

# Installation and Use of LabKey Server for Proteomics

Joshua Eckels,<sup>1</sup> Peter Hussey,<sup>1</sup> Elizabeth K. Nelson,<sup>1</sup> Tamra Myers,<sup>1</sup> Adam Rauch,<sup>1</sup> Matthew Bellew,<sup>1</sup> Brian Connolly,<sup>1</sup> Wendy Law,<sup>2</sup> Jimmy K. Eng,<sup>3</sup> Jonathan Katz,<sup>4</sup> Martin McIntosh,<sup>5</sup> Parag Mallick,<sup>4,6</sup> and Mark Igra<sup>1</sup>

<sup>1</sup>LabKey Software, Seattle, Washington

<sup>2</sup>Marsha Rivkin Center for Ovarian Cancer Research, Seattle, Washington

<sup>3</sup>Department of Genome Sciences, University of Washington, Seattle, Washington

<sup>4</sup>Departments of Medicine and Biomedical Engineering, University of Southern California Center for Applied Molecular Medicine, Los Angeles, California

<sup>5</sup>Fred Hutchinson Cancer Research Center, Seattle, Washington

<sup>6</sup>Department of Radiology, Stanford University, Stanford, California

## ABSTRACT

LabKey Server (formerly CPAS, the Computational Proteomics Analysis System) provides a Web-based platform for mining data from liquid chromatography–tandem mass spectrometry (LC-MS/MS) proteomic experiments. This open source platform supports systematic proteomic analyses and secure data management, integration, and sharing. LabKey Server incorporates several tools currently used in proteomic analysis, including the X! Tandem search engine, the ProteoWizard toolkit, and the PeptideProphet and ProteinProphet data mining tools. These tools and others are integrated into LabKey Server, which provides an extensible architecture for developing high-throughput biological applications. The LabKey Server analysis pipeline acts on data in standardized file formats, so that researchers may use LabKey Server with other search engines, including Mascot or SEQUEST, that follow a standardized format for reporting search engine results. Supported builds of LabKey Server are freely available at <http://www.labkey.com/>. Documentation and source code are available under the Apache License 2.0 at <http://www.labkey.org>. *Curr. Protoc. Bioinform.* 36:13.5.1-13.5.25. © 2011 by John Wiley & Sons, Inc.

Keywords: liquid chromatography • mass spectrometry • proteomics • data analysis • protein identification

## INTRODUCTION

LabKey Server provides an open-source proteomic toolkit that integrates several open-source proteomic data analysis and mining tools into a single easy-to-install software platform (Rauch et al., 2006; Nelson et al., 2011). The system was previously called the Computational Proteomics Analysis System (CPAS). LabKey Server is currently used by a number of research groups at Fred Hutchinson Cancer Research Center (FHCRC) and at other institutions with considerable proteomic research efforts. As of June 2011, the LabKey Server proteomics installation at FHCRC contained over 88,000 liquid chromatography–tandem mass spectrometry (LC-MS/MS) experiments and nearly 700 million collision-induced dissociation (CID) spectra. In a typical week, the system processes thousands of runs and performs tens of millions of peptides identifications and millions of protein identifications.

LabKey Server has been designed for the flexibility required by research environments and the high-throughput processing capacity needed by proteomic facilities. Via

standardized file data formats, LabKey Server accepts data from various instrument types and also supports multiple search engines [e.g., X! Tandem (Craig and Beavis, 2004), Mascot (Perkins et al., 1999), and SEQUEST (Eng et al., 1994; *UNIT 13.3*)]. The Web-based user interface offers users a means to submit searches, manage protein annotations and databases, examine and filter results, compare proteins and peptides across experiments, and share results with others.

LabKey Server currently includes the freely available X! Tandem search engine for identifying peptides by sequence database searching. LabKey Server users can configure X! Tandem to search in its native form or employ a customized scoring algorithm that is compatible with PeptideProphet (Keller et al., 2002) and ProteinProphet (Nesvizhskii et al., 2003). The system also includes support for identifying post-translational modifications, for employing several labeled quantitation tools (XPRESS, Q3, and Libra), and for running the ProteoWizard suite of proteomics tools.

The protocols below offer instructions for using LabKey Server to search and analyze a simple isotope-coded affinity tag (ICAT)-labeled MS/MS sample. Basic Protocol 1 shows how to use the LabKey Server data pipeline to perform an X! Tandem search on the sample, followed by PeptideProphet analysis, XPRESS quantitation (Han et al., 2001), and ProteinProphet analysis. Basic Protocol 2 shows how to use the MS2 Dashboard to view and manipulate the search results within LabKey Server. The Support Protocol describes how to download and install LabKey Server to run on Windows. LabKey Server can also be installed on Linux, Solaris, or MacOS X.

The basic LabKey Server installation described in the Support Protocol runs on a single computer and is sufficient for searching small samples for evaluation purposes. LabKey Server is also fully scalable to a clustered architecture for high-throughput proteomic facilities and can be run in cloud-based configurations.

All components of LabKey Server are open-source development projects. Software components distributed in LabKey Server are under ongoing development, so some of the steps outlined in the protocols below may have changed during the publication period. Check the LabKey Server proteomics documentation at <https://www.labkey.org/wiki/home/Documentation/page.view?name=ms2> for the most up-to-date information.

## **BASIC PROTOCOL 1**

### **SEARCH TANDEM MASS SPECTROMETRY DATA**

The LabKey Server data pipeline searches MS/MS data and makes the results available for viewing and analysis in a Web-based tool. In the instructions given here, LabKey Server searches tandem mass spectra using the X! Tandem search engine. LabKey Server's MS/MS features may then be used to identify peptides based on the search score, but LabKey Server also supports further analysis using methods based on the Institute for Systems Biology's Trans-Proteomic Pipeline (TPP; Keller et al., 2005). Thus, the following protocol makes use of the PeptideProphet and ProteinProphet tools to analyze the evidence of peptide and protein identifications. This protocol also demonstrates how to use the pipeline to optionally perform quantitation using XPRESS (Han et al., 2001).

PeptideProphet is an analysis tool that combines a database score, charge state, and other elements related to the confidence of a peptide identification into a classification score, then maps that classification score to a probability value (Keller et al., 2002). The probability value gives a measure of confidence that the peptide identification is correct; for example, 90% of all peptides with a PeptideProphet score equal to 0.9 should be correctly matched. ProteinProphet is an analysis tool that infers confidence at the protein level by synthesizing these peptide level probabilities (Nesvizhskii et al., 2003). For example, a protein may have a high measure of confidence if it has several only modestly

confident peptides that uniquely match to it, or may have a lower measure of confidence if its confident peptides are also contained in another protein.

## ***Necessary Resources***

### ***Hardware***

LabKey Server is a Web application that runs on Apache Tomcat and accesses a PostgreSQL or Microsoft SQL Server database server. The resource requirements for the Web application itself are minimal, but the computer on which LabKey Server is installed must have sufficient resources to run Apache Tomcat and the database server. The following system resources are recommended:

Processor: a high-performing processor

Physical memory: at least 2 Gb RAM, preferably 4 Gb

Disk space: 1 Gb hard drive space free

### ***Software***

LabKey Server software (see Support Protocol)

### ***Files***

The sample files used in this protocol are available in archived formats for Windows or Unix on the download page at <https://www.labkey.org/wiki/home/Documentation/page.view?name=cpastutorial> and also on the *Current Protocols* Web site at <http://www.currentprotocols.com/protocol/bi1305>. The following files are included in the sample archive (CPASDemo.zip):

Three files containing ICAT-labeled tandem mass spectrometry data, called MM\_c1ICAT11.mzXML, MM\_c1ICAT12.mzXML, and MM\_c1ICAT13.mzXML. The mzXML file format for MS/MS data was developed at the Institute for Systems Biology (<http://www.systemsbiology.org/>).

A FASTA database file named yeast.nci.20070223.fasta.

An analysis definition file for X! Tandem named k.Yeast\_ICAT.xml. This file contains some custom parameters to be passed to X! Tandem for this search operation, which override the default parameters. The custom parameters in this sample file specify information about the search protocol, indicate that the data pipeline should perform quantitation as part of the analysis, specify residue modification mass values, and specify spectrum parameters. The parameters also specify a pluggable scoring algorithm called *k*-score (MacLean et al., 2006).

## ***Getting started with LabKey Server***

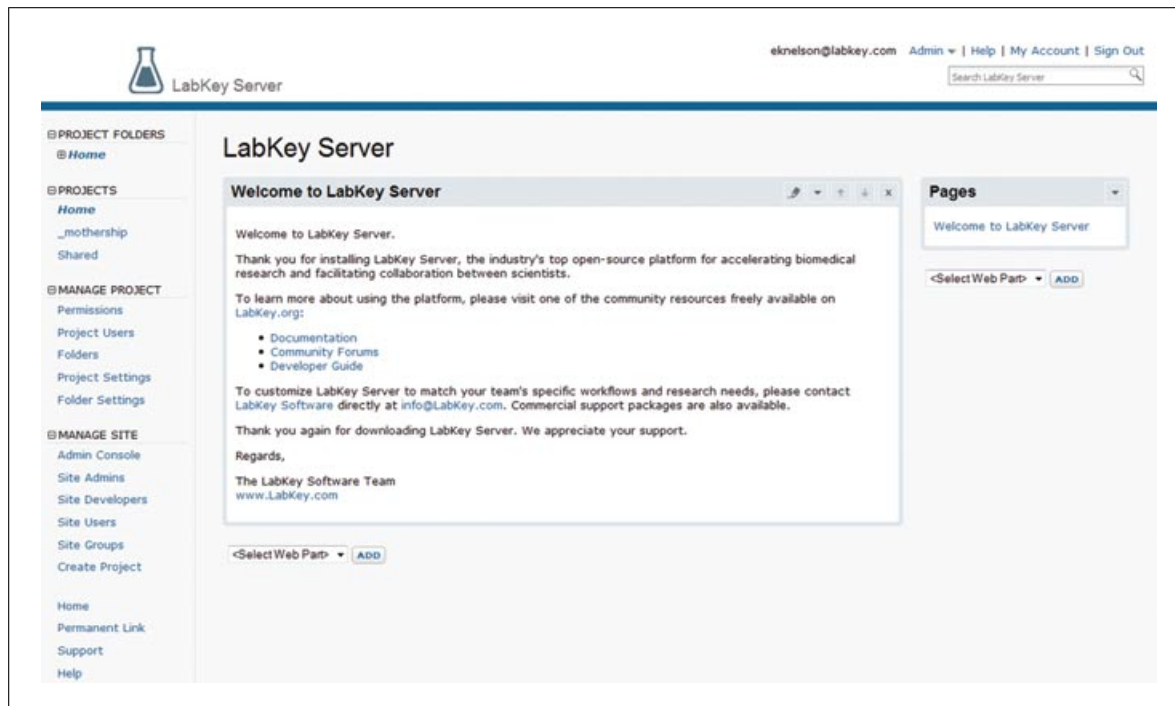
1. Download, install, and configure LabKey Server as described in the Support Protocol.

*As a shortcut, you can skip installing LabKey Server (and all steps in the Support Protocol) by using a preconfigured project on a LabKey Server hosted by LabKey Software. To do so, request a free, limited-time account from <https://hosted.labkey.com>.*

*For a hosted account, all steps in Basic Protocols 1 and 2 are the same, except Import UniProt annotations. That step is unnecessary in a hosted project because it has been done for you by an administrator.*

2. After you have logged into an installation of LabKey Server, a LabKey Server portal page appears in the browser, as shown in Figure 13.5.1.

