A FREELY AVAILABLE SOFTWARE PLATFORM FOR HIGH THROUGHPUT GLYCOMICS MS DATA INTERPRETATION

GRITS: GRits Is The Solution

William York, Brent Weatherly, Khalifeh Al Jadda, Mindy Porterfield, Shahnawaz Khan, Ki Tae Myoung, Michael Tiemeyer and René Ranzinger

NIGMS-funded National Center for Biomedical Glycomics
Complex Carbohydrate Research Center
University of Georgia
Athens GA

National Center for Biomedical Glycomics
NIGMS

GRITS

Software design
Testing
Promotion
Support

Mindy Porterfield

Michael Tiemeyer

Project managment
Software design
Software development

René Ranzinger

Shahnawaz Khan
Brent Weatherly

Software development
GUI development

Khalifeh Al Jadda
Ki Tae Myoung

Software development
GUI development

The Problem

- Glycobiology requires **high throughput** Glycomics analysis
  - Currently, tandem MS is the most effective high-throughput glycomics technology
  - Each glycomics MS run can generate hundreds of spectra, which must be processed and annotated
  - For the data to be useful, it must be associated with meta data describing the biological sample and its manipulation as well as the techniques and parameters used to generate the data
- **Currently avalable tools are not capable of effectively processing such large amounts of data** - Better and more fully automated annotation tools are required
  - Tools like *GlycoWorkbench*, while extremely valuable, require considerable interaction with the human analyst, making their direct application to the annotation of high throughput data impractical
  - The currently available tools for annotation of high throughput MS data are not applicable to glycomics data, are very expensive and/or function as “Black Boxes” with output whose validity is difficult to evaluate
- **Methods for glycoanalysis are diverse and constantly evolving**, requiring a nimble approach to the development of data processing software
  - Software should be able to process data produced by new methods
  - Previously developed algorithms, visualization schemes, data formats and databases should be reused when possible

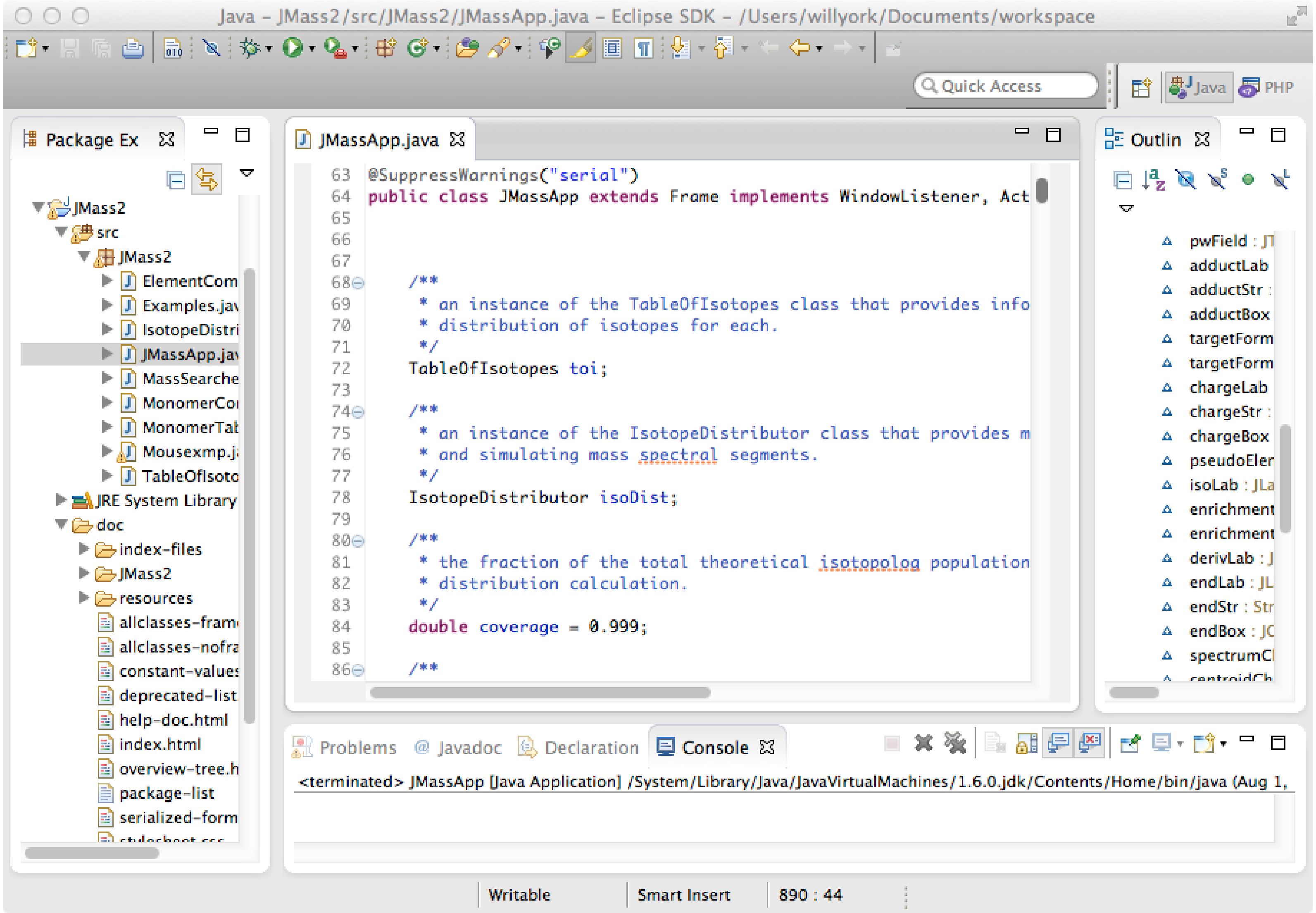
GRITS Is The Solution

- Provides a common, modular framework for glycomics analysis
- Based on the ***Eclipse Platform*** for development of *Integration Ware*
  - Supports development of interactive **plug-ins** to support data processing for diverse analytical methods and protocols
  - Provides a **standard workbench user interface** with well-developed tools (e.g., widgets and interactive tables) to graphically represent and manipulate data and enter diverse types of information
  - Facilitates the use and sharing of **common data representation formats**
  - Provides **a resource management** model
- **Allows rapid software development, as each new plug-in can interact with previously developed plug-ins to utilize the data processing and/or visualization methods that they implement**

http://www.eclipse.org/eclipse/

Many Software Developers Are Already Familiar with Eclipse

National Center for Biomedical Glycomics
NIGMS



```
@SuppressWarnings("serial")
public class JMassApp extends Frame implements WindowListener, Act



    /**
     * an instance of the TableOfIsotopes class that provides info
     * distribution of isotopes for each.
     */
    TableOfIsotopes toi;

    /**
     * an instance of the IsotopeDistributor class that provides m
     * and simulating mass spectral segments.
     */
    IsotopeDistributor isoDist;

    /**
     * the fraction of the total theoretical isotopolog population
     * distribution calculation.
     */
    double coverage = 0.999;

    /**
```

```
<terminated> JMassApp [Java Application] /System/Library/Java/JavaVirtualMachines/1.6.0.jdk/Contents/Home/bin/java (Aug 1,
```

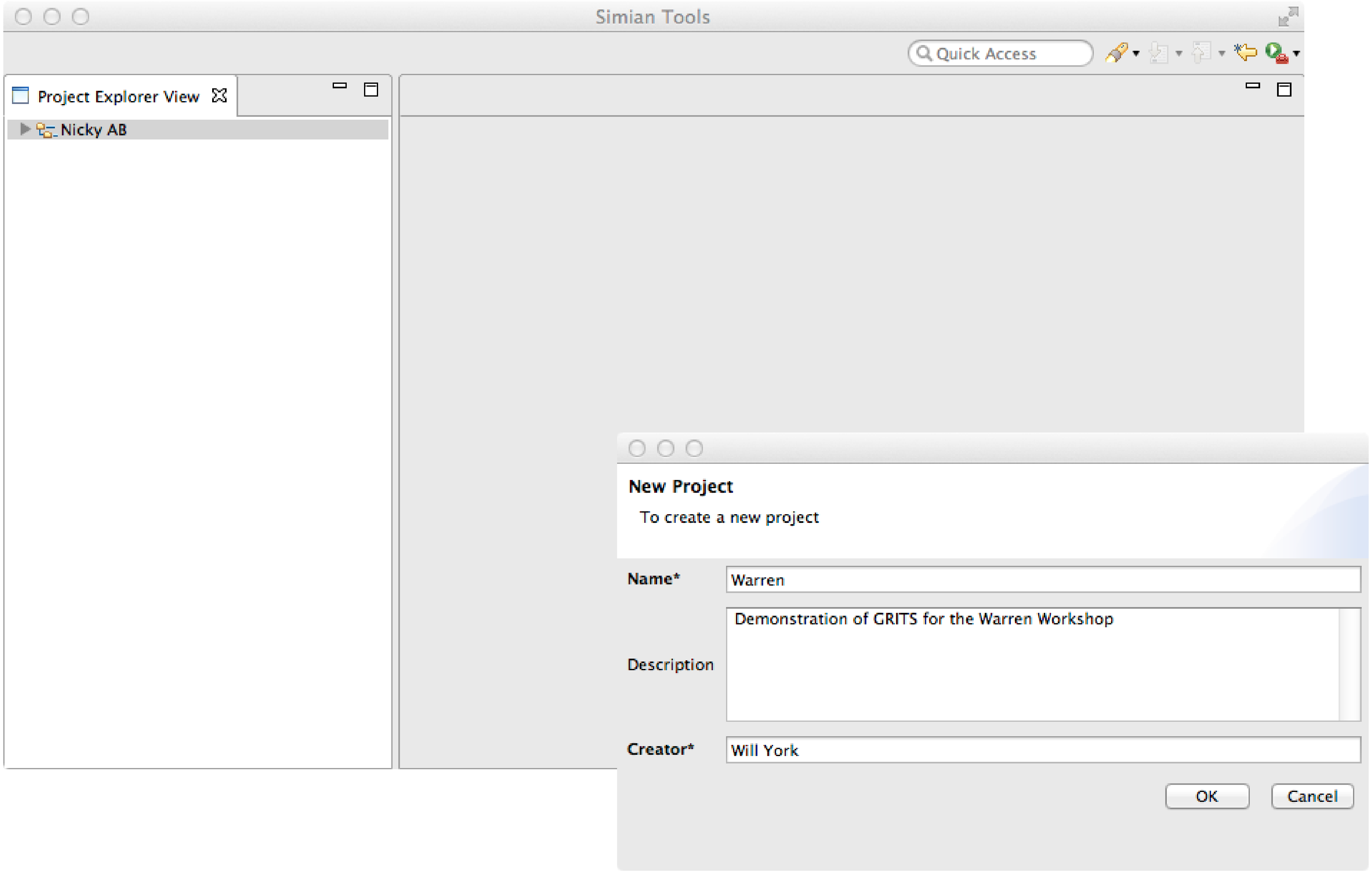
GRITS Is The Solution

Organize Results as Elements of Distinct Projects

- ✓ Attach Multiple Samples to a Project
- ✓ Attach Analytical Data Sets to Each Sample
- ✓ Annotate Analytical Data Using GELATO
- ✓ Verify/Edit Annotations Manually
- ✓ Export Annotated Reports
    - Individual Data Set Annotations
    - Side-By-Side Comparisons of Data Sets

