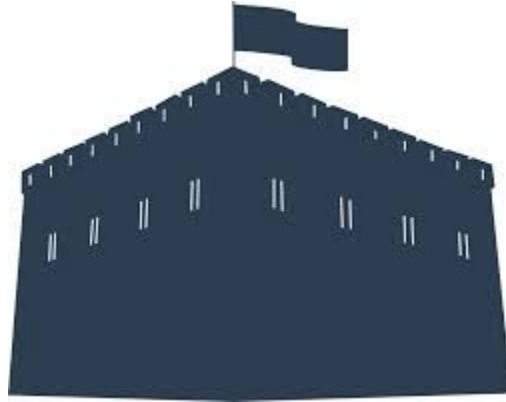


Proteomics Visualization in Galaxy



The Goal

- Provide interactive exploration and visualization of proteomics data
- Deploy as Galaxy Visualization Plugin
 - This eliminates building, configuring and maintaining a server instance for visualization.
 - REST, Ajax and HTML5 compliant.

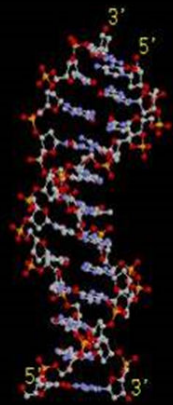


Galaxy Visualization Plugin

- config
 - associates the plugin with datasets
- templates
 - template to generate a webpage
- static (optional)
 - javascript to which can communicate with the galaxy server to interactively retrieve data from a dataset



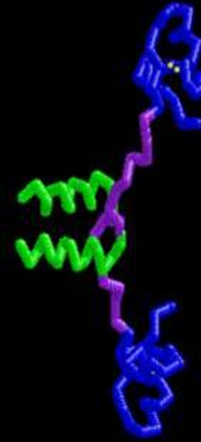
`Omics Technologies



DNA
Genome



RNA
Transcriptome



Protein
Proteome



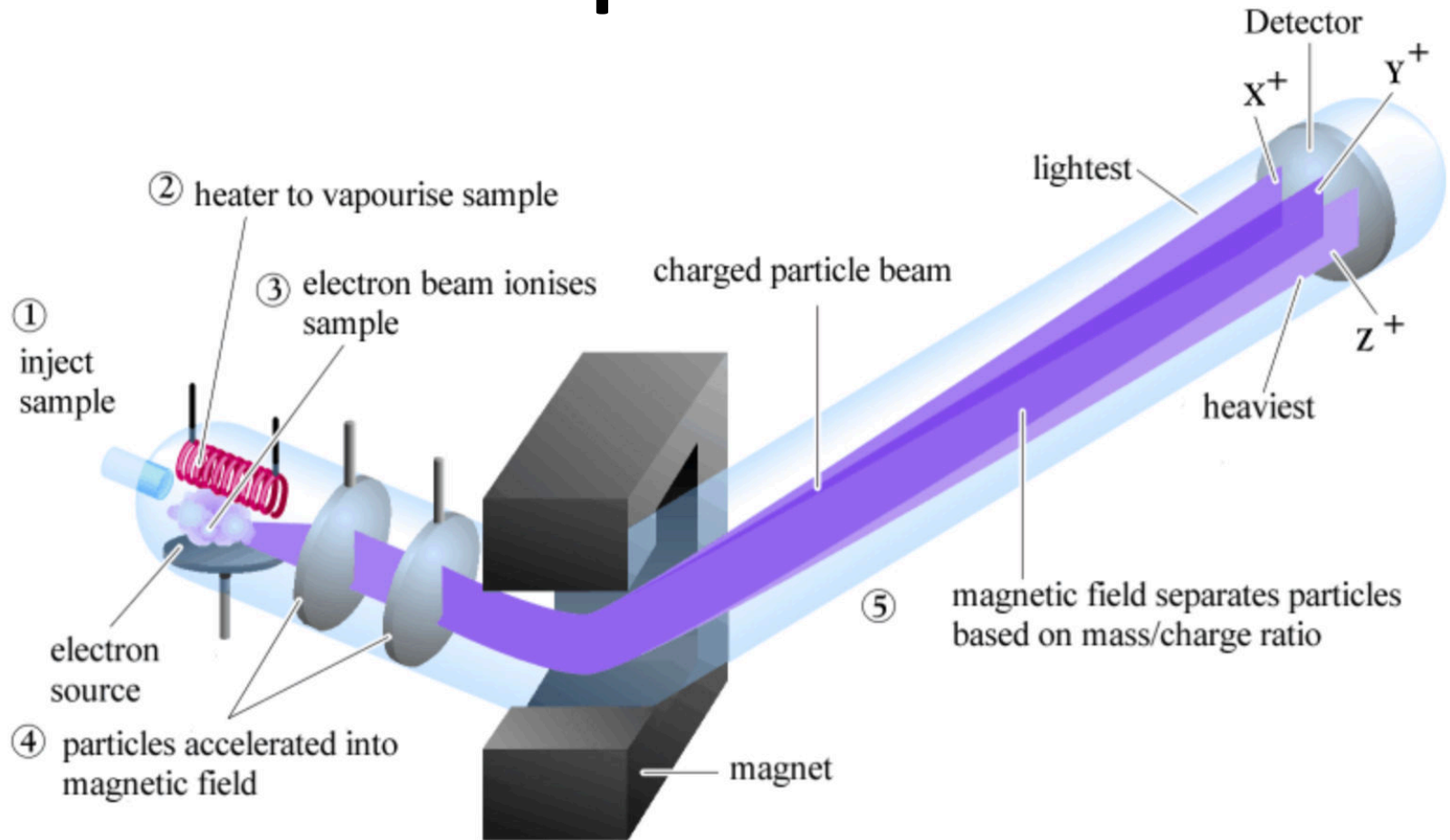
Metabolite
Metabolome

Genomic Sequencing

Mass Spectrometry

- Proteomics complements genomics/transcriptomics
- Direct information on molecular “effectors” of cell functions (enzymes, transporters etc.)
- Properties not predicted by genes/transcripts (PTMs, protein complexes etc.)