*The left navigation pane shows the currently selected project. If you have just installed LabKey Server, this is the Home project. On the hosted server, the name of your project has been customized for you, so it is not called Home.*



**Figure 13.5.1** The LabKey Server welcome page.

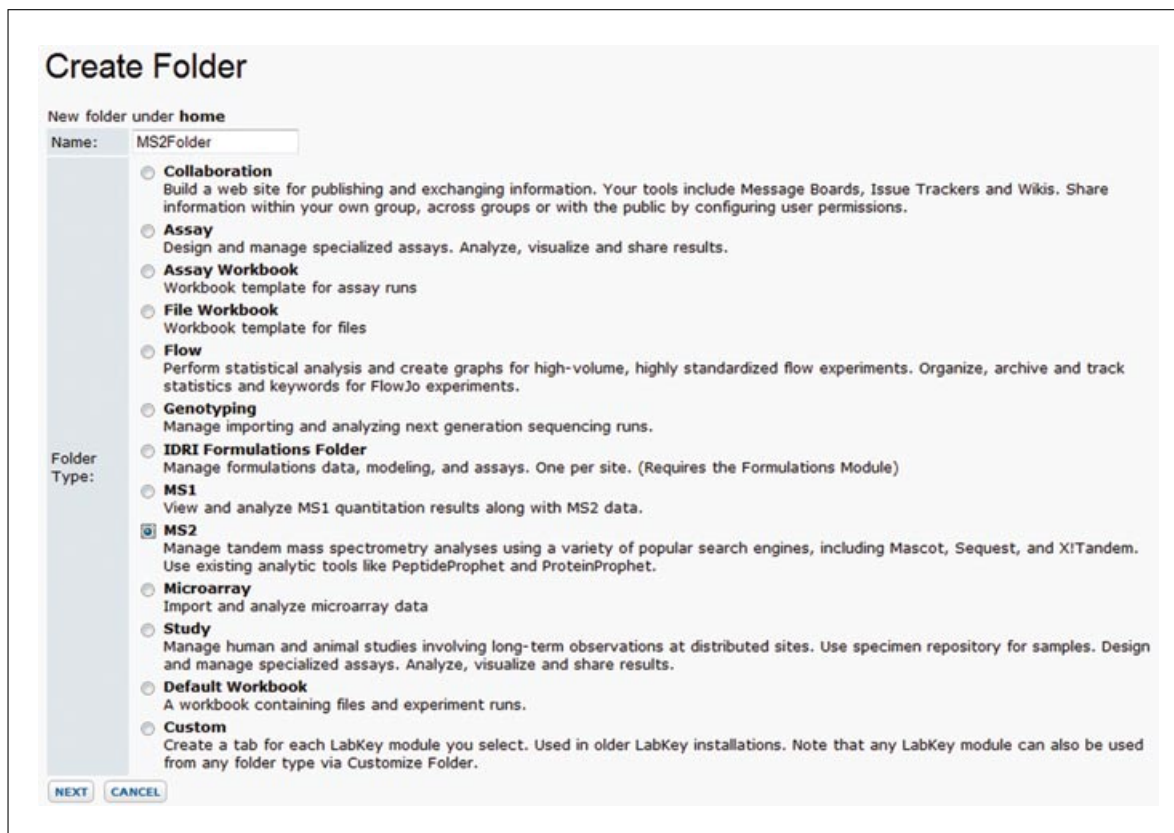
*Subfolders appear beneath a project hierarchically as you add them. Folders and projects help you organize your work and securely share particular workspaces with selected collaborators.*

*Other projects that you have sufficient permissions to view appear beneath the Projects menu. The Manage Project menu shows options for the currently selected project, including security, folder, and settings management. The Manage Site menu shows options for administering the site, including site configuration, user management, and new project creation. The Admin link on the top right of the page provides a shortcut path to all of these options.*

*The Help links at the upper right and lower left lead to the LabKey Server online documentation.*

### **Upload data files**

3. Extract the sample files to a directory on your local computer. You will upload these to your server in a later step.
4. Create a new folder in your current project on your LabKey Server by expanding the Manage Project section in the left navigation pane, then clicking the Folders link. On the Manage Folders page, make sure that the desired project is highlighted, then click Create SubFolder. Enter MS2Folder as the name for the folder, set the folder type to MS2, and then click Next, as shown in Figure 13.5.2.
5. On the Permissions page, click Save and Finish to accept the default permissions.
6. You will now see the portal page for your new folder. For an MS2-type folder, this is called the MS2 Dashboard, as shown in Figure 13.5.3.
7. Click on the Process and Upload Files button under the Data Pipeline section of the page (Fig. 13.5.3). On the Manage Files page, click Upload Files and select the Multiple Files radio button. If prompted by your Web browser, accept running the file upload applet.



**Figure 13.5.2** Screen for creating a new folder in LabKey Server.

8. Separately, use Windows Explorer (or equivalent browser for your platform) to go into the folder where you extracted the demo files. As shown in Figure 13.5.4, drag the files and folders within this folder (e.g., .labkey, databases, etc.) into the drop target box. When prompted, accept replacement of the pre-existing/databases folder with the one you are uploading.

### ***Import UniProt annotations***

9. Go to the Admin link in the top right, select Manage Site, and choose the Admin Console. In the Management section, select Protein Databases. On the Protein Database Admin page, go to the Load Protein Annotations section of the page and select Load New Annot File.

*Importing UniProt annotations is an optional step. It provides greater detail about the proteins you identify, as explored later during run comparison. The annotations used here are a subset of an UniProt XML file available from <http://au.expasy.org/>.*

10. On the Load Protein Annotations page (shown in Fig. 13.5.5), enter the path to the `yeast_a.xml` file on your server. This file is in the databases folder of the CPASDemo folder that you downloaded and unzipped earlier. Leave all other options in their default, including UniProt as the Type. Click Load Annotations.
11. Refresh Protein Database Admin page to see that the job has started. Upload takes several minutes. Refresh again to see that the job has completed, as shown in the Protein Annotations Loaded section of the page, shown in Figure 13.5.6.

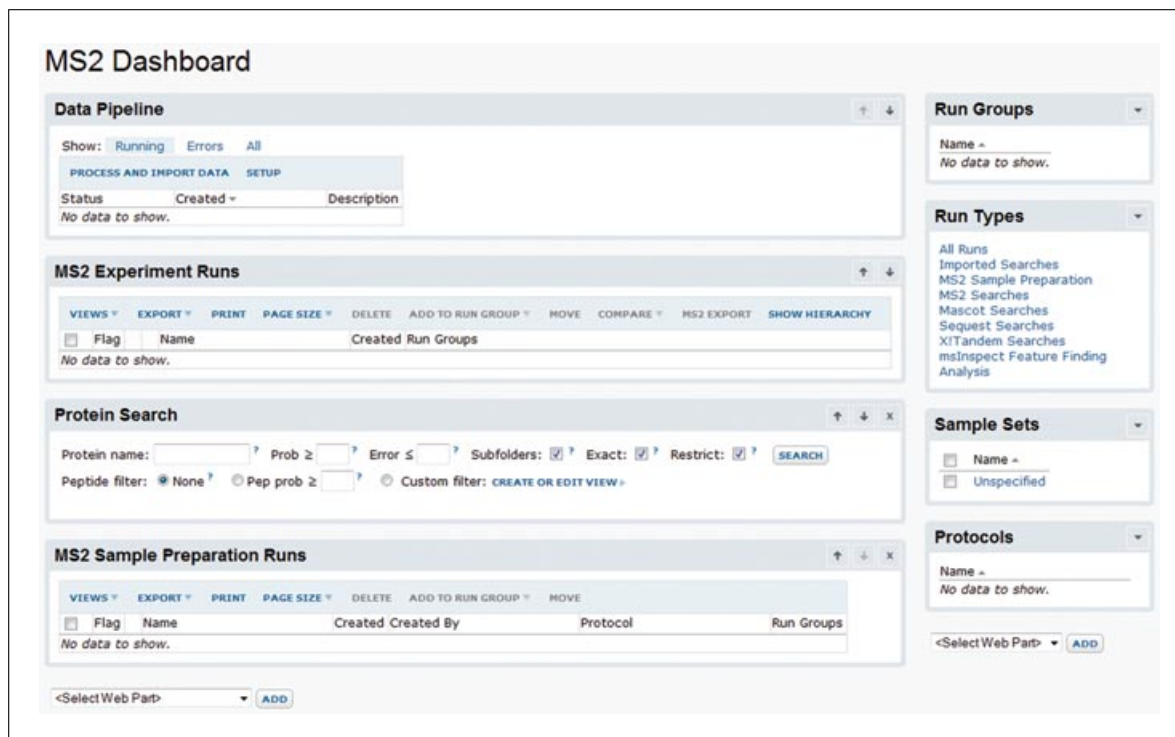


Figure 13.5.3 Screen showing the portal page of the new folder, the MS2 Dashboard.

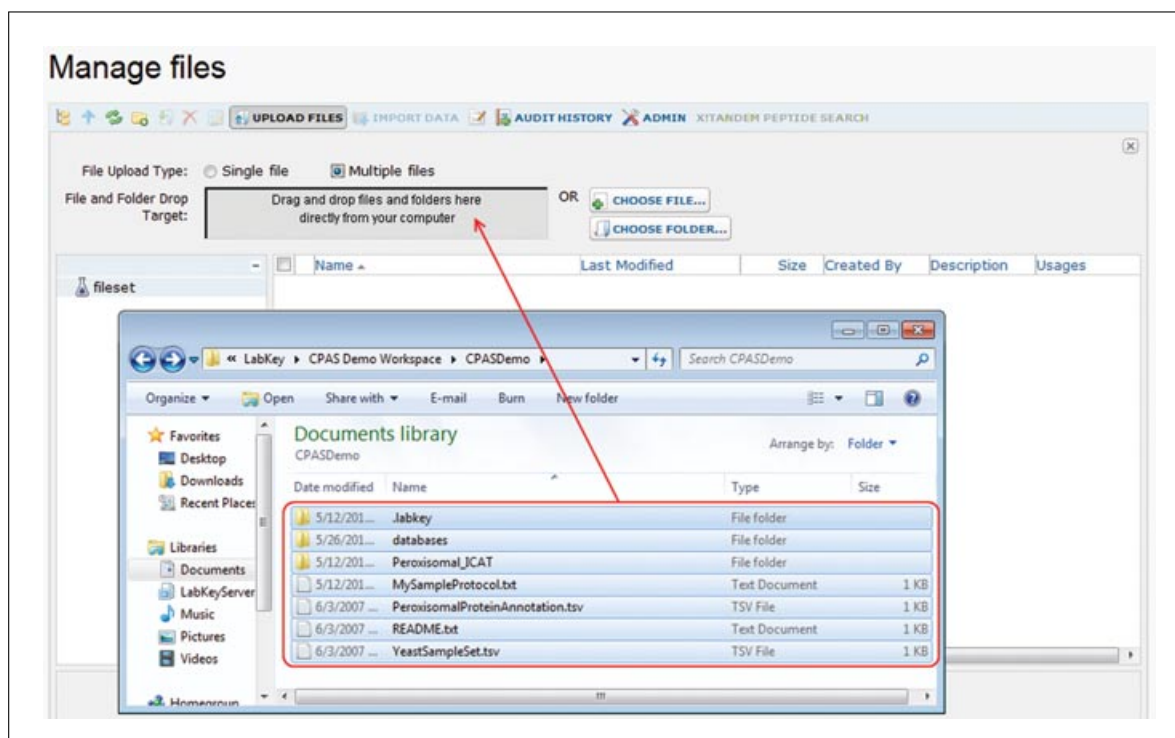


Figure 13.5.4 Screen for uploading files.

### *Perform a search on MS/MS data*

- Return to the MS2 Dashboard and click the Process and Upload Data button in the Data Pipeline section of the page. On the Manage Files page, go into the Peroxisomal\_ICAT folder, select the three .mzXML files, and click X! Tandem Peptide Search, as shown in Figure 13.5.7.

## Load Protein Annotations

Full file path:

Comment:

Type:

Clear existing identifiers and annotations?  ?

**Figure 13.5.5** Load protein annotations.

### Protein Annotations Loaded

<a href="#">DELETE</a> <a href="#">LOAD NEW ANNOT FILE</a> <a href="#">TEST FASTA HEADER PARSING</a> <a href="#">RELOAD SWP ORG MAP</a> <a href="#">LOAD GO</a>							
<input type="checkbox"/>	ID	File Name	File Type	Comment	Insert Date	Completion Date	Records State Processed
<input type="checkbox"/>	1	c:/9.3/sampleddata/xarfiles /ms2pipe/databases /Bovine_mini.fasta	fasta	Mon Oct 05 15:43:45 PDT 2009 c:/9.3/sampleddata /xarfiles/ms2pipe/databases /Bovine_mini.fasta	2009-10-05	2009-10-05	4308 COMPLETE
<input type="checkbox"/>	2	c:/LabKeyServerInstallation /build/deploy/files /MS2Project/@files /databases /yeast.nci.20070223.fasta	fasta	Thu May 26 17:58:44 PDT 2011 c:/LabKeyServerInstallation /build/deploy/files /MS2Project/@files /databases /yeast.nci.20070223.fasta	2011-05-26	2011-05-26	7469 COMPLETE
<input type="checkbox"/>	3	C:\Users\eknelson \Documents\LabKey\CPAS Demo Workspace\CPASDemo \databases\yeast_a.xml	uniprot		2011-05-31	2011-05-31	6109 COMPLETE

**Figure 13.5.6** Screen for monitoring the loading of protein annotations.

- Specify a search protocol that defines the X! Tandem search parameters. The sample files include a saved search protocol named `ICATSrch.xml`. To use this protocol, select `k.Yeast_ICAT` from the Analysis Protocol list. The details of the protocol appear as shown in Figure 13.5.8.

