

Glycan pattern search

Davide Alocci, Julien Mariethoz and [Frédérique Lisacek](#)

Swiss Institute of Bioinformatics

Abstract. Glycans are branched tree-like molecules composed by building blocks linked together by chemical bonds. The molecular structure of a glycan can be encoded into a direct acyclic graph where each node represents a building block and each edge serves as a chemical linkage between two building blocks. In this context RDF is a possible software solution for storing structures and SPARQL can be directly used to perform a substructure search. Glycan pattern searching is an important database feature for querying structure and experimental databases.

To perform a glycan pattern search, two questions need to be solved: (i) the automatic generation of a relevant SPARQL query and (ii) the import of known glycan structures into a triple store. First we developed a software solution that reads a structure encoded in a widely used standard in glycomics (GlycoCT), and inserts it into a Virtuoso triple store using an ontology that we specially defined for glycan structures. Then we implemented the automatic translation of a pattern into a SPARQL query using the same ontology.

In the end the program is presented as a web interface. The user inputs the glycan pattern encoded in the GlycoCT format and the software retrieves all the matching full structures in the triple store. This software is integrated and operational to search patterns in the appropriate glycan-related databases (e.g., SugarBindDB: sugarbind.expasy.org).