Mass Spectrometer



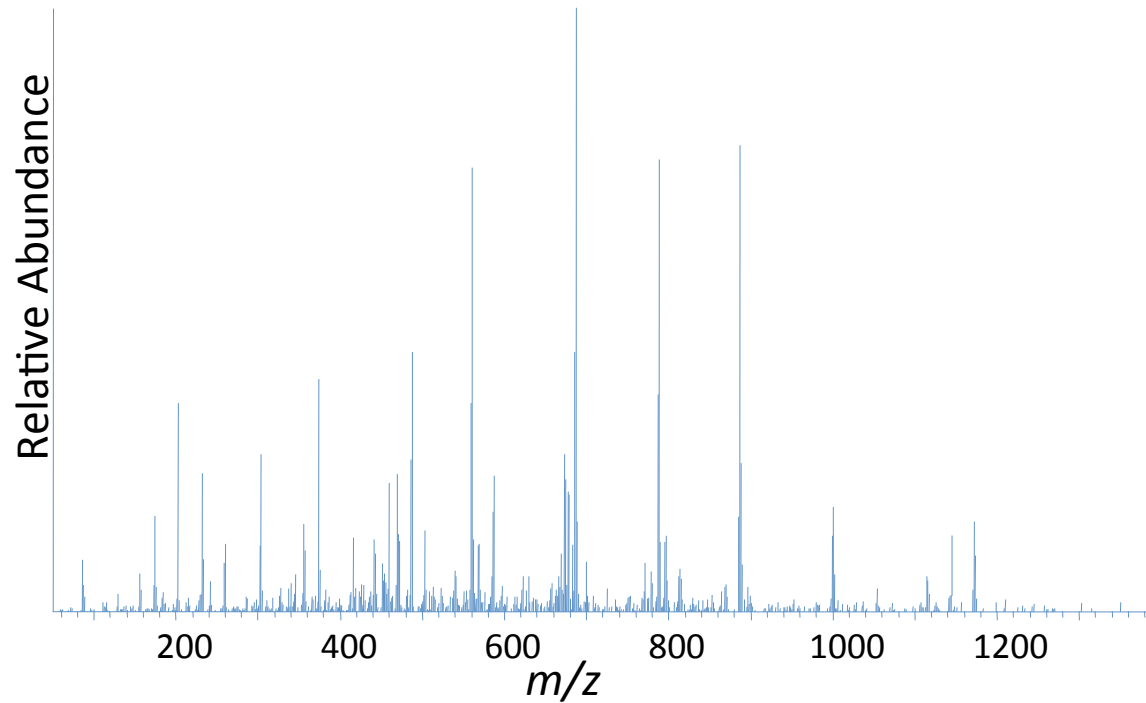
Record over time:
Intensity of particles detected at Mass/Charge
Ratios (m/z)

Mass Spectrometer Output

Peak Lists

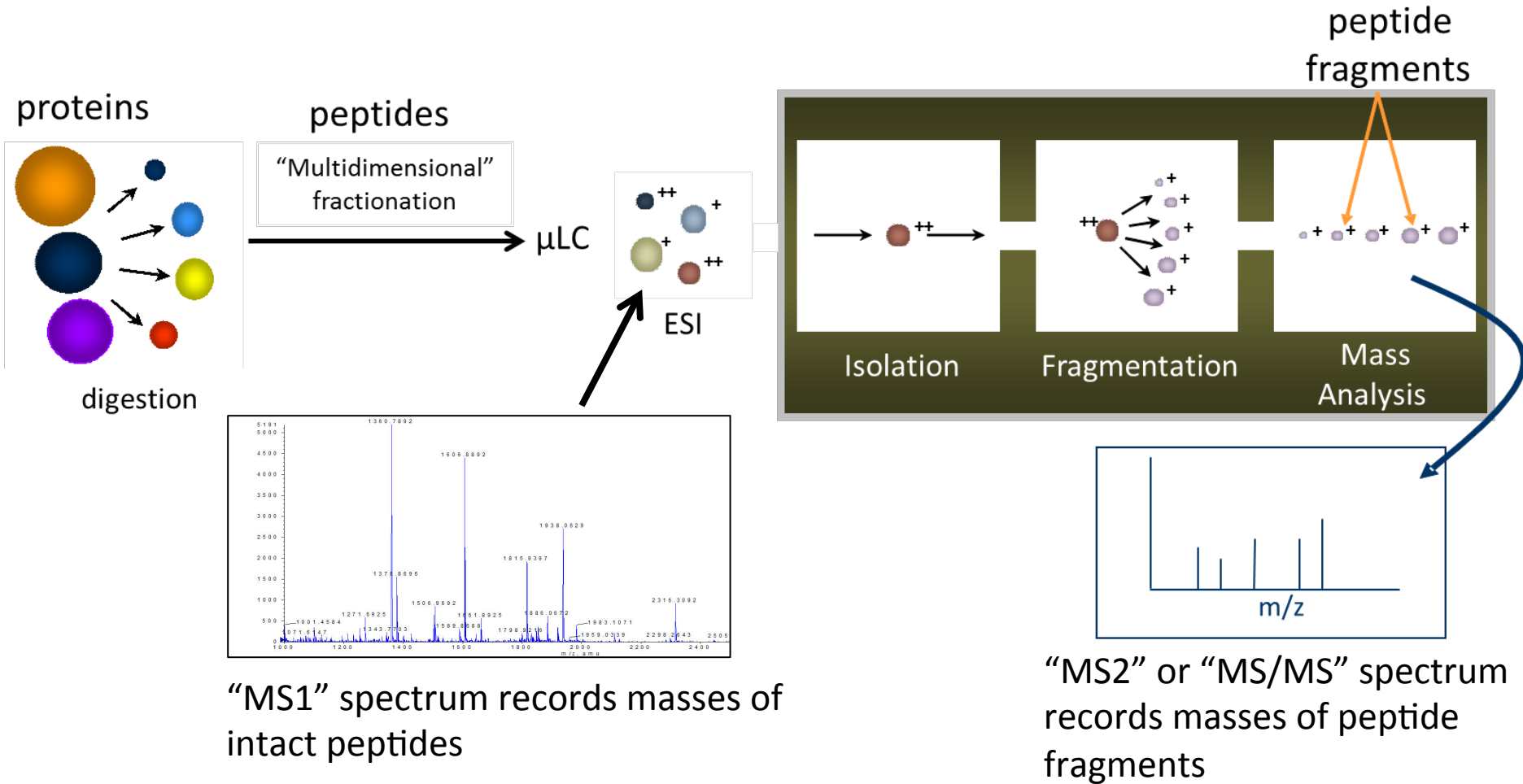
Mass/Charge	Intensity
112.087204	72.720093
115.086487	217.396378
120.080681	63.673885
121.927612	47.203938
126.055283	176.274216
127.086617	229.260223
129.067535	70.479492
129.102478	2281.949707

