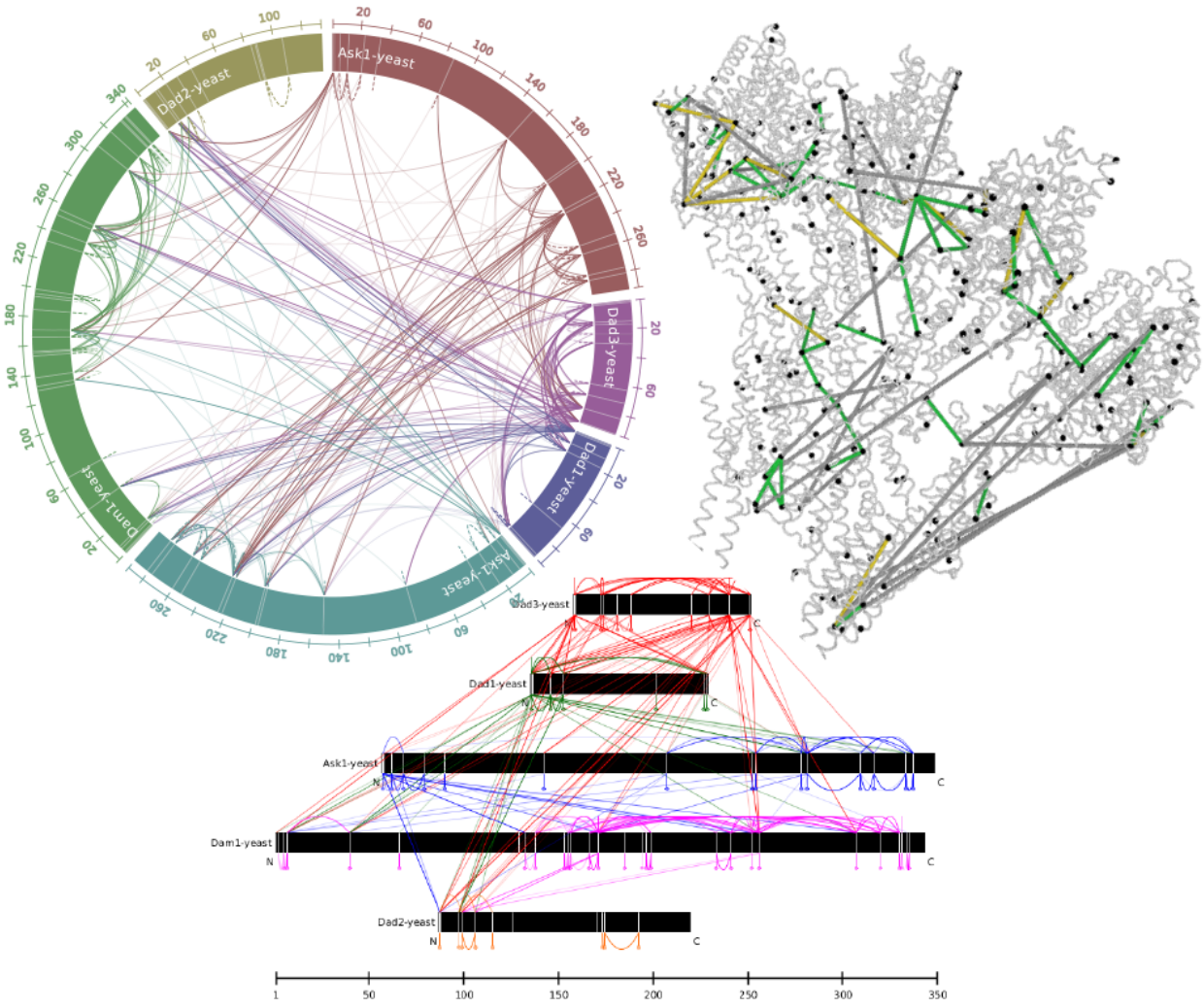
proxl Documentation

Release 1.0

Michael Riffle

Dec 15, 2021

Using proxl



Proxl is a web application and database for storing, viewing, analyzing, and sharing protein cross-linking mass spectrometry data.

Proxl is available to use free-of-charge at https://www.yeastrc.org/proxl_public/

Here you will find documentation for both using and installing proxl locally.

CHAPTER 1

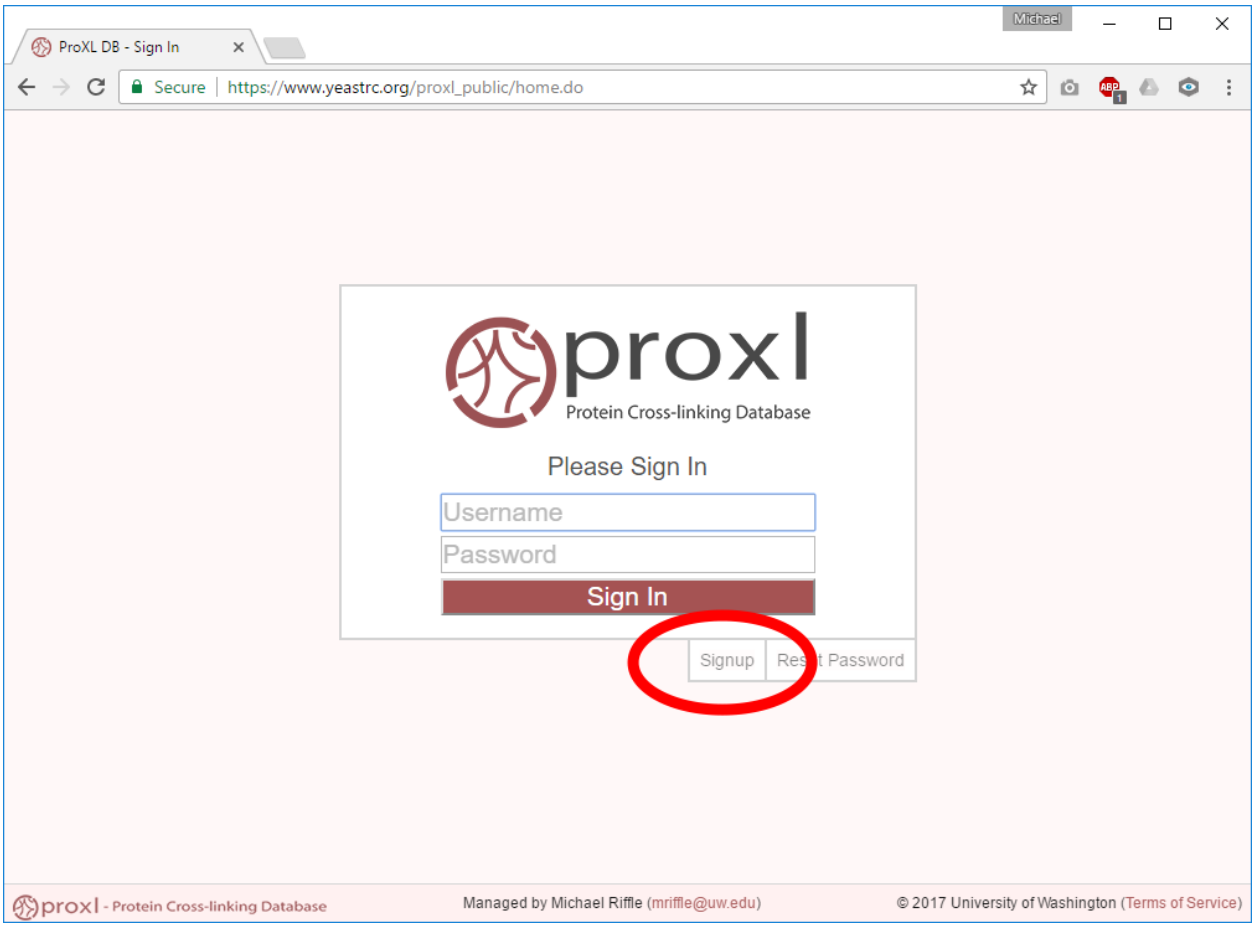
Getting Started

Proxl is a database and web application for viewing, analyzing, and sharing bottom-up proteomics data resulting from chemical crosslinking and mass spectrometry analysis. It includes tools for viewing, downloading, and visualizing data; including tools for comparing data between different searches and pipelines.

1.1 Gaining Access

1.1.1 Public Installation

A publicly-available version of proxl is available at https://www.yeastrc.org/proxl_public/. To create an account, click the "Signup" tab on the signin screen:

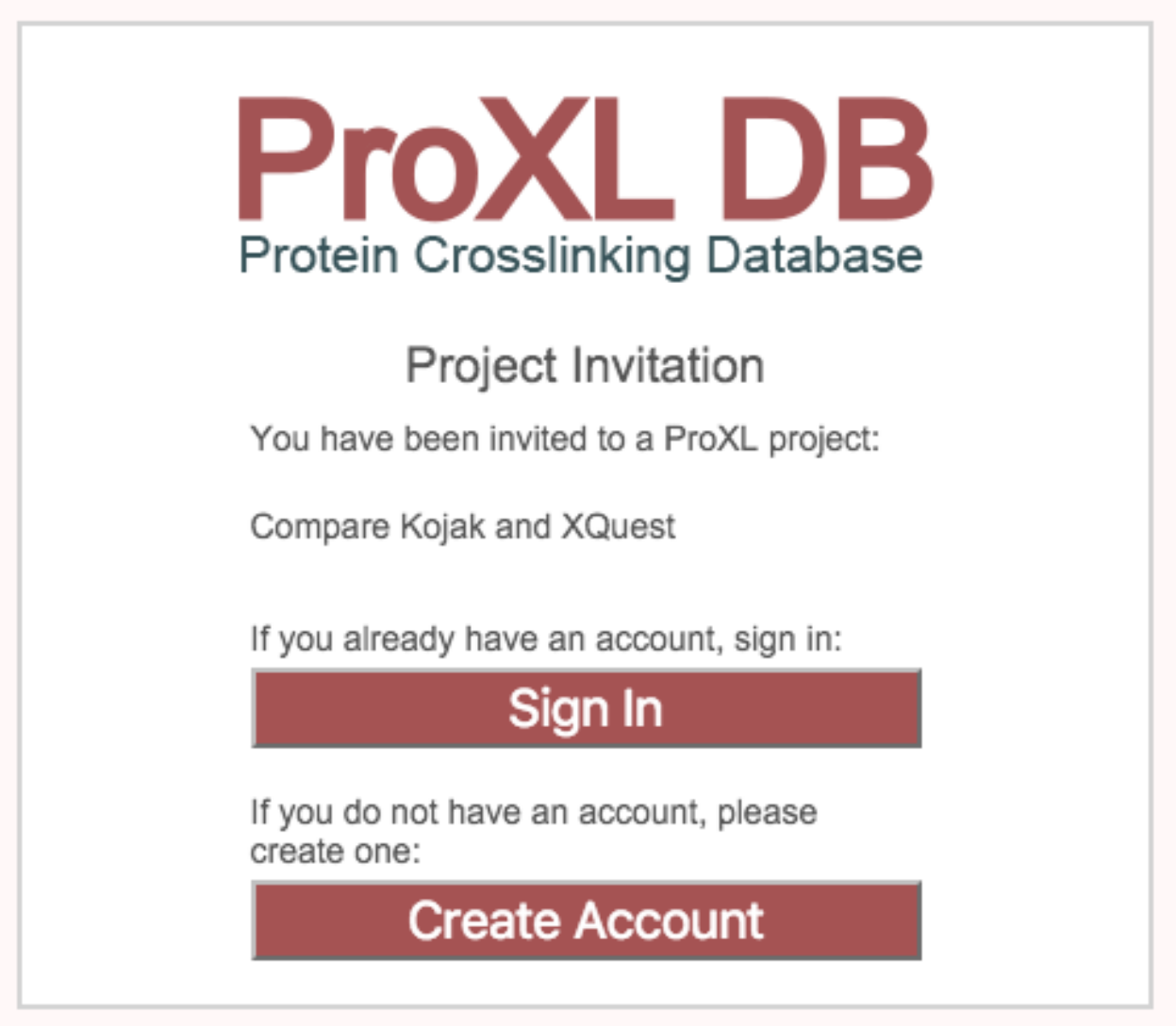


1.1.2 Private Installations

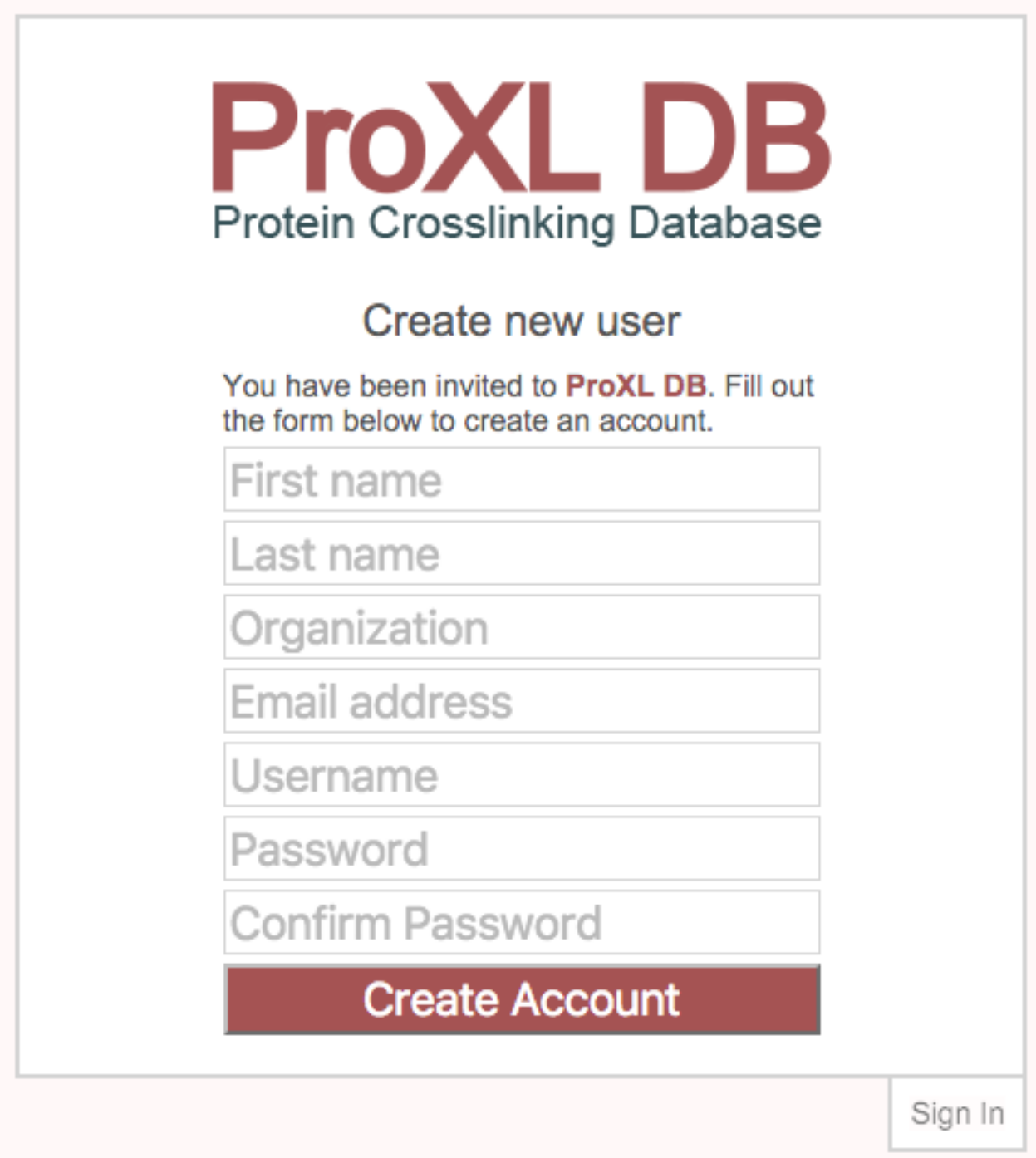
If downloaded and installed privately, the site administrator may choose to allow public registration (see above) or restrict access to only those who are invited to projects.

If restricted to invites-only, the only way to gain access to a proxl installation is to be invited to a project by an existing user. Only users associated with a project may invite users to the project. (To learn how to invite a user to a project, see *Researchers*.) Invitations will appear in your email with a link to register as a user of proxl. **Important**: Be sure to check your SPAM folder for proxl invitations if they do not appear.

Unless you are already logged into proxl, following the link in the invitation email brings you to the following screen. (If you are already logged in, you will be taken to a list of you projects, which now includes the one to which you were invited.)



