BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 21 2014, pages 3131–3133
doi:10.1093/bioinformatics/btu425

Structural bioinformatics

Advance Access publication July 10, 2014

GlycoDigest: a tool for the targeted use of exoglycosidase digestions in glycan structure determination

Lou Gotz[1], Jodie L. Abrahams[2], Julien Mariethoz[1], Pauline M. Rudd[3], Niclas G. Karlsson[4], Nicolle H. Packer[2], Matthew P. Campbell[2,*] and Frederique Lisacek[1,5,*]

[1]Proteome Informatics Group, Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland, [2]Biomolecular Frontiers Research Centre, Macquarie University, North Ryde, NSW 2109, Australia, [3]National Institute for Bioprocessing Research and Training, GlycoScience Group, Dublin, Ireland, [4]Department of Medical Biochemistry and Cell Biology, University of Gothenburg, 40530 Gothenburg, Sweden and [5]Section of Biology, University of Geneva, 1211 Geneva, Switzerland

Associate Editor: Janet Kelso

ABSTRACT

Summary: Sequencing oligosaccharides by exoglycosidases, either sequentially or in an array format, is a powerful tool to unambiguously determine the structure of complex *N*- and *O*-link glycans. Here, we introduce GlycoDigest, a tool that simulates exoglycosidase digestion, based on controlled rules acquired from expert knowledge and experimental evidence available in GlycoBase. The tool allows the targeted design of glycosidase enzyme mixtures by allowing researchers to model the action of exoglycosidases, thereby validating and improving the efficiency and accuracy of glycan analysis.

Availability and implementation: http://www.glycodigest.org.

Contact: matthew.campbell@mq.edu.au or frederique.lisacek@isb-sib.ch

Received on February 17, 2014; revised on April 29, 2014; accepted on June 25, 2014

1 INTRODUCTION

Protein glycosylation plays a key role in a wide range of biological processes. Advances in glycansequencing technology show that glycoproteins and glycolipids exist in many glycosylated variants, or glycoforms, which can differ substantially in their biochemical properties and functions. Glycans are structurally diverse, incorporating a wide range of monosaccharide residues and glycosidic linkages. Given that protein glycosylation is involved in numerous cellular processes and is implicated in disease progression, the ability to accurately characterize glycan structures and substructures and the determination of the function of these modifications, is increasingly important. However, the difficulty associated with oligosaccharide sequencing has resulted in researchers using only their knowledge of the biosynthetic pathways for protein glycosylation to report proposed glycan structures.

Various analytical platforms are available for the sensitive and efficient analysis of *N*- and *O*-glycans. Such strategies include chromatographic and electrophoretic separations (with fluorescence detection), mass spectrometry (MS) and a combination of both. However, each particular methodology has its strengths and weaknesses, especially in terms of the information content (structural features) that can be readily measured. For example, limitations in material, the lack of fragments and resolution/separation of peaks may yield only partial structures whereby the linkage, the identity or sequence of residue(s) cannot be fully identified. The use of orthogonal methodologies for structural and functional glycomics overcomes these constraints but often necessitates the integration of results from multiple complementary tools.

1.1 Exoglycosidase digestions of oligosaccharides

One of the most commonly used methods for determining the sequence and structure of oligosaccharides is the enzymatic analysis of oligosaccharides using highly specific exoglycosidases, either sequentially or in array form. Enzymatic methods have the ability to determine, in many cases unambiguously, full and accurate sequence and linkage information. As described by (Kobata, 1979), they are highly specific to anomeric configuration and often to monosaccharide type and linkage. Some are also specific to other structural features, such as local and distant saccharide branching. This approach has been extensively reported for analysis of fluorescently labeled *N*-linked oligosaccharides (and to a lesser degree for *O*-links) in which treatment with known enzymatic specificities, followed by chromatographic separation and observation of shifts in retention time, is used to validate individual glycan structural features (Guile *et al.*, 1996). Furthermore, exoglycosidases combined with MS have been demonstrated to be effective for yielding structural information (Ali *et al.*, 2012).

GlycoDigest is a tool that simulates the action of these enzymes on released oligosaccharides. It has been developed to assist glycobiologists design mixtures of exoglycosidases that can be used to guide the precise determination of glycan structures.

2 DESIGN AND IMPLEMENTATION

GlycoDigest is available in three formats: (i) stand-alone Java application, (ii) web-based Java applet (http://glycodigest.org/digest.html) and (iii) an integrated feature of UniCarbKB (http://unicarbkb.org/glycodigest) and GlycoBase databases; links are provided for each fully defined structure. The application is written in Java 7.0, and the source code is available at

*To whom correspondence should be addressed.

© The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

https://bitbucket.org/sib-pig. The stand-alone version is a Swing-based application built with the MigLayout layout manager, which can be run on all operating systems. GlycoDigest is available as a Java library that is integrated with UniCarbKB, allowing users to theoretically digest *N*- and *O*-linked glycan structures stored in this curated database. GlycanBuilder (Damerell *et al*., 2012) acts as the interface for all GlycoDigest variants, allowing users to create structures by sequentially adding monosaccharides, modifications or reducing-end markers. The application integrates libraries developed by EUROCarbDB and GlycomeDB including universal formats for encoding of glycan structures (Campbell *et al*., 2014a).