Information “currency” of MS: mass spectra



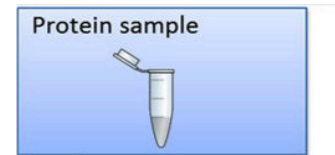
Protein identification by MS

Peptide fractionation coupled to tandem mass spectrometry (MS/MS)

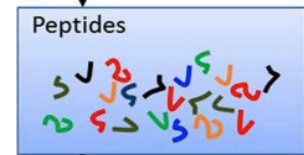


Proteomics Identification Workflow

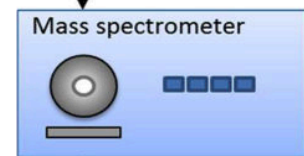
Experimental Sample



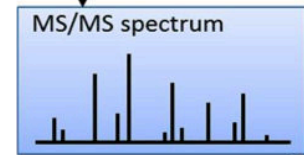
Enzymatic digestion



Separate and ionize



Acquire data



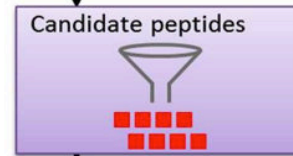
Experimental Spectra



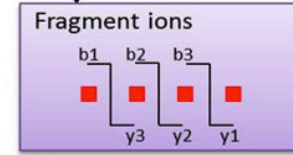
Enzymatic digestion



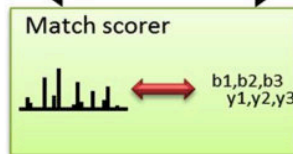
Peptide mass filter



Calculate fragment ions



Protein Fasta Search DB

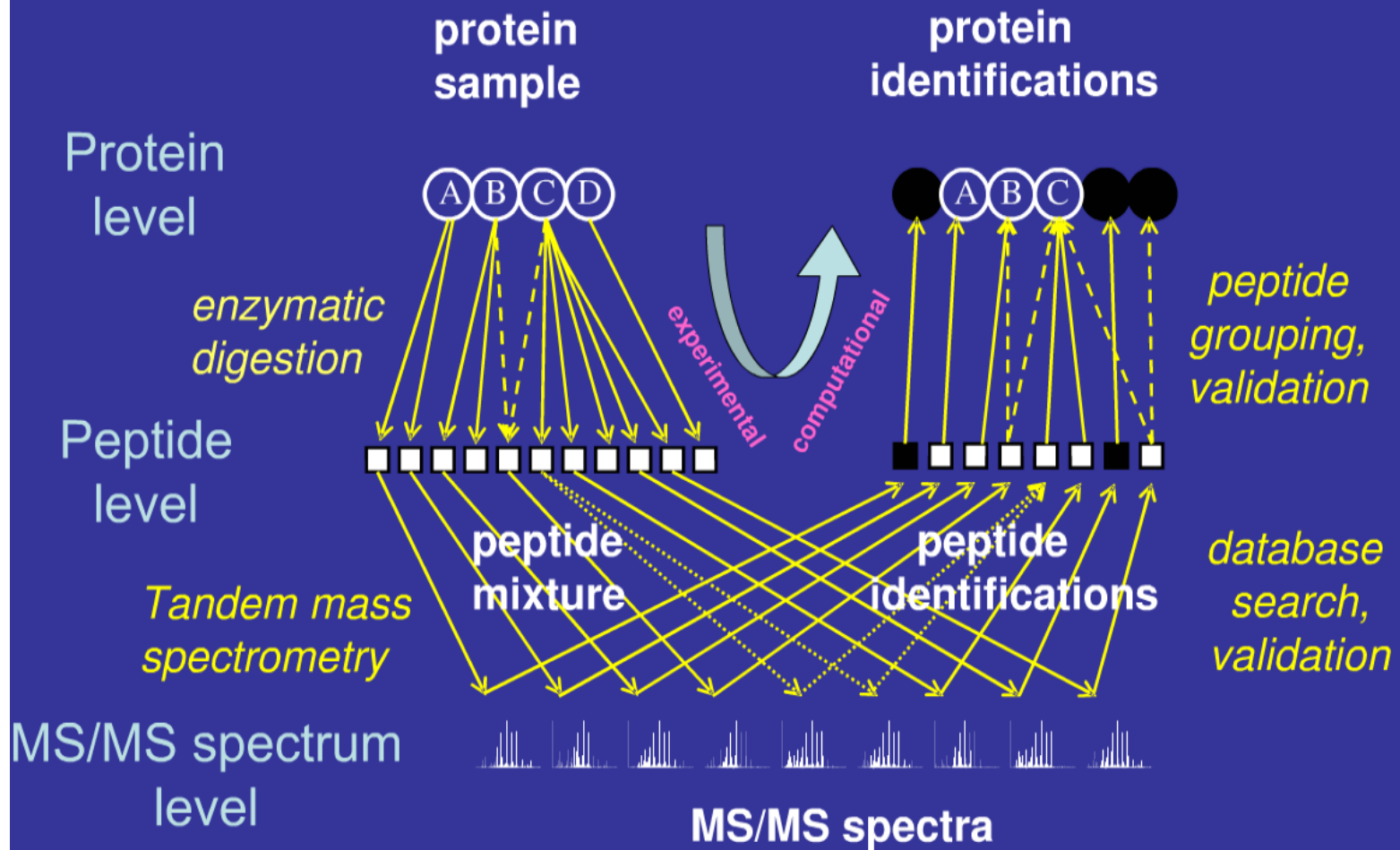


Theoretical Spectra

Peptide Spectral Matching Applications

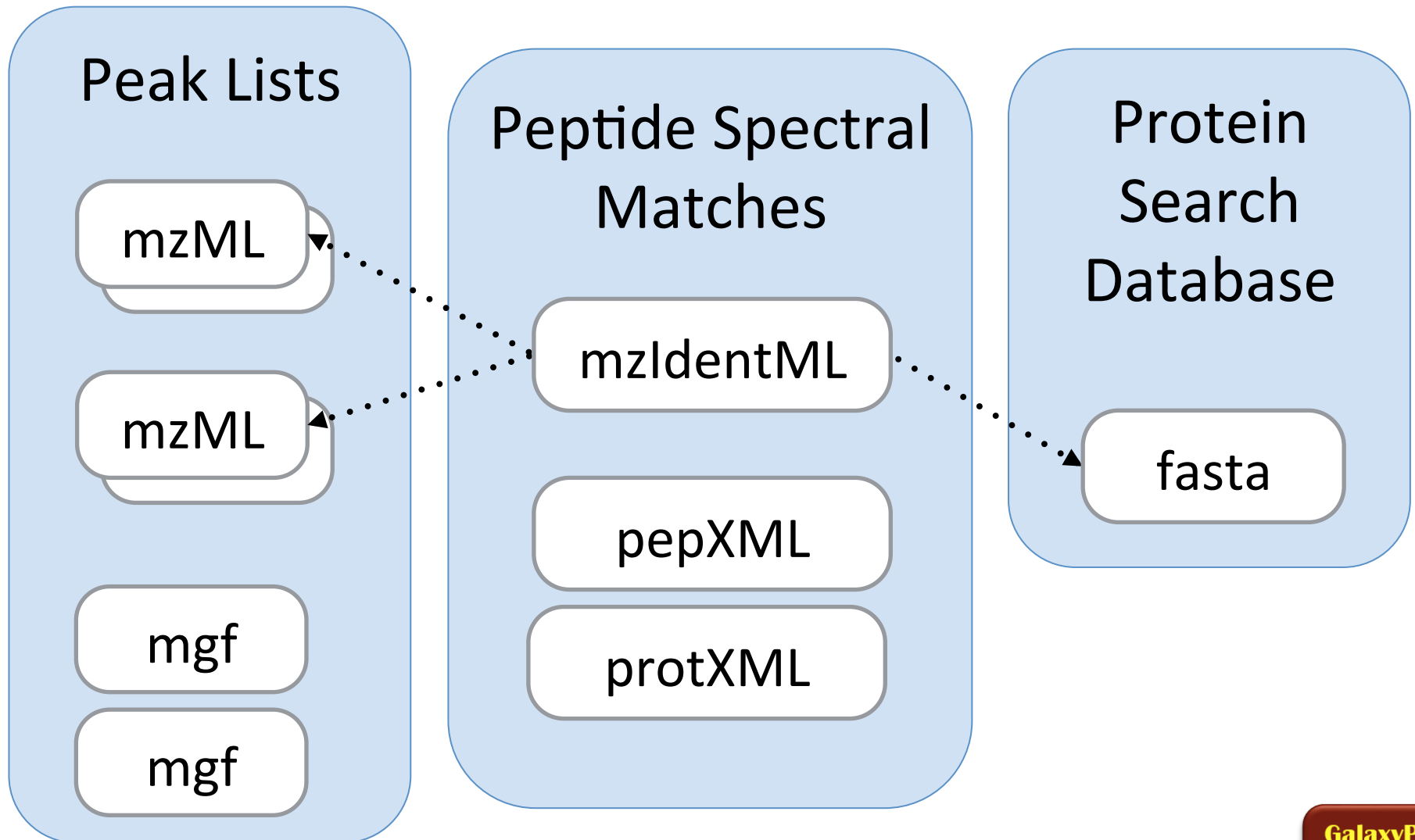


Shotgun Protein Identification



Slide from Alexey Nesvizshkii talk at <http://www.scivee.tv/node/12671>

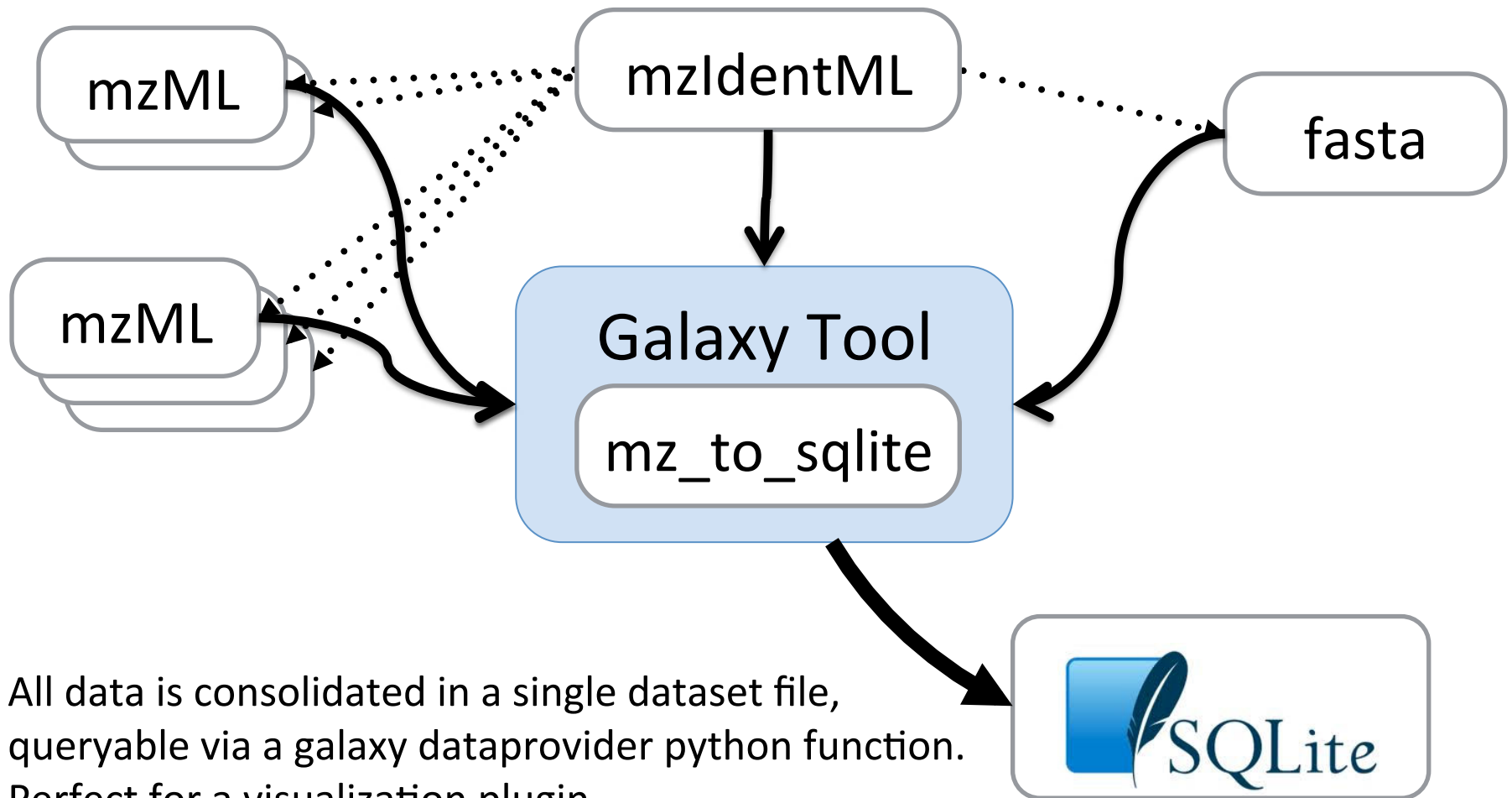
Proteomics Formats



Proteomics Visualization Plugin Issues

- Proteomics file formats are often large XML files
- Data spread across multiple datasets
 - Identification formats use references to data in other files:
e.g. their file paths in the temporary job working directory on the compute node which ran the application
- Relatively slow to examine, not interactive, especially for filtering and aggregating values

Solution: Aggregate into SQLite DB



All data is consolidated in a single dataset file, queryable via a galaxy dataprovider python function. Perfect for a visualization plugin.

Proteomics MzSQLite Datatype

- New Galaxy datatype: mz.sqlite
 - Extends generic SQLite datatype, with a proteomics specific schema
 - Schema based on MzIdentML and mzML format specifications
- Can use the REST-like SQLite dataproviders which return table results for SQL query requests from the browser

Proteomics Visualization Plugin

- Provides tables to view proteins, peptides, and PSMs (Peptide Spectral Match)
- Want to search, filter, and sort tables
- Selecting items in a table, filters the other tables to only display related items
- Selecting a PSM opens a lorikeet spectral viewer for examining the match.
- Hovering over a protein accession opens a protein coverage display.