GRITS

Organize Data as Elements of a Project



New Project

To create a new project

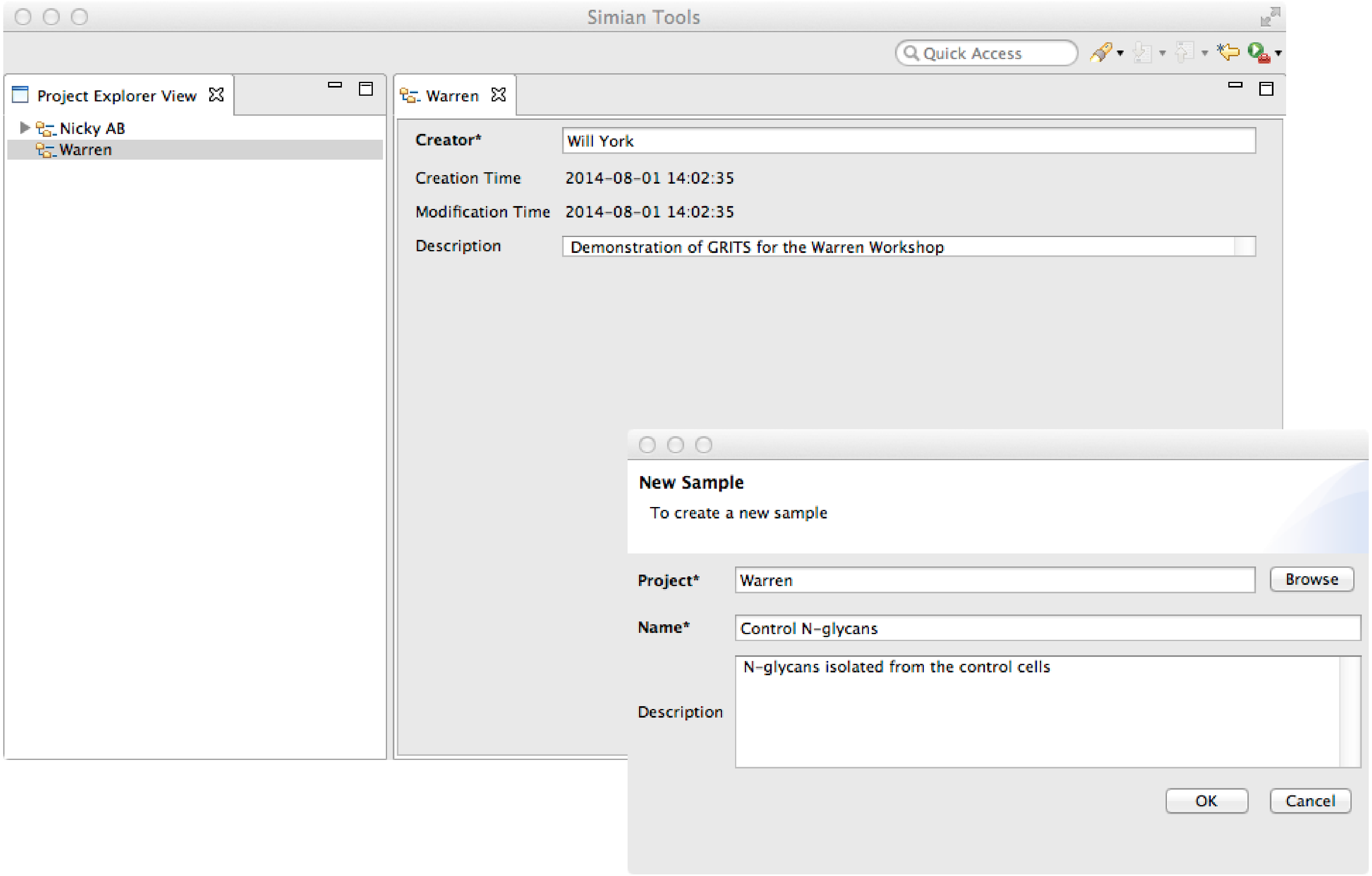
Name* Warren

Description Demonstration of GRITS for the Warren Workshop

Creator* Will York

GRITS

Add a Sample to a Project



Creator* Will York

Creation Time 2014-08-01 14:02:35

Modification Time 2014-08-01 14:02:35

Description Demonstration of GRITS for the Warren Workshop

New Sample

To create a new sample

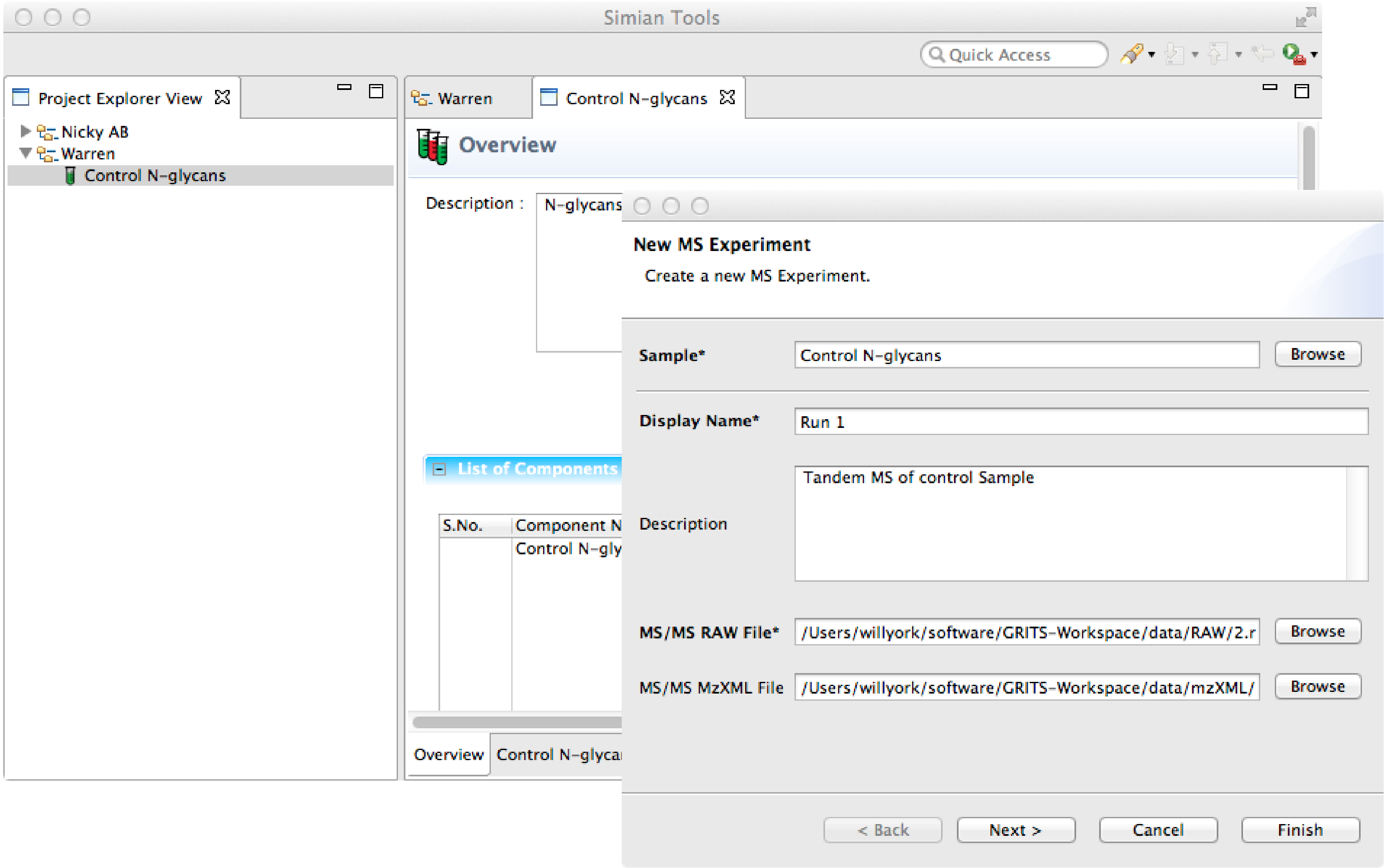
Project* Warren

Name* Control N-glycans

Description N-glycans isolated from the control cells

GRITS

Add Analytical Data for a Sample



