Development of a software suite for the interpretation of glycomics high-throughput MS/MS and CRM-based data

René Ranzinger, Brent Weatherly, Mindy Porterfield, Michael Tiemeyer, Lance Wells, William York, and Michael Pierce

National Center for Biomedical Glycomics, Complex Carbohydrate Research Center, University of Georgia, Athens, GA, USA

http://glycomics.ccrc.uga.edu Email: hawkeye@uga.edu

Ongoing technical advancements have made mass spectrometry (MS) the dominant experimental technique used for the analysis and identification of glycan structures, whether as purified glycans or as complex mixtures of glycans extracted from biological samples. Currently, most interpretation and annotation of the high throughput MS/MS data generated in this context is done manually, due to lack of software support. The freely available tools (e.g. GlycoWorkbench, GlycoPeakfinder, GlycoMOD) and commercial systems (e.g. SimGlycan®) that have been developed over the last decades are not suited for analysis of datasets that include hundreds or thousands of MS^n spectra. Thus, the development of new software tools capable of processing such high throughput datasets and assisting in their interpretation and annotation is crucial to keep up with the rapid pace of technological development and data generation.

We have, therefore, developed the GRITS Toolbox, a modular software suite for the processing, interpretation and storage of glycomics data with a focus on MS data. After loading MS data into the software (as RAW and/or mzXML files), users can invoke various integrated data processing modules, including the Glycomics Elucidation and Annotation Tool (GELATO), which associates spectral features in the data sets with structures supplied by customizable databases. Alternatively, annotations generated by SimGlycan® can be imported into GRITS. An extensive graphical user interface allows the annotated data to be browsed, visualized, manually modified and exported to Excel for further processing. The results of several different experiments can also be merged and displayed side by side to identify differences in the glycosylation patterns of analyzed samples.

The validity of the data annotation depends critically on the quality of the databases that provide structures that are associated with spectral features. Therefore, GRITS provides several of its own N-glycan, O-glycan and glycolipid databases that have been generated using knowledge from a highly curated glycan structure ontology. We have populated this ontology using a web-based curation system, called Qrator (http://glycomics.ccrc.uga.edu/qrator/). The current version of the software suite is freely available via the project website: http://www.grits-toolbox.org.

In addition, software has been developed that utilizes Consecutive Reaction Monitoring MS^2 data (iCRM) that allows the identification and quantification of O-linked glycans from a complex mixture released from glycoproteins or cells. The next steps are to automate the collection and processing of these data and incorporate them into the GRITS software suite, as well as to expand coverage to N-linked glycans and glycolipids using shotgun proteomics. This technology has been applied to identify potential diagnostic and prognostic cancer markers in sera

Please view the sections in numerical order

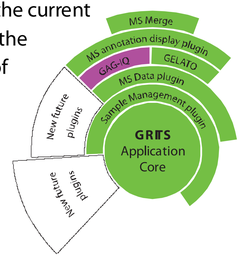
1

Ongoing technical advancements have made tandam mass spectrometry (MS/MS) the dominant experimental technique used to identify structures of purified glycans as well as complex mixtures of glycans extracted from biological samples. Currently, most interpretation of the generated high throughput MS/MS data are done manually, due to the lack of software support. The freely available tools and commercial systems that have been developed over the last decades are not suited for analysis of large datasets that include hundreds or thousands of MS/MS spectra. **GRITS toolbox is a freely available software suite for the processing of these datasets.**

http://www.grits-toolbox.org

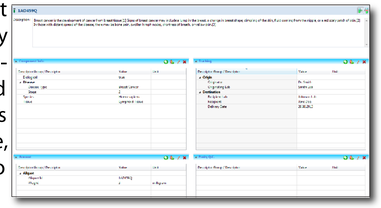
Extendable Software Platform 2

GRITS toolbox is a platform independent software based on the Eclipse software framework (https://www.eclipse.org) and the Java programming language. Although the current version of GRITS is primarily focused on the processing, interpretation and storage of glycomics MS data, its **modular architecture based on plugins** can reduce development time by allowing **easy extension of the software** with new functionality while reusing existing program components.



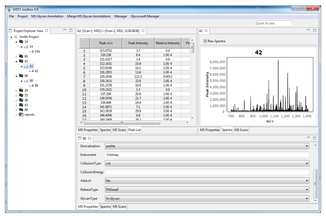
Sample Information 3

To allow a comprehensive overview of the processed data when exchanging GRITS projects with colleagues, the software allows **storing of meta-data about the analyzed sample**. This can happen as a free text description or a set of defined terms for origin, species, disease, data about sample amount and quality control. Providing this optional information and using the defined terms simplifies data exchange, and data submission to repositories.



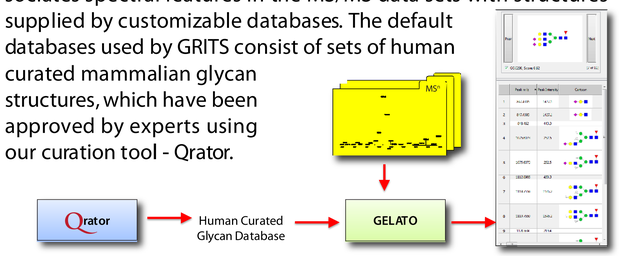
Mass Spectrometry Data 4

MS data in **mzXML format** can be loaded into GRITS toolbox. Thermo RAW file format is supported as well but auto generation of corresponding mzXML files using a public web service is restricted to files smaller than 10 MB. Uploaded files are copied and become part of the GRITS project to facilitate full access for colleagues when sharing projects. MS data can be **displayed as peak lists or in a spectral viewer**. Meta data about sample preparation and instrument setup can be stored as well.