*The X! Tandem parameters shown in Figure 13.5.8 override the default parameters defined for searches performed on this data pipeline. The default X! Tandem parameters can be changed on the pipeline setup page; for more information on changing the default X! Tandem parameters, see the LabKey Server documentation.*

- Click the Search button to begin the search. LabKey Server provides status information as the search progresses. Refresh the page when the search is finished to see results appear as MS2 Experiment Runs. The search against the sample data file should be complete in a few minutes, depending on the speed of the computer being used.

*While the search is in progress, click on the Status message to display the Job Status page. The Job Status page displays details about the job, including the process log as it is being generated. Click on the link with the extension `.log` to see the log. Refresh the page to see changes to the log file during the search and analysis operation.*

*Note that if a large data file is searched or a large FASTA file is used, the process can take several hours. While the X! Tandem search is underway, the log file may not appear to update.*

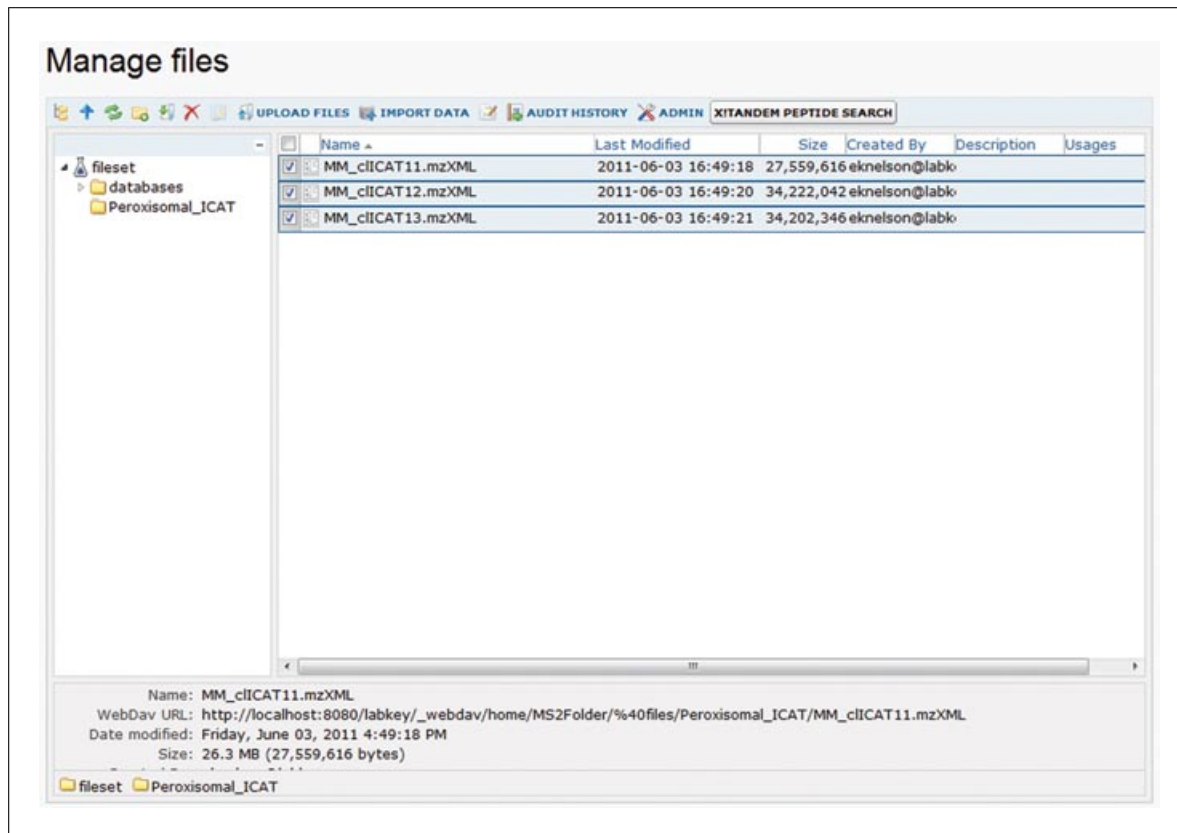


Figure 13.5.7 Select the mzXML data files to search for peptides.

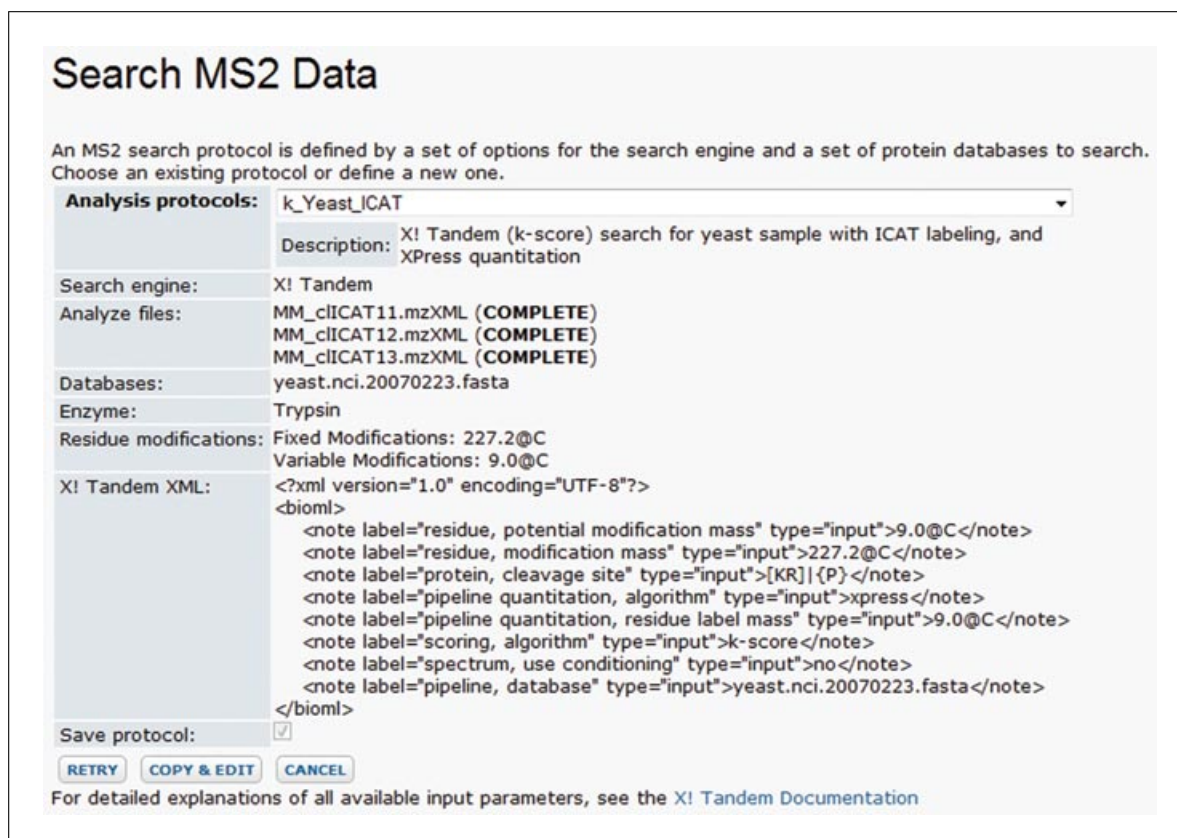


Figure 13.5.8 Specify a search protocol.

## 13.5.8

## VIEWING AND ANALYZING MS/MS DATA IN LabKey SERVER

After the LabKey Server data pipeline searches and processes the MS/MS data, the results are uploaded into the LabKey Server database, where they can be viewed and analyzed. The MS2 Dashboard page displays links to the results of the X! Tandem search, PeptideProphet scoring, quantitation calculations, and ProteinProphet analysis, as well as sensitivity and error rate information. The data displayed in the MS2 Dashboard can be manipulated in various ways—by sorting and filtering, by showing and hiding data columns, by grouping peptides by the proteins in which they are found, by saving custom views, and by exporting to other formats.

This protocol explores many of the analysis features accessible from the MS2 Dashboard. The MS2 Dashboard provides many features beyond those covered here; for example, it supplies a search box that allows users to locate results matching a particular protein or peptide of interest.

### *Necessary Resources*

#### *Hardware*

LabKey Server is a Web application that runs on Apache Tomcat and accesses a PostgreSQL or Microsoft SQL Server database server. The resource requirements for the Web application itself are minimal, but the computer on which LabKey Server is installed must have sufficient resources to run Apache Tomcat and the database server. The following system resources are recommended:

Processor: a high-performing processor

Physical memory: at least 2 Gb RAM, preferably 4 Gb

Disk space: 1 Gb hard drive space free

#### *Software*

LabKey Server (see Support Protocol)

#### *Files*

The sample files used in this protocol are available in archived formats for Windows or Unix on the download page at <https://www.labkey.org/wiki/home/Documentation/page.view?name=cpastutorial> and also on the *Current Protocols* Web site at <http://www.currentprotocols.com/protocol/bi1305>. The following files are included in the sample archive (CPASDEMO.zip):

Three files containing ICAT-labeled tandem mass spectrometry data, called MM\_c1ICAT11.mzXML, MM\_c1ICAT12.mzXML, and MM\_c1ICAT13.mzXML. The mzXML file format for MS/MS data was developed at the Institute for Systems Biology (<http://www.systemsbiology.org/>).

A FASTA database file named yeast.nci.20070223.fasta.

An analysis definition file for X! Tandem named k.Yeast\_ICAT.xml. This file contains some custom parameters to be passed to X! Tandem for this search operation, which override the default parameters. The custom parameters in this sample file specify information about the search protocol, indicate that the data pipeline should perform quantitation as part of the analysis, specify residue modification mass values, and specify spectrum parameters. The parameters also specify a pluggable scoring algorithm called *k*-score (MacLean et al., 2006)

### *View MS/MS sample data*

1. Complete Basic Protocol 1.

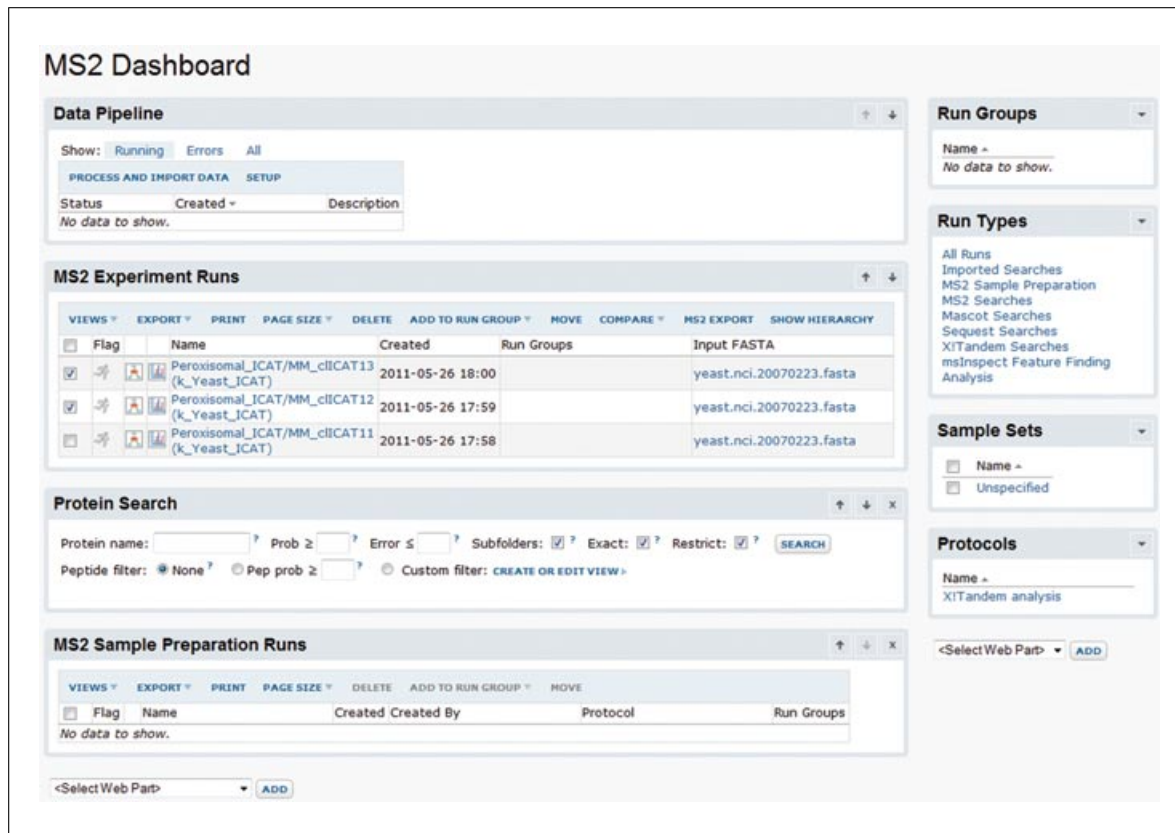


Figure 13.5.9 View of uploaded runs on the MS2 Dashboard.