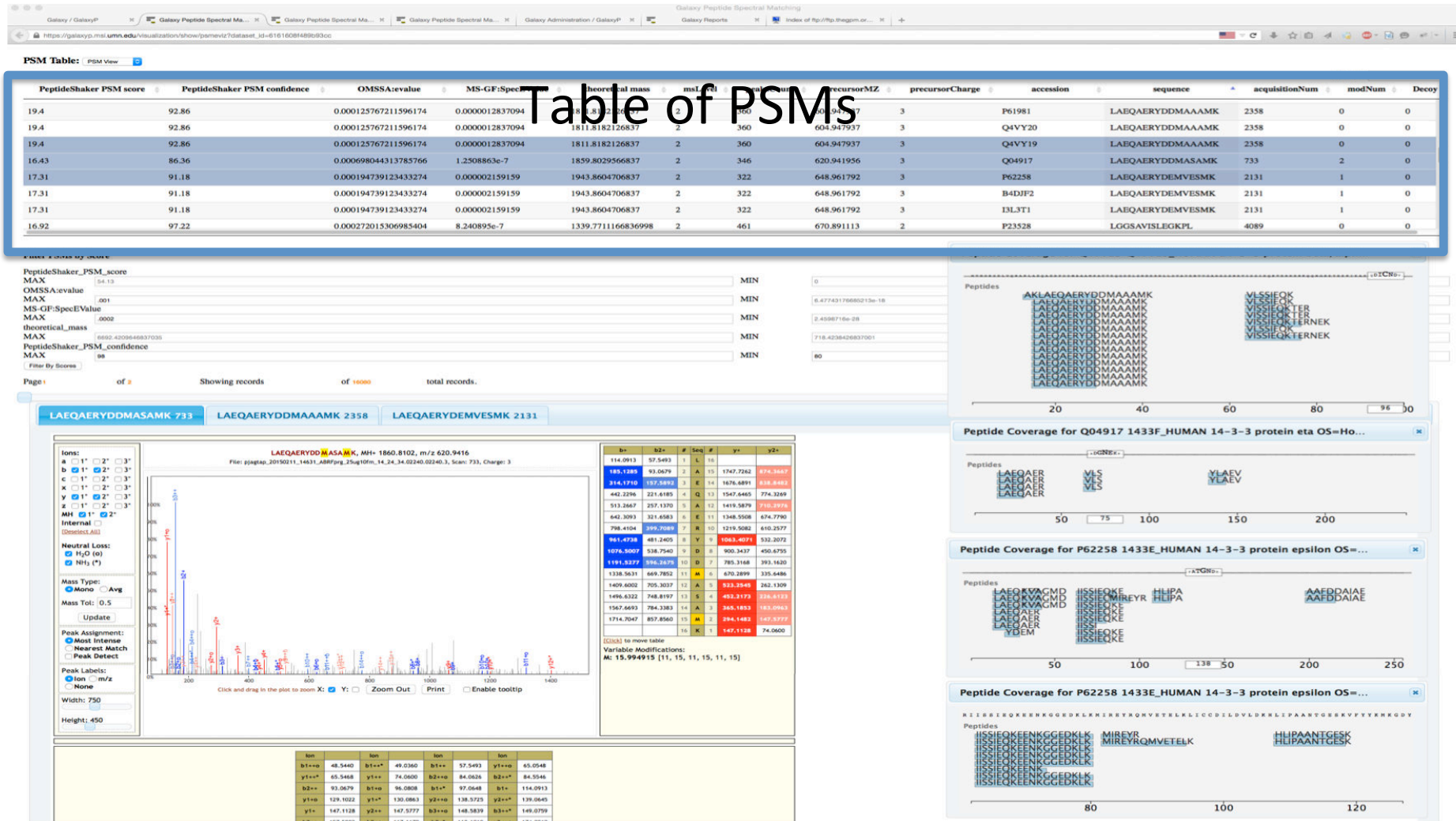


Galaxy Peptide Spectral Matching

Filter PSMs by Score			Peptide Coverage for Q4VY19 Q4VY19_HUMAN 14-3-3 protein beta/alpha...		
PeptideShaker_PSM_score			Peptides		
MAX	54.13	MIN	AKLAEGAEYDDMAAAAMK		
OMSSA/evalue		MIN	LAEGAEYDDMAAAAMK		
MAX	.001	MIN	LAEGAEYDDMAAAAMK		
MS-GF+SpecEValue		MIN	LAEGAEYDDMAAAAMK		
MAX	.0002	MIN	LAEGAEYDDMAAAAMK		
theoretical_mass		MIN	LAEGAEYDDMAAAAMK		
MAX	6892.4209546837035	MIN	LAEGAEYDDMAAAAMK		
PeptideShaker_PSM_confidence		MIN	LAEGAEYDDMAAAAMK		
MAX	95	MIN	LAEGAEYDDMAAAAMK		
Filter By Scores			LAEGAEYDDMAAAAMK		
Page 1	of 2	Showing records	LAEGAEYDDMAAAAMK		
		of 10000	LAEGAEYDDMAAAAMK		
		total records.	LAEGAEYDDMAAAAMK		



