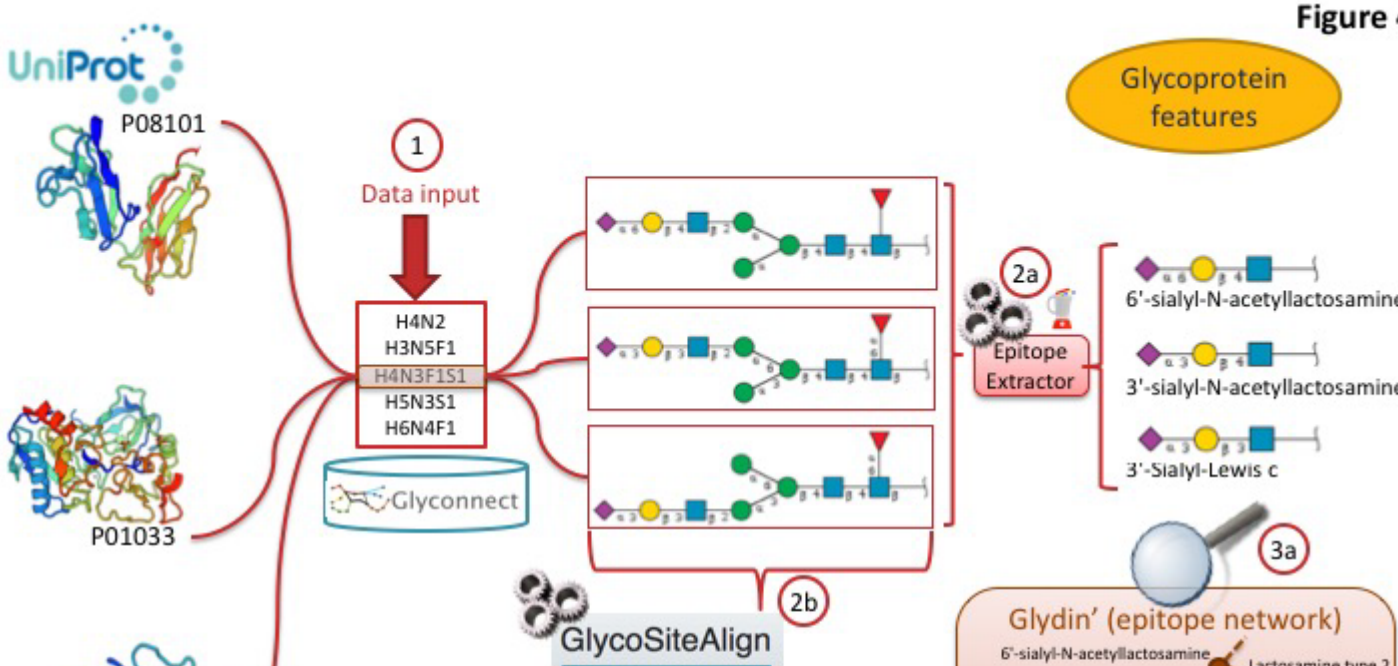
From MS to glycoprotein features.
This toolset is designed to match the expected boost of glycoproteomics (glycan composition at specific sites on complex mixtures of glycoproteins) data that is currently just reaching high throughput level. An example of how to integrate some of these dedicated tools for extracting glycoprotein features from MS data is shown below



From Mass Spectrometry data to glycoprotein profile. A typical scenario of the possible combination of PepSweetener and Glynsight to support the manual annotation of MS1 mass spectra of intact N-glycopeptides and integrate quantitative information when available. Users can process MS1 Spectra using PepSweetener to identify all the possible N-glycan compositions on a single human protein. Entire glycopeptide masses are broken into the respective contributions of the peptide and the glycan masses. Glynsight can be used to identify specific glycosylation patterns. The procedure can be repeated with a second protein and Glynsight will automatically generate the differential analysis of glycan profiles on the proteins. The integration with Glyconnect leads to displaying the potential glycan structures known to match the differentially expressed monosaccharide compositions.