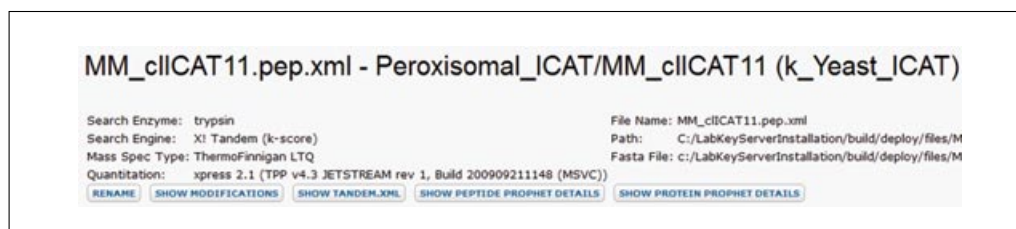


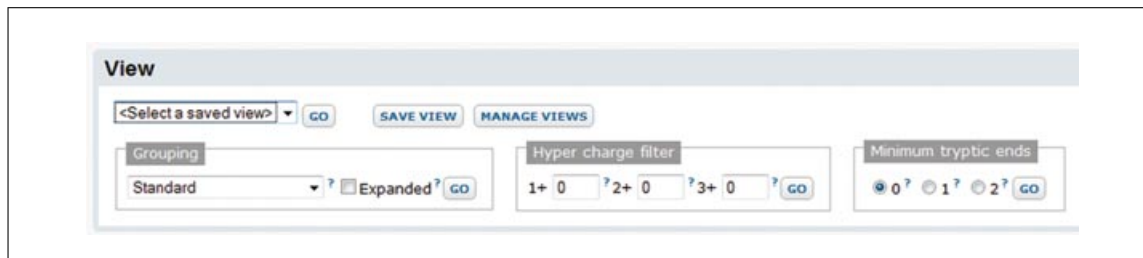
Figure 13.5.10 View of run details in the MS2 Viewer.

- To view the processed MS/MS sample data, go to the MS2 Dashboard, the portal page in the MS2Folder. Under MS2 Experiment Runs, click the link that reads `Peroxisomal_ICAT/MM_cIICAT11 (k_Yeast_ICAT)` to display the data from the sample run. Details about the run appear near the top of the page, as shown in Figure 13.5.9.

*Some of the peptides identified in the run details page may have special characters within the peptide string, for example, the single apostrophe in the peptide, `K.TC'SASWPR.A`. These special characters indicate the presence and position of a post-translational modification, one you asked the search engine to consider in the X! Tandem search parameters file. To see translations of these special characters, click Show Modifications.*

*To see the X! Tandem parameters used for the search, click the "Show tandem.xml" button. To see sensitivity and error rate information for the PeptideProphet and ProteinProphet analyses, click the respective Details button, shown in Figure 13.5.10.*

- In the View section of the screen, shown in Figure 13.5.11, make sure that the Grouping option is set to Standard, as it is by default. When Grouping is set to Standard, the grid below is labeled Peptides and displays peptides and their associated scores.



**Figure 13.5.11** Specify how to view peptide and protein data.

Scan	Z	Hyper	Next	B	Y	Expect	Ion%	CalcMH+	dMass	PepProphet	Peptide	SeqHits Protein
1812	2+	615.000	299.000	1.000	2.000	0.002	100%	1335.6720	+1.6250	0.9996	R.CTVEYDDSR.I	1 UPSP: PXA2_YEAST
2668	2+	636.000	331.000	1.000	2.000	0.001	73%	1720.9530	+1.1450	0.9996	R.GDLVLDVGC'GVGGPAR.E	1 UPSP: ERG6_YEAST
3599	2+	474.000	210.000	1.000	1.000	0.000	59%	2065.1260	+1.9260	0.9995	R.LNSASAVAC'GYLQAFVSK.T	1 UPSP: YB80_YEAST
3658	2+	556.000	271.000	1.000	2.000	0.001	63%	2412.2270	+1.0200	0.9995	R.QGYGLTETFAGVCLSEPFK.D	1 UPSP: LCF2_YEAST
3790	2+	563.000	276.000	1.000	2.000	0.001	43%	2936.4250	+1.0220	0.9995	K.GYDDVVVQC'TWEGDNNVLSLTSAK.S	1 UPSP: ACOX_YEAST
4082	2+	582.000	298.000	1.000	2.000	0.004	59%	2100.2660	+1.2250	0.9995	R.VLPQLCVPYLVSPVAFK.L	1 UPSP: ACOX_YEAST
1885	2+	577.000	317.000	0.000	2.000	0.002	91%	1751.8310	+1.6110	0.9994	K.YYINNEEDCAYK.T	1 UPSP: TCB1_YEAST
2129	2+	608.000	359.000	2.000	1.000	0.001	54%	1739.9110	+1.9280	0.9994	R.Q^C'SSGLTAVNDIANK.I	1 UPSP: THIK_YEAST
2188	2+	526.000	298.000	1.000	2.000	0.004	80%	1829.9790	+1.7370	0.9992	R.SNCIAPGAIDNTEGLK.R	1 UPSP: SPS19_YEAST

**Figure 13.5.12** View of peptide results by PeptideProphet score.

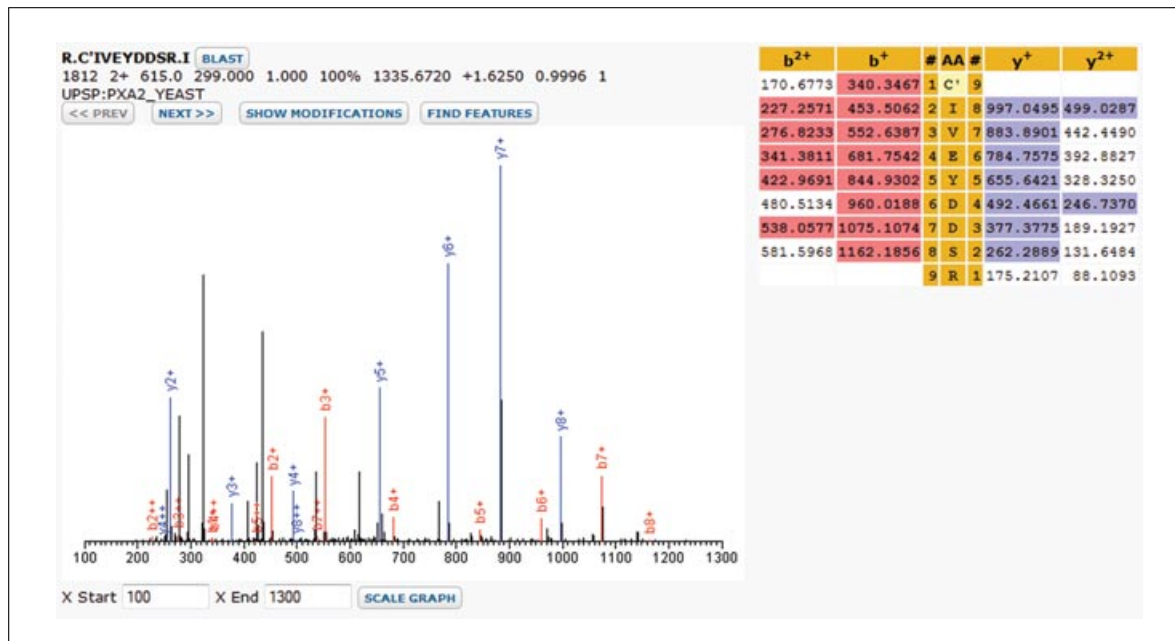
**Figure 13.5.13** Screen for customizing view of peptide results.

*The other Grouping options include legacy options (for compatibility with views in previous releases of LabKey Server) and a Protein Groups option that shows ProteinProphet results without the peptides data. Most users will want to leave the Grouping option set to Standard.*

*The View section of the screen also includes options for selecting hyper charge filters and setting the minimum number of tryptic ends.*

### Analyze MS/MS sample data

- Examine the results from the PeptideProphet analysis. Peptide data for the run appears in the Peptides section. To view the peptides scored with the highest confidence



**Figure 13.5.14** View of spectrum information for an individual peptide. For color version of this figure see <http://www.currentprotocols.com/protocol/bi1305>.

by PeptideProphet, locate the PepProphet column in the Peptides section. Click the column heading link and select Sort Descending to sort in descending order.

- Once the search results are sorted in descending order, the peptides that were scored by PeptideProphet with high confidence levels (ranging from 0.9996 downward) appear at the top of the section. Figure 13.5.12 shows a portion of these results.
- In addition to sorting and filtering a grid by clicking on column headers, you can incorporate sorts, filters and custom column arrangements into temporary or saved custom views for any data grid. Choose the Views button at the top of any grid and select the Customize View option. Use the tabs at the left to toggle between the Column, Filter, and Sort options, then click View Grid when you are finished to return to the grid view.

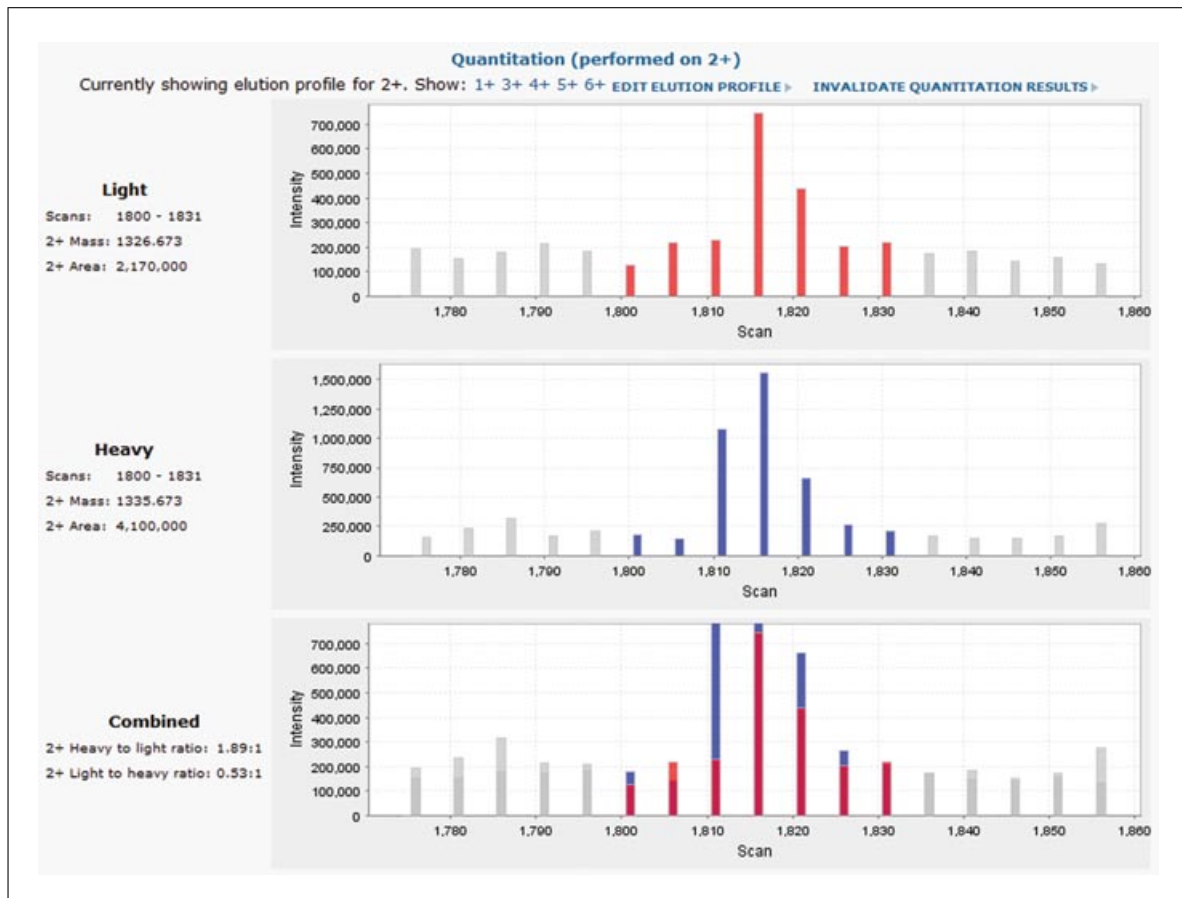
*Since you have already added one sort to the displayed grid, this sort setting appears on the Sort tab of the view customization interface, as shown in Figure 13.5.13.*

*Note that some useful columns have been hidden to display this data set, but they can be displayed again. On the Column tab, select the columns to add and click View Grid.*

*A specific combination of column layout, sorting, and filtering settings can be saved as a named view that can subsequently be accessed through a drop-down list. To save a new named view, select Save instead of View Grid in the view designer. The saved view will be available under the Views menu for the grid. The saved view can be applied to viewing, comparing, and exporting other runs or groups of runs. With this feature, consistent analyses can be performed across datasets.*

- To view detailed results for a given peptide, click the link for the peptide sequence in the Peptides data grid. Spectra information and quantitation results appear as shown in Figures 13.5.14 and 13.5.15, respectively.