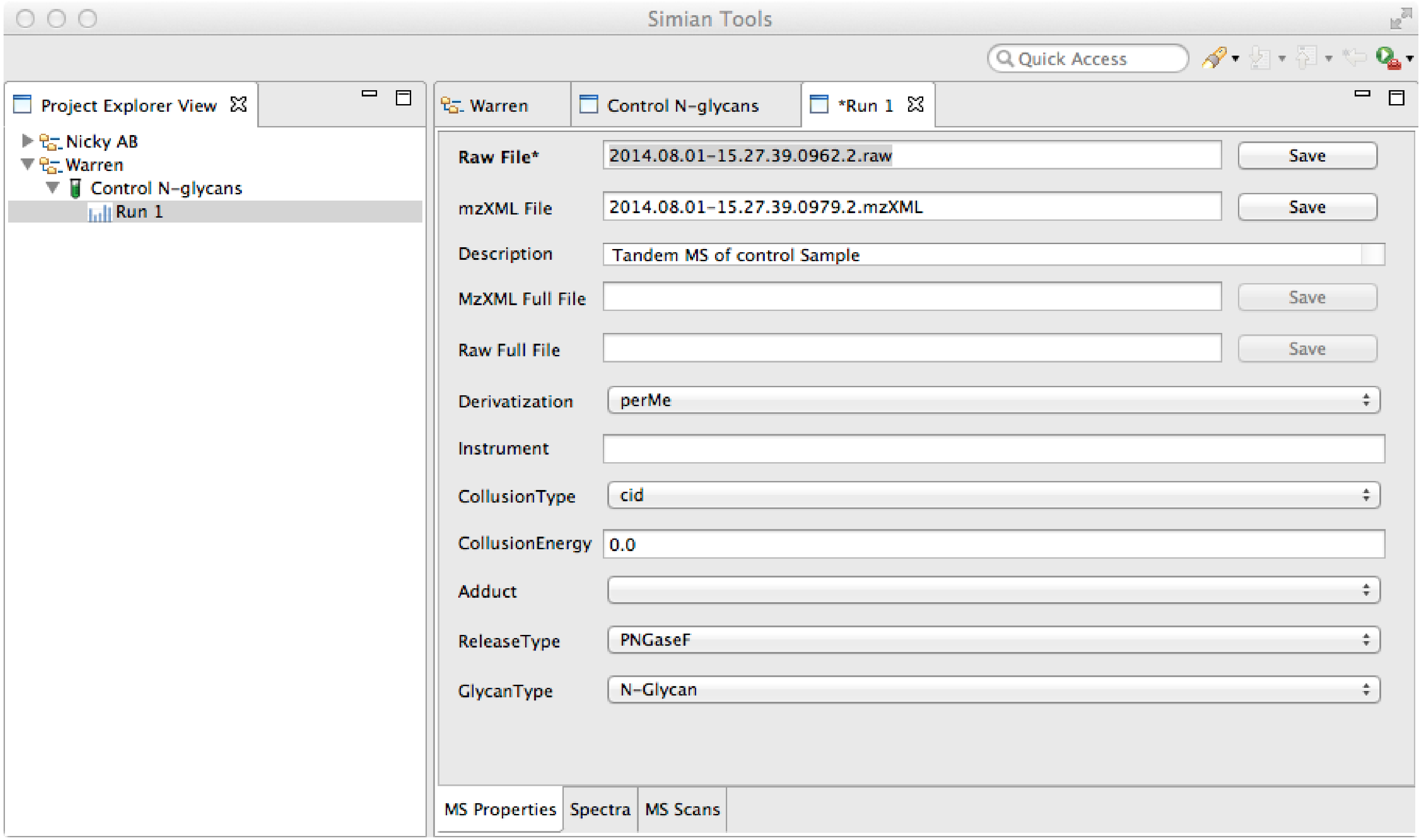
New MS Experiment

Create a new MS Experiment.

Sample*	Control N-glycans
Display Name*	Run 1
Description	Tandem MS of control Sample
MS/MS RAW File*	/Users/willyork/software/GRITS-Workspace/data/RAW/2.r
MS/MS MzXML File	/Users/willyork/software/GRITS-Workspace/data/mzXML/

GRITS

Add Analytical Data for a Sample



Raw File*: 2014.08.01-15.27.39.0962.2.raw

mzXML File: 2014.08.01-15.27.39.0979.2.mzXML

Description: Tandem MS of control Sample

MzXML Full File:

Raw Full File:

Derivatization: perMe

Instrument:

CollusionType: cid

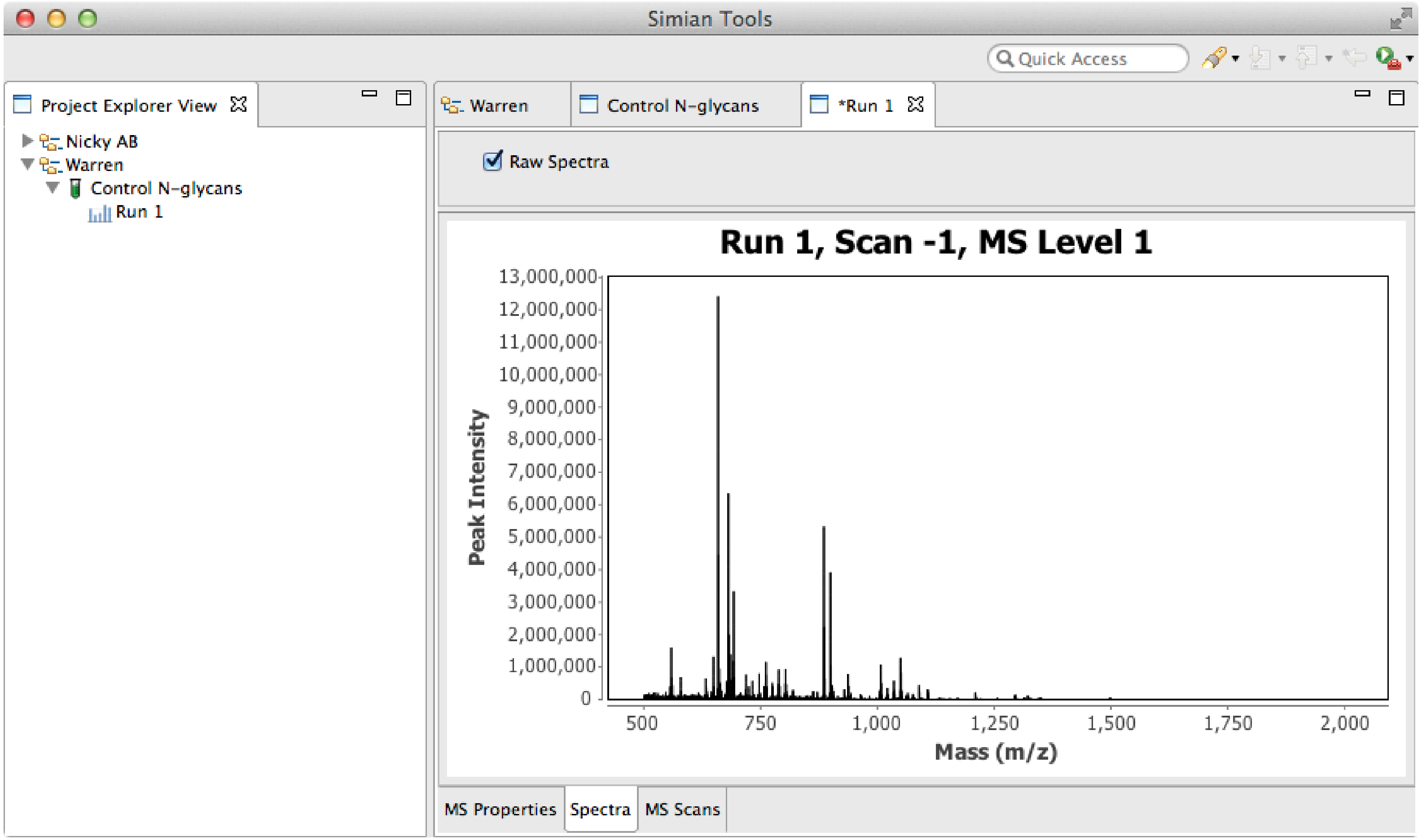
CollusionEnergy: 0.0

Adduct:

ReleaseType: PNGaseF

GlycanType: N-Glycan

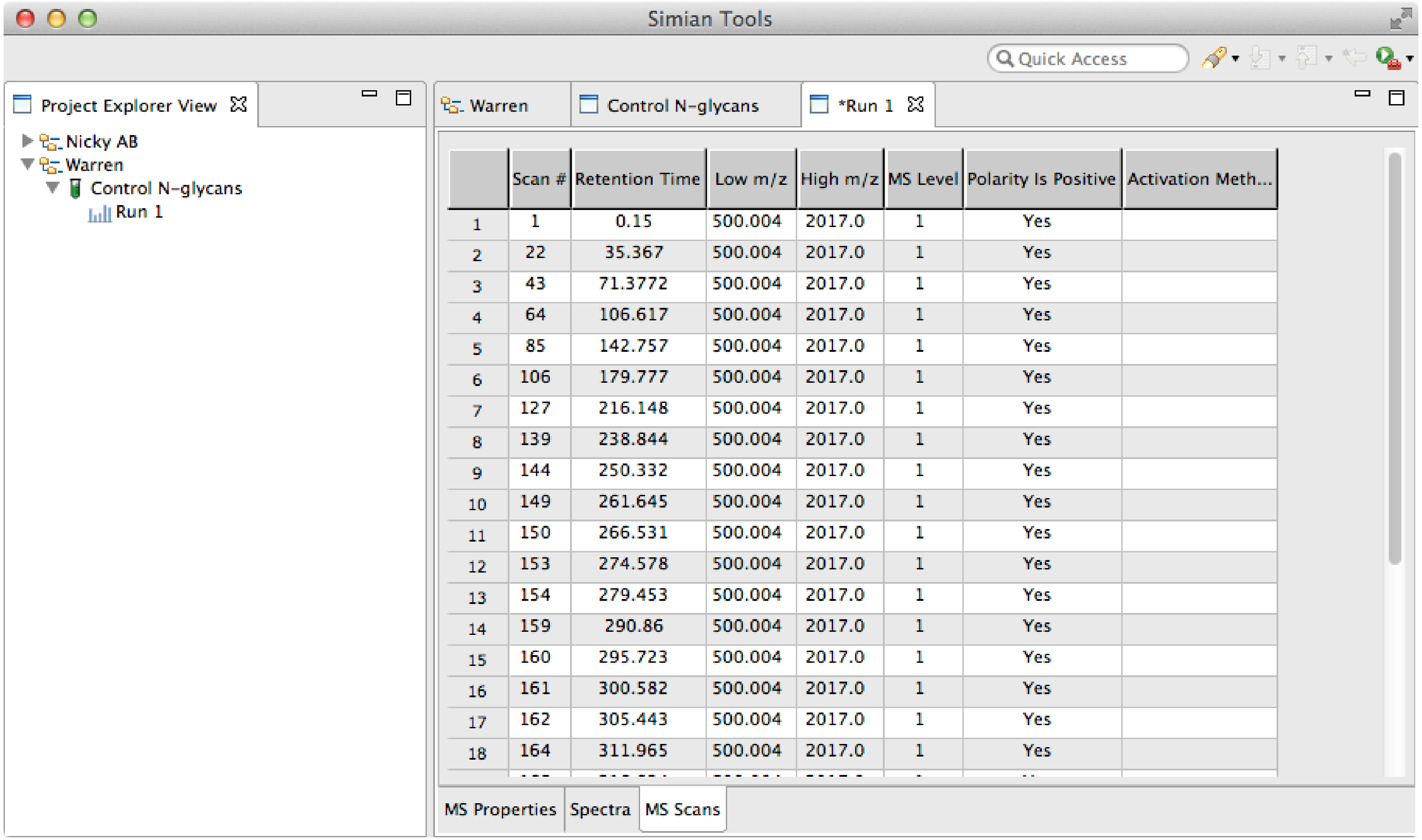
GRITS
Add Analytical Data for a Sample



☑ Raw Spectra

GRITS

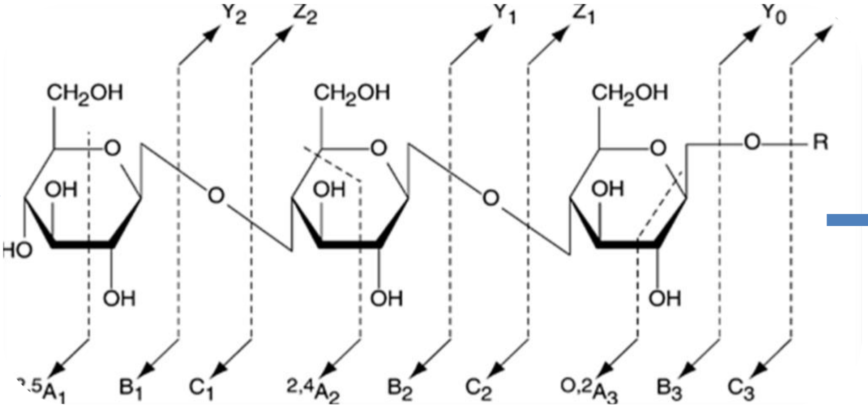
Add Analytical Data for a Sample

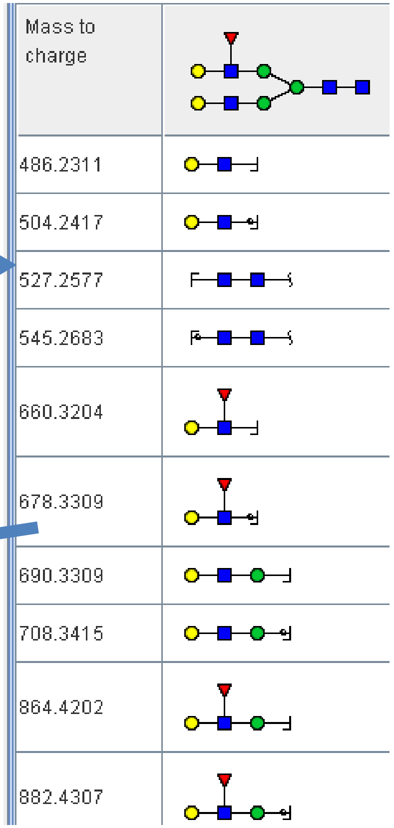


	Scan #	Retention Time	Low m/z	High m/z	MS Level	Polarity Is Positive	Activation Meth...
1	1	0.15	500.004	2017.0	1	Yes	
2	22	35.367	500.004	2017.0	1	Yes	
3	43	71.3772	500.004	2017.0	1	Yes	
4	64	106.617	500.004	2017.0	1	Yes	
5	85	142.757	500.004	2017.0	1	Yes	
6	106	179.777	500.004	2017.0	1	Yes	
7	127	216.148	500.004	2017.0	1	Yes	
8	139	238.844	500.004	2017.0	1	Yes	
9	144	250.332	500.004	2017.0	1	Yes	
10	149	261.645	500.004	2017.0	1	Yes	
11	150	266.531	500.004	2017.0	1	Yes	
12	153	274.578	500.004	2017.0	1	Yes	
13	154	279.453	500.004	2017.0	1	Yes	
14	159	290.86	500.004	2017.0	1	Yes	
15	160	295.723	500.004	2017.0	1	Yes	
16	161	300.582	500.004	2017.0	1	Yes	
17	162	305.443	500.004	2017.0	1	Yes	
18	164	311.965	500.004	2017.0	1	Yes	

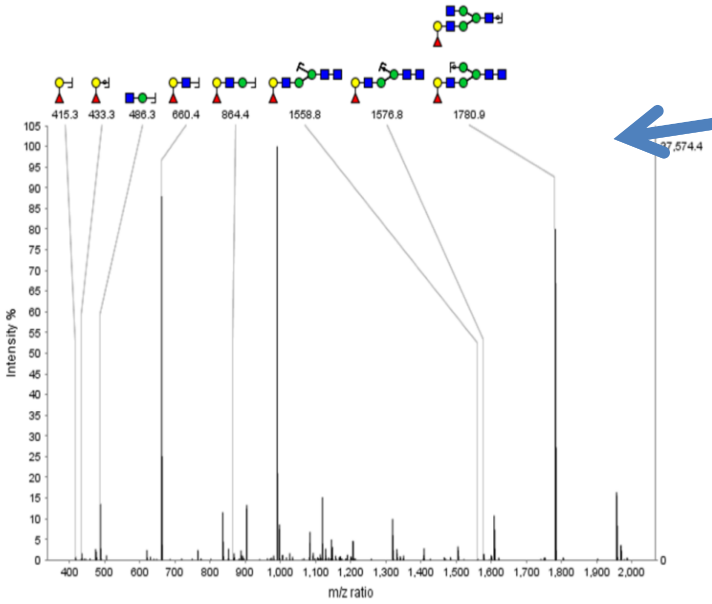
Automated MS-Annotation: Match Theoretical Glycan Ion Fragments With Experimental Mass Spectra

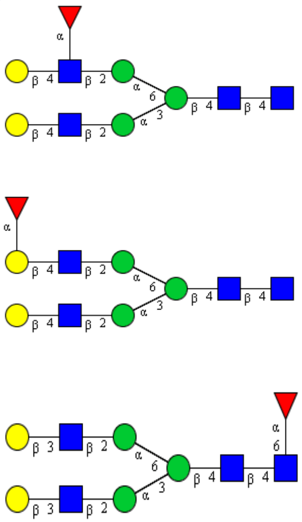
Database structures





Mass to charge	
486.2311	
504.2417	
527.2577	
545.2683	
660.3204	
678.3309	
690.3309	
708.3415	
864.4202	
882.4307	





Problems with Commercial MS Annotation Software

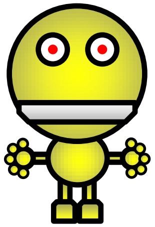
Not High-Throughput

Results are a dead end – Provides NO ability to

- Export annotated spectra
- Extract signal intensities
- Capture fragment intensities
- Manipulate cartoons
- Visualize the entire profile

Quality of structural assignment limited by the databases

- Redundancies
- Omissions
- Inaccuracies



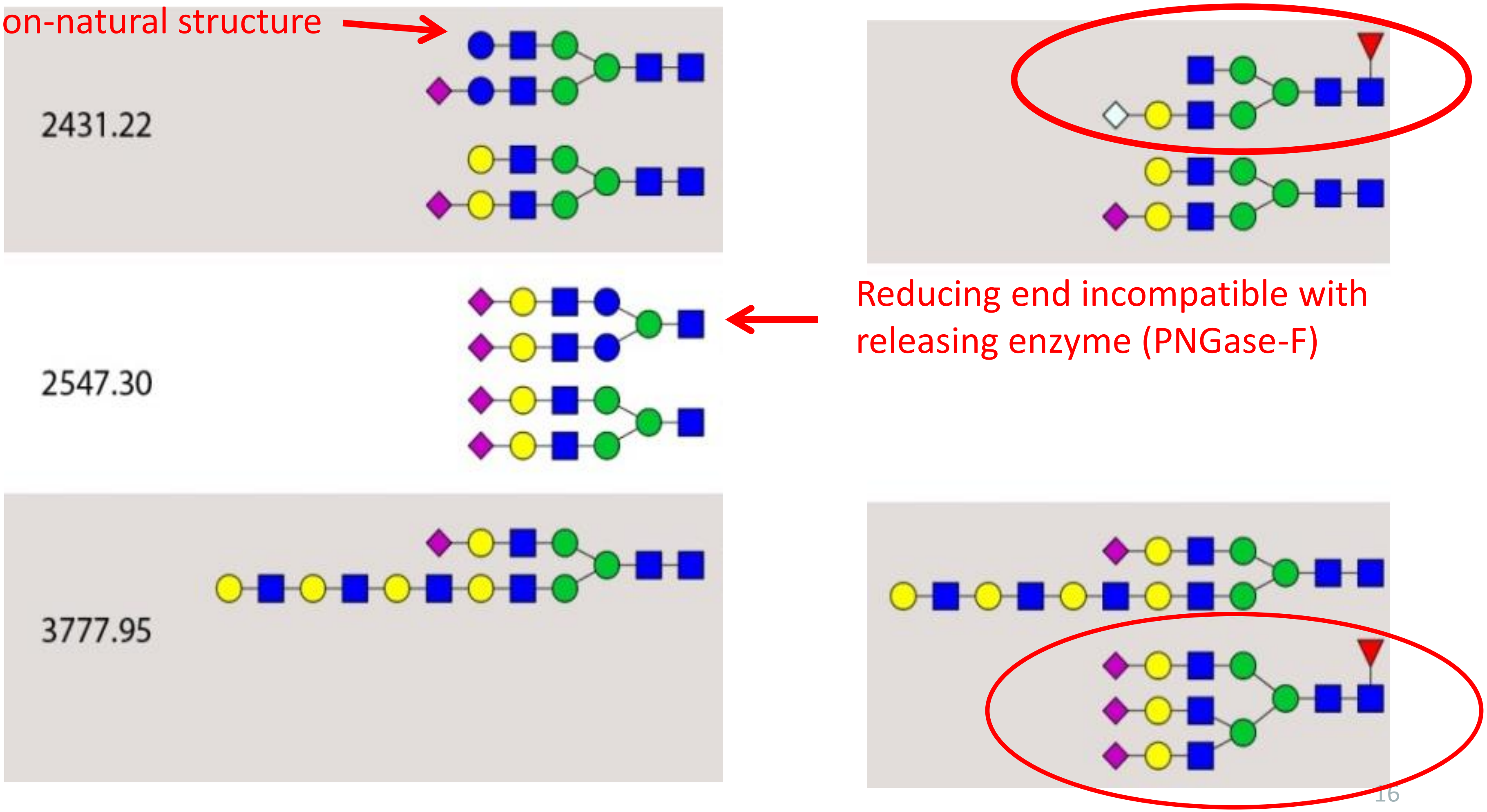
Database Bot

Bot transforms databases into SG compatible

- *Rene Ranzinger* developed Database Bot – Generates databases compatible with SimGlycan using structures from a highly curated source
- *Matthew Eavenson*, *Krys Kochut* and *Will York* developed **Qrator** - a curation tool enabling experts to assemble valid glycan structures into gold standard databases
- *Mike Tiemeyer* curated 888 N-linked and 152 O-Linked structures using **Qrator**
- Bot was used to make "**Sweetie-N** and **Sweetie-O**"
- Can make other specialized databases