Proteomics Visualization Plugin

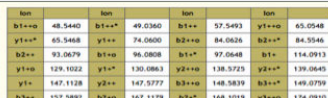


PeptideShaker PSM score	PeptideShaker PSM confidence	OMSSA:eValue	MS-GF+SpecEValue	theoretical mass	msLevel	peaksCount	precursorMZ	precursorCharge	accession	sequence	acquisitionNum	modNum	Decoy
19.4	92.86	0.000125767211596174	0.0000012837094	1811.8182126837	2	360	604.947937	3	P61981	LAEQAERYDDMAAAAMK	2358	0	0
19.4	92.86	0.000125767211596174	0.0000012837094	1811.8182126837	2	360	604.947937	3	Q4VY20	LAEQAERYDDMAAAAMK	2358	0	0
19.4	92.86	0.000125767211596174	0.0000012837094	1811.8182126837	2	360	604.947937	3	Q4VY19	LAEQAERYDDMAAAAMK	2358	0	0
16.43	86.36	0.000698044313785766	1.2508863e-7	1859.8029566837	2	346	620.941956	3	Q04917	LAEQAERYDDMASAMK	733	2	0
17.31	91.18	0.000194739123433274	0.000002159159	1943.8604706837	2	322	648.961792	3	P62258	LAEQAERYDEMVESMK	2131	1	0
17.31	91.18	0.000194739123433274	0.000002159159	1943.8604706837	2	322	648.961792	3	B4DJF2	LAEQAERYDEMVESMK	2131	1	0
17.31	91.18	0.000194739123433274	0.000002159159	1943.8604706837	2	322	648.961792	3	D3L3T1	LAEQAERYDEMVESMK	2131	1	0
16.92	97.22	0.000272015306985404	8.240895e-7	1339.7711166836998	2	461	700.891113	2	P23528	LGGSAVISLEGLPKL	4089	0	0