*These figures both show results for labeled quantitation using XPRESS. LabKey Server also supports using Q3 for labeled quantitation, or using spectrum counting to perform label-free quantitation.*



**Figure 13.5.15** View of quantitation information for an individual peptide.

- To explore protein identifications, go back to the run view for the `MM_clICAT11.pep.xml` file and look at the Peptides grid. Two different mechanisms have been used to identify proteins from peptides: (1) the search engine and (2) ProteinProphet. The search engine chooses one protein from the database that “best” matches the peptide. This protein is shown in the “Protein” column in the default view of the Peptides grid. The ProteinProphet tool uses a more sophisticated algorithm and considers multiple possible proteins that match a given set of peptides. To view the results of the ProteinProphet analysis, go to the Peptides grid, click Views, and select ProteinProphet. You will see scored results for proteins found in the sample, as shown in Figure 13.5.16. The proteins that have high probabilities are very likely to have been present in the original sample.

*Search engine protein identifications are often easier to interpret than ProteinProphet results because they form a simple one-to-many relationship with their corresponding peptides. ProteinProphet results, on the other hand, provide a more sophisticated model of sample components by forming a probability-weighted, many-to-many relationship between proteins and peptides.*

- To view the peptides that were found in the sample and that contributed to the identification of a protein, expand the protein group by clicking the “+” button. Figure 13.5.17 shows the expanded protein group. There will often be more than one peptide per group.
- To see how the individual peptides identified map to the protein sequence, click the protein name link. The peptides found are indicated in colored font in the sequence shown in the UPSP:ACOX\_YEAST section, as shown in Figure 13.5.18. The number

Peptides										
VIEWS		PAGE SIZE		EXPORT ALL		EXPORT SELECTED		1 - 72 of 72		
Group	Prob	Spectrum	Protein	Sequence	Mass	Peptides	Unique	AACoverage	Best Name	Best Gene Name Description
			Ids%							
1	1.0	0.0572	UPSP:ACOX_YEAST	84042.1739	4	3	0.076	UPSP:ACOX_YEAST	P13711	saccharomyces cerevisiae
2	1.0	0.0287	UPSP:ERG6_YEAST	43430.8297	2	2	0.089	UPSP:ERG6_YEAST	P25087	saccharomyces cerevisiae
3	1.0	0.0690	UPSP:FOX2_YEAST	98703.1764	5	4	0.062	UPSP:FOX2_YEAST	Q02207	saccharomyces cerevisiae
4	1.0	0.0286	UPSP:HEMH_YEAST	44596.1003	2	2	0.076	UPSP:HEMH_YEAST	P16622	saccharomyces cerevisiae
5	1.0	0.0715	UPSP:LCF2_YEAST	83437.9588	5	5	0.102	UPSP:LCF2_YEAST	P39518	saccharomyces cerevisiae
6	1.0	0.0416	UPSP:POR15_YEAST	172255.7338	3	3	0.024	UPSP:POR15_YEAST	Q04182	saccharomyces cerevisiae

Figure 13.5.16 View of ProteinProphet results.

Peptides																																																																											
VIEWS		PAGE SIZE		EXPORT ALL		EXPORT SELECTED		1 - 72 of 72																																																																			
Group	Prob	Spectrum	Protein	Sequence	Mass	Peptides	Unique	AACoverage	Best Name	Best Gene Name Description																																																																	
			Ids%																																																																								
1	1.0	0.0572	UPSP:ACOX_YEAST	84042.1739	4	3	0.076	UPSP:ACOX_YEAST	P13711	saccharomyces cerevisiae (ba																																																																	
<table border="1"> <thead> <tr> <th>Scan</th> <th>Expect</th> <th>Hyper</th> <th>Next</th> <th>B</th> <th>Y</th> <th>Ion%</th> <th>CalcMH+</th> <th>dMass</th> <th>PepProphet</th> <th>Peptide</th> <th>SeqHits</th> <th>Protein</th> </tr> </thead> <tbody> <tr> <td>3790</td> <td>0.001</td> <td>563.000</td> <td>276.000</td> <td>1.000</td> <td>2.000</td> <td>43%</td> <td>2936.4250</td> <td>+1.0220</td> <td>0.9995</td> <td>K.GYDDWVYVQCTWEGDNNVLSLSAK.S</td> <td>1</td> <td>UPSP:ACOX_YEAST</td> </tr> <tr> <td>4082</td> <td>0.004</td> <td>582.000</td> <td>298.000</td> <td>1.000</td> <td>2.000</td> <td>59%</td> <td>2100.2660</td> <td>+1.2250</td> <td>0.9995</td> <td>R.VLPQLCVPLVSPVAFK.L</td> <td>1</td> <td>UPSP:ACOX_YEAST</td> </tr> <tr> <td>1892</td> <td>0.000</td> <td>546.000</td> <td>281.000</td> <td>1.000</td> <td>1.000</td> <td>73%</td> <td>1922.8970</td> <td>+0.9740</td> <td>0.9991</td> <td>R.Q^TTCGGHGYSQYNGFGK.G</td> <td>1</td> <td>UPSP:ACOX_YEAST</td> </tr> <tr> <td>1893</td> <td>0.034</td> <td>385.000</td> <td>276.000</td> <td>1.000</td> <td>1.000</td> <td>47%</td> <td>1913.8970</td> <td>+1.6480</td> <td>0.9803</td> <td>R.Q^TCGGHGYSQYNGFGK.G</td> <td>1</td> <td>UPSP:ACOX_YEAST</td> </tr> </tbody> </table>											Scan	Expect	Hyper	Next	B	Y	Ion%	CalcMH+	dMass	PepProphet	Peptide	SeqHits	Protein	3790	0.001	563.000	276.000	1.000	2.000	43%	2936.4250	+1.0220	0.9995	K.GYDDWVYVQCTWEGDNNVLSLSAK.S	1	UPSP:ACOX_YEAST	4082	0.004	582.000	298.000	1.000	2.000	59%	2100.2660	+1.2250	0.9995	R.VLPQLCVPLVSPVAFK.L	1	UPSP:ACOX_YEAST	1892	0.000	546.000	281.000	1.000	1.000	73%	1922.8970	+0.9740	0.9991	R.Q^TTCGGHGYSQYNGFGK.G	1	UPSP:ACOX_YEAST	1893	0.034	385.000	276.000	1.000	1.000	47%	1913.8970	+1.6480	0.9803	R.Q^TCGGHGYSQYNGFGK.G	1	UPSP:ACOX_YEAST
Scan	Expect	Hyper	Next	B	Y	Ion%	CalcMH+	dMass	PepProphet	Peptide	SeqHits	Protein																																																															
3790	0.001	563.000	276.000	1.000	2.000	43%	2936.4250	+1.0220	0.9995	K.GYDDWVYVQCTWEGDNNVLSLSAK.S	1	UPSP:ACOX_YEAST																																																															
4082	0.004	582.000	298.000	1.000	2.000	59%	2100.2660	+1.2250	0.9995	R.VLPQLCVPLVSPVAFK.L	1	UPSP:ACOX_YEAST																																																															
1892	0.000	546.000	281.000	1.000	1.000	73%	1922.8970	+0.9740	0.9991	R.Q^TTCGGHGYSQYNGFGK.G	1	UPSP:ACOX_YEAST																																																															
1893	0.034	385.000	276.000	1.000	1.000	47%	1913.8970	+1.6480	0.9803	R.Q^TCGGHGYSQYNGFGK.G	1	UPSP:ACOX_YEAST																																																															
2	1.0	0.0287	UPSP:ERG6_YEAST	43430.8297	2	2	0.089	UPSP:ERG6_YEAST	P25087	saccharomyces cerevisiae (ba																																																																	
3	1.0	0.0690	UPSP:FOX2_YEAST	98703.1764	5	4	0.062	UPSP:FOX2_YEAST	Q02207	saccharomyces cerevisiae (ba																																																																	

Figure 13.5.17 View of peptides by protein assignments.

UPSP:ACOX_YEAST	
Sequence Mass:	84,042
AA Coverage:	8% (57 / 748)
Mass Coverage:	7% (6,224 / 84,042)
MTRRTTINPD SVVLPQKFI QKERADSKIK VDQVNTFLES SPERRTLTHA LIDQIVNDPI LKTDTDYYDA KMQEREITA KKIARLASYM EHDIKTVRKH FRDSDLKMKEL QANDPDKASP LTNKDLFIFD KRLSLVANID PQLGTRVGVH LGLFGNCIRG NGTDEQIRYW LQERGATLMK GIYGCFAFTE LGHGSNVAQL QTRAVYDKQN DTFVIDPDL TATRWIGGA AHSATHAAVY ARLIVEGKDY GVKTFVVPLR DPSTFQLLAG VSIQDYGAKM GRDGDINGWI QFRNVVPIRE FMLSRFTKVV RSPDGSVTVK TEPQLDQISG YSALLSGRVN MVMDSFRFGS KFATIAVRYA VGRQQFAPRK GLSETQLIDY FLHQYRVLPQ LCVPLVLSFV AFKLMNDNYS TLDELYNASS SAYKAALVTV SKKLNLFID SASLKATNW LIATLIDELR <b>QTCGGHGYSQ YNGFGRGYDD WVVQCTWEGD NNVLSTSAK</b> SILKFKIDSA TKGRFDNTLD VDSFSYLKQP YIGSVVSGEI KSGLKEGDDY TEIWSITLIK LLAHIGTLVE KRSRIDSVSK LLVLVSKFHA LRCLMKTYYD KLNRSRSHIS DEITRESMWN VYKLFSLYFI DKHSGEFQQF KIFTPDQISK VVQQLLALL PIVRKDCIGL TDSFELPDAM LNSPIGYFDG DIYHNYFNEV CRNNPVEADG AGKPSYHALL SSMLGRGFEP DQKLGGAANA EILSKINK	

Figure 13.5.18 View of protein sequence information. For color version of this figure see <http://www.currentprotocols.com/protocol/bi1305>.

of times these peptides were observed is shown in the colored bars in the Protein Sequence section, as shown in Figure 13.5.19.

Expand the Annotations section of the page to see annotations associated with the protein, as shown in Figure 13.5.20. The annotations that appear here came from both the yeast.nci.20070223.fasta file and the yeast.a.xml UniProt annotation file. In this case, the description and Swiss-Prot Names came from the FASTA file, while the Swiss-Prot Accessions and Ensembl information came from the UniProt XML.

In addition to loading protein annotations from public data sources (e.g., UniProt XML annotations), LabKey Server can also load custom protein annotations. Both custom protein annotations and annotation from public sources can be joined into grid views of experimental results that show protein-level information. Custom annotations are commonly used to select subsets of proteins for inclusion or exclusion in the results displayed.