Advantages of Customized Databases

Mass SimGlycan™ SweetieN



Non-natural structure

Reducing end incompatible with releasing enzyme (PNGase-F)

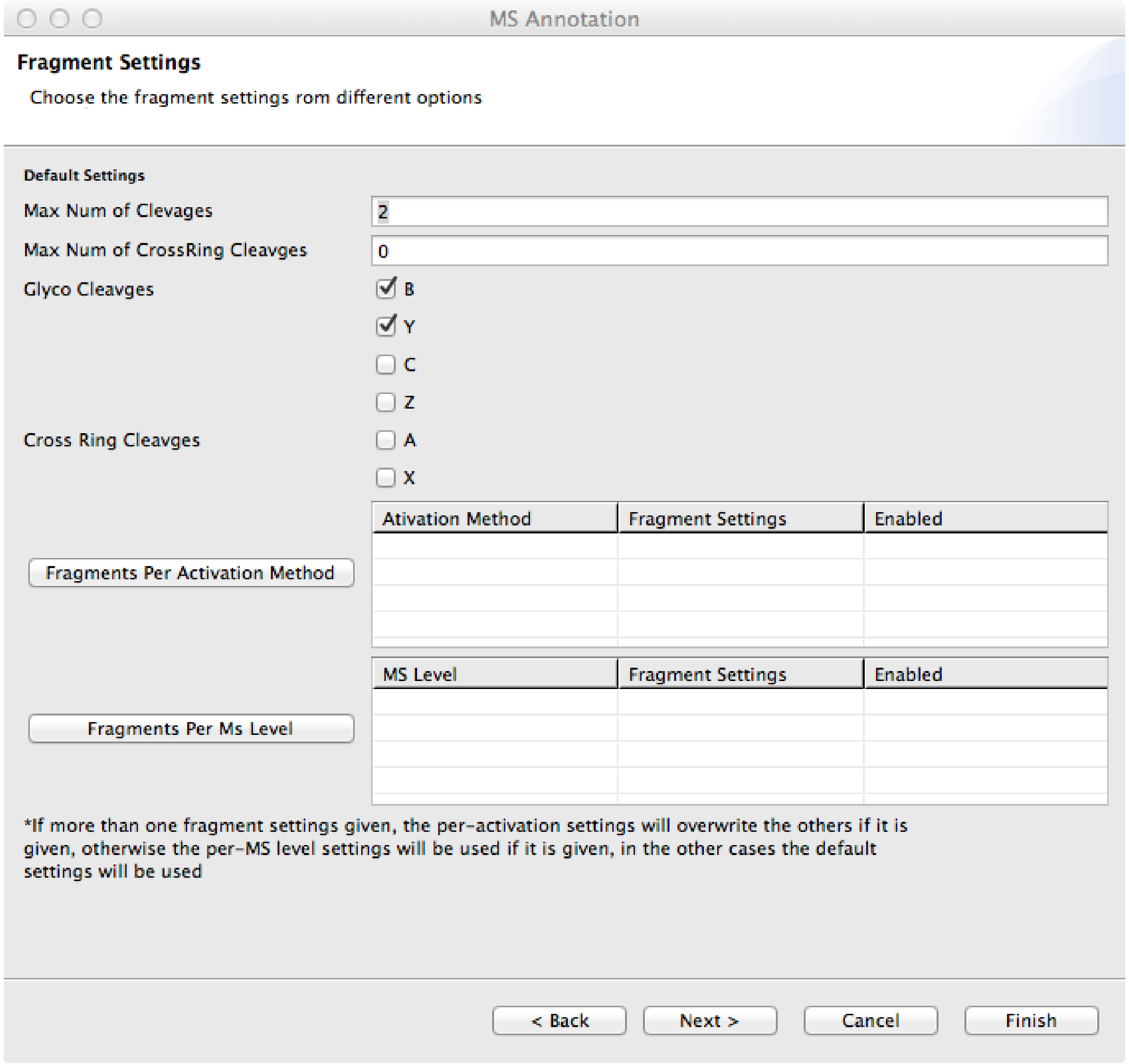
GELATO - Glycan ELucidation and Annotation Tool

- Developed by *Khalifeh Al Jadda* and *Rene Ranzinger*
- In-silico generation of theoretical fragment ions for each glycan supplied by a customized database – Uses **GlycoWorkbench*** API and fragmentation engine
- Matches theoretical fragment ions to fragment ions detected in real mass spectra
- Provides scores based on the number of matched ions and their intensity – allows ranking of glycan annotations of mass spectral features
- Results are comparable to those produced by SimGlycan
- No expensive license fees - **Free Distribution**
- Probability-based scoring under development for future versions

*A.Ceroni, K. Maass, H.Geyer, R.Geyer, A.Dell and S.M.Haslam (2008) *Journal of Proteome Research*, **7**: 1650

GRITS

Invoke GELATO to Annotate Analytical Data



MS Annotation

Fragment Settings

Choose the fragment settings rom different options

Default Settings

Max Num of Clevages: 2

Max Num of CrossRing Cleavges: 0

Glyco Cleavges: ☑ B ☑ Y ☐ C ☐ Z

Cross Ring Cleavges: ☐ A ☐ X

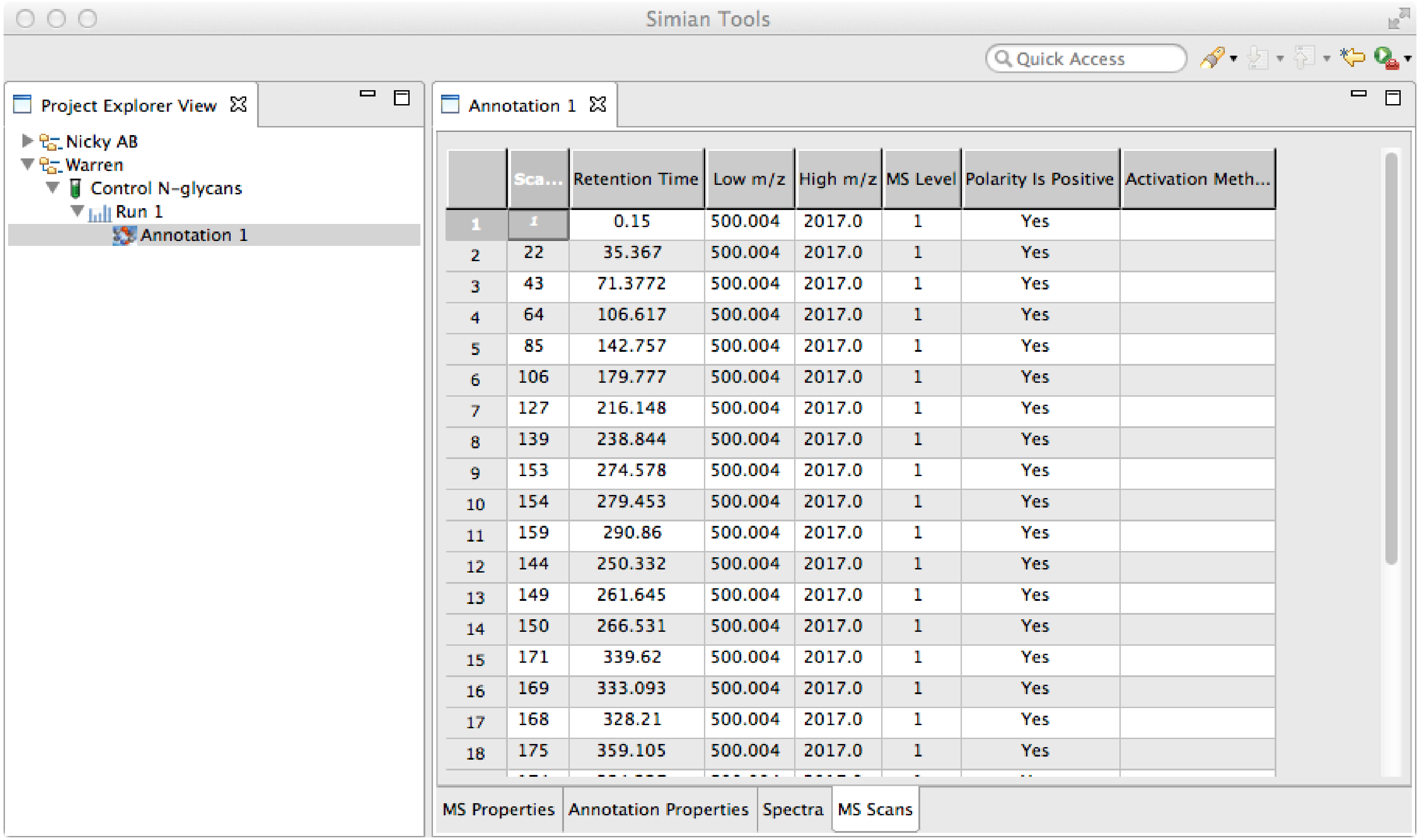
Ativation Method	Fragment Settings	Enabled

MS Level	Fragment Settings	Enabled

*If more than one fragment settings given, the per-activation settings will overwrite the others if it is given, otherwise the per-MS level settings will be used if it is given, in the other cases the default settings will be used