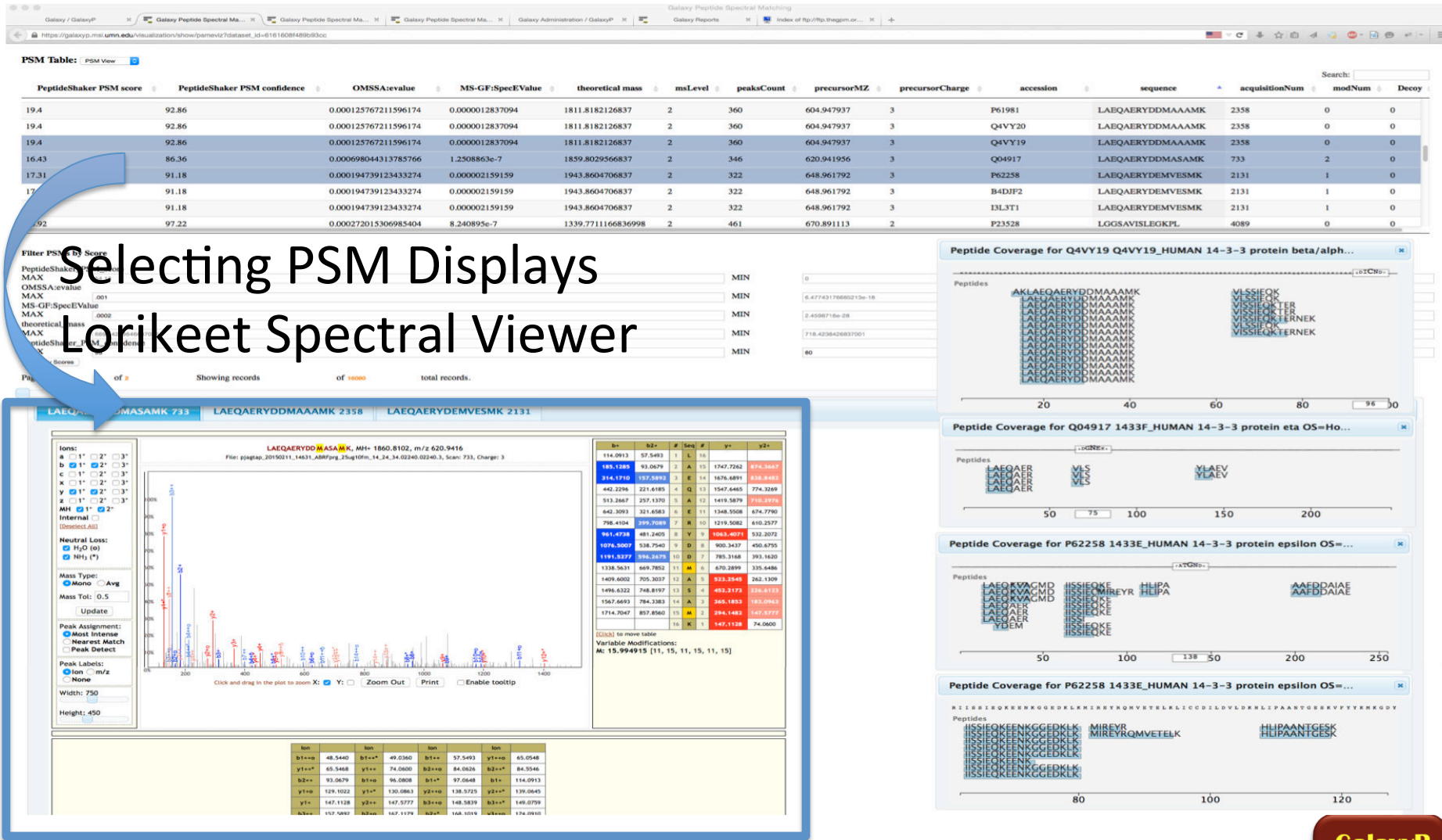
Filter PSMs by Score

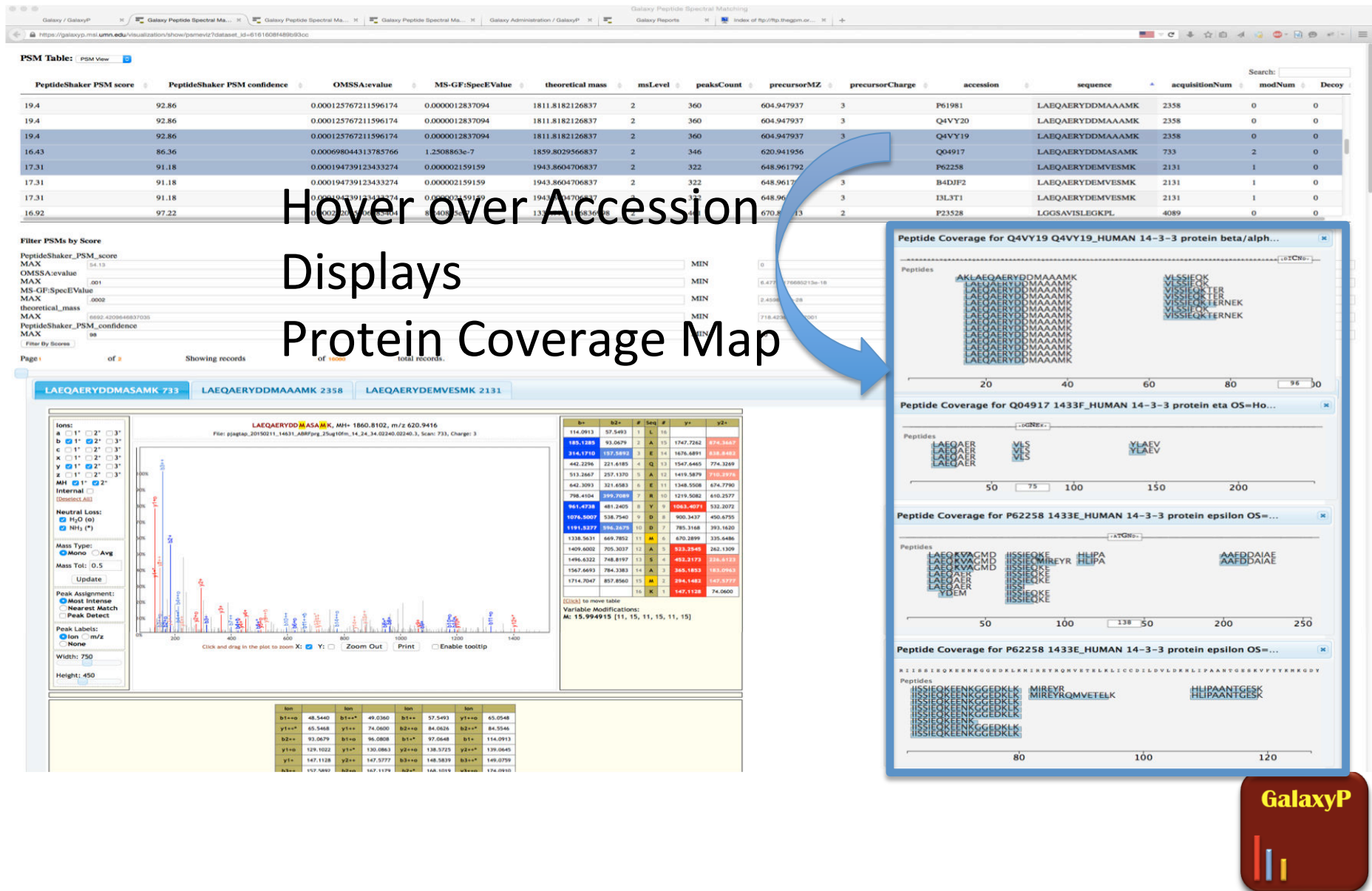
PeptideShaker_PSM_score			
MAX	54.13	MIN	0
OMSSA:evalue			
MAX	.001	MIN	6.47743176680213e-18
MS-GF:SpecEValue			
MAX	.0002	MIN	2.4558716e-28
theoretical_mass			
MAX	6092.4200646937035	MIN	718.4238426837001
PeptideShaker_PSM_confidence			
MAX	99	MIN	80

Page 1 of 2 Showing records of 16000 total records.



Proteomics Visualization Plugin





The Client Code

- A JQuery Javascript library plugin.
 - This minimizes Javascript library bloat. We have JQuery, JQuery-UI, and JQuery **Data Tables** libraries for the core functionality
 - We use additional libraries for add-on 'widgets'. For example, **Lorikeet** MS/MS spectrum viewer.

The Client Code Design Decisions

- Generate SQL queries to retrieve data
- Manage paging, sorting, and filtering in the construction of the SQL queries
- Use DataTables to display query results
 - Popular – used in RStudio
 - Active developer community
- Provide multiple starting views on the data
- Selecting items in a display generates events to display or filter other items



Scaling Up

- The web browser is not happy managing 500K rows of data. Who knew? Right?