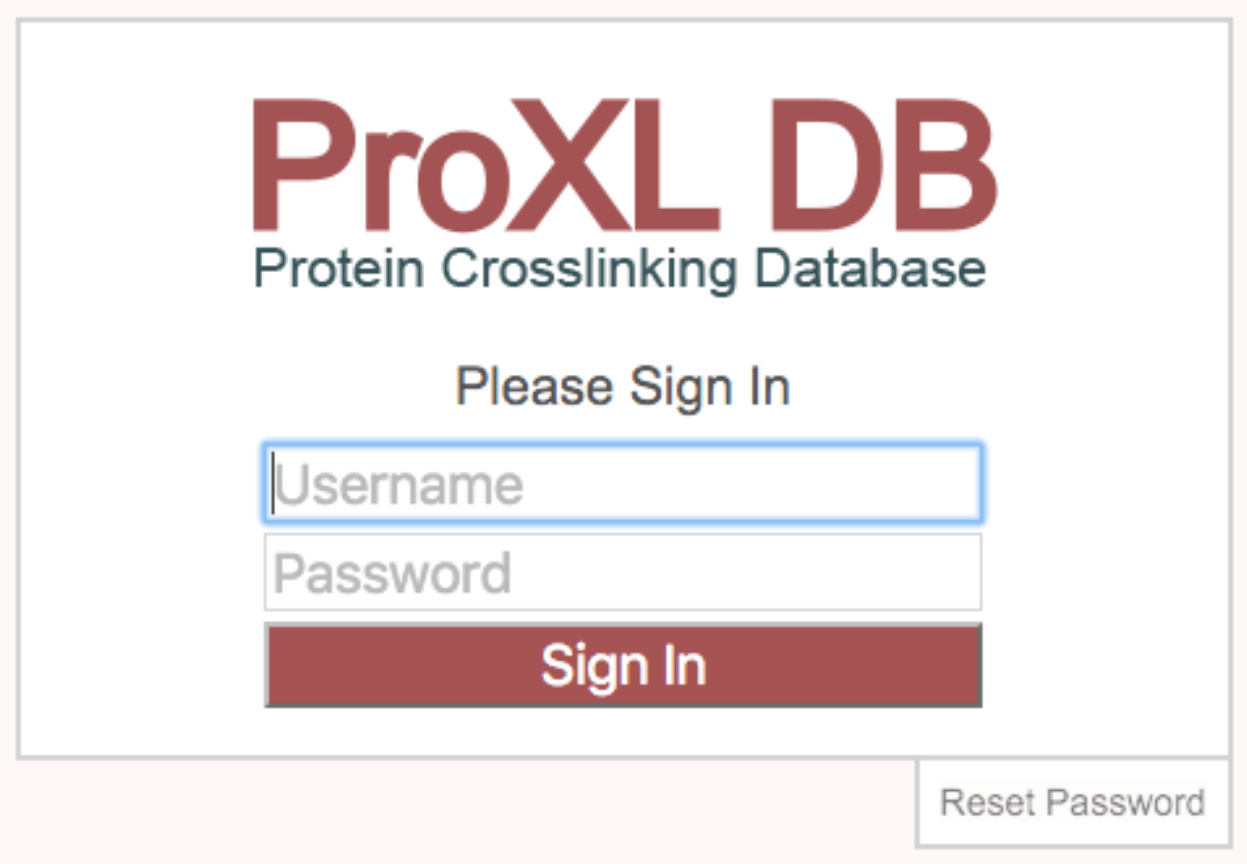
Choosing "Sign In" will sign you into proxl, and the new project will be listed in your project list. Choosing "Create Account" will present you with the following registration form:



Filling out this form will create a new account, log you in, and display the project to which you were invited in your project list.

1.2 Sign In

Accessing the URL for any non-public data or the URL for the home page of proxl will produce the password prompt below. Provide your username and password to proceed.



1.3 Forgot Password

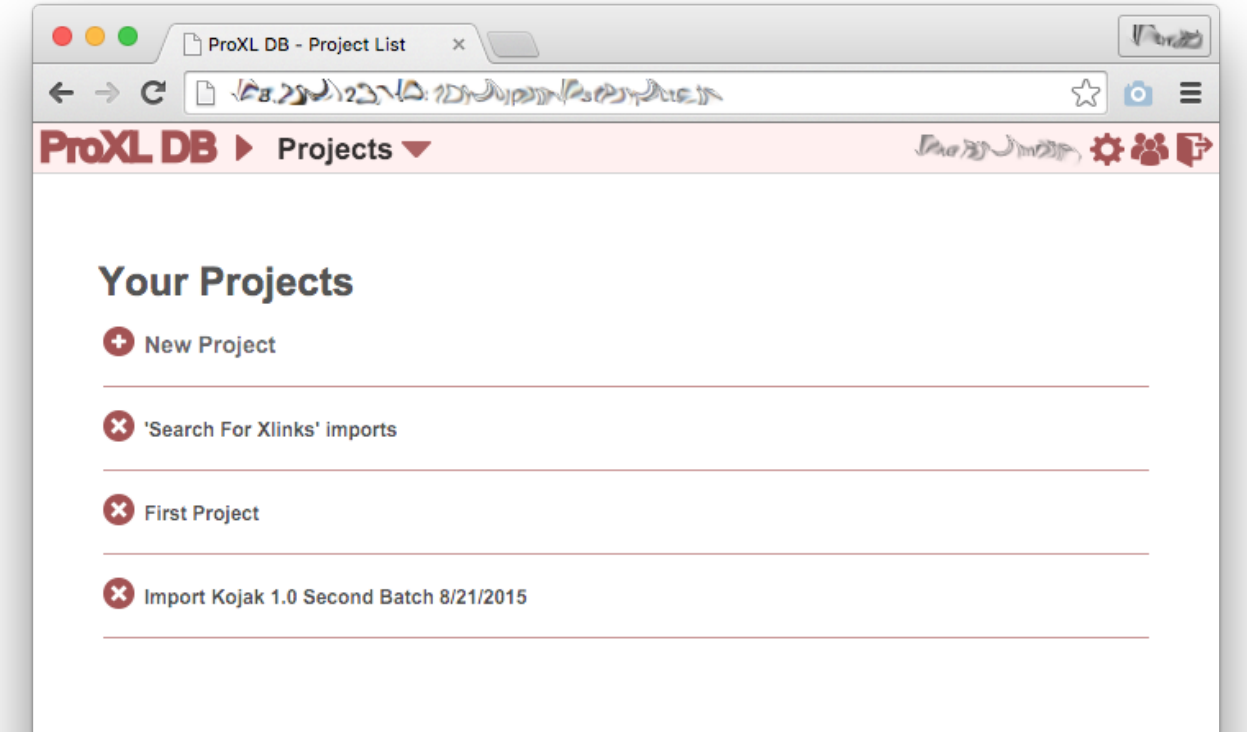
If you are unable to remember your password, click the "Reset Password" button at the bottom-right of the sign-in form. Entering either your username or email address on file for your account will send a link to your email address that may be used to reset your password. For security reasons, this link is only valid for 24 hours.

1.4 Projects and Access Control

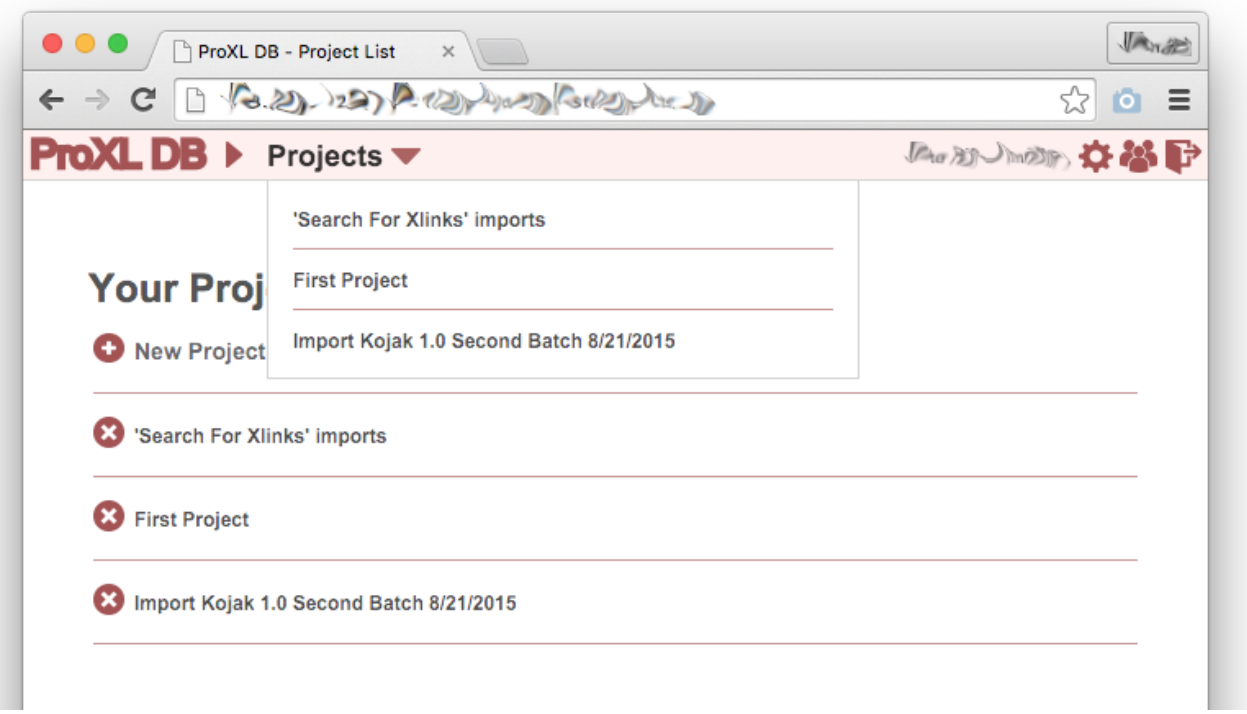
Projects in proxl are the core means of organizing data and controlling access to those data. While projects serve to logically organize data by theme or aim, they also serve to organize the data by who you would like to have access to those data. All data are associated with projects and, by default, only the researchers associated with a project may view data associated with that project. (These permissions can be changed, see *Public Access*.)

1.5 Accessing Projects and Data

Once logged in, you will be presented with a list of projects that you may currently access. For example:



In this example, the user has access to three projects. Clicking on the titles of any of these projects will navigate to that project and its data. From any page, the user may also mouse over (or tap on mobile devices) the "Projects" text at the top page to see a drop-down list of project titles. Clicking on any of these will navigate to that project's page and data:



1.6 Return to Project List

To return to the project list (required for adding or deleting projects), click the proxl icon or the "Projects" text at the top-left of any page in proxl.

1.7 Adding New Projects

Click on "(+) New Project" above the project list to add a new project by supplying a title and (optionally) an abstract for your project. Once the project is added, it will appear in the project list with the supplied title. Click on the project title to navigate to that project's overview page to manage access and add researchers to the project. (See "*Researchers*.)

1.8 Deleting Projects

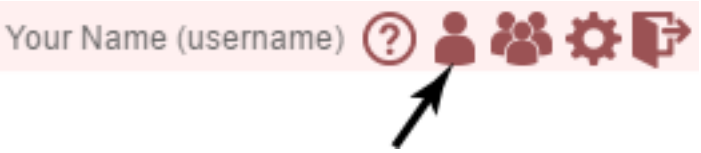
Click the red (X) icon to the left of the project title on the project list to delete that project. Only project owners may delete projects. Once a project is deleted, the project and all associated data will be removed.

1.9 Get Help

To view documentation for proxl, click the (?) icon on the top-right nagivation bar present on all pages.

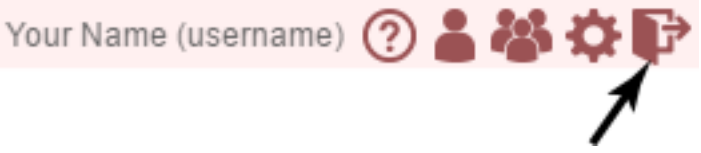
1.10 Manage Account

You may change your name, email address, organization, username or password by clicking the manage account icon (gear shape) at the top-right of the page:



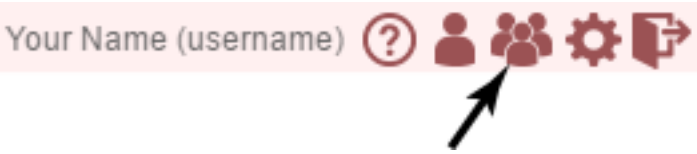
1.11 Sign Out

To securely sign out, click the sign-out icon at the top-right of the page and close your web browser.



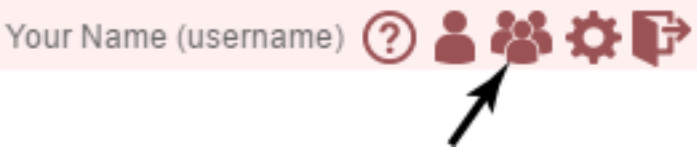
1.12 Manage Users (Admin Only)

If you are an administrator, you may manage or invite users by clicking the manage user icon at the top-right of the page:



1.13 System Configuration (Admin Only)

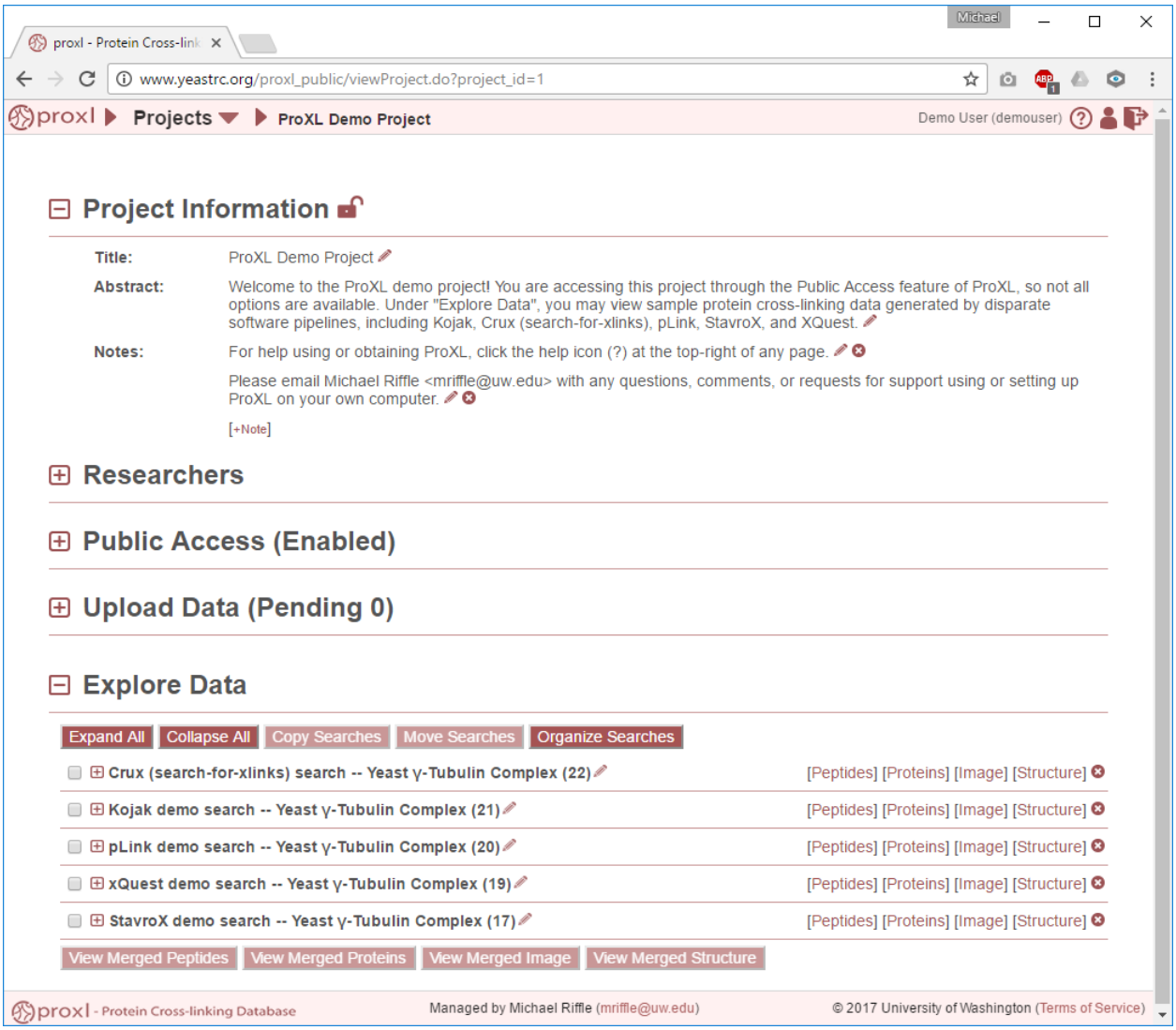
If you are an administrator, you may configure proxl by clicking this link. See *Proxl Configuration*.