GRITS

Invoke GELATO to Annotate Analytical Data

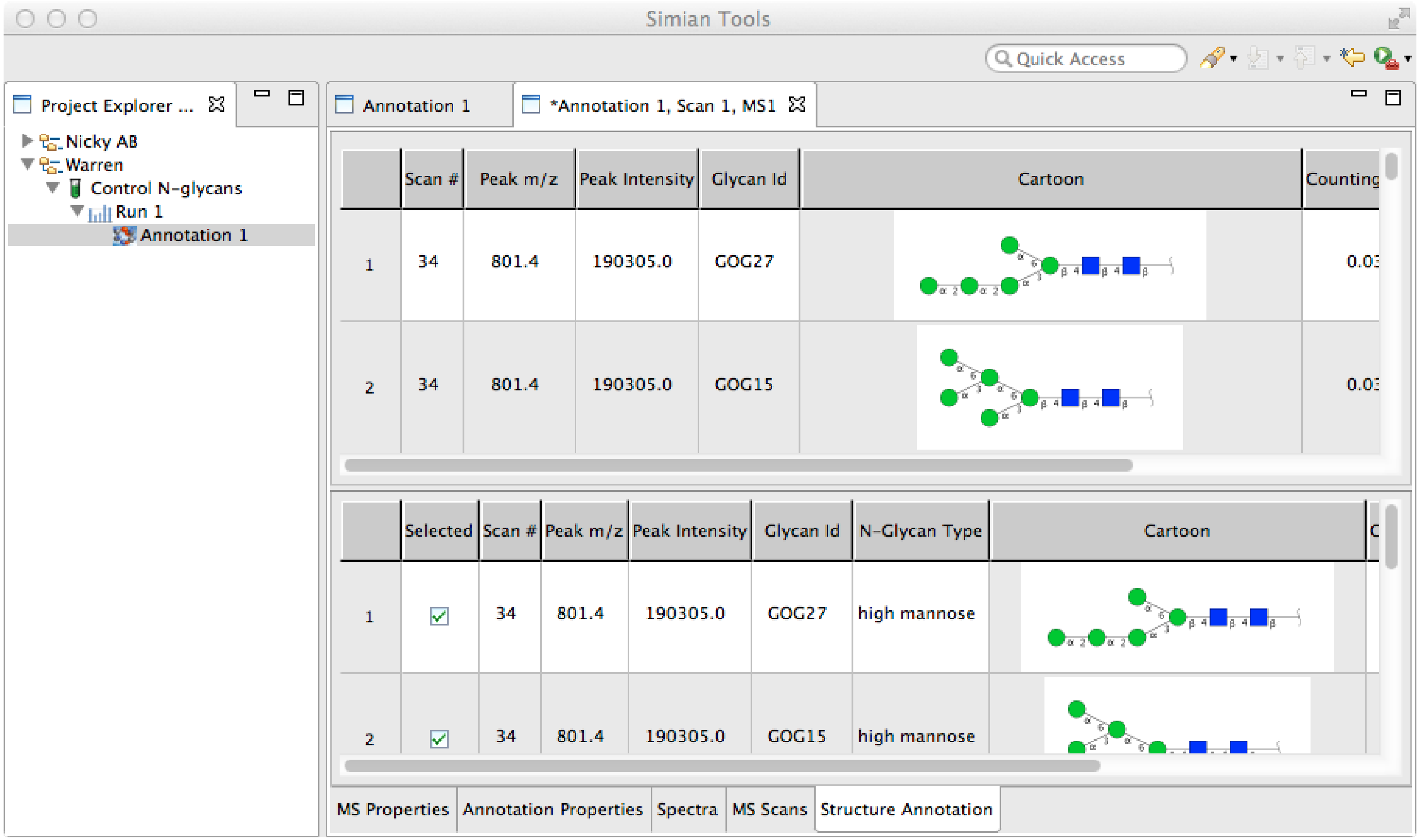


- Nicky AB
- Warren
  - Control N-glycans
    - Run 1
      - Annotation 1

	Sca...	Retention Time	Low m/z	High m/z	MS Level	Polarity Is Positive	Activation Meth...
1	1	0.15	500.004	2017.0	1	Yes	
2	22	35.367	500.004	2017.0	1	Yes	
3	43	71.3772	500.004	2017.0	1	Yes	
4	64	106.617	500.004	2017.0	1	Yes	
5	85	142.757	500.004	2017.0	1	Yes	
6	106	179.777	500.004	2017.0	1	Yes	
7	127	216.148	500.004	2017.0	1	Yes	
8	139	238.844	500.004	2017.0	1	Yes	
9	153	274.578	500.004	2017.0	1	Yes	
10	154	279.453	500.004	2017.0	1	Yes	
11	159	290.86	500.004	2017.0	1	Yes	
12	144	250.332	500.004	2017.0	1	Yes	
13	149	261.645	500.004	2017.0	1	Yes	
14	150	266.531	500.004	2017.0	1	Yes	
15	171	339.62	500.004	2017.0	1	Yes	
16	169	333.093	500.004	2017.0	1	Yes	
17	168	328.21	500.004	2017.0	1	Yes	
18	175	359.105	500.004	2017.0	1	Yes	

GRITS

Manual Verification of Annotations



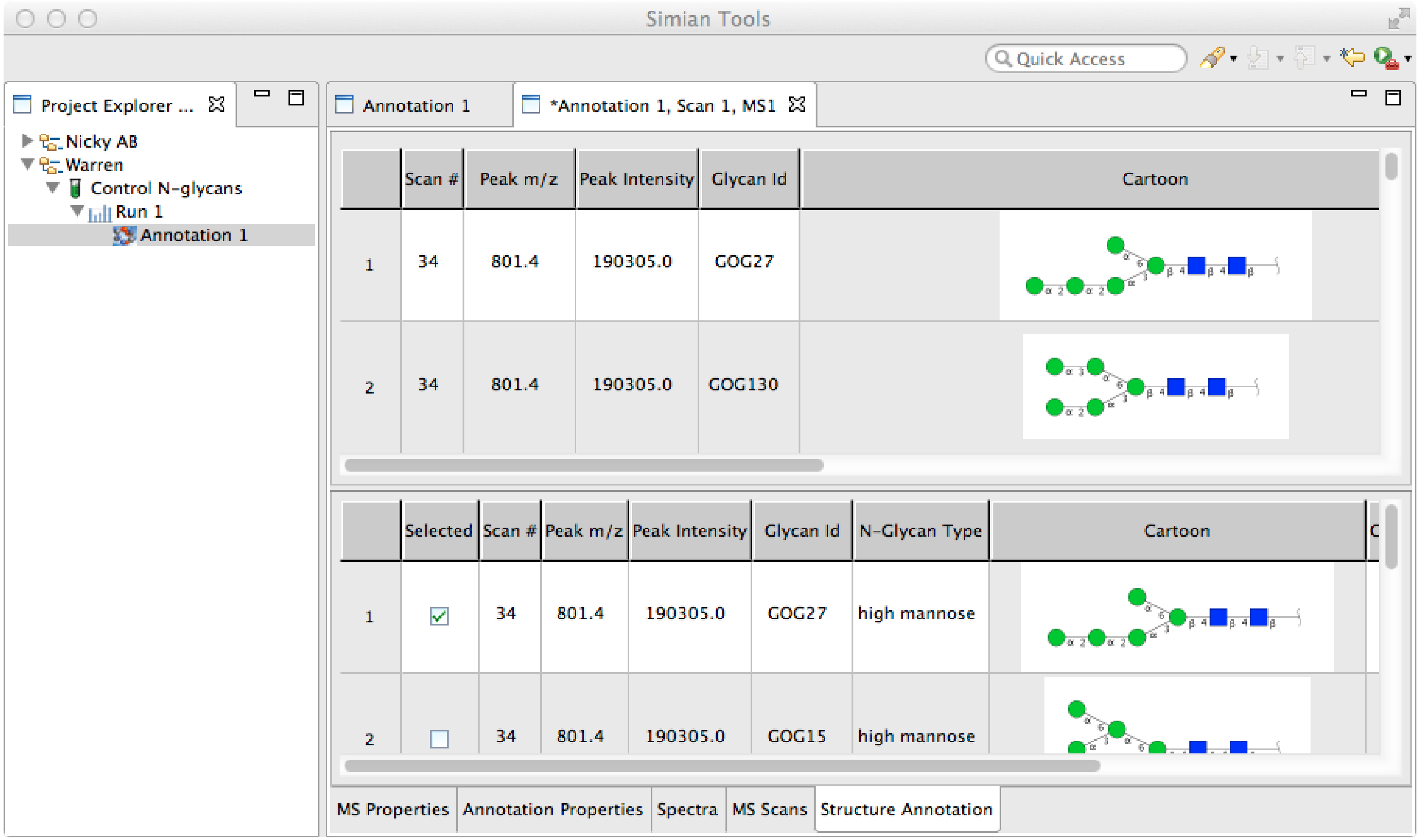
- Nicky AB
- Warren
  - Control N-glycans
    - Run 1
      - Annotation 1

	Scan #	Peak m/z	Peak Intensity	Glycan Id	Cartoon	Counting
1	34	801.4	190305.0	GOG27		0.0
2	34	801.4	190305.0	GOG15		0.0

	Selected	Scan #	Peak m/z	Peak Intensity	Glycan Id	N-Glycan Type	Cartoon	C
1	☑	34	801.4	190305.0	GOG27	high mannose		
2	☑	34	801.4	190305.0	GOG15	high mannose		