Predominant precursor masses in the MS spectra can be input into PepSweetener. This software supports the manual annotation of intact glycopeptides, using custom web visualization regardless of the instrument that produced the data. An interactive heat-map chart displays the results; it features the combined mass contributions of theoretical (usually tryptic) peptides and attached glycan compositions. The variations in tile colours correspond to ppm deviations from the query precursor mass. Annotation can be refined through glycan composition filtering, sorting by mass and tolerance, and checking MS-MS data consistency via an in silico peptide fragmentation diagram (in-house fragmentation tool common with that of UniCarb-DB). PepSweetener is mainly designed as a complement or extension to software being developed for automatic analysis of glycoproteomics MS data and avoiding their dependency on a set workflow or type of instrument. The outcome of this study will guide the presentation of the Glycomics@ExPASy toolbox towards a more informative and instructive section on MS-based glycoproteomics data analysis tools.

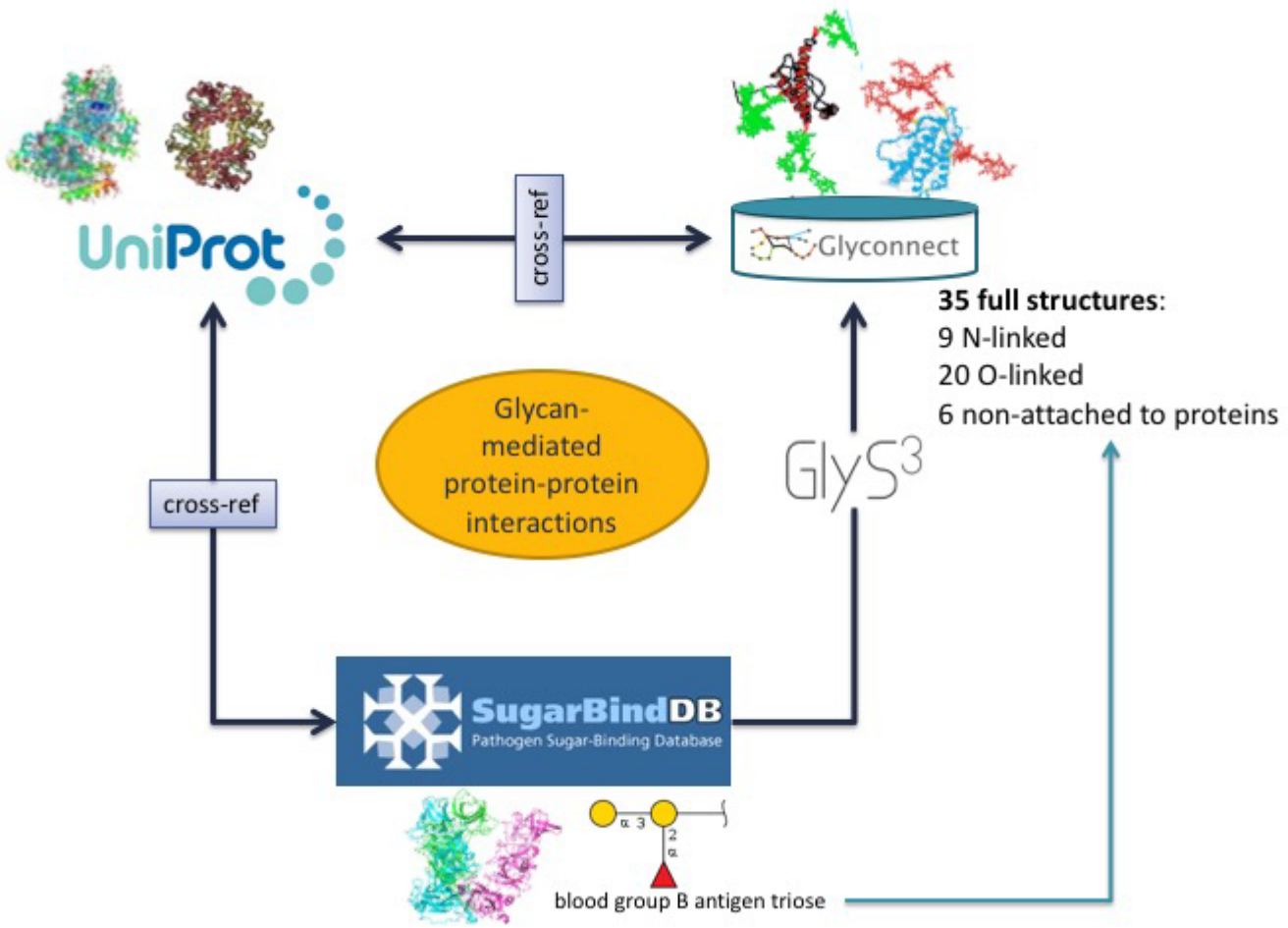
Exploring glycoprotein features.
Current global glycome profiling experiments generate one or more set(s) of glycan compositions and structures with their respective expression on a protein, in a tissue or a cell. Tools and databases in Glycomics@ExPASy can be combined to explore distinctive glycan features that characterise glycoproteins as shown in figure 3. In this case, the entry point of the workflow is GlyConnect to which a list of glycan compositions is submitted. The GlyConnect search tool will retrieve the possible related glycan structures and the proteins that have been reported to have these compositions/structures attached and stored in the databases. A conceptual map displays the results; the compositions sit in the middle and connect glycan structures and associated glycoproteins, respectively on the right and the left sides on the Figure below.