CHAPTER 2

Project Overview Page

Clicking on the title of a project on the project list or project pull-down list at the top of the page will bring you to that project's overview page. **Important**: Only users listed as researchers on the project may access this page. Example project overview page:



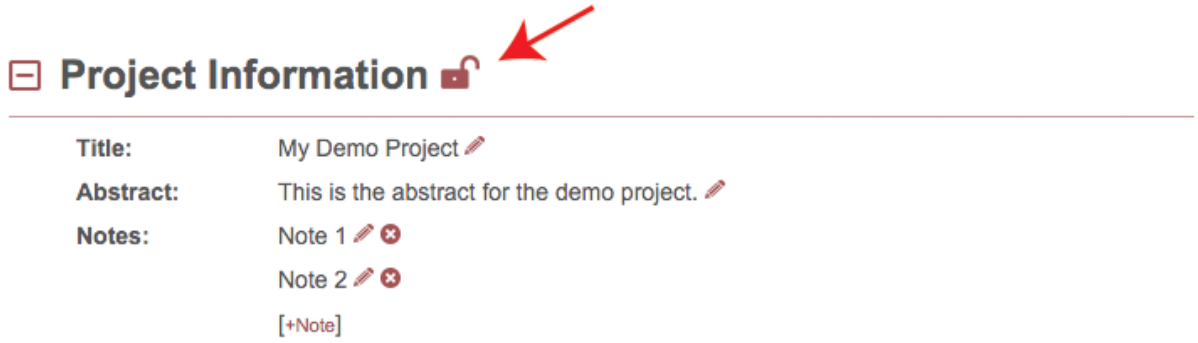
From here, you may edit the project information, add or remove researchers, toggle public access, explore or manage the data, and lock the project. Click the [+] icon next to each section to expand and view that section. Please see below for help on all of the sections.

2.1 Project Information

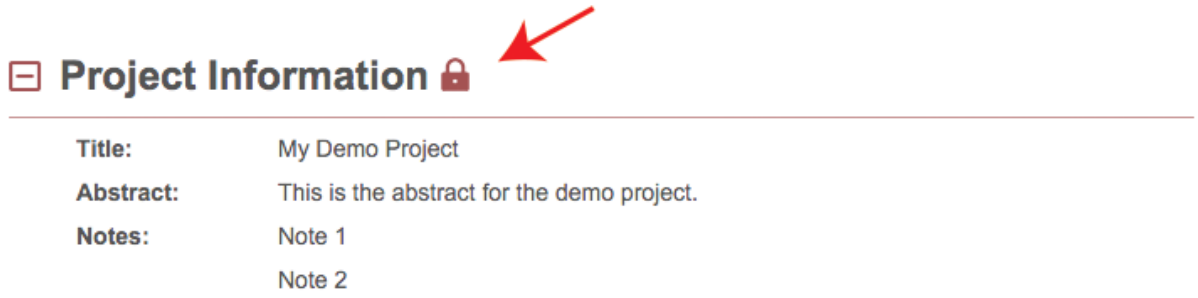
Project members may edit the title or abstract by clicking the pencil icon next to the respective item. Project members may add notes to the project by clicking "[+Note]". Notes may be edited or deleted by clicking on the associated pencil or delete (X) icons. The title, abstract, and notes are visible to public users (if public access is enabled, see below).

2.1.1 Lock Project

The project may be locked by clicking the lock icon next to the "Project Information" section header:

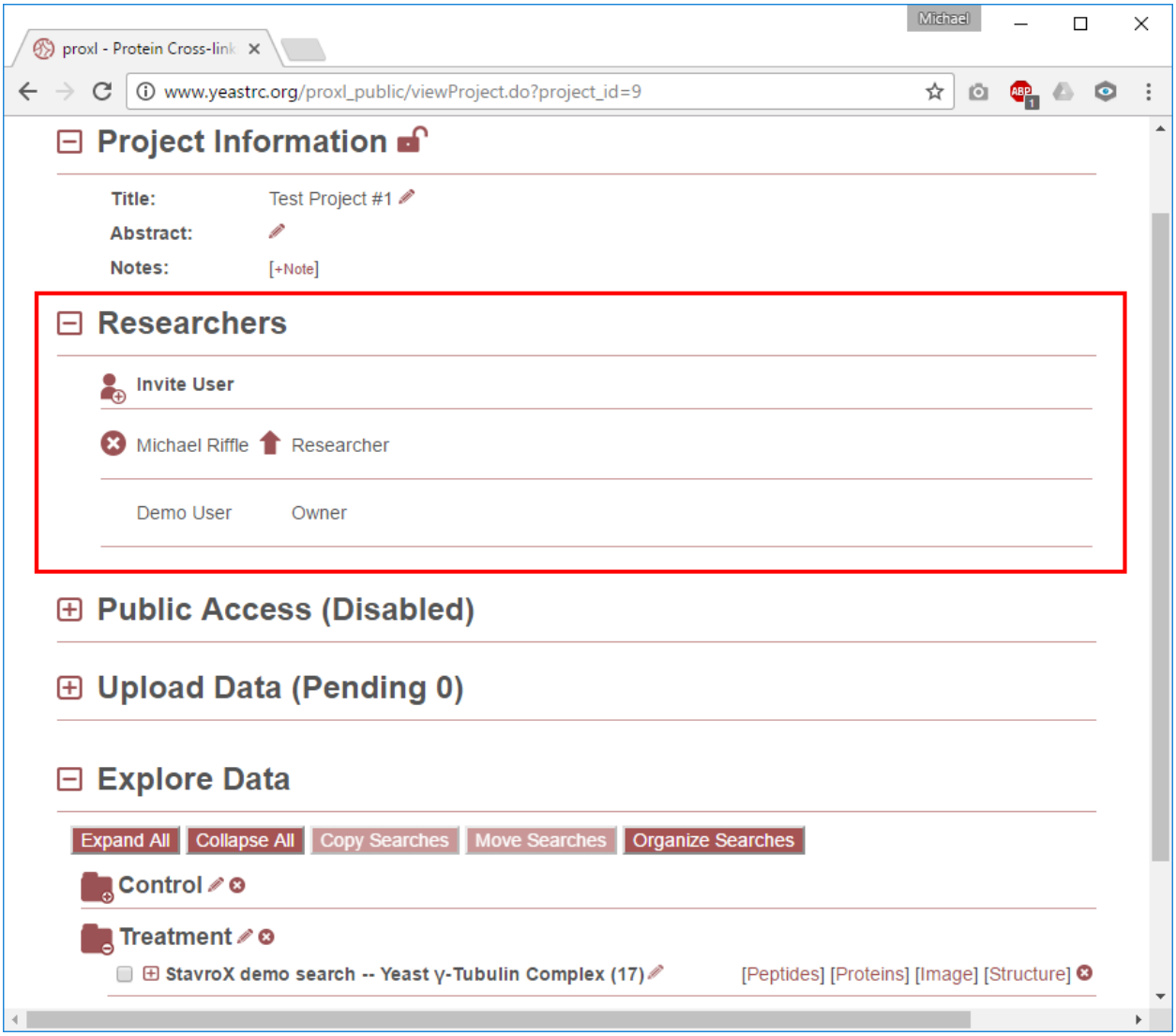


Clicking the icon changes the icon to a locked state:



Projects that are locked may not be changed in any way until they are unlocked. This includes associated users, public access, uploading or annotating data, and so on. This is meant to accompany projects that are supporting publications, and so should not change. To unlock a project, the project owner may click on the lock icon next to the "Project Information" section header.

2.2 Researchers

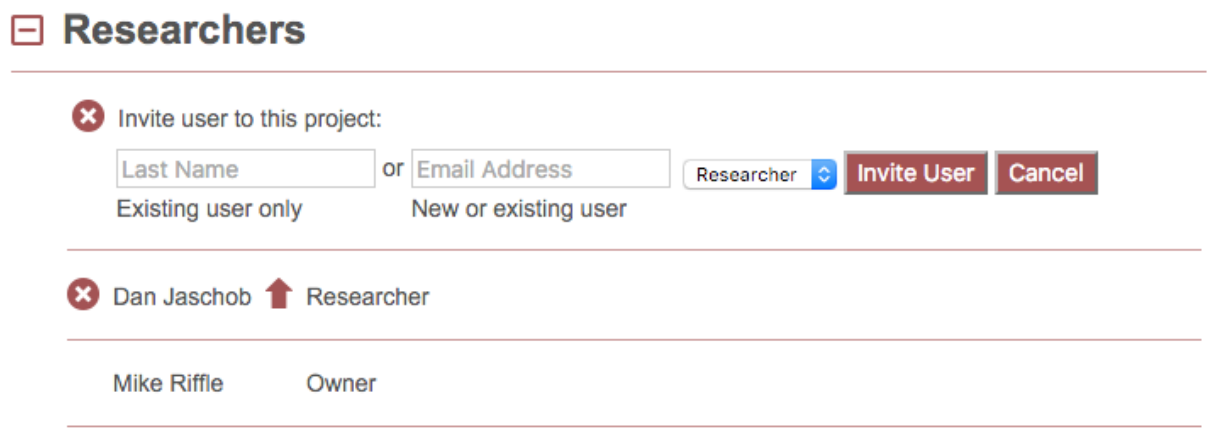


This section defaults to not being expanded, click the [+] icon next to this section title to expand it. This section lists the users associated with this project. Some important notes about the researchers section:

- Only users listed here may access the project overview page or the project's data. (Except if public access is enabled, see below.)
- The "Researchers" section of this page is not visible to public users.
- Any user may invite other users to the project.

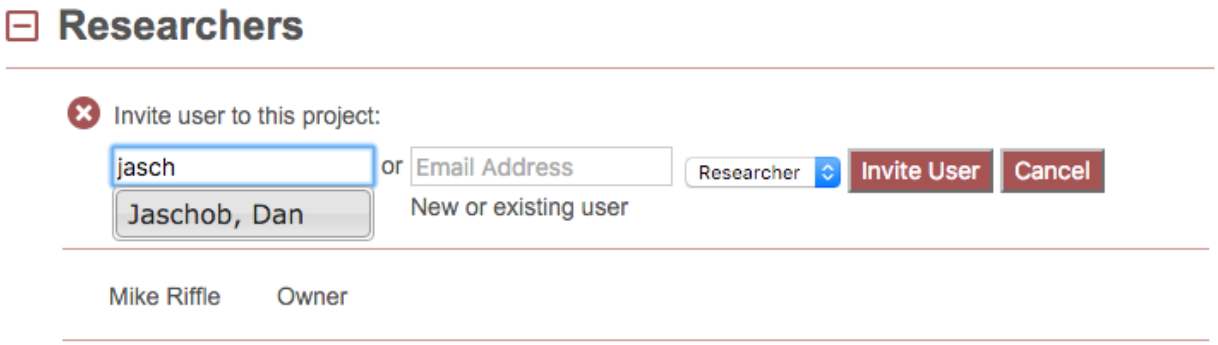
2.2.1 Invite User

To invite a user to the project, click the "Invite User" text or associated icon at the top of the user list. This will open the following dialogue.

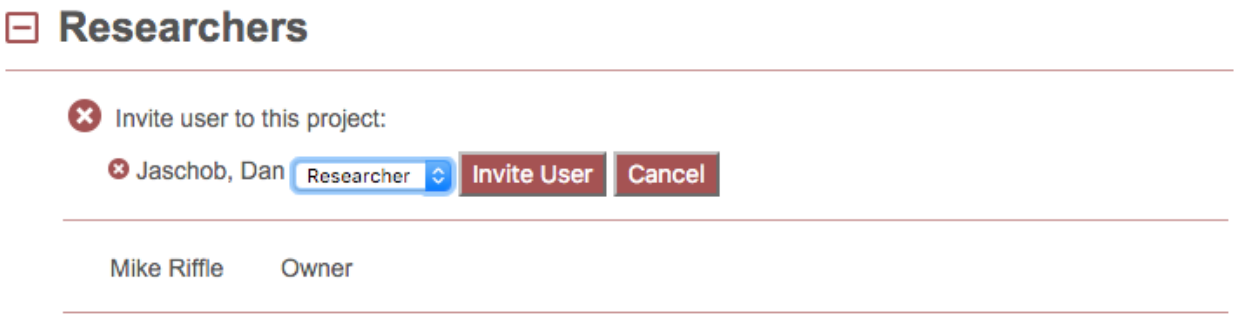


For Existing Users

To invite an existing user to this project, type their last name or email address into the respective text box. If they are found in the database, their name will appear as you type:



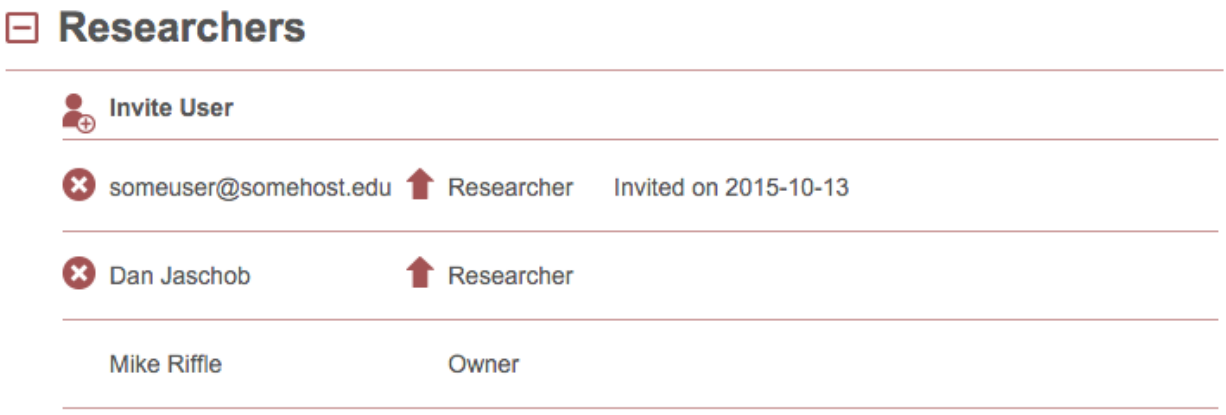
Clicking on the name produces:



Here you may designate their level of access (if you are a project owner) and click “Invite User” to invite that user to the project. The user will immediately have access to the project, and this project will appear in their list of projects. Alternatively, click “Cancel” to cancel the process.

For New Users

To invite a new user to proxl and provide them access to this project, type in their email address and click "Invite User". They will receive an email with a link for creating an account in proxl and this project will appear in their project list. New users invited to the project will appear as:



Once they have created an account, their name will appear in the user list instead of an email address, and the "Invited on…" text will no longer be present.