- DataTable can handle data pagination client-side or server-side.
- But, we want to keep the Galaxy Data Provider simple
- So where to handle server-side pagination?



DataTables Server on the Client

- DataTables allow us to point to a function for server-side Ajax calls: `DataManager.ajaxDataProvider()`.
- Our client code adheres to the DataTable server-side communication protocol, queries the Galaxy data provider via the API then packages a response adhering to the DataTable communication response protocol. (<https://datatables.net/manual/server-side>)
- The client code is now highly responsive even when scaled to large data sets.



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

Biochemistry, Molecular Biology &
Biophysics

Tim Griffin

UNIVERSITY OF MINNESOTA
SUPERCOMPUTING
INSTITUTE

Tom McGowan

**COMMUNITY BASED
SOFTWARE DEVELOPMENT**

The Galaxy Community

Center for Mass Spectrometry and
Proteomics

Pratik Jagtap

Ira Cooke

*La Trobe University, Melbourne ,
Australia*

John Chilton

*Galaxy Team
Penn State University*

Where's the code?

<https://github.com/galaxyproteomics/tools-galaxyp>

<http://uwpr.github.io/Lorikeet>

<https://github.com/Genentech/pviz>

<https://code.google.com/p/compomics-utilities/>



Funding
NSF, NIH

