Linking Glycan and Protein Data

Catherine Hayes[1,*], Vincenzo Daponte[1] and Frederique Lisacek[1,2,3]

[1]*Proteome Informatics Group, SIB Swiss Institute of Bioinformatics, Geneva, 1211, Switzerland*

[2]*Department of Computer Science, University of Geneva, Geneva, 1227, Switzerland*

[3]*Section of Biology, University of Geneva, Geneva, 1211, Switzerland.*

Abstract

Glycosylation is a common and important post-translational modification. Using federated queries it is possible to position glycosylation in useful contexts such as glycoproteins, disease or cellular location.

We present example queries to query glycan structures, glycosylated proteins and cellular location using GlySTreeM, GlyConnect and neXtProt.

1. Introduction

Glycan molecules are often found as post-translational modifications on proteins or lipids [1]. They play roles in many biological processes, such as digestion and mucosal protection [2, 3], maintenance of structural integrity of proteins [4], homing of lectins in the inflammation process [5] and the shielding of pathogenic viruses from the human immune system [6].

The biosynthesis of glycans is not template-driven; rather, it can depend on expression of hydrolases, availability of monosaccharide donors or correct protein folding. Therefore, glycan structure elucidation is quite important in understanding the role that glycans play in binding events [1]. However, comprehensive structural analysis is not always possible, and results yield heterogeneous information, from simple monosaccharide composition (monosaccharide analysis) to fully characterised structures (NMR). Liquid chromatography in tandem with mass spectrometry (LC-MS), as a standard methodology yields sequence, topology and linkage information about glycans. Even though it is often supplemented by biosynthetic knowledge and orthogonal techniques such as binding assays there may still be uncertainty in parts of the glycan.

Incomplete structures present challenges to the understanding of the precise role of glycans in binding events; structural uncertainty rules out simple structure matching. The GlySTreeM ontology was developed to describe glycan structures [7] allowing for various levels of ambiguity, in tandem with a method to translate the search input into SPARQL queries, GlycoQL [8]. We present here federated queries that show the power of GlySTreeM when used with GlyConnect and neXtProt, for querying glycosylation data in relation to proteins and disease states.

---

2. Queries

All queries can be accessed at GitHub

2.1. Query 1

From GlySTreeM give me the IDs of all the structures that contain O-linked Core 2 and at least one sialic acid and one fucose residue

2.2. Query 2

From GlySTreeM and federated query with GlyConnect, list all glycan structures associated with IgG and the disease myositis (DOID: 633).

2.3. Query 3

From GlyConnect for protein "Beta-2-glycoprotein 1", list glycosylation sites, glycan ids found on these sites, and from neXtProt show cellular locations.

References

[1] A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, P. H. Seeberger, Essentials of glycobiology (2022). doi:10.1101/9781621824213.

[2] C. Josenhans, J. Müthing, L. Elling, S. Bartfeld, H. Schmidt, How bacterial pathogens of the gastrointestinal tract use the mucosal glyco-code to harness mucus and microbiota: New ways to study an ancient bag of tricks., International journal of medical microbiology : IJMM 310 (2020) 151392. doi:10.1016/j.ijmm.2020.151392.

[3] D. Qu, G. Wang, L. Yu, F. Tian, W. Chen, Q. Zhai, The effects of diet and gut microbiota on the regulation of intestinal mucin glycosylation., Carbohydrate polymers 258 (2021) 117651. doi:10.1016/j.carbpol.2021.117651.

[4] M. R. Wormald, P. M. Rudd, D. J. Harvey, S. C. Chang, I. G. Scragg, R. A. Dwek, Variations in oligosaccharide-protein interactions in immunoglobulin g determine the site-specific glycosylation profiles and modulate the dynamic motion of the fc oligosaccharides., Biochemistry 36 (1997) 1370–1380. doi:10.1021/bi9621472.

[5] A. Hermenean, D. Oatis, H. Herman, A. Ciceu, G. D'Amico, M. C. Trotta, Galectin 1-a key player between tissue repair and fibrosis., International journal of molecular sciences 23 (2022). doi:10.3390/ijms23105548.

[6] Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, Site-specific glycan analysis of the sars-cov-2 spike., Science (New York, N.Y.) 369 (2020) 330–333. doi:10.1126/science.abb9983.

[7] V. Daponte, C. Hayes, J. Mariethoz, F. Lisacek, Dealing with the ambiguity of glycan substructure search., Molecules (Basel, Switzerland) 27 (2021). doi:10.3390/molecules27010065.

[8] C. Hayes, V. Daponte, J. Mariethoz, F. Lisacek, This is glycoql., Bioinformatics (Oxford, England) 38 (2022) ii162–ii167. doi:10.1093/bioinformatics/btac500.