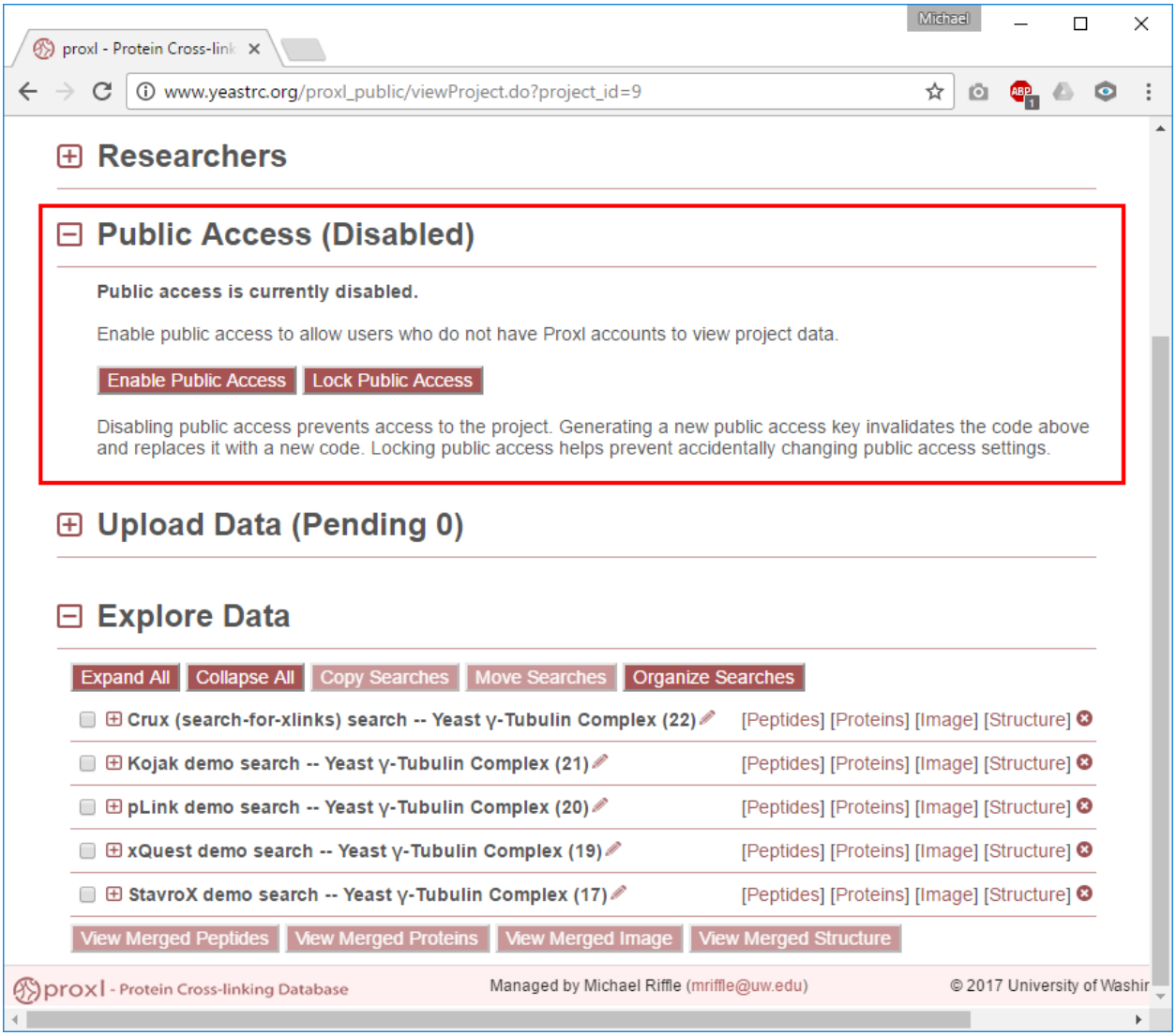
2.2.2 Remove User

Click the red (X) icon to the left of a listed user to remove that user from this project. That user will immediately lose access to the project and its associated data. This will not remove existing users, only revoke their access to this project. For invited email addresses, this will invalidate the access code included in the invitation email and they will not be able to use it to create a proxl account.

2.2.3 Promote or Demote User

Clicking the up or down arrow next to the access level of a user will either promote that user to an owner or demote that user to a researcher. Owners have complete complete access to a project, including the ability to lock or unlock it, enable or disable public access, promote or demote other users, define default views of data, or delete data.

2.3 Public Access

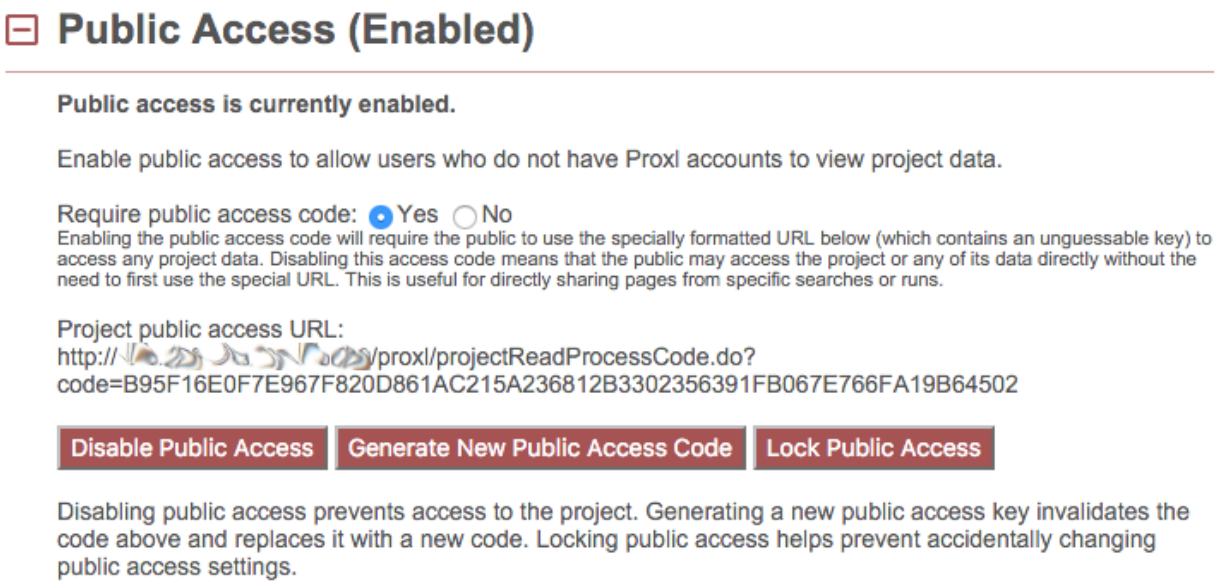


This section defaults to not being expanded, click the [+] icon next to this section title to expand it. This section controls whether or not public access is enabled and, if it is, whether or not a public access code is required to access the project and its data. Some important notes about the public access section:

- Only project owners may change public access settings.
- The "Public Access" section of this page is not visible to public users.

2.3.1 Enable or Disable Public Access

Enabling public access allows access to the data without requiring that users have proxl accounts. Clicking "Enable Public Access" enables public access and changes the display of this section to indicate that public access is enabled.



Public Access Code

By default, public access is enabled in a way that requires a specially-formatted URL that contains an unguessable public access code. This URL is listed here as "Project public access URL." This exact URL must be used to access the project before the user may access any of the data. This is useful for semi-private sharing of data with select users (or reviewers) without making the data completely publicly accessible.

The requirement for the public access code may be removed by clicking "No" next to "Require public access code." If "No" is selected, URLs for the project or any of the data pages may be directly shared without the need of the user to first use the public access code. This is useful for truly public sharing of the data, such as in the case of publication.

The "Generate New Public Access Code" button will generate and replace the current unguessable public access code with a new code. This will revoke access to users that have used the previous code.

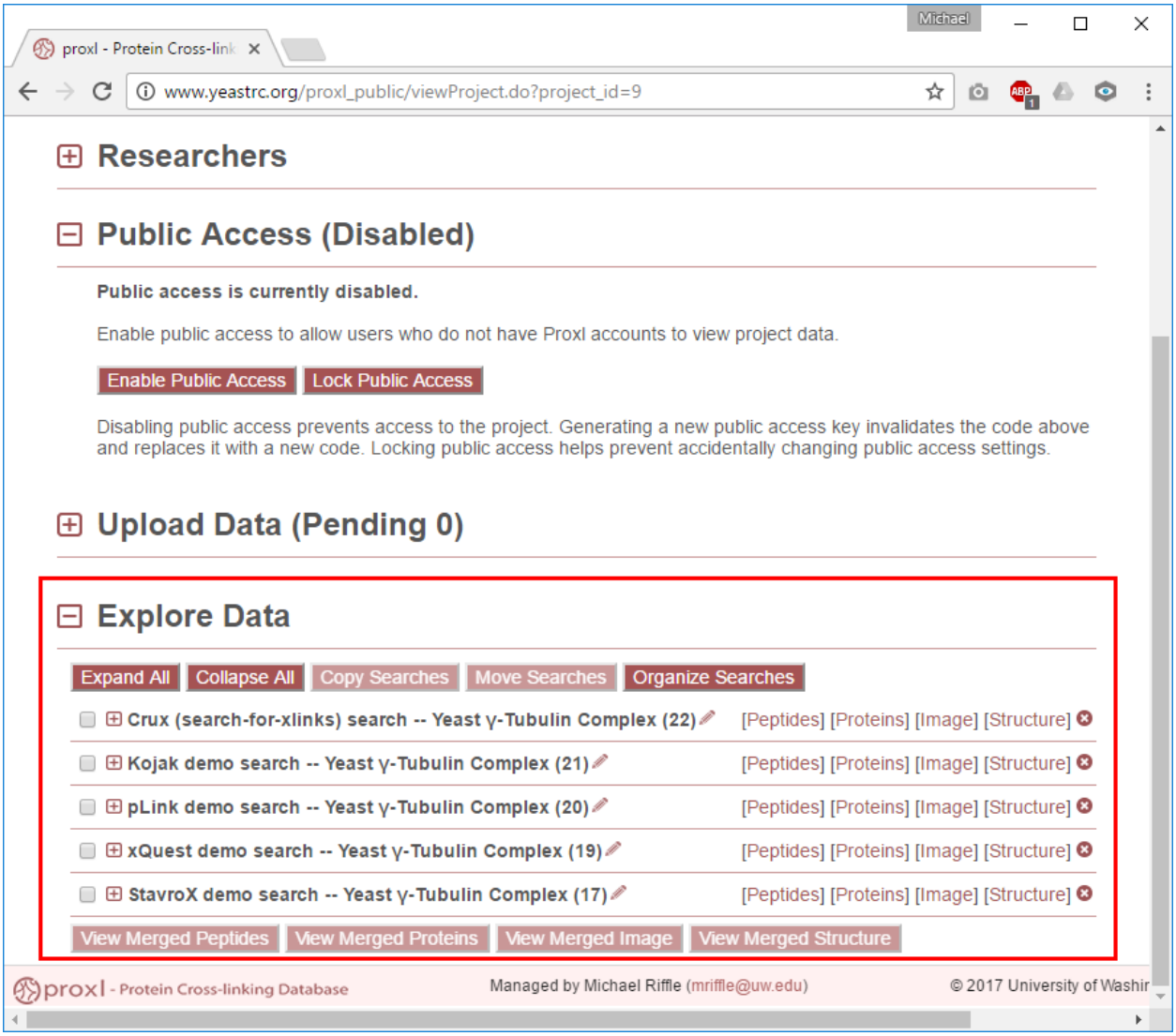
Lock Public Access

Clicking the "Lock Public Access" button makes it impossible to change public access code settings without first clicking "Unlock Public Access." This is meant to prevent accidental disabling of public access or generation of new public access codes, which would revoke previously-granted access to the public, colleagues or reviewers.

2.4 Upload Data

We have set up a separate page describing uploading data. Please see *Upload Data*.

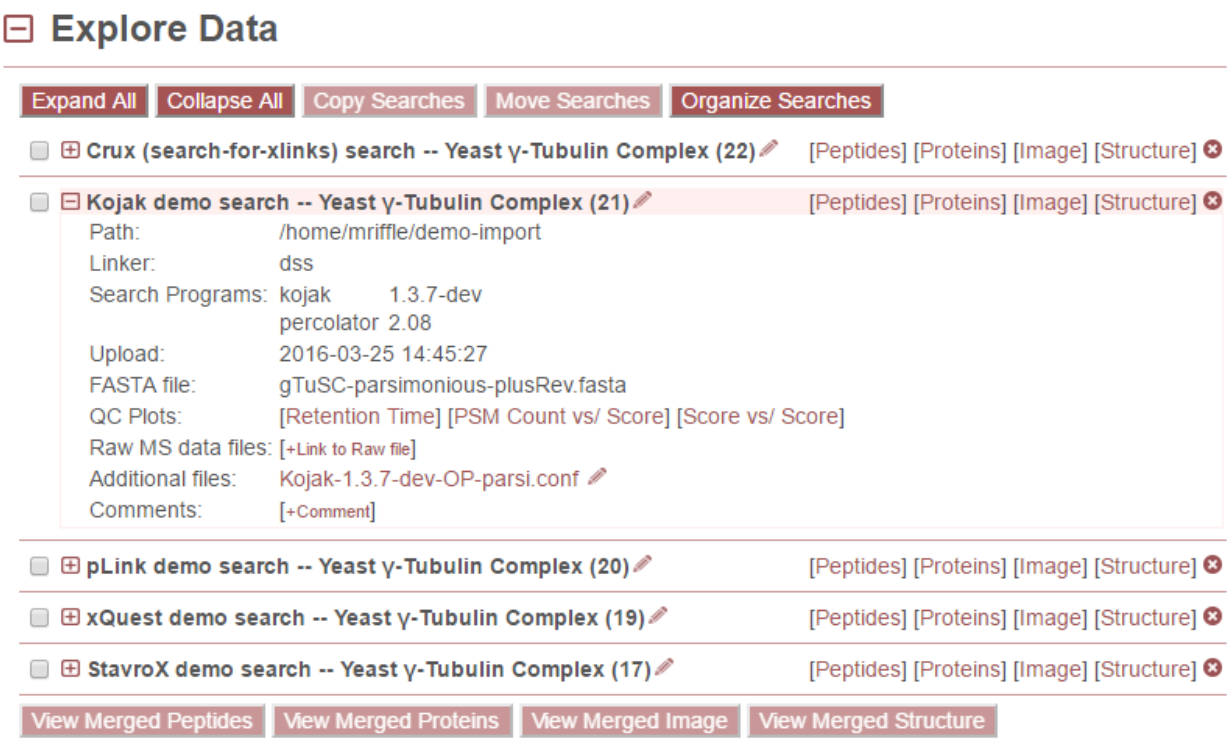
2.5 Explore Data



This section lists each of the searches associated with this project. A "search" in this context are all the data resulting from running a software pipeline (e.g., Kojak or xQuest) against spectra data (e.g., a mzML file). Project researchers may change the name of these searches by clicking the pencil icon to the right of the current search name. To the right of the search name the search ID number is listed in parentheses as a standard way to refer to specific searches.

2.5.1 View Search Information

To view information about a search, click the [+] icon to the left of the search name. This will display the following information. (Alternatively, click the "Expand All" button at the top of the to see all information about all searches.)

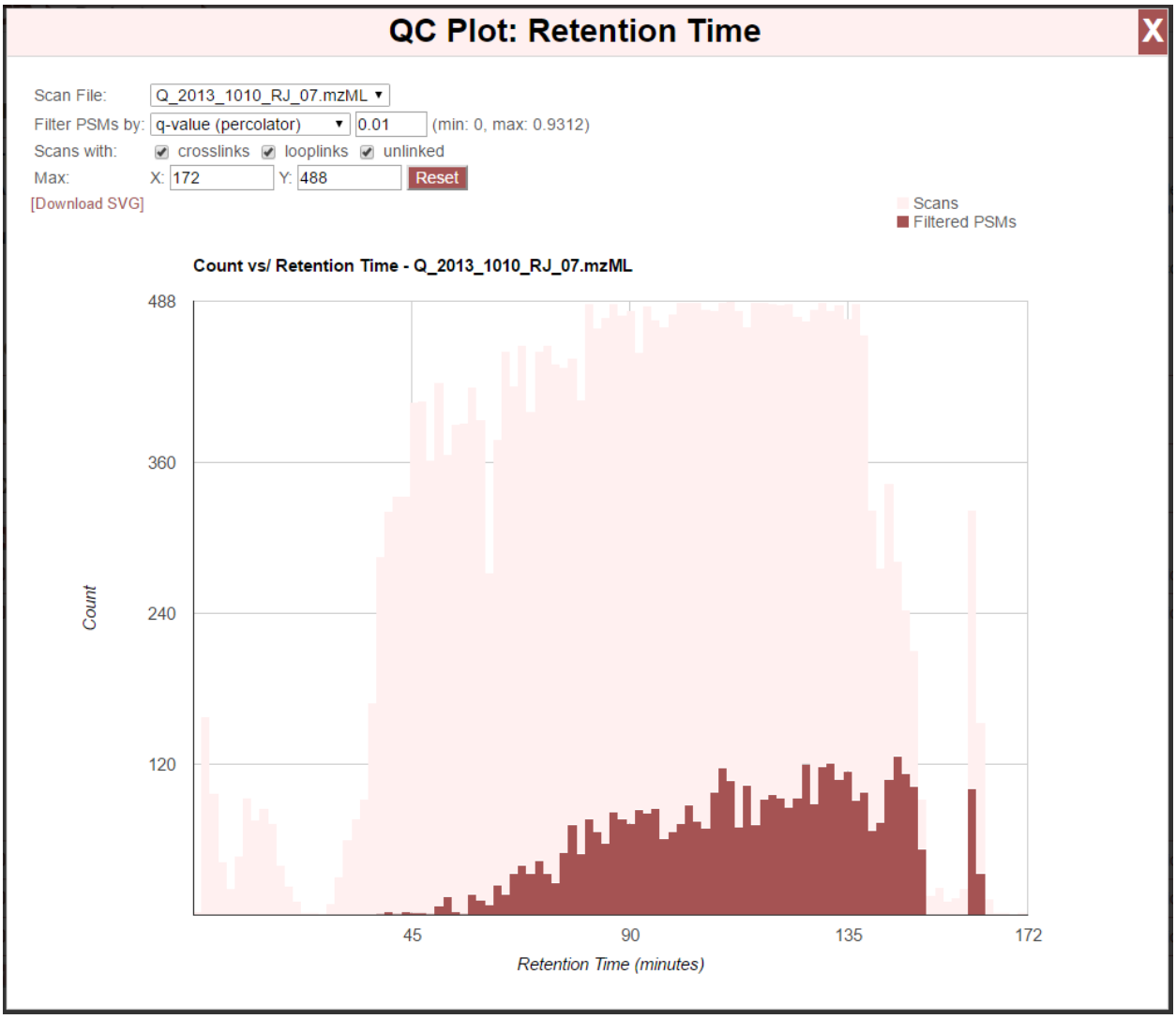


Search information includes the following information for each search:

- Path - The path the data were in when imported into proxl.
- Linker - the crosslinker(s) used in this experiment
- Upload - the date the data were uploaded to proxl
- QC Plots - links to a retention time QC plot. See below.
- Raw MS data files - Links to RAW files that contain the raw machine output for this experiment. Project owners may add URL links to RAW files by clicking [+Link to Raw file].
- Additional files - Links to additional files associated with this search, such as the configuration or parameters files for the respective search program.
- Comments - Lists the comments that have been added to the search. Comments may be deleted by clicking the red (X) to the left of the comment, or edited by clicking the pencil icon to the right.

