UniCarbKB: a knowledge platform for glycoproteomics

M.P. Campbell[1], J. Mariethoz[2], C.A Hayes[3], M. Thaysen-Andersen[1], P.M. Rudd[4], N.G Karlsson[3], N.H. Packer[1], F. Lisacek[2]

[1]Biomolecular Frontiers Research Centre, Macquarie University, Sydney, Australia
[2]Proteome Informatics Group, SIB Swiss Institute of Bioinformatics, Geneva, Switzerland
[3]Department of Biomedicine, Gothenburg University, Sweden
[4]NIBRT, Dublin, Ireland

Background

Glycans, both in the form of polysaccharides or glycoconjugates are increasingly recognised as being implicated in human health. Glycosylation is probably the most important post-translational modification in terms of the number of proteins modified and the diversity generated. Since glycoproteins, glycolipids and glycan-binding proteins are frequently located on the cell's primary interface with the external environment many biologically significant events can be attributed to glycan recognition. In other words, glycans mediate many protein-protein interactions. In spite of such a central role in biological processes, the study of glycans remains isolated, protein-carbohydrate interactions are rarely reported in bioinformatics databases and glycomics is lagging behind other –omics.

Recent progress in method development for characterising the branching structures of complex carbohydrates has now enabled high throughput technology. Automation then calls for software development. Adding meaning to large data collections requires bioinformatics means. Current glycobioinformatics resources do cover information on the structure and function of glycans, their association with proteins or their enzymatic generation. However, this information is partial, scattered and often inaccessible to non-glycobiologists.

In partnership with expert international research groups we are involved with the development of the UniCarb KnowledgeBase (UniCarbKB), an effort to develop and provide an informatic framework for the storage and the analysis of high-quality data collections on glycoconjugates, including informative meta-data and MIRAGE-compliant annotated experimental datasets (Minimum Information Required for A Glycomics Experiment). UniCarbKB is an initiative designed to support research in systems biology by complementing proteomics with glycomics (Campbell et al., 2011).

Results

UniCarbKB is based on the reengineering of GlycoSuiteDB and EUROCarbDB and built on the foundations of lightweight Java Rails architecture implementing new search features to explore the wealth of new data now available. The new version will be on-line late 2012. The framework adopts agreed standards to store structural and metadata content including the translation of GlycoSuiteDB structure entries into the GlycoCT format offering a comprehensive structure database (Herget et al., 2008). Significant improvements to the data schema have enabled the merger of these two databases in particular the rational adoption of taxonomic, tissue and disease ontologies. Besides, a standard Resource Description Framework (RDF)

representation for carbohydrate structure, biological and bibliographic annotations and experimental evidence will be released in 2013.

More generally, in the framework of UniCarbKB we aim to: 1) organise data to enable user-friendly interaction and querying by adopting standardisation and ontology guidelines; 2) build a platform that will support the inclusion of new data mining tools and connect disparate existent glycobiology resources including the UniCarb-DB annotated glycan structure MS/MS repository; and 3) integrate functional data through cross-linking with sugar-binding information. At this point, the latter is achieved with mapping SugarBind - a database collecting protein-carbohydrate binding pairs associated with initial stages of infection by human pathogens - to GlycoSuiteDB, in the prospect of connecting the pathogen adhesion information to oligosaccharide structures attached to proteins and/or lipids.

Conclusion

UniCarbKB offers a unique approach to access the most comprehensive biocurated overview of existing glycoinformation associated with proteins in a site-specific manner both from the attachment and the recognition perspective. The initiative is driven as a community endeavour to promote data sharing in glycobiology and ensure its future development and growth.

Acknowledgements

References

1. Campbell MP et al. UniCarbKB: putting the pieces together for glycomics research. Proteomics (2011) 11(21): 4117-21.
2. MIRAGE (Minimum Information Required for A Glycomics Experiment):
   http://glycomics.ccrc.uga.edu/MIRAGE/index.php/Main_Page
3. UniCarbKB: http://unicarbkb.org/
4. GlycosuiteDB: http://glycosuitedb.expasy.org/glycosuite/glycodb
5. EUROCarbDB: http://www.eurocarbdb.org/
6. UniCarbDB: http://www.unicarb-db.org/
7. SugarBind: http://sugarbind.expasy.org/sugarbind/
8. Herget S, Ranzinger R, Maass K, Lieth CW. GlycoCT-a unifying sequence format for carbohydrates. Carbohydr Res. (2008) 343(12):2162-71.