From composition to glycoprotein features. An interactive way of extracting glycoprotein features from glycan compositions combining published data and ad hoc tools. A list of compositions is input in GlyConnect, which retrieves all the proteins reported as having these compositions attached to them (on the left) and reported glycan structures corresponding to this composition (on the right) annotated in this knowledge base. Glycan structures can be further processed to extract contained glycan epitopes using EpitopeXtractor. Glycoepitope results can be mapped on Glydinâ??, an interactive epitope network. Glydinâ?? aggregates glycan epitopes from four different sources (databases and literature reviews) and provides links to the original information. When epitopes are taken from SugarBindDB, further information on the pathogens can be browsed.

This visualization is well suited for understanding the potential relations between proteins and glycans. Activating the integrated EpitopeXtractor function provides a selection of glycan structures.

Glycan-mediated protein-protein interactions.
Using another combination of tools and databases in Glycomics@ExPASy, potential correlations between a glycan-binding protein (GBP) of a pathogen a host glycoprotein and a glycan structure can be made.
In the first scenario, the starting point is a glycoepitope recognised by a specific GBP, a bacterial lectin described in SugarBindDB.

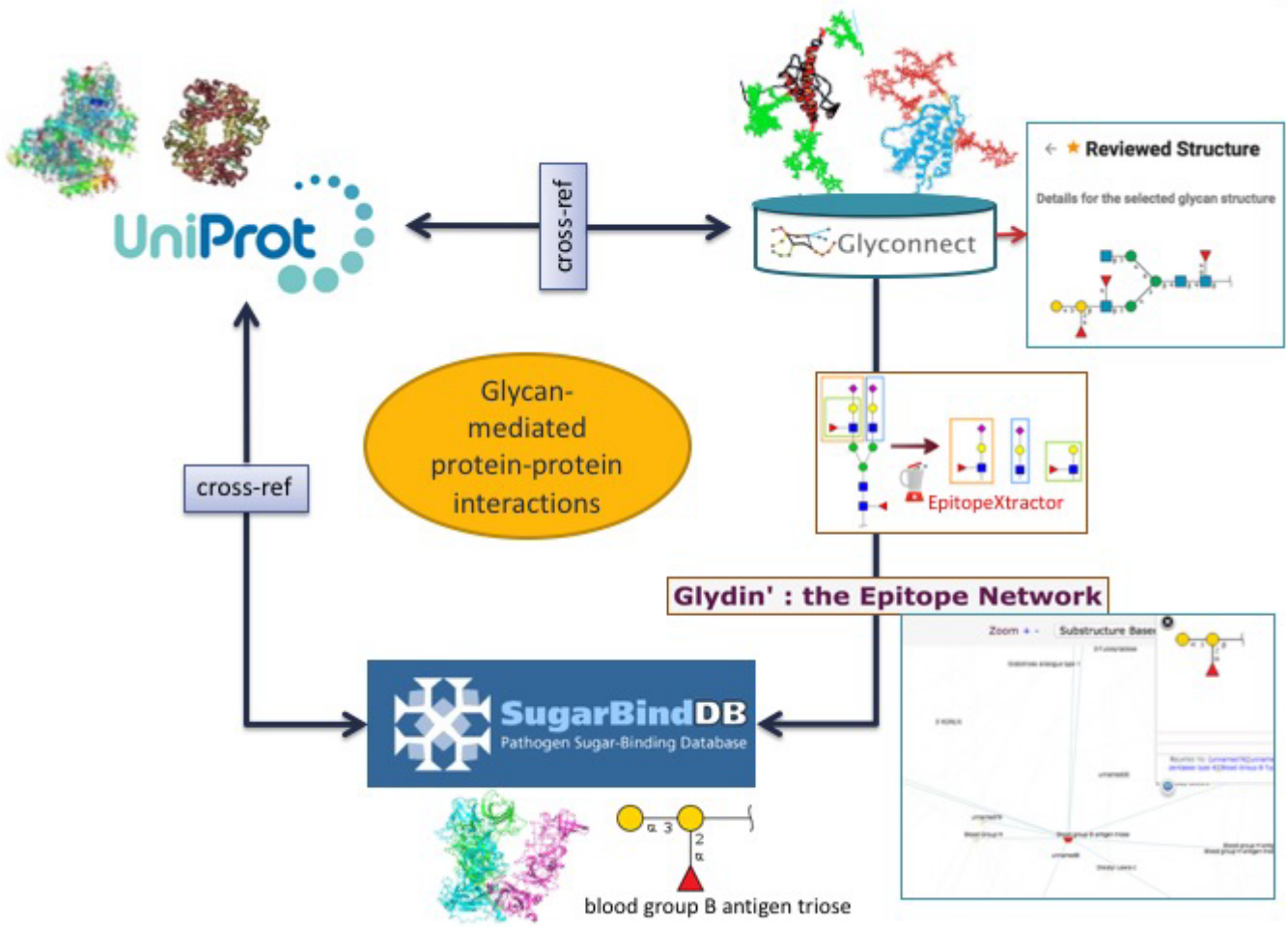


The blood group B antigen triose illustrates this point. A binding event in this database is always formed by A pair composed of a GBP/lectin and a glycoepitope part of a glycan present on the host surface defines a binding event in the database. Whenever possible, further information of a GBP/lectin is available via cross-reference to UniProt. The glycoepitope can be used as an input of the GlyS3 substructure search tool to match the full structures stored in GlyConnect that contain this specific ligand. The list of glycan structures retrieved by GlyS3 can be explored in GlyConnect that reports relationships between glycans and glycoproteins.

The second scenario starts from a glycan structure in GlyConnect and relies on its reported relationships with glycoproteins.

Glycan mediated protein-protein interactions