Retention Time QC Plot

Clicking the "[Retention Time]" link next to "QC Plots:" produces the retention time QC plot:



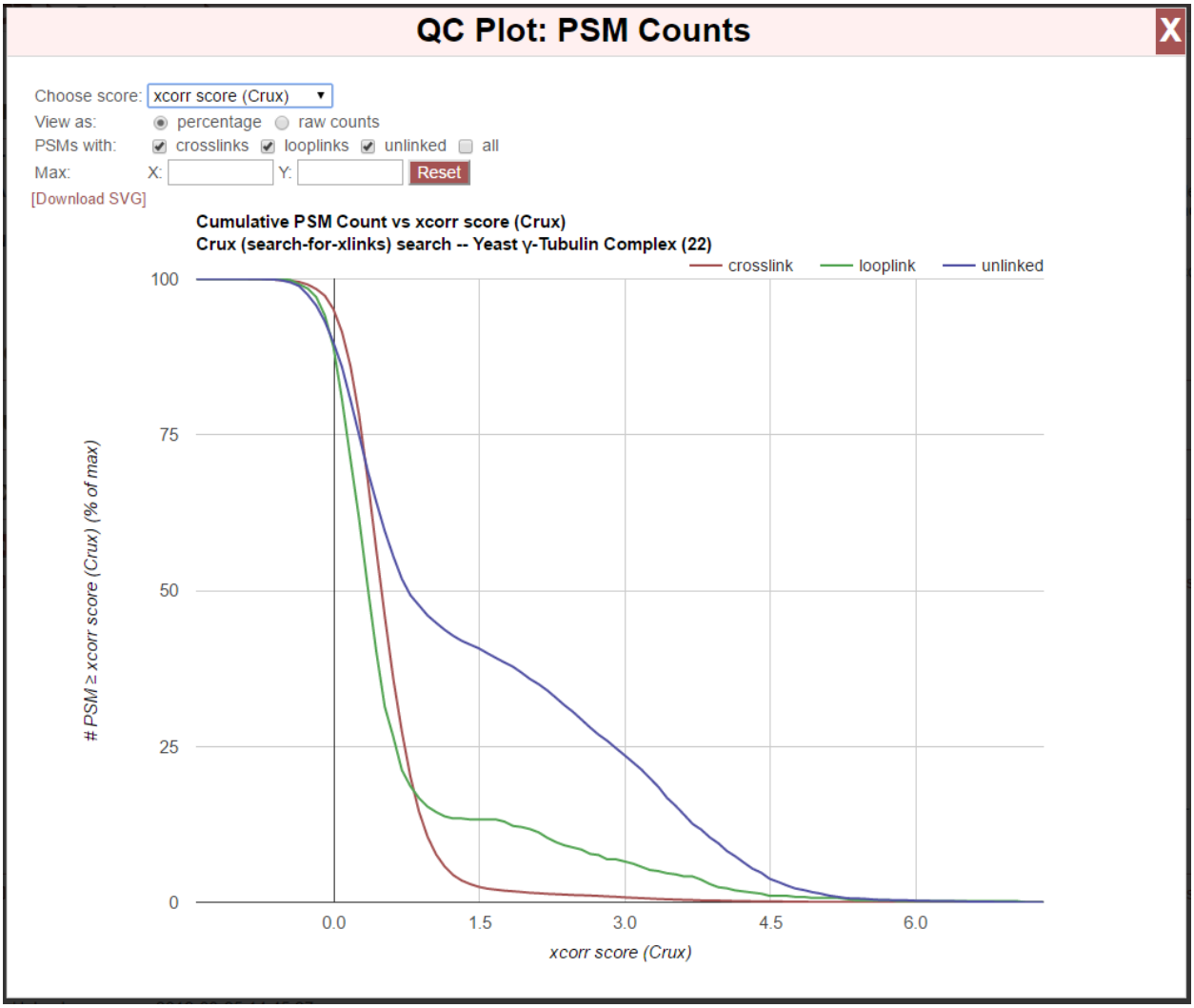
This is a histogram showing the number of MS2 scans taken versus retention time. The pink bars show all scans, and the dark red bars show the number of those scans that meet our filtering criteria at the top of the plot–or, "Filtered PSMs". To close the chart, click the "X" in the top right of the window, or anywhere in the browser outside the chart window.

The chart options are:

- Scan File - If multiple spectral files were searched, each will be listed here. The data in the chart reflect the selected spectral file.
- Filter PSMs by - Select the score type and cutoff value for that score to be used to plot the number of PSMs meeting those filtering criteria (red bars). The minimum and maximum values for the selected score type are given in parentheses.
- Scans with - Counts for "Filtered PSMs" will only include scans that resulted in a PSM where the peptide was of a type that is checked here. E.g., if only "crosslinks" is checked, only scans that resulted in crosslinked peptides will be used to for "Fitlered PSMs" counts. If "looplinks" and "crosslinks" are checked, only PSMs resulting in crosslinked or looplinked peptides will be counted.
- Max - Values entered here will be the maximum value on the X or Y axis–used for rescaling the chart.

Cumulative PSM Count vs/ score QC plot

Clicking the "[PSM Count vs/ Score]" link next to "QC Plots:" produces a plot of cumulative PSM count vs/ a chosen score type:



Possible values for the chosen score type are presented along the x-axis. Scores for which lower values are more significant, the y-value represents the number of PSMs with the value on the x-axis or lower. Score for which higher values are more significant, the y-value represents the number of PSMs with the value on the x-axis or higher.

Each class of PSM is presented as a separate line: crosslinks, looplinks, and unlinked. A line for all PSM types combined may optionally be displayed by checking "all" in the "PSMs with:" options.

The "Choose score:" option allows choosing which score from the search is used to generate the plot.

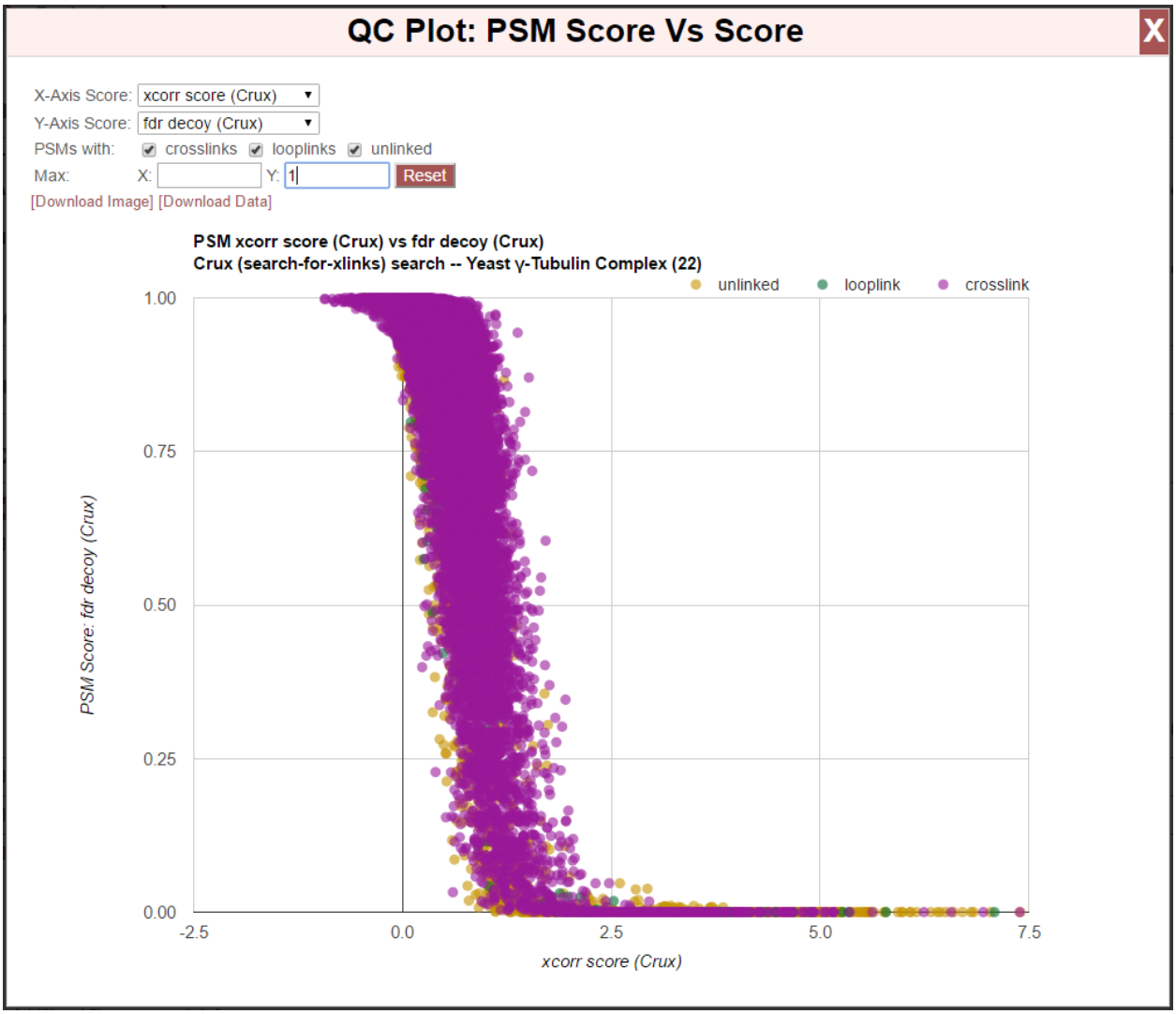
The "View as:" option allows switching between raw counts (default) and percentage. The "raw counts" option uses the raw PSM counts for the respective types. The percentage option displays the PSM counts as the percentage the total PSMs of the respective type, and so applies the same scale to all lines, which will always move from 0 to 100 in the plot.

The "PSMs with:" options toggle the visibility of the lines corresponding to the respective class of PSM.

The "Max:" options allow for a user-supplied maximum for the X- and Y-axes. "Reset" clears the user-supplied maximum values.

PSM Score Vs Score QC Plot

Clicking the "[Score vs/ Score]" link next to "QC Plots:" produces a plot of any PSM-level score vs/ any PSM-level score:

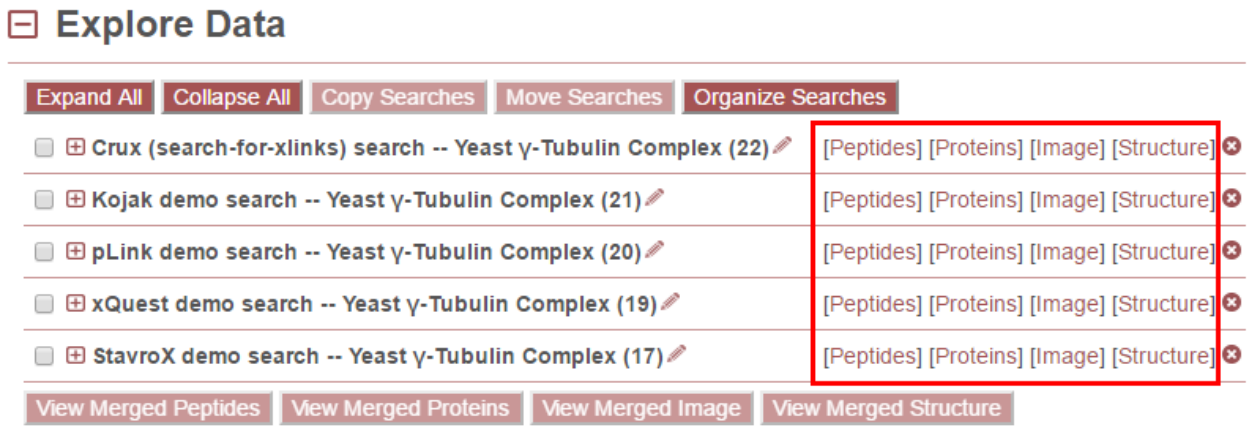


This is a scatter plot showing how PSM-level scores correlate with other PSM-level scores. For example, the above image shows calculated false discover rates (FDR) for PSMs vs/ the underling Xcorr score calculated by Crux. This can be used to discover unexpected relationships between scores, or to understand the effects of post processing statistical tools.

The chart options are:

- X-Axis Score - The score to use for the x-axis.
- Y-Axis Score - The score to use for the y-axis.
- PSMs with - Select the type of PSMs to plot (cross-links, loop-links, and/or unlinked).
- Max - Values entered here will be the maximum value on the X or Y axis–used for rescaling the chart.

2.5.2 View Data



Use the links to the right of the search names (in red box above) to view the data. There are four views currently available:

- **Peptides** - Provides a table view of the identified peptides and associated data. See *Peptides View Page* for more details.
- **Proteins** - Provides a table view of the crosslinks and looplinks at the protein level. See *Proteins View Page* for more details.
- **Image** - Provides a graphical view of the data where proteins are represented as proportionately-sized bars that are annotated with link locations within and between proteins in the context of sequence annotation data. See *Image View Page (Protein Bar View)* for more details.
- **Structure** - Provides a view of crosslinking data on 3D protein structures, including an interface for uploading PDB files and mapping sequences from the FASTA file onto those PDB files using pairwise sequence alignment. See *Structure View Page* for more details.

2.5.3 View Merged Data

Proxl allows for comparing and contrasting multiple searches, even if those searches were analyzed with different software pipelines. Proxl refers to this as merging data. To merge data from multiple searches, click the check boxes to the left of the searches of interest and click either "View Merged Peptides", "View Merged Proteins", "View Merged Image", or "View Merged Structure."

For more information please see: *Merged Peptides View Page*, *Merged Protein View Page*, *Image View Page (Protein Bar View)*, and *Structure View Page*.

2.5.4 Copy Searches

Project owners may copy searches to a different project (where they must also be an owner). To copy data, check the checkbox to the left of the search name(s) you wish to copy and click the "Copy Searches" button above the search list to see the following dialogue:

Projects to which you have permission to copy data are listed. Click on the project name to which the searches should be copied. Copied searches are treated independently with regards to the search name, comments, and other search metadata. For example, adding a comment or changing the name of the search in the new project will not affect the original search's name or comments.

