Glycomics goes visual and interactive

Alessandra Gastaldello

Abstract

Glycomics@ExPASy the glycomics tab of the Swiss Institute of Bioinformatics server (www.expasy.org/glycomics) was created in 2016 to centralise web-based glycoinformatics resources developed within an international network of glycoscientists. The philosophy of this toolbox is to be {glycoscientist AND protein scientist}???friendly with the aim of popularising (a) the use of bioinformatics in glycobiology and (b) the relation between glycobiology and protein-oriented bioinformatics resources. The scarcity of bridging data led us to design tools as interactive as possible based on database connectivity in order to facilitate data exploration and support hypothesis building. The current set of resources is mostly built on top of curated or experimental data relative to glycan structures, glycoproteins, host-pathogen interactions and mass spectrometry data. The information is provided to the user with dedicated services embedded in web applications. Query tools support compositional, substructure, epitope queries to match structures and related features. This presentation will cover the latest release of Glycomics@ExPASy and focus on the newly developed tools most of which are visual and interactive.

Knowledge of glycoproteins, their site-specific glycosylation patterns, and the glycan structures that they present to their recognition partners in health and disease is gradually being built on using a range of experimental approaches. The data from these analyses are increasingly being standardized and presented in various sources, from supplemental tables in publications to localized servers in investigator laboratories. Bioinformatics tools are now needed to collect these data and enable the user to search, display, and connect glycomics and glycoproteomics to other sources of related proteomics, genomics, and interactomics information. We here introduce GlyConnect (https://glyconnect.expasy.org/), the central platform of the Glycomics@ExPASy portal for glycoinformatics. GlyConnect has been developed to gather, monitor, integrate, and visualize data in a user-friendly way to facilitate the interpretation of collected glycoscience data. GlyConnect is designed to accommodate and integrate multiple data types as they are increasingly produced.

Glycomics aims to identify the whole set of functional glycans of glycoconjugates (attached to proteins or lipids) in biological samples. Glycoproteomics aims to characterize the complete structure of all glycoproteins in biological samples, including the glycosylation sites of proteins and the various glycan structures attached to each of these sites. Mass spectrometry (MS) and microarray are high-throughput technologies that are commonly used in glycomics and glycoproteomics, which often result in the generation of large experimental datasets. Bioinformatics approaches play an essential role in automated analysis and interpretation of such data. This unit describes and discusses the computational tools currently available for these analyses, and their glycomics and glycoproteomics applications.

A key point in achieving accurate intact glycopeptide identification is the definition of the glycan composition file that is used to match experimental with theoretical masses by a glycoproteomics search engine. At present, these files are mainly built from searching the literature and/or querying data sources focused on posttranslational modifications. Most glycoproteomics search engines include a default composition file that is readily used when processing MS data. We introduce here a glycan composition visualizing and comparative tool associated with the GlyConnect database and called GlyConnect Compozitor.

It offers a web interface through which the database can be queried to bring out contextual information relative to a set of glycan compositions. The tool takes advantage of compositions being related to one another through shared monosaccharide counts and outputs interactive graphs summarizing information searched in the database. These results provide a guide for selecting or deselecting compositions in a file in order to reflect the context of a study as closely as possible. They also confirm the consistency of a set of compositions based on the content of the GlyConnect database. As part of the tool collection of the Glycomics@ExPASy initiative, Compozitor is hosted at where it can be run as a web application. It is also directly accessible from the GlyConnect database.

Glycomics@ExPASy is the glycomics tab of ExPASy, the server of SIB Swiss Institute of Bioinformatics. It was created in 2016 to centralize web-based glycoinformatics resources developed within an international network of glycoscientists. The hosted collection currently includes mainly databases and tools created and maintained at SIB but also links to a range of reference resources popular in the glycomics community. The philosophy of our toolbox is that it should be {glycoscientist AND protein scientist}-friendly with the aim of (1) popularizing the use of bioinformatics in glycobiology and (2) emphasizing the relationship between glycobiology and protein-oriented bioinformatics resources. The scarcity of data

bridging these two disciplines led us to design tools as interactive as possible based on database connectivity to facilitate data exploration and support hypothesis building.

Glycomics@ExPASy was designed, and is developed, with a long-term vision in close collaboration with glycoscientists to meet as closely as possible the growing needs of the community for glycoinformatics.

This work is partly presented at 4th Glycobiology World Congress 2018, September 17-19, 2018

Alessandra Gastaldello

Proteome Informatics Group-SIB Swiss Institute of Bioinformatics, Switzerland, E-mail: Alessandra.Gastaldello@sib.swiss