GRITS

Manual Verification of Annotations



- Nicky AB
- Warren
  - Control N-glycans
    - Run 1
      - Annotation 1

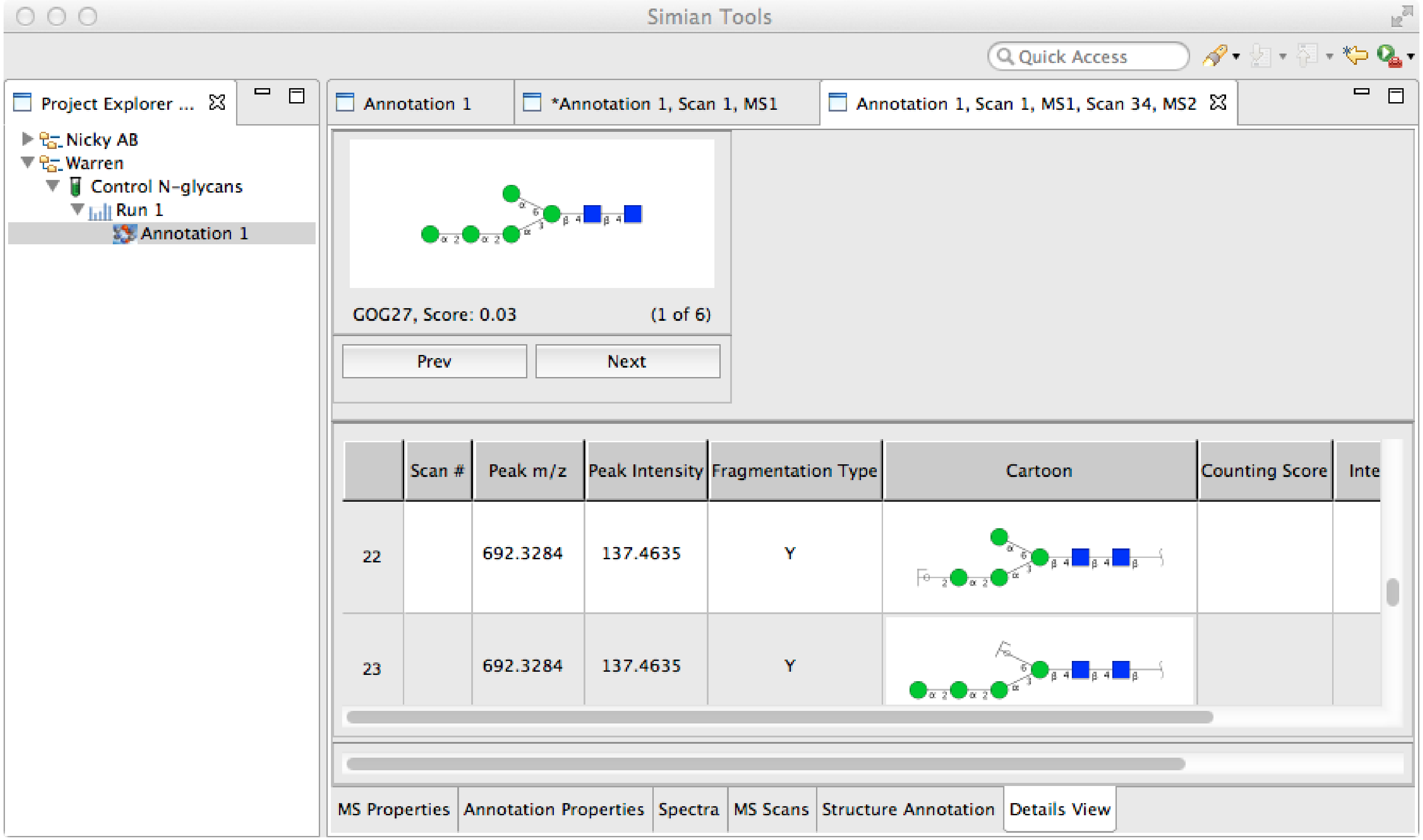
	Scan #	Peak m/z	Peak Intensity	Glycan Id	Cartoon
1	34	801.4	190305.0	GOG27	
2	34	801.4	190305.0	GOG130	

	Selected	Scan #	Peak m/z	Peak Intensity	Glycan Id	N-Glycan Type	Cartoon	C
1	☑	34	801.4	190305.0	GOG27	high mannose		
2	☐	34	801.4	190305.0	GOG15	high mannose		

GRITS

Display Fragment Ions That Are Used as the Basis for Annotation

National Center for Biomedical Glycomics

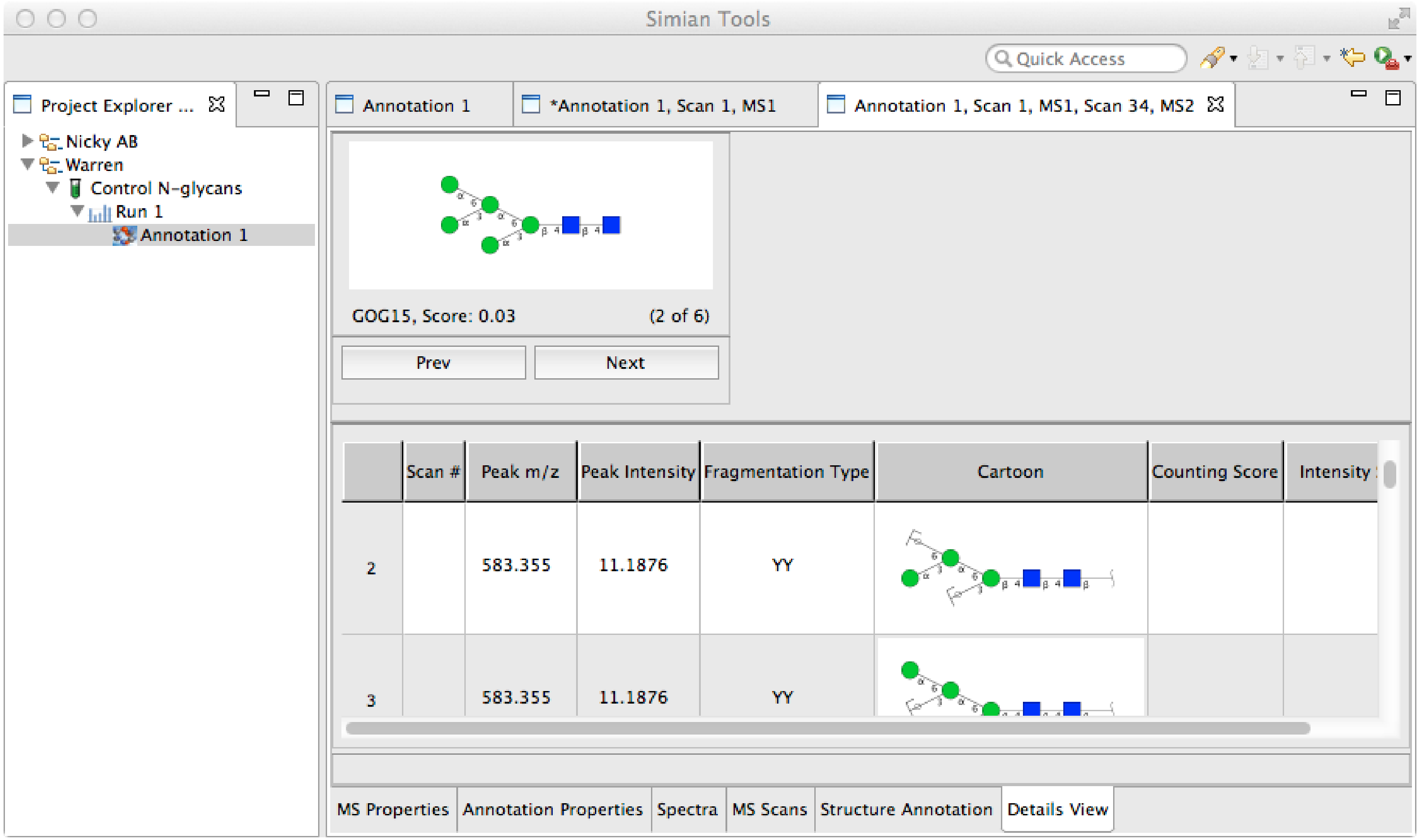


GOG27, Score: 0.03 (1 of 6)

	Scan #	Peak m/z	Peak Intensity	Fragmentation Type	Cartoon	Counting Score	Inte
22		692.3284	137.4635	Y			
23		692.3284	137.4635	Y			

GRITS

Display Fragment Ions That Are Used as the Basis for Annotation



- Nicky AB
- Warren
  - Control N-glycans
    - Run 1
      - Annotation 1

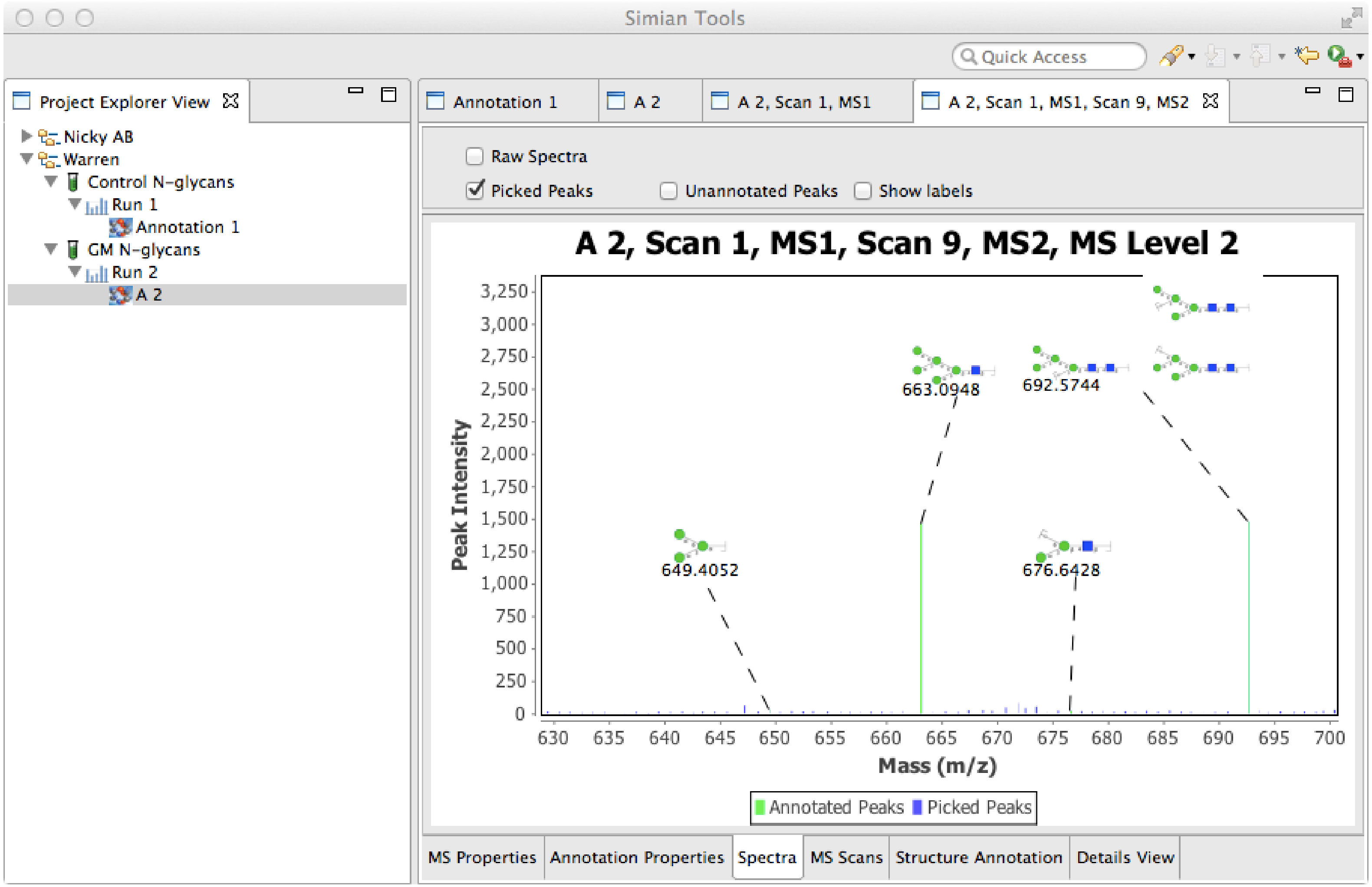
GOG15, Score: 0.03 (2 of 6)