2.5.5 Move Searches

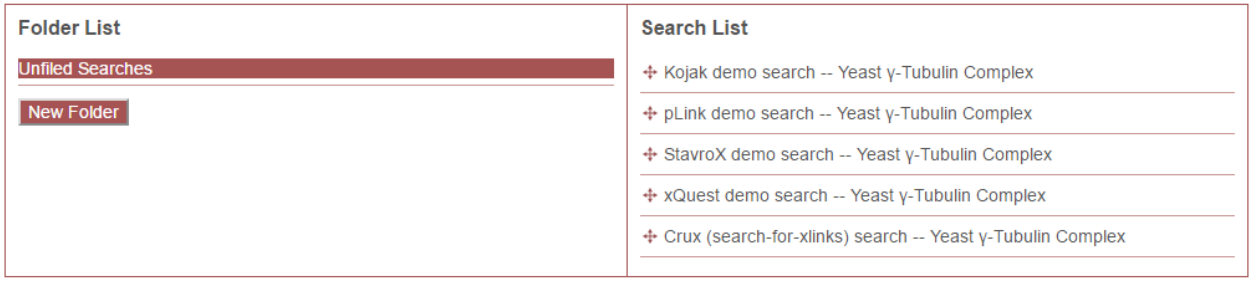
Project owners may move searches to a different project (where they must also be an owner). To move data, check the checkbox to the left of the search name(s) you wish to copy and click the "Move Searches" button above the search list to see the following dialogue:

Projects to which you have permission to copy data are listed. Click on the project name to which the searches should be copied. Moving a search will remove it from the current project and place it in the new project.

2.5.6 Organize Searches

The "Organize Searches" button opens an interface for rearranging the search list and creating and placing searches inside of "folders" for improved organization when the project contains many searches.

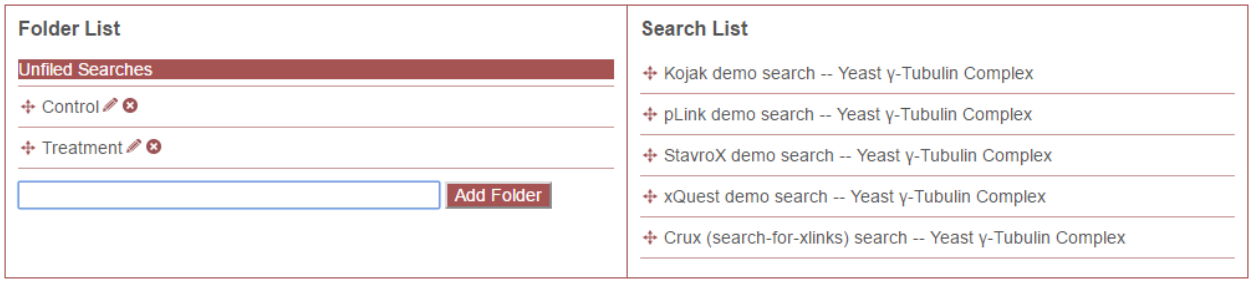
Clicking the "Organize Searches" button opens the following dialog:



The right-hand panel is labeled "Search List". This lists the searches contained in the currently-selected folder. To re-arrange the order in which searches are listed, click and drag the search to the desired order in the list. To see and re-arrange searches in another folder, click on the name of the folder in the left-hand panel.

The left panel is labeled "Folder List". This is a list of the folders that have been created for organizing searches. Note: Searches not in any folder are listed here under a special folder named "Unfiled Searches". Any searches in "Unfiled Searches" will not be placed into a folder when shown to users in the web application.

To create a new folder, click the "New Folder" button, type in the name of the new folder, and click "Add Folder." In the following example, two folders have been created, "Control" and "Treatment."

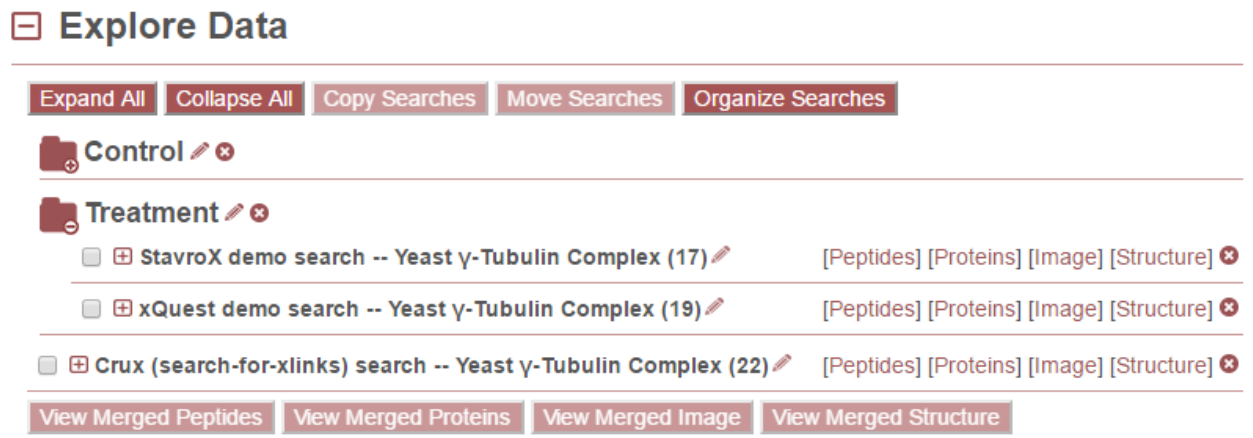


Folders may be deleted by clicking the small red "X" icon next to the folder name. Folder names may be edited by clicking the pencil icon. And folders may be re-arranged by clicking and dragging them to the desired position.

To place a search in a folder, first click on the folder name that currently contains the search (remember, searches not in folders are under "Unfiled Searches"). Then click the name of the search in the right-hand panel and drag it onto the row containing the folder in the left-hand panel and release the mouse button. This will "drop" that search into that folder.

When done, click the "Done Organizing Searches" button above the folder list to return to the normal interface.

In the following example, the StavroX and xQuest demo searches were added to the "Treatment" folder. The Crux demo search was left unfiled. And the user has clicked on the "Treatment" folder to view the searches listed under it.



2.5.7 Delete Searches

Project owners may delete searches by clicking the red (X) icon to the right of the view data links. This will remove all data associated with that search from the database–use with care. Note that if this search was copied to another project, that copy search will not be deleted.

CHAPTER 3

Upload Data

To upload data into proxl, the data must first be converted into proxl XML then uploaded via the proxl web site.

3.1 Convert output to proxl XML

Converters have been developed for several cross-linking proteomics pipelines. Click on the name of the software, below, to download and learn more about the respective converter.

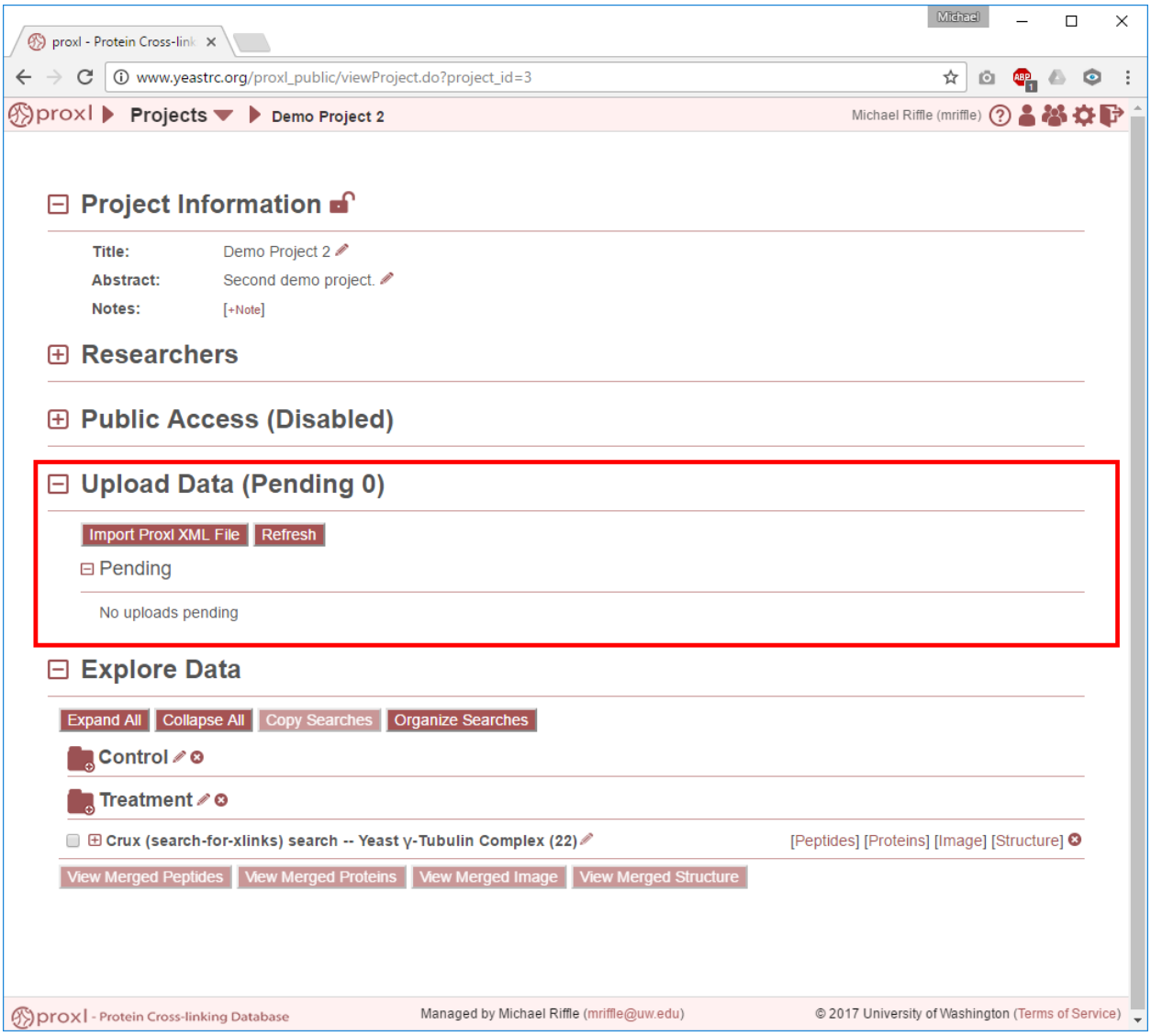
- iProphet (TPP)
- Kojak
- Crux
- pLink 1.x pLink 2.x
- StavroX
- MetaMorpheus
- xQuest

If you are using one of the software packages listed above, download the runnable file and follow its instructions to convert your data to proxl XML. If you encounter any issues or have any questions running any of these software, please email us at mriffle@uw.edu.

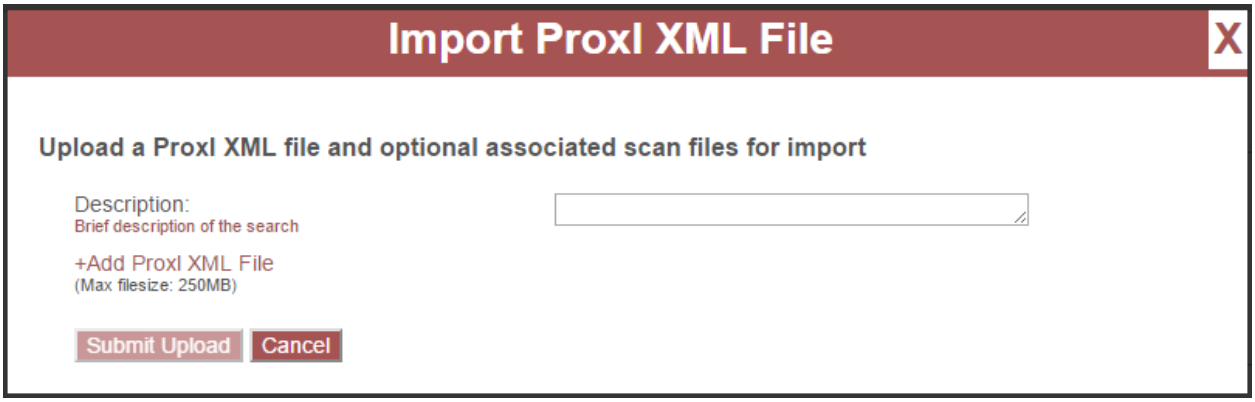
If you are not using one of the software packages listed above, please visit our *Proxl XML Converter Development Guide* page for information about how to develop a converter. We are happy to answer any questions, or work with you directly on the development of any new converter.

3.2 Import Data

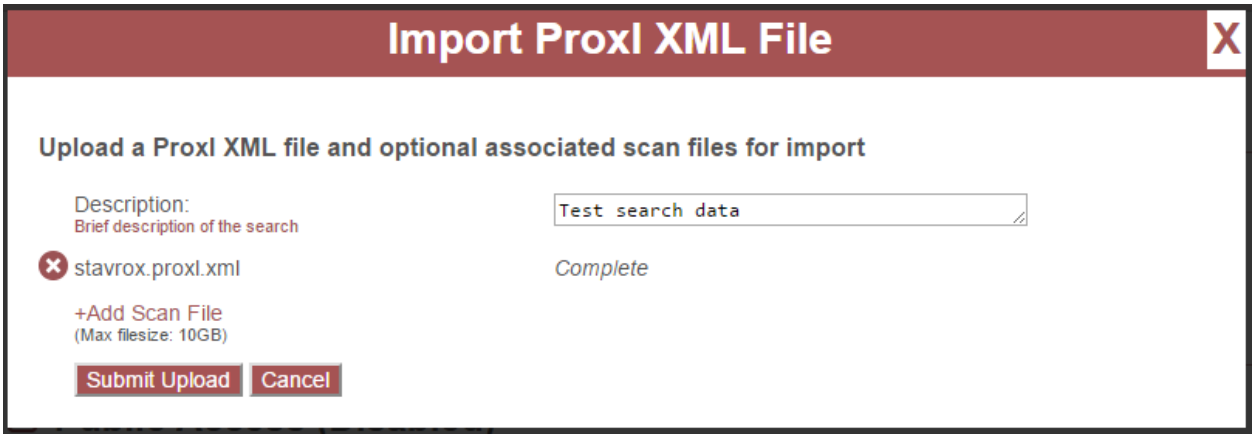
Click the [+] icon next to "Upload Data" on the project overview page to expand the data upload section and view upload status.



To upload a proxl XML file, click the "Import Proxl XML File" button:



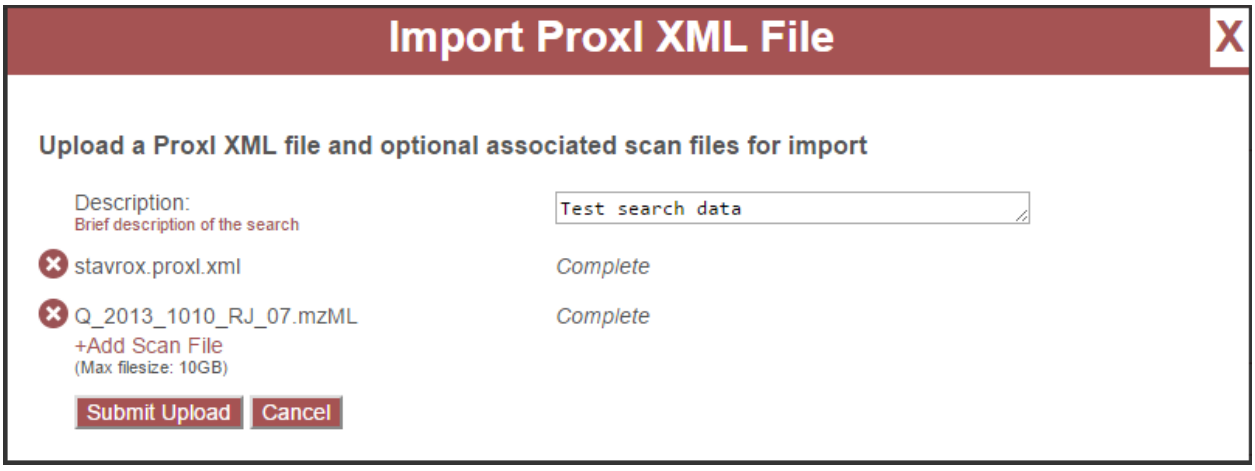
Add a description for the search (may be edited later), and click the "+Add Proxl XML File" link to initiate a file selection dialog. Select the proxl XML file you would like to upload. You will see the following:



Below the description, the file name of the uploaded proxl XML file is given. The red "X" icon may be clicked to delete that file and upload a different one. To the right of the file name is the upload status.

Below the file name is a link labeled "+Add Scan File". This may be used to optionally upload a mzML or mzXML file (or multiple files) containing the scan data that was searched. Viewing spectra associated with PSMs is only available if a scan file is uploaded with the proxl XML file.