11. Protein data from multiple runs can be compared to show how many unique peptides contributing to a protein assignment were found in each run as well as the total number of peptides found.



## Compare ProteinProphet

### Comparison Overview

### Comparison Details

VIEWS EXPORT PRINT PAGE SIZE 1 - 91 of 91

Protein Information		Runs					
		MM_clcICAT11.pep.xml - Peroxisomal_ICAT/MM_clcICAT11 (k_Yeast_ICAT)		MM_clcICAT12.pep.xml - Peroxisomal_ICAT/MM_clcICAT12 (k_Yeast_ICAT)		MM_clcICAT13.pep.xml - Peroxisomal_ICAT/MM_clcICAT13 (k_Yeast_ICAT)	
Protein	Num Runs	Group	Prob	Group	Prob	Group	Prob
UPSP:AATC_YEAST	3	21	0.9956	24	0.9969	29	0.9957
UPSP:ACOX_YEAST	3	1	1.0	1	1.0	1	1.0
UPSP:ADT2_YEAST	3	11-1	1.0	2	1.0	2	1.0
UPSP:DHSB_YEAST	3	24	0.9928	43	0.9669	20	0.9993
UPSP:ECI1_YEAST	3	30	0.9762	5	1.0	4	1.0

Figure 13.5.21 A comparison of proteins from multiple runs.

## Compare ProteinProphet

### Comparison Overview

### Comparison Details

VIEWS EXPORT PRINT PAGE SIZE 1 - 91 of 91

Columns Available Fields

Filter

Sort

- Protein
- Num Runs
- Protein Group
- Group
- Prob
- Unique
- Peptides
- Spectrum Ids%
- AACoverage
- Protein Prob
- Error
- Protein Prophet

Show Hidden Fields

DELETE REVERT VIEW GRID SAVE

Selected Fields

- Protein
- Num Runs
- Group
- Prob
- Unique

Protein Information		Runs					
		MM_clcICAT11.pep.xml - Peroxisomal_ICAT/MM_clcICAT11 (k_Yeast_ICAT)		MM_clcICAT12.pep.xml - Peroxisomal_ICAT/MM_clcICAT12 (k_Yeast_ICAT)		MM_clcICAT13.pep.xml - Peroxisomal_ICAT/MM_clcICAT13 (k_Yeast_ICAT)	
Protein	Num Runs	Group	Prob	Group	Prob	Group	Prob
UPSP:AATC_YEAST	3	21	0.9956	24	0.9969	29	0.9957
UPSP:ACOX_YEAST	3	1	1.0	1	1.0	1	1.0
UPSP:ADT2_YEAST	3	11-1	1.0	2	1.0	2	1.0

Figure 13.5.22 Screen for customizing your view of proteins from multiple runs and adding a column.

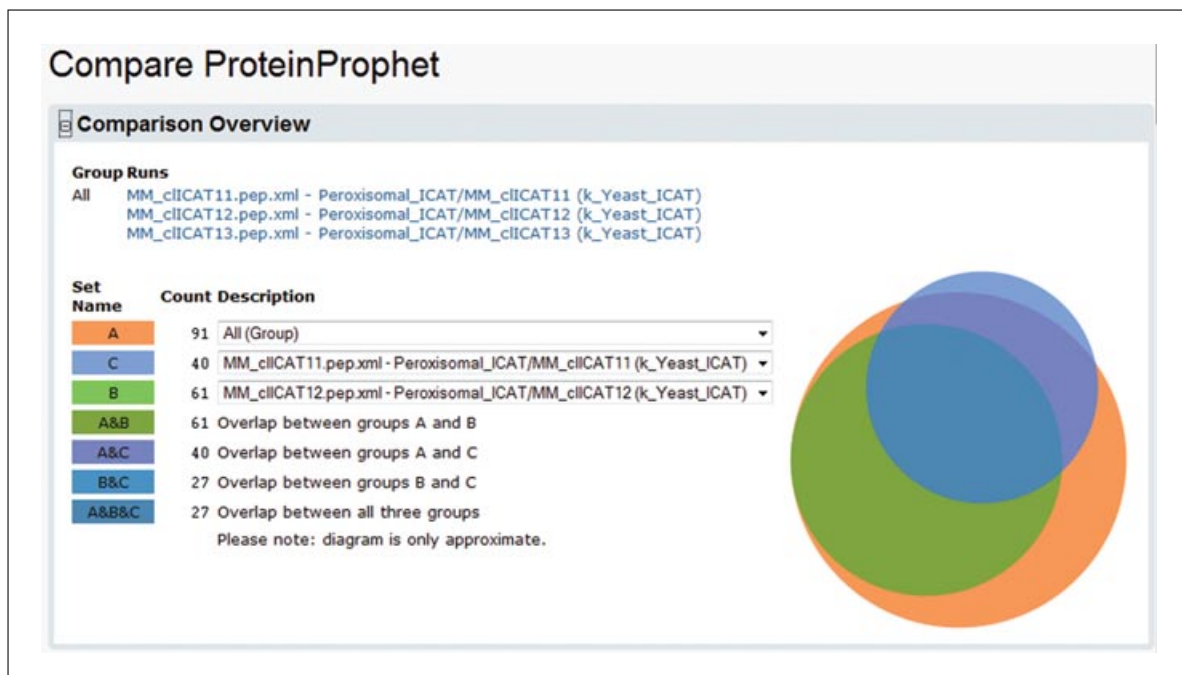
14. Data can be exported in several formats. Users can export a single run, a subset of the data from a single run, or data from multiple runs.

*To export data from a single run, navigate to the MS2 Dashboard and click on a single run in the MS2 Experiment Runs section of the page. Arrange the view to display the data in the desired configuration. Select the rows to export, click Export Selected, and select the file format (Excel, TSV or AMT). Alternatively, click Export All and select the format.*

*To export data from more than one run, return to the MS2 Dashboard. Select the runs to export from the MS2 Experiment Runs section of the page, then click the MS2 Export button. Apply the desired saved view to specify how the data should be filtered and sorted and which columns should be exported. Choose the format for exporting (Excel, TSV, DTA, PKL or AMT), and click the Go button to export to that format.*

Comparison Details										
Views EXPORT PRINT PAGE SIZE 1 - 91 of 91										
The current <default> view is unsaved. Revert, Edit, Save										
Protein Information		Runs								
		MM_clICAT11.pep.xml - Peroxisomal_ICAT/MM_clICAT11 (k_Yeast_ICAT)			MM_clICAT12.pep.xml - Peroxisomal_ICAT/MM_clICAT12 (k_Yeast_ICAT)			MM_clICAT13.pep.xml - Peroxisomal_ICAT/MM_clICAT13 (k_Yeast_ICAT)		
Protein	Num Runs	Group	Prob	Unique	Group	Prob	Unique	Group	Prob	Unique
UPSP:AATC_YEAST	3	21	0.9956	1	24	0.9969	1	29	0.9957	1
UPSP:ACOX_YEAST	3	1	1.0	3	1	1.0	3	1	1.0	3
UPSP:ADT2_YEAST	3	11-1	1.0	2	2	1.0	2	2	1.0	2
UPSP:DHSB_YEAST	3	24	0.9928	1	43	0.9669	1	20	0.9993	1
UPSP:ECI1_YEAST	3	30	0.9762	1	5	1.0	1	4	1.0	2
UPSP:ERG6_YEAST	3	2	1.0	2	6	1.0	2	34	0.994	2

**Figure 13.5.23** A comparison of proteins from multiple runs with an added column, Unique.



**Figure 13.5.24** A Venn diagram of runs compared with ProteinProphet.

## INSTALL AND CONFIGURE LabKey SERVER

LabKey Server runs on Windows, Linux, MacOS X or Solaris. A binary installer is available for installing LabKey Server on a PC running Microsoft Windows, which is described in this protocol.

LabKey Server can be installed manually on any supported operating system. For detailed information on installing LabKey Server manually, see the LabKey Server documentation at <https://www.labkey.org/wiki/home/Documentation/page.view?name=manualInstall>.

### Necessary Resources

#### Hardware

LabKey Server is a Web application that runs on Apache Tomcat and accesses a PostgreSQL or Microsoft SQL Server database server. The resource requirements for the Web application itself are minimal, but the computer on which LabKey Server is installed must have sufficient resources to run Apache Tomcat and the database server. The following system resources are recommended:

## SUPPORT PROTOCOL

### Using Proteomics Techniques

## 13.5.17

Processor: a high-performing processor  
Physical memory: at least 2 Gb RAM, preferably 4 Gb  
Disk space: 1 Gb hard drive space free

### Software

The LabKey Server open source software source code is freely available for download at <http://www.labkey.org> under the terms of the Apache License 2.0[21].

Compiled binaries for Windows, Unix, Linux, or Macintosh installation are available for free through LabKey Software at <http://www.labkey.com>. A graphical installer is available for computers running Windows XP or later.

Installation of LabKey Server v11.1 using the graphical Windows binary installer will install the following components on the computer:

- The Apache Tomcat Web server, version 5.5.29
- The PostgreSQL database server, version 8.3.7
- The Java Runtime Environment, version 1.6.0-22
- The LabKey Server Web application components, version 11.1

Additional third-party components required by LabKey Server, including the Institute for Systems Biology's Trans-Proteomic Pipeline. These components are described in the online documentation

### Files

No local files are required

**NOTE:** LabKey Server can also be run against Microsoft SQL Server 2005, 2008 or 2008 R2. To do so, the LabKey Server configuration file requires modification after the installation is complete. Instructions for modifying the configuration file are available in the LabKey Server online documentation. Running against PostgreSQL 8.3, 8.4 and 9.0 is also supported, but no modifications to the configuration file are required for these platforms. For further information on supported versions of Web servers and database servers, see the LabKey Server online documentation.

### **Install LabKey Server using the Windows graphical installer**

1. Download the graphical LabKey Server installer for Windows from <http://labkey.com/forms/register-to-download-labkey-server>. One of two installers can be selected, depending on whether the computer already has an existing installation of the Java Runtime Environment (JRE). If you are unsure if you already have the JRE, choose the option that includes the JRE.
2. Run the installer. Accept the warning that the installation will allow your machine to act as a Web server and a database server. Accept the license.
3. Provide connection information for an outgoing (SMTP) mail server. The mail server is used to send e-mail generated by the LabKey Server system, including e-mails sent to new users when they are given accounts on LabKey Server. The installer will prompt specification of an SMTP host, port number, and address from which automated system e-mails are to be sent.

*LabKey Server can operate without an SMTP server connection, but cannot send system e-mails. Even if there is no SMTP server available, it is necessary to specify an e-mail address for the From address field during the installation process.*

*Internet Information Server, included with recent versions of Windows, includes a basic SMTP server which can be configured locally. For more information, see the LabKey Server online documentation or the Windows documentation.*

4. Provide a user name and password for the database superuser for PostgreSQL, the database server which is installed by the installer.

*The PostgreSQL superuser is a user who is allowed all rights, in all databases, including the right to create user accounts. It may be prudent to write down the user name and password for safekeeping.*
5. Provide a user name and password for the Windows service user.

*Note that LabKey Server is installed as a Windows service and must run under a unique Windows user account; an existing user account cannot be specified.*
6. Complete the installation, then open a Web browser and navigate to <http://localhost:8080/labkey> to begin using LabKey Server.
7. Set up a new account on LabKey Server when prompted by entering an e-mail address and a password. The new account is created with global administrative privileges. Remember this password as it is required to log into LabKey Server.
8. Click Install to configure all LabKey Server modules. When all modules are updated, click Next.
9. Use the default site-level file root to set up a standard directory structure. Click Next.

*Once the site-level file root is set, LabKey Server configures corresponding subdirectories in the file system for each project and folder on the server. This allows you to upload, store, and process data files, including proteomics files.*
10. Customize settings for the LabKey Server installation. Click the Save button to save the settings and continue.