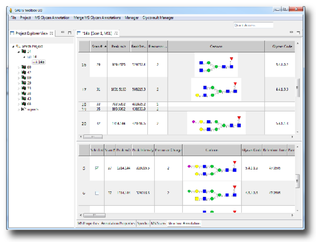
GELATO 5

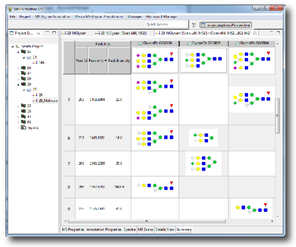
The Glycomics Elucidation and Annotation Tool (GELATO) is the **integrated MS/MS annotation module** within GRITS, which associates spectral features in the MS/MS data sets with structures supplied by customizable databases. The default databases used by GRITS consist of sets of human curated mammalian glycan structures, which have been approved by experts using our curation tool - Qrator.



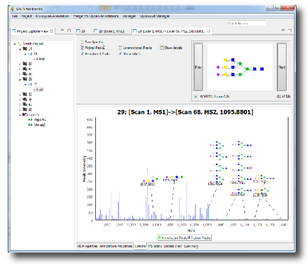
MS Annotation 6

After automatic annotation of the MS/MS data the annotation results can be **manually inspected using diverse display options**. The MS^1 annotation overview shows the MS^2 precursor annotation of each sub spectrum in the upper part of the application and the different candidate structures for each peak in the lower part.





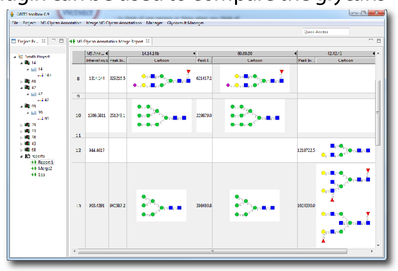
Double clicking the MS^2 precursor annotation, opens the fragment summary view to show all **precursor candidate structures side by side** with the MS^2 fragment ions supporting each structure. The same procedure can be used to study MS^n annotations.



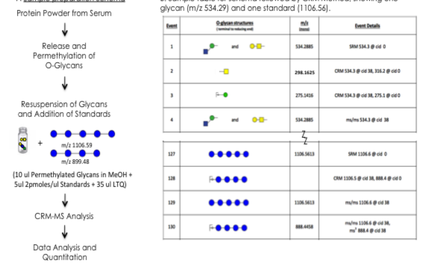
The annotation **spectrum viewer** shows the raw spectrum and highlights peaks that are annotated or are not annotated. In addition the viewer shows both the annotation of **MS^1 spectra with intact structures** and annotation of **MS^n spectra with fragment ions**.

MS Merge 7

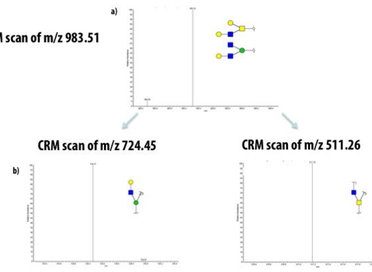
The MS annotation merge plugin provides a visual **side-by-side comparison** of the user curated **MS annotations of multiple samples** or MS runs. The plugin can be used to compare the glycans structures present in one sample with the glycans in another sample. Intensity values are provided alongside the structures, making it easy to find changes in the glycan expression level for different samples.



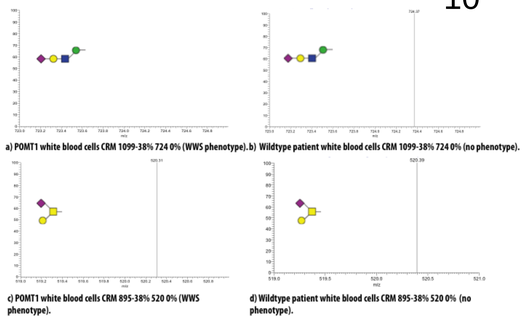
Schematic for Intelligent Consecutive Reaction Monitoring (iCRM) method for the quantitative analysis of O-glycans. 8



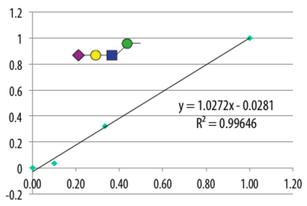
Application of the iCRM method for the analysis of an isobaric O-glycan structure. 9



Analysis of O-glycans from Walker-Warburg Syndrome patient using iCRM (DBP). A blood test for CMD 10



Linearity of the O-man tetrasaccharide from mouse brain following serial dilutions. 11



Linearity of response from serial dilution of O-Man tetrasaccharide structure. **O-Glycans were prepared from mouse brain and analyzed via CRM for generation of diagnostic fragment at varying concentrations (1, .33, .1). Theoretically the line should be y=1x+0. We observed y=1.03x-0.03 with an R^2=.996. Data was normalized to a spiked in control and the intensity of the most concentrated sample was set to 1.**

iCRM 12

- Targeted approach to glycan assignment and quantification
- Front-end can be automated
- Separation of Isobaric Glycans for Quantification
- Quantification is in MSn (IceCream Software)
- Automated data extraction for Assignment
- Automated Normalized Quantification
- Automated Comparative Glycomics