Clicking the "+Add Scan File" link opens a file selection dialog. Upon selecting a file, it will begin uploading to the server–a progress bar will be visible. Once the upload is complete, you will see the following:



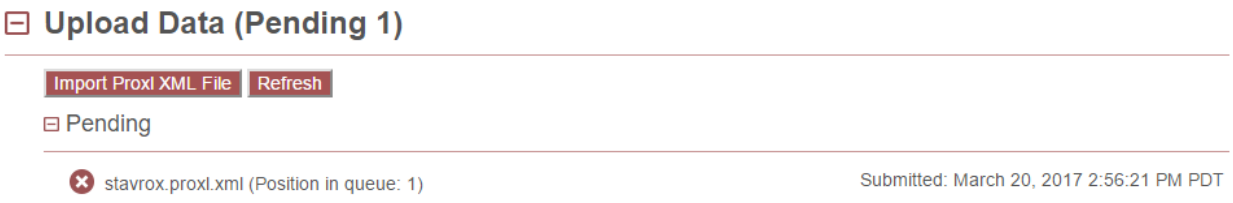
To deleted an uploaded scan file, click the "X" icon next to its file name. If multiple scan files were used in the search, you may continue to upload additional scan files by clicking the "+Add Scan File" link.

To submit the data to proxl for processing and import, click the "Submit Upload" button.

3.3 Import Status

If not expanded, click the [+] icon next to "Upload Data" on the project overview page to expand the data upload section and view upload status.

The number of pending queued uploads for this project is listed next to the "Upload Data" section header. The pending uploads are listed individually under the "Pending" section of the "Upload Data" section:

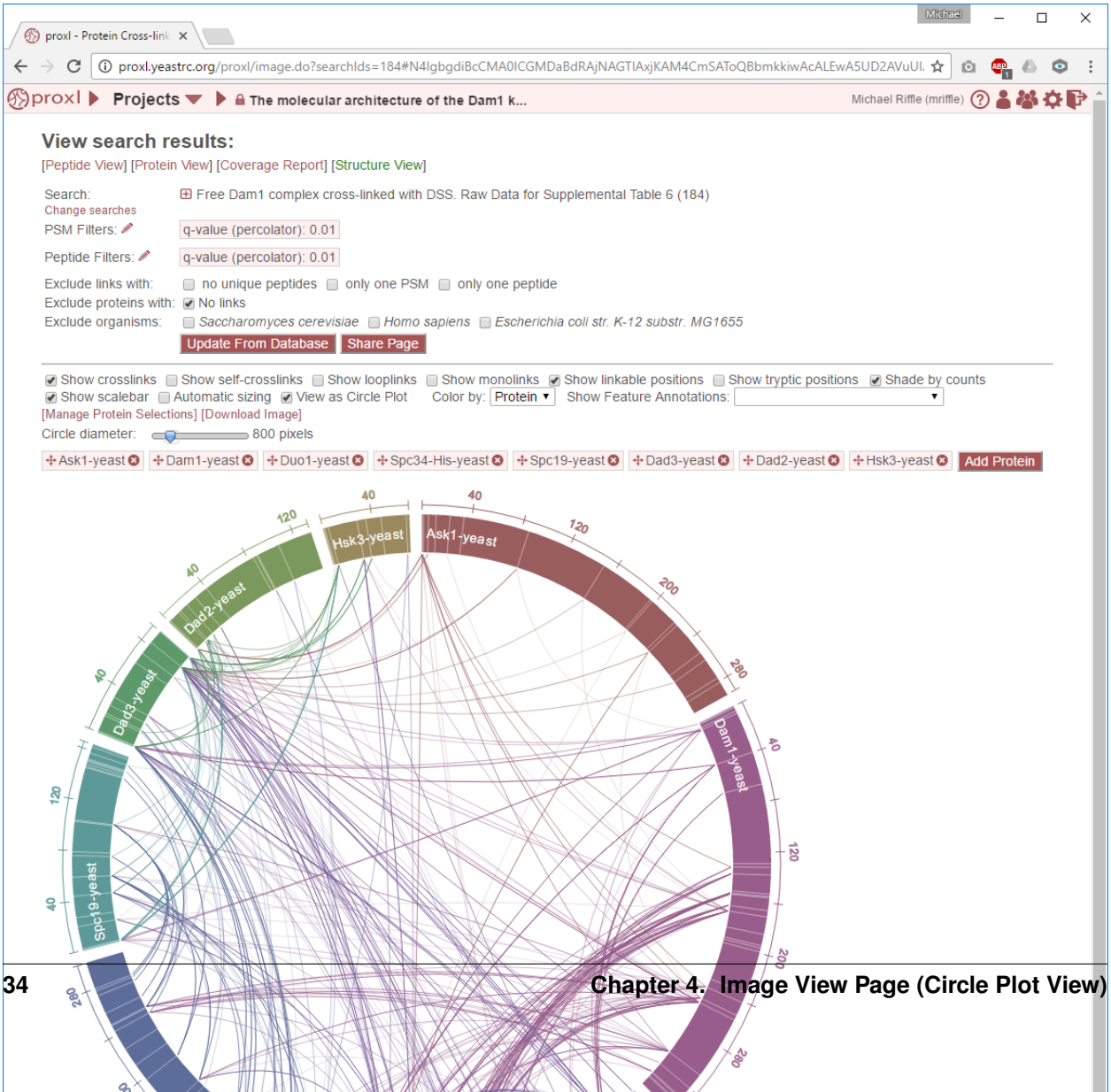


Clicking the "X" icon will remove this upload from the queue, effectively canceling the upload. This is only available if processing of this upload request has not yet begun.

Once completed, the upload will be moved from the "Pending" list to the "History" list, and the data for the new upload will be available in the "Explore Data" section:

CHAPTER 4

Image View Page (Circle Plot View)

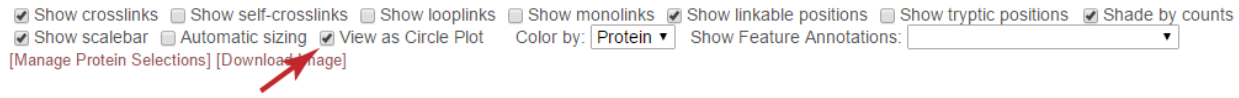


The circle plot image viewer displays a SVG rendering of select proteins as curved bars around the perimeter of a circle. The bars are annotated with positions of monolinks, looplinks, crosslinks, and other biological sequence annotations. The viewer is highly interactive and contains many options for customization (see below).

4.1 Viewer Basics

4.1.1 Switch to "Protein Bar" View

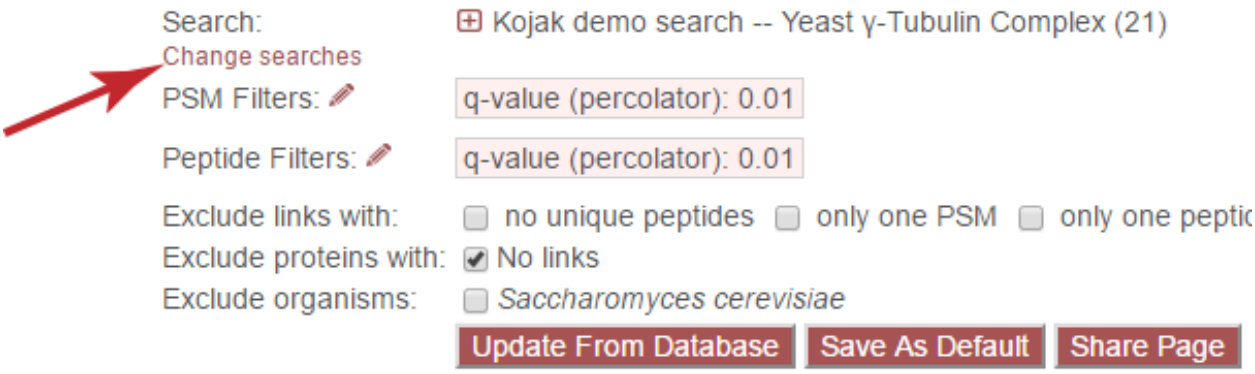
To switch to the "Protein Bar" view, click the checkbox next to "View as Circle Plot" in the viewer options. See *Image View Page (Protein Bar View)* for more information.



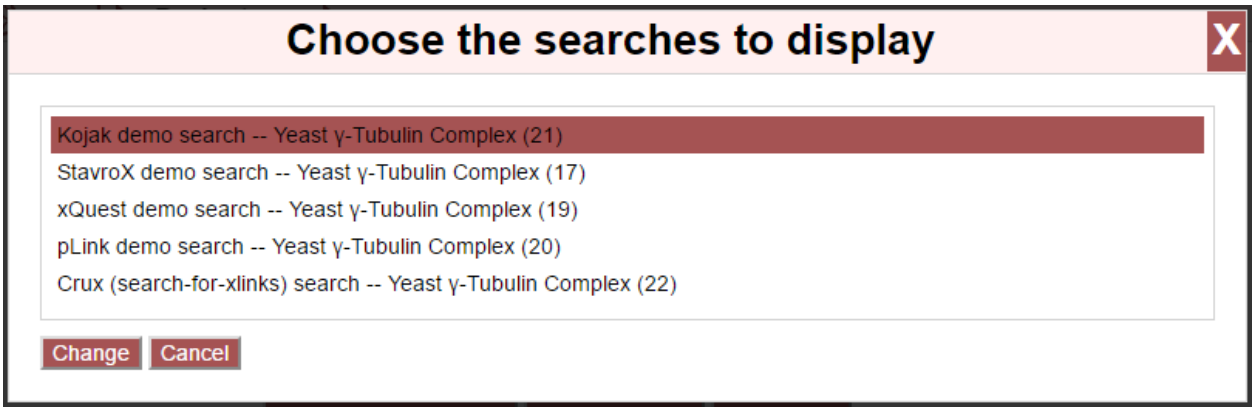
4.1.2 URL Captures State of Page

The URL of the page is dynamically updated at all times to reflect the complete state of the viewer–including filter parameters, protein bar positions, and all viewer options. As such, the URL may be bookmarked or shared to link to a viewer with the same content and appearance as the current viewer–just copy and paste from the address bar. For complicated visualizations, this is a convenient way to save and share you work. (Note, this link only works for other users who are listed on this project unless public access is enabled.)

4.1.3 Change Searches

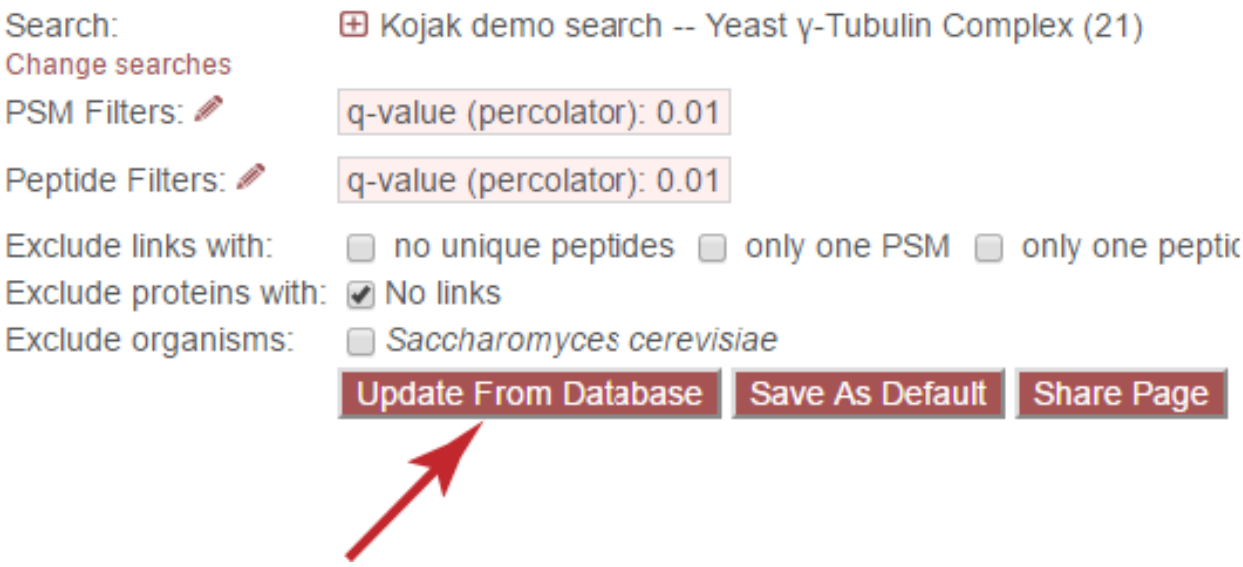


The "Change searches" link allows the user to change which searches are currently being displayed. None of the other options on the page (such as viewer settings) will be changed, only the searches from which data are displayed. Clicking the link causes the following overlay to be displayed:



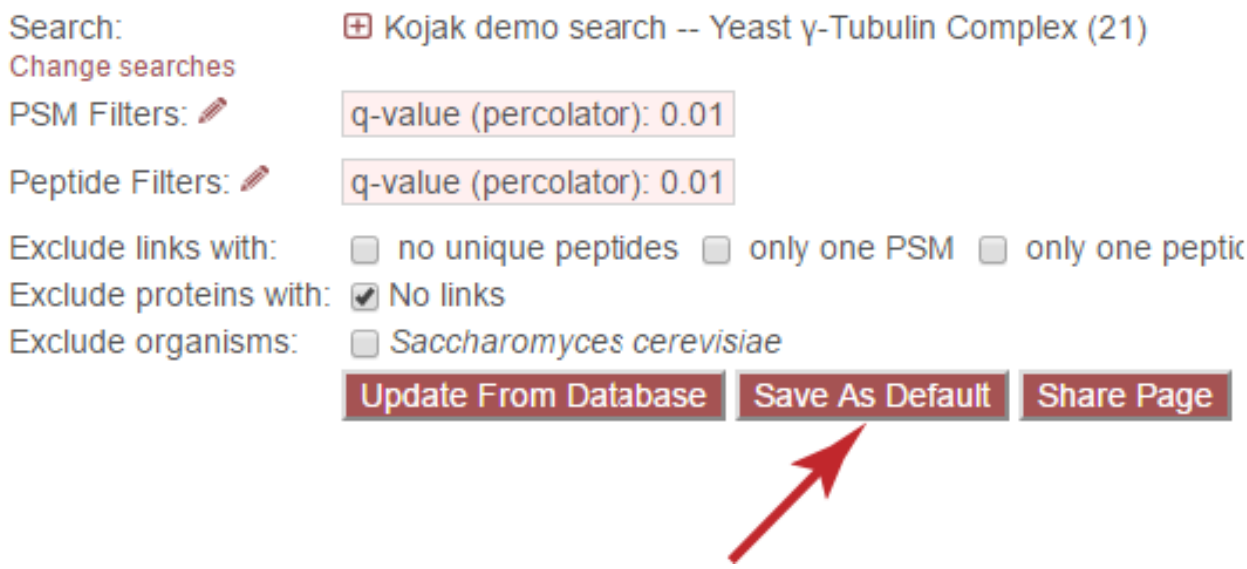
Select or de-select searches by clicking on them in the list. Once done, click "Change" to update the page with the new data or "Cancel" to close the overlay.

4.1.4 Update From Database



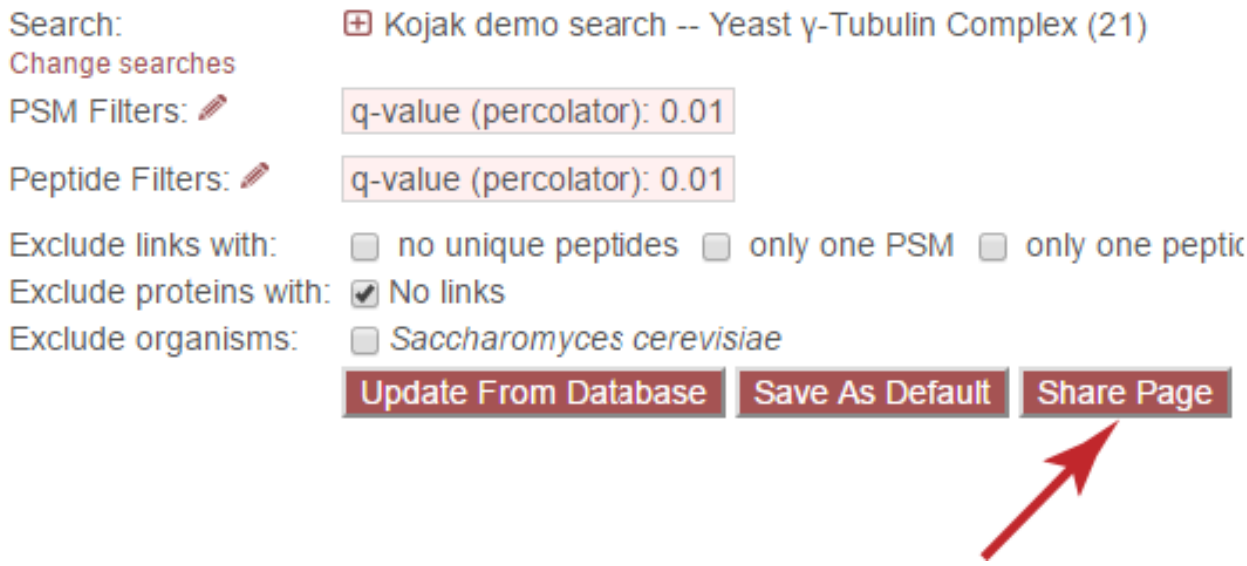
If the user changes any filter parameters–such as PSM/peptide score cutoffs, link type exclusions, protein exclusions, or taxonomic exclusions–this button must be clicked to reflect the new filter choices.