	Scan #	Peak m/z	Peak Intensity	Fragmentation Type	Cartoon	Counting Score	Intensity
2		583.355	11.1876	YY			
3		583.355	11.1876	YY			

GRITS

Display Fragment Ions That Are Used as the Basis for Annotation

National Center for Biomedical Glycomics
NIGMS

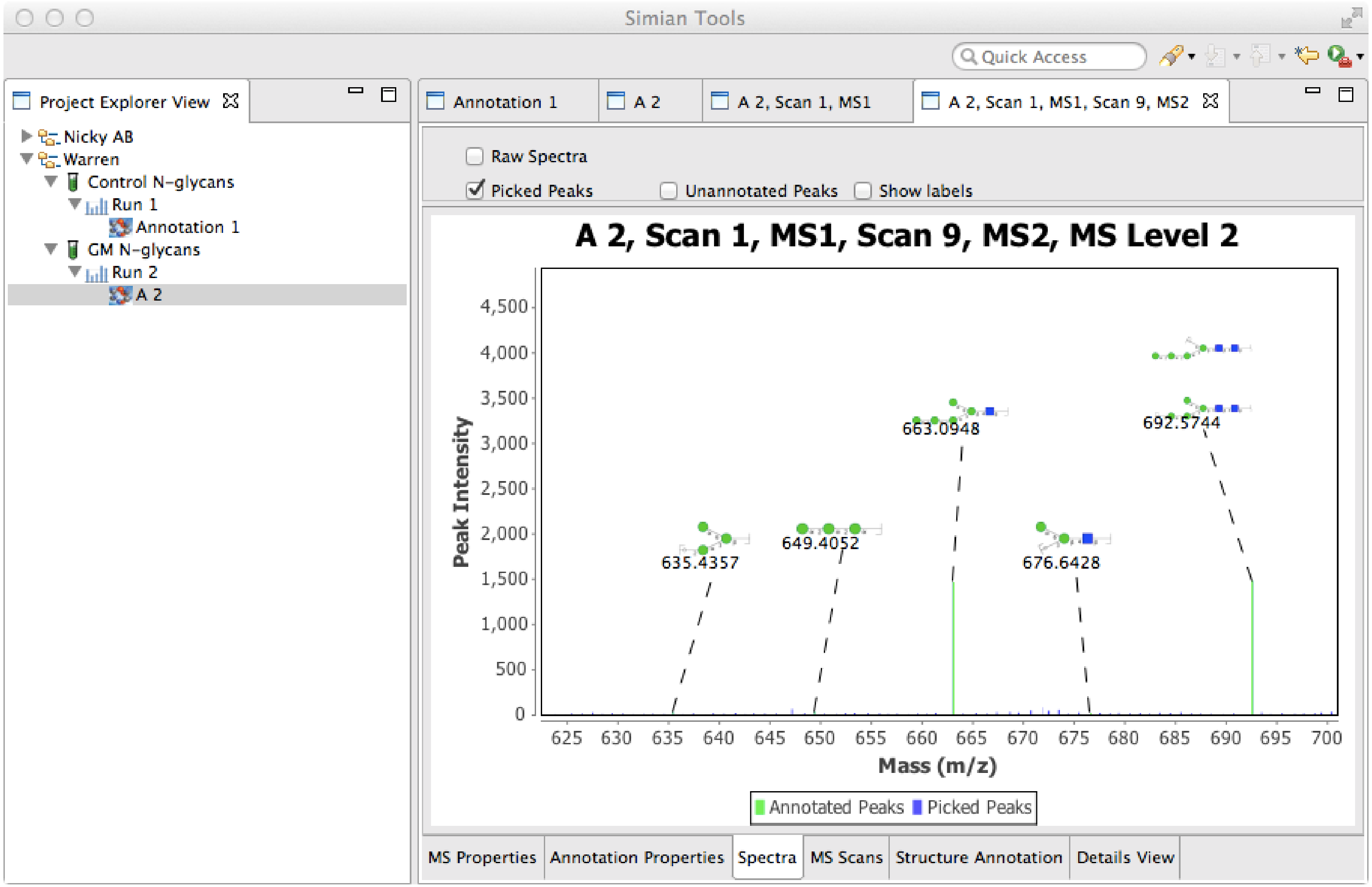


☐ Raw Spectra

☑ Picked Peaks ☐ Unannotated Peaks ☐ Show labels

GRITS

Display Fragment Ions That Are Used as the Basis for Annotation



Simian Tools

Project Explorer View

- Nicky AB
- Warren
  - Control N-glycans
    - Run 1
      - Annotation 1
  - GM N-glycans
    - Run 2
      - A 2

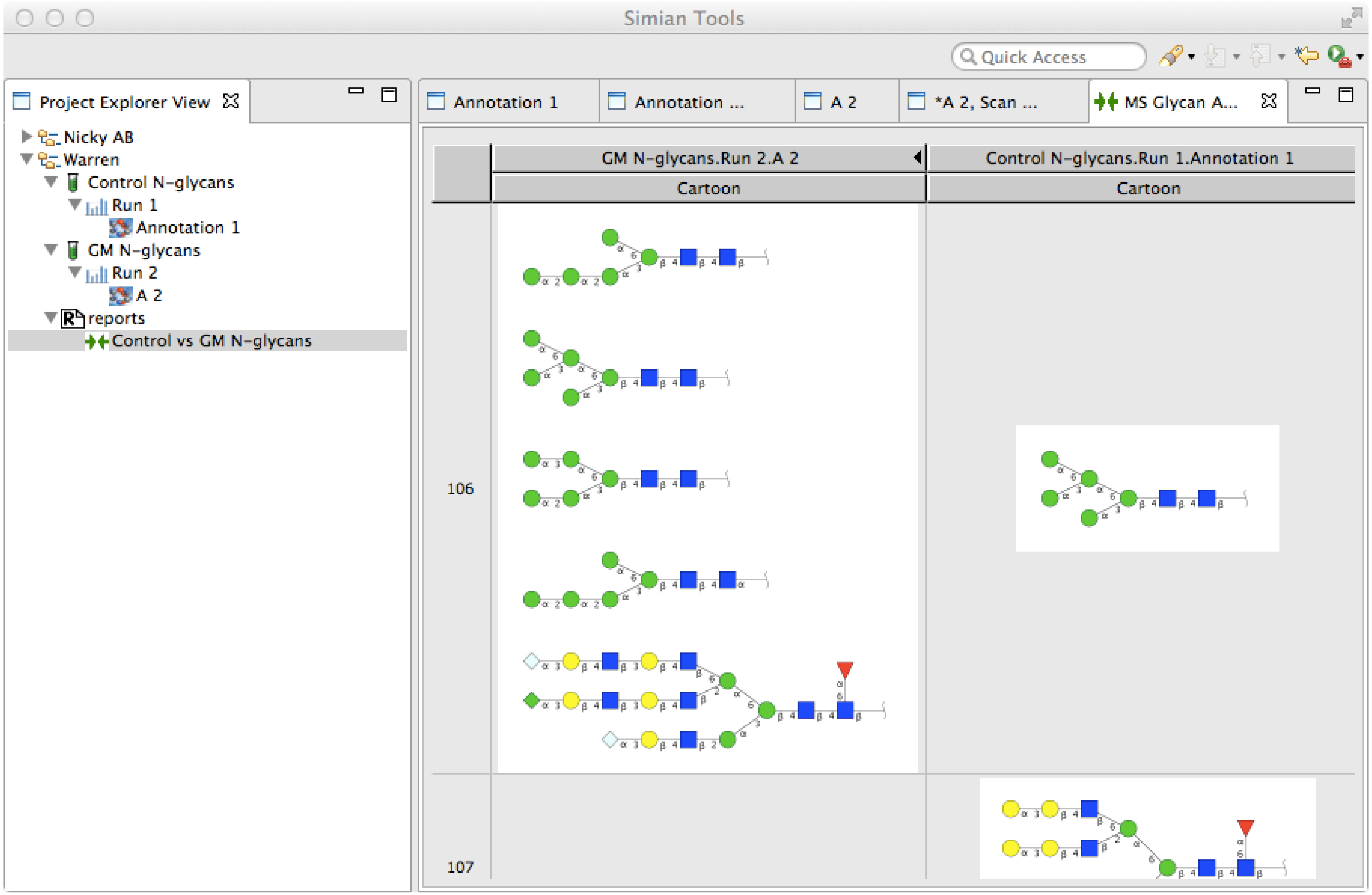
Annotation 1 | A 2 | A 2, Scan 1, MS1 | A 2, Scan 1, MS1, Scan 9, MS2

☐ Raw Spectra

☑ Picked Peaks ☐ Unannotated Peaks ☐ Show labels

GRITS

Provides Exportable Reports – Including Side-By-Side Comparisons of Data Sets



Thanks!