The figure shows an example of a reviewed N-linked glycan structure. GlyConnect also offers the option of running EpitopeXtractor to generate a selection of glycoepitopes contained in this starting glycan. Leveraging the binding data in SugarBindDB, the obtained glycoepitopes can be associated with a collection of GBPs/lectins that recognize one or more of these glycoepitopes. In the end, the workflow allows the selection of GBPs that could possibly interact with the glycoproteins on which the starting glycan has been reported to be attached. Cross-references of both glycoproteins in GlyConnect and GBPs/lectins in SugarBindDB to UniProt can be used to further rationalise potential interacting partners.



Glycan mediated protein-protein interactions. This figure shows how a new hypothesis on glycan mediated protein-protein interaction can be built using published data in Glyconnect and SugarBindDB: (a) In this scenario glycan binding protein (GBP) is selected in SugarBind. The information on the glycan ligand recognized by the GBP is used to perform a substructure search on all the structures in Glyconnect with the GlyS3 glycan substructure search tool. The structures identified by GlyS3 are used in Glyconnect to create a list of target proteins that can interact with the initial GBP. In the example of the blood group B antigen triose, there are 35 full structure types in GlyConnect that contain this glycoepitope (b) In this scenario a glycoprotein in GlyConnect is selected with its list of associated glycan structures. Glycans are processed with EpitopeXtractor to single out all the glycan epitopes contained. The Glydinâ?? interactive map of structurally related glycoepitopes helps visualising the potential common substructures in the complete set of glycoepitopes. Then, extracted epitopes are used in SugarBind to identify all the reported GBPs that can possibly interact with the initial glycoprotein. In this example, the VP1 capsid protein of the Norwalk virus is known to bind the blood group B antigen triose. Note that protein structures shown above UniProt and those shown above GlyConnect are not related to the example but simply illustrating the difference in the information that is stored on the unglycosylated protein in contrast with the stored information on intact glycoproteins

Category

1. News