*The Customize Settings page provides properties for modifying the look and feel of the LabKey Server site; setting system properties and notification options; and configuring Lightweight Directory Access Protocol (LDAP) and Secure Socket Layer (SSL). These properties can be set from LabKey Server at any time by clicking Manage Site in the left navigation pane, clicking Admin Console, then clicking the Site Settings link.*

*The LabKey Server home page should appear in the browser as shown in Figure 13.5.1.*

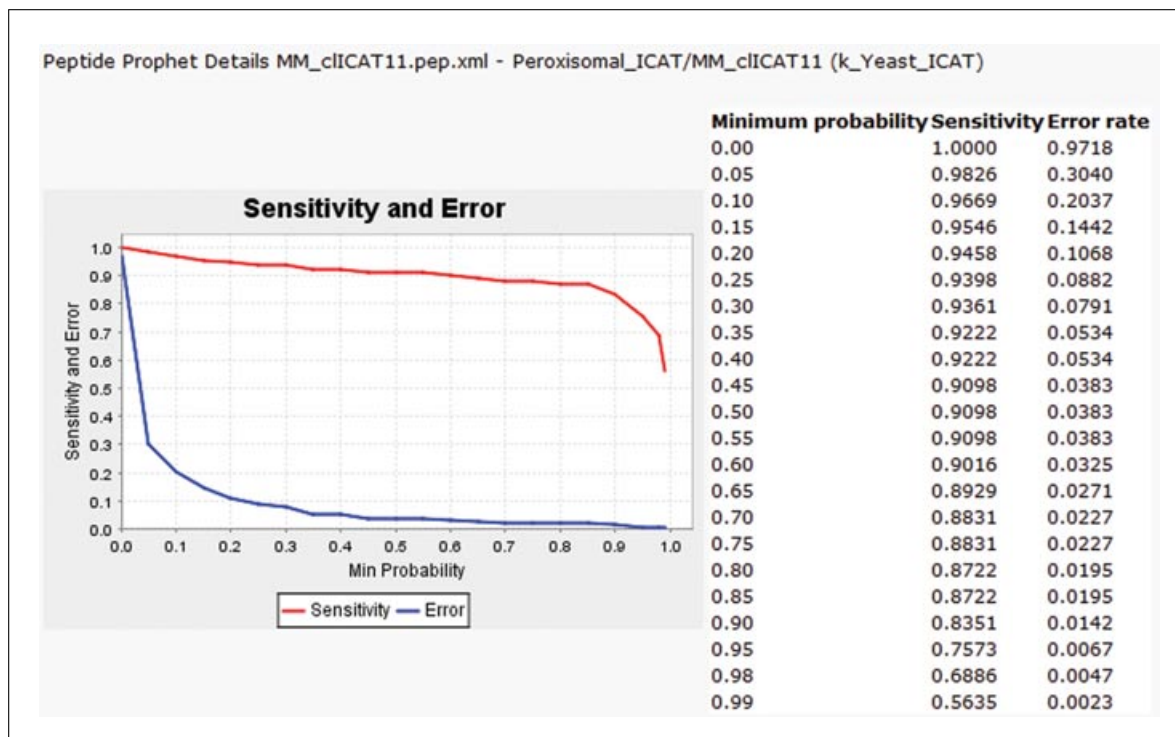
## GUIDELINES FOR UNDERSTANDING RESULTS

Both peptide- and protein-centric results are available for analysis in LabKey Server. For peptide level analysis, a list of positive identifications can be determined by filtering based on native database search scores as well as by the PeptideProphet probability scores. ProteinProphet probability scores are used to determine statistical significance at the protein level.

### Peptide Level Analysis

The PeptideProphet algorithm (Keller et al., 2002) takes an entire search result set as input, assumes a mixture model of positive and negative identifications are present, models positive and negative distributions from the dataset, and assigns probabilities based on the modeled distributions. Figure 13.5.25 displays a sample PeptideProphet details view within LabKey Server.

To filter results to achieve a 5% error rate, the details view indicates that a probability cutoff of 0.60 should be applied. An alternate analysis strategy is to apply a stringent filter such as a cutoff of 0.95 probability; the details view indicates that this corresponds to an error rate of approximately 0.32%.



**Figure 13.5.25** View of sensitivity and error rate information for the PeptideProphet analysis. For color version of this figure see <http://www.currentprotocols.com/protocol/bi1305>.

The PeptideProphet details view also shows graphs that compare the observed data with the positive and negative identifications predicted by the model for each charge state (+1, +2, and +3), as shown in Figure 13.5.26.

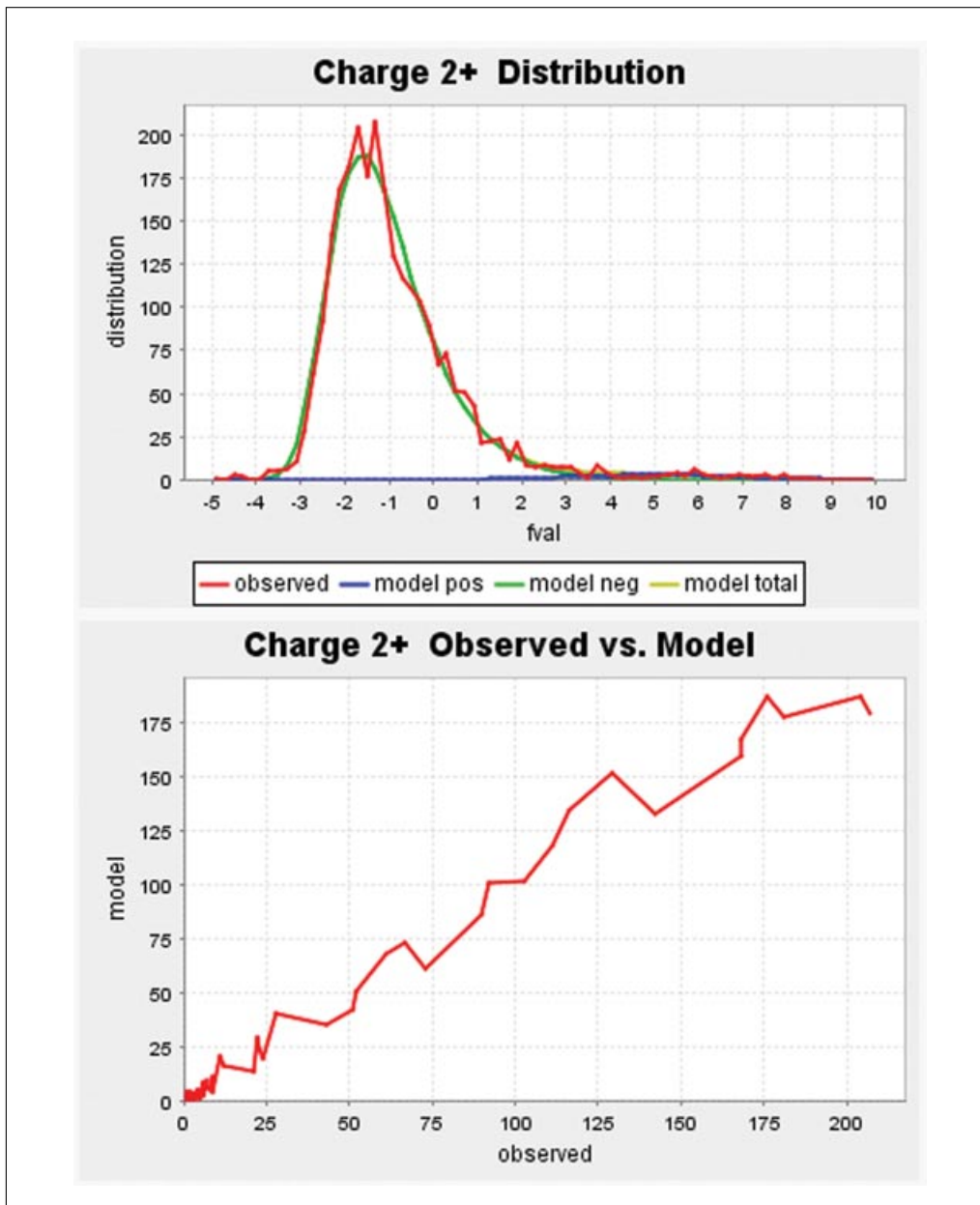
For a typical LC-MS/MS dataset usually only a small subset (10% to 20%) of the tandem mass spectra return positive identifications. Given that mass spectrometers quickly acquire many thousands of spectra, it is imperative to apply reasonable search filters to minimize false positives.

### Protein Level Analysis

A more relevant way to interrogate results, especially for large scale MS/MS datasets, is to skip the peptide level analysis completely and focus directly on the protein level. LabKey Server's details view for ProteinProphet (Nesvizhskii et al., 2003) displays identified proteins with their confidence level in terms of a probability score as well as links to the underlying peptide identifications associated with each protein (or protein group). As with the peptide analysis strategies outlined above, users can select a specific filter based on a minimum protein probability score and choose a targeted error rate that defines the protein probability cutoff to be applied.

By applying simple protein level filters, one can rapidly isolate a list of putative protein identifications. These protein identifications can be viewed with or without the supporting peptide sequences.

Note that MS/MS database search results should be interpreted as possible or likely peptide and protein identifications and not treated as final confirmation of identity. Ideally, orthogonal validation methods, such as an immunoblot, can be applied to validate putative protein identifications when warranted by the experiment. LabKey Server does provide links back to the raw spectrum to assist in validating any individual match at the spectrum level.



**Figure 13.5.26** View of graphs showing distribution of PeptideProphet scores. For color version of this figure see <http://www.currentprotocols.com/protocol/bi1305>.

## COMMENTARY

### Background Information

LabKey Server is an open-source, Web-based system that provides an extensive set of tools for processing, storing, integrating, and analyzing MS/MS experiment results and other types of data (Nelson et al., 2011). It is operating system- and database-independent, easy to install, and compliant with multiple standard file formats. It provides a securable platform for mining, sharing, and publishing data from high-throughput experimental environments. A wide range of data types, such as experimental results and clinical records,

can be integrated into common views. As covered in Basic Protocol 2, LabKey Server allows users to draw together columns from related data sources into unified grid views, as well as to apply filters and sorts to these grids, all without knowing SQL. SQL-fluent users can also produce more advanced analyses using LabKey Server's SQL dialect. The platform is highly extensible and provides client libraries in a variety of languages, including JavaScript, Java, SAS, Perl, and R, plus a built-in interface for R scripting (Nelson et al., 2011).

LabKey Server incorporates open source analysis tools, such as X! Tandem and the Trans-Proteomic Pipeline (TPP), and makes use of open data formats, such as mzXML, pepXML, protXML, and FuGE. Alternative proteomic tools and data systems are typically either commercial in nature and lack related functionality, are not available, or are not easily accessed by life science researchers who do not have extensive computational expertise. Additionally, a LabKey Server installation provides an entire framework for proteomic analysis including primary identification using sequence database search, validation of peptide and protein identifications, and quantitation of stable isotope-labeled samples.

### Critical Parameters and Troubleshooting

LabKey Server itself does not perform analysis but rather interfaces with the other installed tools (X! Tandem, SEQUEST, Mascot and TPP) to analyze proteomic data. The primary analysis involves MS/MS sequence database searching. The critical parameters for any database search analysis include specifying: (1) which sequence database to use for searching, (2) relevant post-translational modifications to consider, and (3) precursor ion mass tolerance.

These parameters can be specified as part of executing an X! Tandem search (see step 13 in Basic Protocol 1). If you choose to create a new protocol, you can select a sequence database from FASTA files you have uploaded. The graphical interface also allows you to choose from pre-defined residue modifications or define your own. Modifications can be considered variable, so the search checks for both the standard and modified mass, or fixed, so the search looks for the modified mass. The interface also provides an entry field for XML that can define precursor ion mass tolerance and other search engine parameters, which are described here: <https://www.labkey.org/wiki/home/Documentation/page.view?name=pipelineParams>.

Correct proteins will not be identified if their sequence is not in the sequence database being queried, so it is imperative that an appropriate sequence database be specified. Typically a species-specific database (e.g., human sequence database for a human sample) is the optimal choice if the organism of the input sample is known. Otherwise a nonredundant protein database composed of all species sequences can be searched.

The choice of precursor mass tolerance is an instrument-specific setting. For low mass accuracy data, such as that acquired on an ion trap mass spectrometer, typical settings are  $\pm 1$  Da. High mass accuracy instruments, such as time-of-flight or Fourier transform instruments, will afford much smaller precursor mass tolerances if desired. The key with this parameter is to specify a mass tolerance that is not too narrow for the data being searched.

The LabKey Server data pipeline displays the status of jobs that are running, and also indicates when an error has occurred in the analysis. When pipeline errors do occur, diagnostic information is available in LabKey Server that helps to determine the nature of the problem. The pipeline status page provides a link to further information about each pipeline job as shown in Figure 13.5.27. Click on the link to view additional details about the pipeline job, including any log files generated.

