

On-line knowledge of glycans involved in infection

Julien Mariethoz¹, Matthew Campbell², Khaled Khatib³, Catherine A. Hayes⁴, Pauline M. Rudd⁵, Niclas G. Karlsson⁴, Nicolle H. Packer², Frederique Lisacek¹

1. Proteome Informatics Group, Swiss Institute of Bioinformatics, Geneva, Switzerland

2. Biomolecular Frontiers Research Centre, Faculty of Science, Macquarie University, Sydney, Australia

3. Glycoinformatics, Inc., Great Falls, Virginia, USA

4. Department of Biomedicine, Gothenburg University, Sweden

5. NIBRT, Dublin, Ireland

Human epithelial cells and secreted fluids display a dense, heterogeneous array of cell specific glycan structures to invading microorganisms. Glycosylation is probably the most important post-translational modification in terms of the number of proteins modified and the diversity generated. 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 branched structures of complex carbohydrates has now enabled higher throughput technology. Automation then calls for software development. Adding meaning to large data collections requires corresponding bioinformatics methods and tools. Current glycobioinformatics resources do cover information on the structure and function of glycans, their interaction with proteins or their enzymatic synthesis. However, this information is partial, scattered and often difficult to get to for 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. UniCarbKB (<http://unicarbk.org>) is an initiative designed to support research in systems biology by complementing proteomics with glycomics (Campbell et al., 2011). It aims to: 1) organise data to enable user-friendly interaction and querying by adopting standardisation and controlled vocabulary guidelines; 2) build a platform that will support the inclusion of new data mining tools and connect disparate existent glycobiology resources; 3) integrate functional data through cross-linking with sugar-binding information. 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.

We show that the bioinformatics tools we are developing in the context of the UniCarbKB initiative can efficiently support analytical strategies. For example, interactions between pathogen lectins and their glycan targets described in the medical and biochemical literature for several decades are collected, standardised, and organised into a searchable format in the SugarBind Database, released by the MITRE Corporation in 2005 until 2008. The database transferred in 2010 to the ExPASy server (<http://sugarbind.expasy.org/sugarbind>) is under major revision and soon extended to cross-link to UniCarbKB entries. This will ease the exploration of glycan structures involved in infection. To support this statement, we will give examples of how specific structures affect the adhesion of different microbes.

We acknowledge Elaine Mullen of the MITRE Corporation for expert advice on the SugarBind database.