3 EXOGLYCOSIDASE KNOWLEDGEBASE

Careful preparation and formulation of exoglycosidase arrays is critical to accurate glycan structural assignments, and strategies for designing digestions are detailed (Royle *et al*., 2006). Because of the structural complexity of glycans, i.e. degree of branching and steric hindrance, the sequencing of particular motifs is often dependent upon the correct combination of exoglycosidases. For example, the Fuc(α1-3) residue of the Lewis x epitope can only be cleaved after treatment with β-galactosidase. Knowledge of exoglycosidase reactions has been acquired by extensive analytical experience and the creation of suitable experimental databases. GlycoBase is an High Performance Liquid Chromatography (HPLC) retention time database of experimentally characterized glycan structures, which provides comprehensive information on known exoglycosidase digestions (Campbell *et al*., 2008). This wealth of information builds upon the methods used by the reagent array analysis method (Prime and Merry, 1998). From this base, the rules included in GlycoDigest are experimental observations on the behavior and mode of action of exoglycosidases.

4 USING GLYCODIGEST

4.1 GlycoDigest desktop application

The stand-alone release offers a user-friendly interface consisting of three components: (i) glycan sequences can be directly entered into the 'load' panel in genuine or Consortium for Functional Glycomics (CFG) modified, International Union of Pure and Applied Chemistry (IUPAC) condensed or GlycoCT format, or built using the embedded version of GlycanBuilder; (ii) a combination of exoglycosidase enzymes can be selected from the middle panel, and the input structure is processed *in silico* using the rules explained above; (iii) the theoretical results of the enzymatic digestion are returned in GlycoCT format and displayed in the 'Simulation Results' panel (Fig. 1). Graphical representations of the glycan structures can be displayed using the image rendering Application Programming Interface (API), which are linked to UniCarbKB entries.

4.2 GlycoDigest web version

A Java applet version of GlycoDigest uses the same functionality and interface described above. However, the structure encoding entries in Panels 1 and 3 are displayed in the same page as opposed to being displayed in new windows. As the stand-alone version, known structures are linked to UniCarbKB entries.

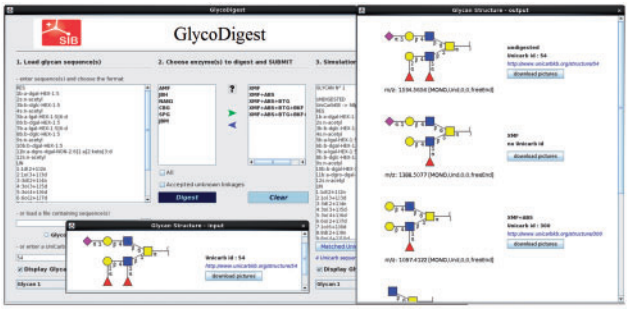

Fig. 1. Interface for the desktop version of GlycoDigest with examples of input/output glycan structures displayed in the essentials of Glycobiology/CFG format

4.3 Integrating GlycoDigest with UniCarbKB

UniCarbKB offers public access to a growing, curated database of information on the glycan structures on glycoproteins. For each *N*- and *O*-link glycan structure with defined linkage information, a 'GlycoDigest' link provides users the option to select a combination of enzymes for simulating exoglycosidase treatment(s). Proposed structures for each undigested and digested glycan resulting from the action of the exoglycosidases can be searched in UniCarbKB (Campbell *et al*., 2014b) for associated published and experimental information, and for combined exoglycosidase-MS/MS characterization of oligosaccharides (Ali et al., 2012) using UniCarb-DB spectra (Hayes et al., 2011). Alternatively, digestions can be run on structures built using GlycanBuilder.

5 CONCLUSIONS

GlycoDigest fills the need for a robust and intuitive means for glycobiologists to simulate and design mixtures of exoglycosidases with which to experimentally validate proposed glycan structures. The tool minimizes any chance of user bias in interpreting what may be ambiguous results and ensures that all structures compatible with the experimental data are considered. The availability of the GlycoDigest Java library will allow bioinformaticians the option of integrating this tool for managing the input and output of exoglycosidase-treated glycan structures.

Funding: Swiss National Science Foundation (31003A_141215); Macquarie University Research Excellence Scheme Postgraduate Scholarship; The Australian National eResearch Collaboration Tools and Resources project (RT016).

Conflict of interest: none declared.

REFERENCES

Ali,L. *et al*. (2012) Structural identification of O-linked oligosaccharides using exoglycosidases and MSn together with UniCarb-DB fragment spectra comparison. *Metabolites*, **2**, 648–666.

Campbell,M.P. *et al*. (2008) GlycoBase and autoGU: tools for HPLC-based glycan analysis. *Bioinformatics 2008*, **24**, 1214–1216.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016

Campbell,M.P. *et al.* (2014a) Toolboxes for a standardised and systematic study of glycans. *BMC Bioinformatics*, **15** (**Suppl 1**), S9.

Campbell,M.P. *et al.* (2014b) UniCarbKB: building a knowledge platform for glycoproteomics. *Nucleic Acids Res*, **42**, D215–D221.

Damerell,D. *et al.* (2012) The GlycanBuilder and GlycoWorkbench glycoinformatics tools: updates and new developments. *Biol. Chem.*, **393**, 1357–1362.

Guile,G.R. *et al.* (1996) A rapid high-resolution high-performance liquid chromatographic method for separating glycan mixtures and analyzing oligosaccharide profiles. *Anal Biochem*, **240**, 210–226.

Hayes,C.A. *et al.* (2011) UniCarb-DB: a database resource for glycomic discovery. *Bioinformatics*, **27**, 1343–1344.

Kobata,A. (1979) Use of endo- and exoglycosidases for structural studies of glycoconjugates. *Anal. Biochem.*, **100**, 1–14.

Prime,S. and Merry,T. (1998) Exoglycosidase sequencing of N-linked glycans by the reagent array analysis method (RAAM). *Methods Mol. Biol.*, **76**, 53–69.

Royle,L. *et al.* (2006) Detailed structural analysis of N-glycans released from glycoproteins in SDS-PAGE gel bands using HPLC combined with exoglycosidase array digestions. *Methods Mol. Biol.*, **347**, 125–143.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 30, 2016