### Advanced Parameters

#### *Quantitation in LabKey server*

In addition to identifying peptides and proteins, LabKey Server interfaces with XPRESS (Han et al., 2001) for the quantitation of proteins. Quantitation analysis involves specifying additional pipeline parameters for X! Tandem. Quantitation parameters define: (1) the residues which are isotopically labeled, (2) the mass difference between light and heavy isotopic forms, and (3) the mass tolerance to apply when reconstructing ion chromatograms for peak area quantitation. These parameters are specified with the “pipeline quantitation” tags in the analysis parameters.

Example parameters for an ICAT analysis (which labels cysteine residues and has a 9.0 Da mass difference between the light and heavy forms of the reagent) are shown in Figure 13.5.28.

Example parameters for a SILAC analysis (where both lysine and arginine residues are labeled) are shown in Figure 13.5.29.

Note the specification of different mass differences for the respective lysine and arginine labels.

#### *Custom scoring algorithms*

Protein identifications from tandem mass spectra rely on matching of observed mass spectra to protein sequence databases. Different search engines generate scores according

Data Pipeline		
Show: Running Errors All		
PROCESS AND IMPORT DATA SETUP		
Status	Created	Description
SEARCH WAITING	2011-06-01 15:16	Peroxisomal_ICAT/MM_cIICAT11 (k_Yeast_ICAT)
SEARCH RUNNING	2011-06-01 15:16	Peroxisomal_ICAT/MM_cIICAT13 (k_Yeast_ICAT)
SPLIT WAITING	2011-06-01 15:16	Peroxisomal_ICAT (k_Yeast_ICAT)

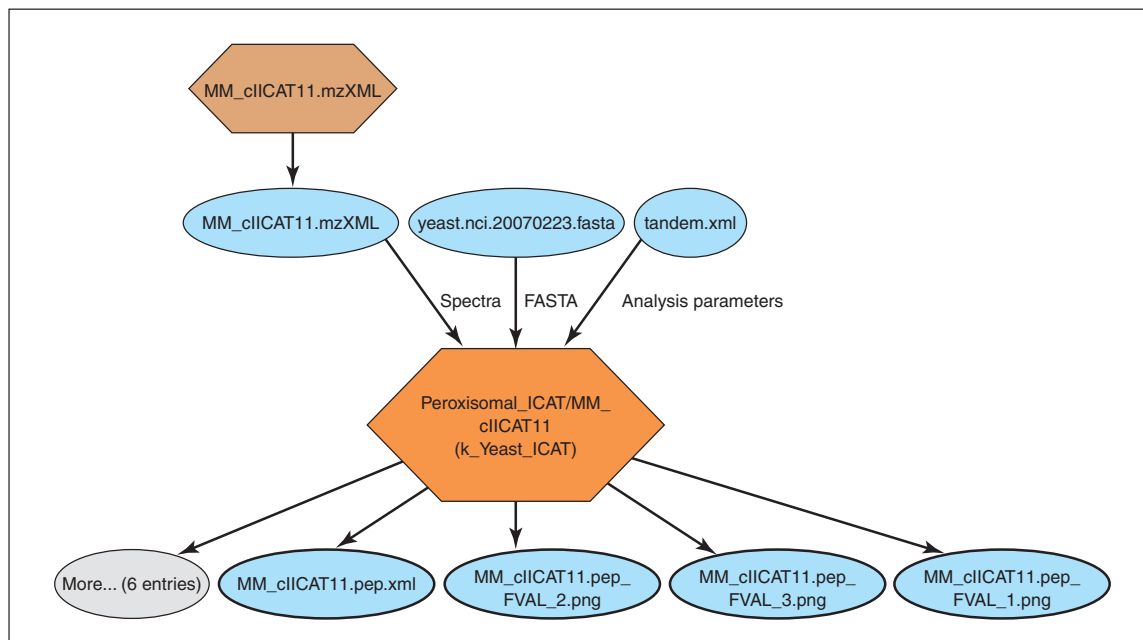
**Figure 13.5.27** View of job status and error information for the data pipeline.

```
<note label="pipeline quantitation, algorithm" type="input">xpress</note>
<note label="pipeline quantitation, residue label mass" type="input">9.00c</note>
<note label="pipeline quantitation, mass tolerance" type="input">1.0</note>
```

**Figure 13.5.28** Example parameters for an ICAT analysis.

```
<note label="pipeline quantitation, algorithm" type="input">xpress</note>
<note label="pipeline quantitation, residue label mass" type="input">6.00K,10.00R</note>
<note label="pipeline quantitation, mass tolerance" type="input">1.0</note>
```

**Figure 13.5.29** Example parameters for a SILAC analysis.



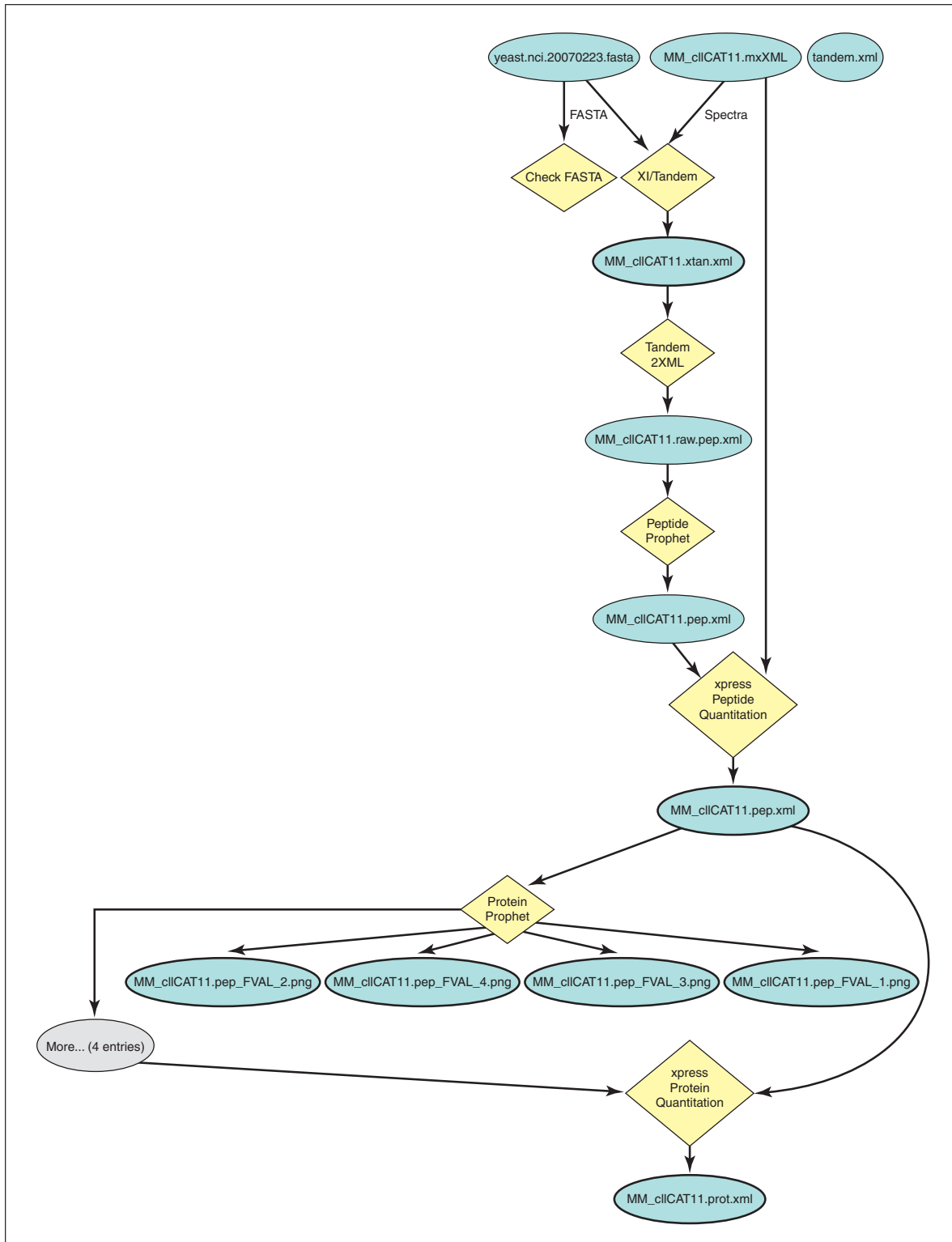
**Figure 13.5.30** Example of an experimental run graph.

to their respective algorithms that gauge the likelihood that a positive match has been made. A user-defined pluggable scoring algorithm (MacLean et al., 2006) has been written as an extension to the X! Tandem application (Craig and Beavis, 2004) to allow novel scoring algorithms to be tested. The search protocol can be configured to use a custom X! Tandem scor-

ing algorithm using the “scoring, algorithm” setting.

### Experiment descriptions

LabKey Server also provides an experimental annotation framework for describing the inputs (samples) and outputs (peptides, proteins, and genes) of MS2 experiments, and



**Figure 13.5.31** Example of an experimental run graph details view.

for describing the exact steps of a processing chain. It supports the association of inputs, protocol steps, and results with each other and with the samples on which the research is being conducted, with the ability to extend the information model to accommodate additional properties at many levels.

Figures 13.5.30 and 13.5.31 show an example of a processing chain used for the MS2 data employed in Basic Protocols 1 and 2. Figure 13.5.31 is the details view of the Peroxisomal\_ICAT/MM\_clcAT11 (k.Yeast\_ICAT) node in Figure 13.5.30. Nodes in the processing tree are clickable

and link directly to further details or source files. The steps required to set up and view Figures 13.5.30 and 13.5.31 are covered in the LabKey Server proteomics tutorial: <https://www.labkey.org/wiki/home/Documentation/page.view?name=cpastutorial>.

### Acknowledgments

The authors wish to thank the following individuals for their assistance: Andrea Detter, Brendan MacLean, George Snelling, Karl Lum, Stephen Hanson, Britt Piehler, Trey Chadick, Ren Lis, Kevin Krouse, Cory Nathe, Nick Arnold, Kristin Fitzsimmons, and Kim Carson.

### Literature Cited

- Craig, R. and Beavis, R.C. 2004. TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* 20:1466-1467.
- Eng, J. K., McCormack, A.L., and Yates, J.R. 3rd. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5:976-989.
- Han, D.K., Eng, J., Zhou, H., and Aebersold, R. 2001. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* 19:946-951.
- Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74:5383-5392.
- Keller, A., Eng, J., Zhang, N., Li, X.J., and Aebersold, R. 2005. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* 1:2005.0017.
- MacLean, B., Eng, J.K., Beavis, R.C., and McIntosh, M. 2006. General framework for developing and evaluating database scoring algorithms using the TANDEM search engine. *Bioinformatics* 22:2830-2832.
- Nelson, E.K., Piehler, B., Eckels, J., Rauch, A., Bellew, M., Hussey, P., Ramsay, S., Nathe, C., Lum, K., Krouse, K., Stearns, D., Connolly, B., Skillman, T., and Igra, M. 2011. LabKey server:

An open source platform for scientific data integration, analysis and collaboration. *BMC Bioinformatics* 12:71.

- Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. 2003. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75:4646-4658.
- Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 18:3551-3567.
- Rauch, A., Bellew, M., Eng, J., Fitzgibbon, M., Holzman, T., Hussey, P., Igra, M., MacLean, B., Lin, C.W., Detter, A., Fang, R., Faca, V., Gafken, P., Zhang, H., Whiteaker, J., States, D., Hanash, S., Paulovich, A., and McIntosh, M. 2006. Computational Proteomics Analysis System (CPAS): An extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. *J. Proteome Res.* 5:112-121.

### Internet Resources

- <https://www.labkey.org>  
*The LabKey Software Foundation site provides source code, documentation and community support forums for LabKey Server.*
- <https://www.labkey.com>  
*The LabKey Software site provides installers for LabKey Server and related components. LabKey Software develops LabKey Server.*
- <https://hosted.labkey.com>  
*LabKey Software's hosting site provides free, limited-time accounts for trial usage of LabKey Server.*
- <http://www.systemsbiology.org/>  
*The Institute for Systems Biology distributes a variety of proteomics tools, including XPRESS, PeptideProphet, ProteinProphet, and many mzXML conversion tools.*
- <http://www.thegpm.org/TANDEM/>  
*Open source software for matching tandem mass spectra with peptide sequences.*
- <https://proteomics.fhcr.org/CPAS/>  
*The Fred Hutchinson Cancer Research Center's Computational Proteomics Laboratory (CPL) in-stance of LabKey Server.*