4.1.5 Save as Default

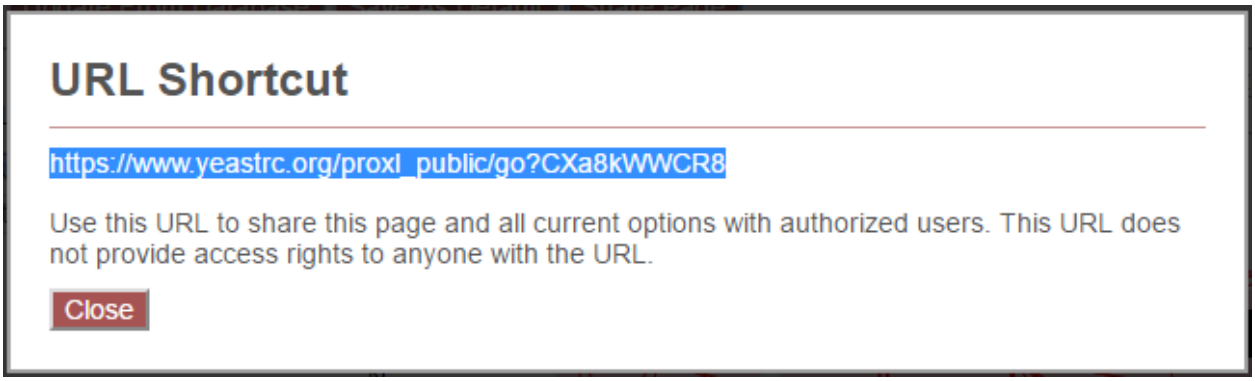


Project owners may click "Save as Default" to save the current URL (see above) as the default view of the "Image View" for this project. This default view will be populated with the same proteins, positions, and viewer options as when the button is clicked. This is a convenient way to share data with collaborators or the public that does not require that they manipulate the image viewer to see the data.

4.1.6 Share Page



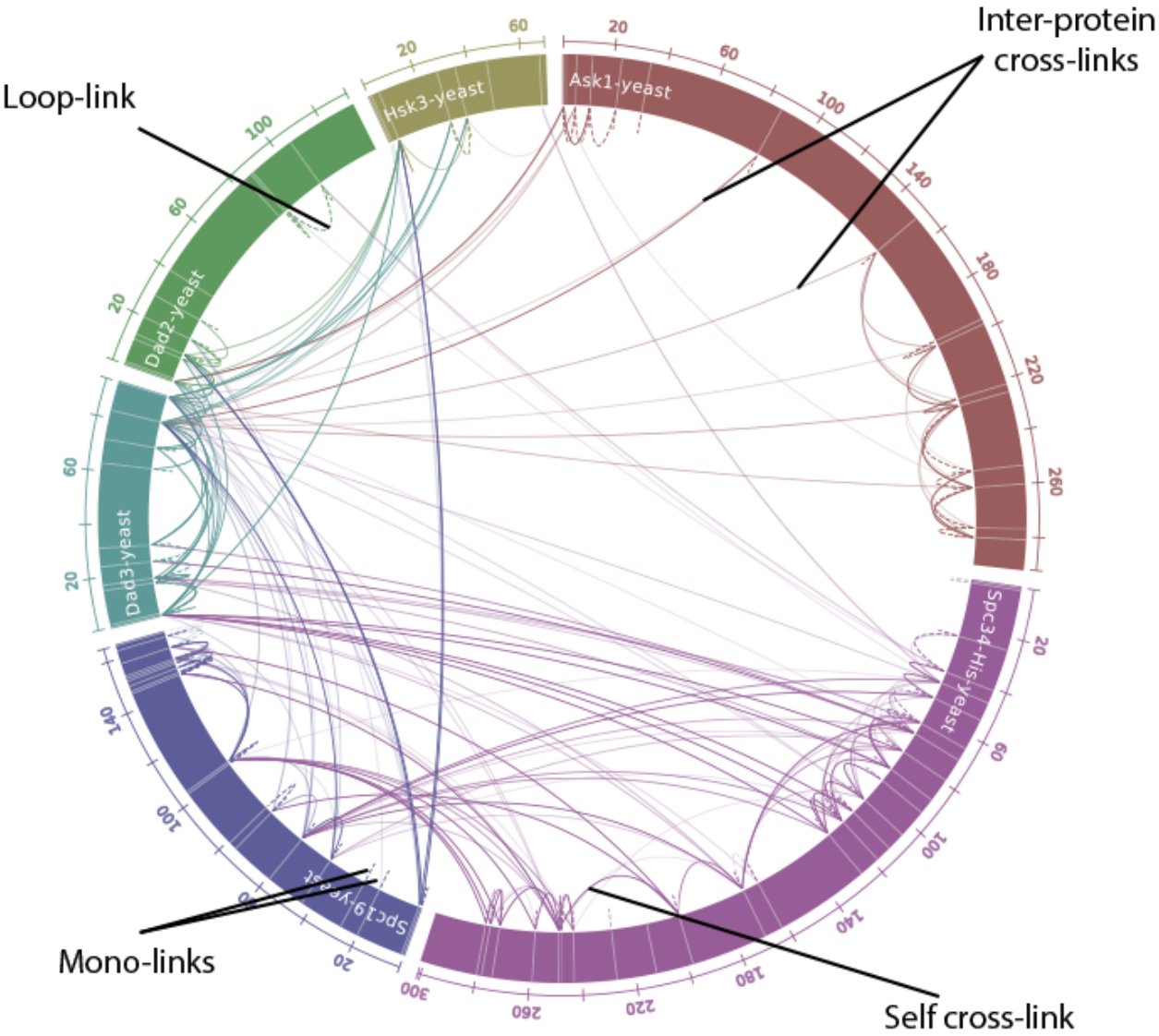
Clicking the "Share Page" button will generate a shortcut URL for viewing the current page, including all viewer settings, filter options, and so forth. The shortened URL will appear in an overlay as:



Copying and sharing the highlighted URL will direct users to the view of the page when the URL was generated. Note that this URL does not grant access to the page to any user that would not otherwise have access.

4.1.7 Viewer Display

Below is a labeled example of the circle plot viewer. Inter-protein cross-links appear as arcs between proteins. Self cross-links (cross-links within the same protein) appear as solid-colored loops on the interior of the circle. Loop-links appear as dashed loops on the interior of the circle. And mono-links appear as short dashed lines sticking into the interior of the circle.



4.1.8 Coloring

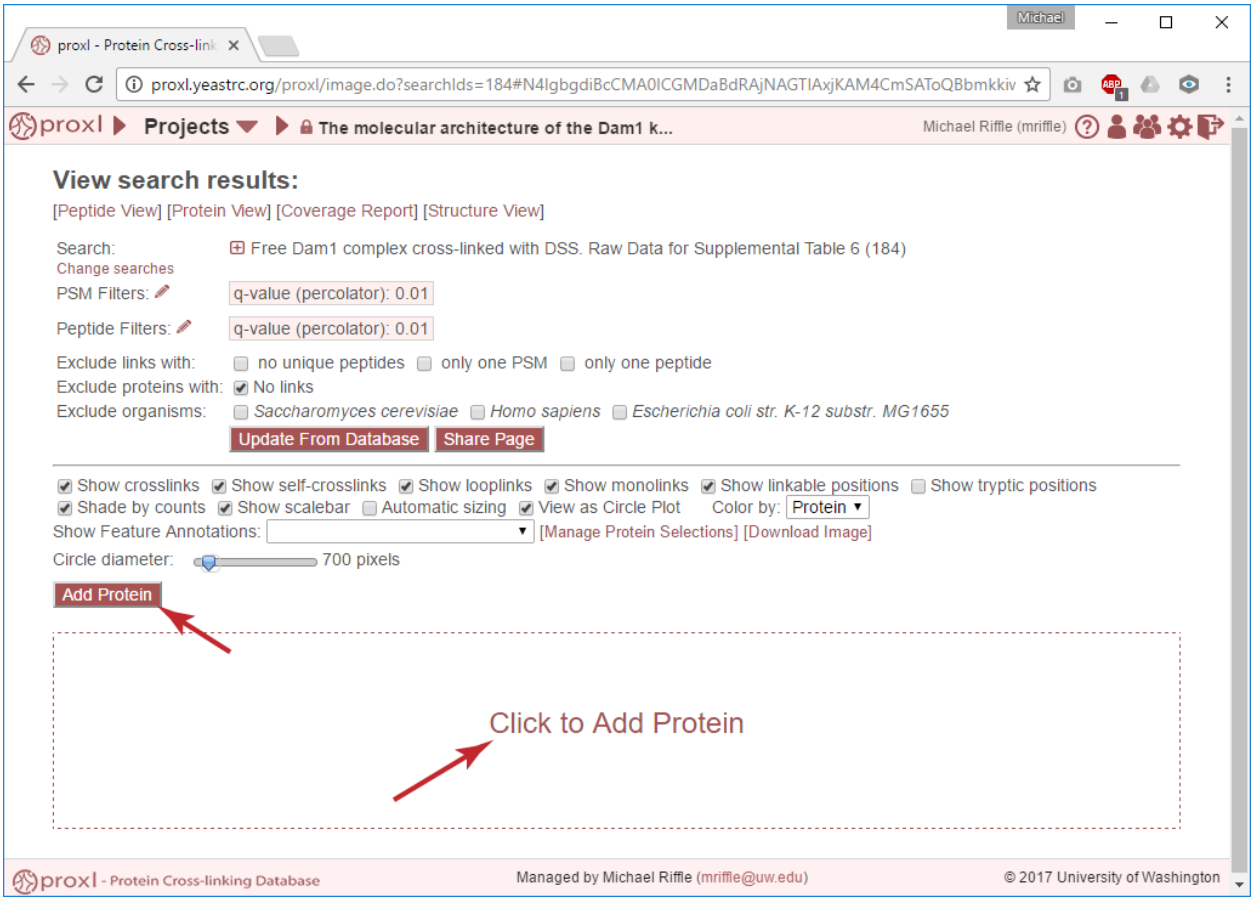
By default, the coloring of the links corresponds to the "originating" protein in the viewer. For example, if the first protein is colored red, all of its self-crosslinks, looplinks, and monolinks will be red. And all inter-protein cross-links containing this protein will be red. The originating protein of a crosslink will be the protein which appears first in the viewer (from top-down). Similarly, if the second protein is colored green, all of its links and originating inter-protein cross-links will be colored green (except for those involving the first protein, which are red). This coloring scheme is intended to ease confusion about which links involve which proteins, farther down the list of proteins.

This coloring scheme can be changed to show in which search(es) the link appears (See *Color by search*). And, specific proteins in the viewer can be highlighted to only color links involving those proteins (See *Highlight Proteins*).

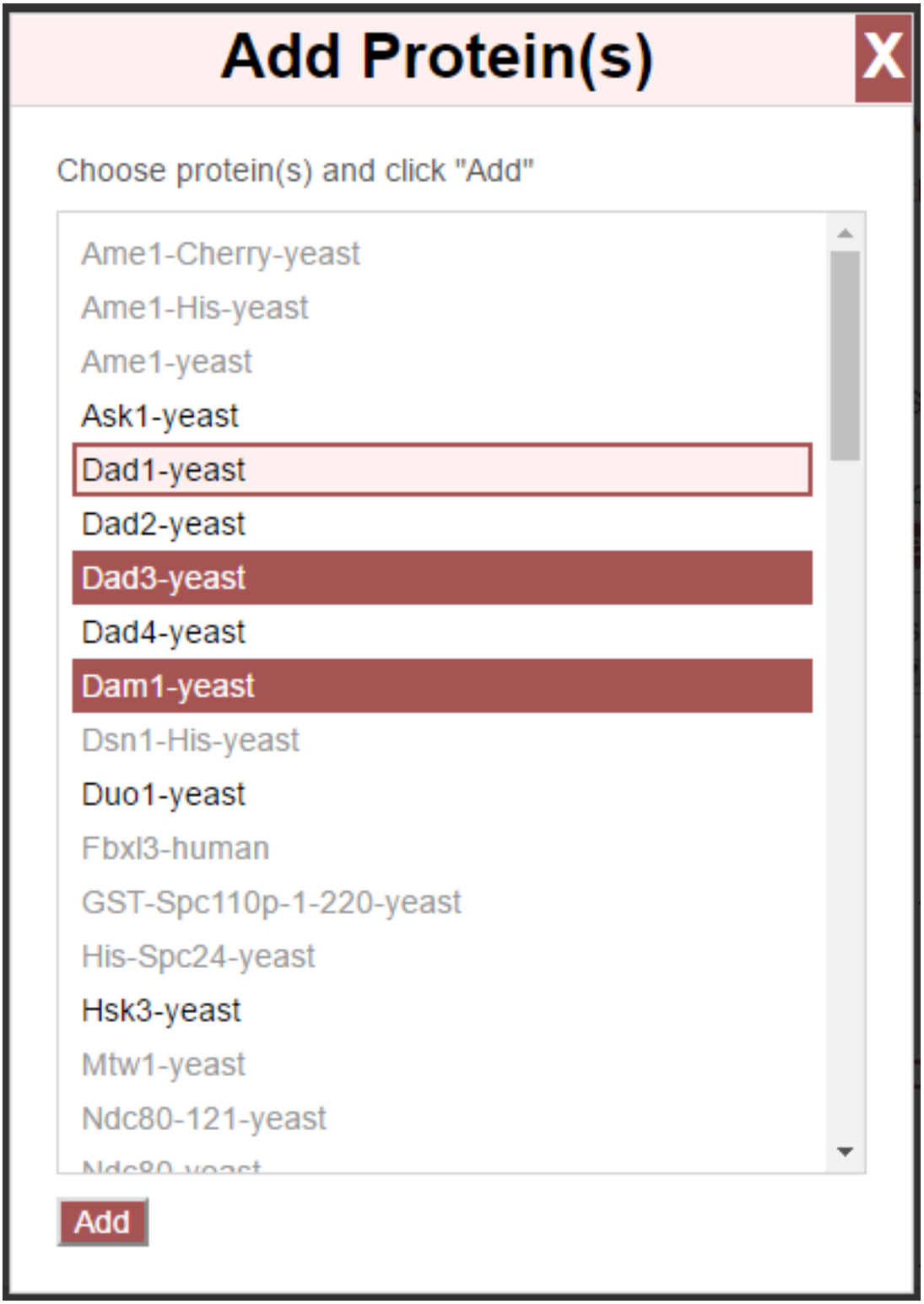
4.2 Viewer Interaction

4.2.1 Add a Protein

To add a protein to an empty viewer (no proteins visible), click either the "Add Protein" button or the "Click to Add Protein" text in the viewer area:



This will open the "Add Protein(s)" overlay, where one or more proteins may be selected:



Cick “Add” to add the selected proteins to the viewer. Click anywhere outside the overlay, or on the “X” in the top-right corner, to close the overlay.

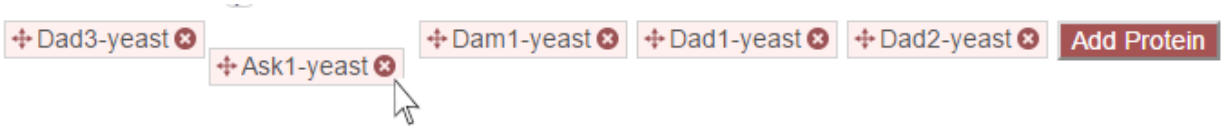
To add more proteins to the viewer, click the “Add Protein” button above the viewer.

4.2.2 Remove a Protein

To remove a protein from the viewer, click the small red (X) next to the name of the protein above the image area:

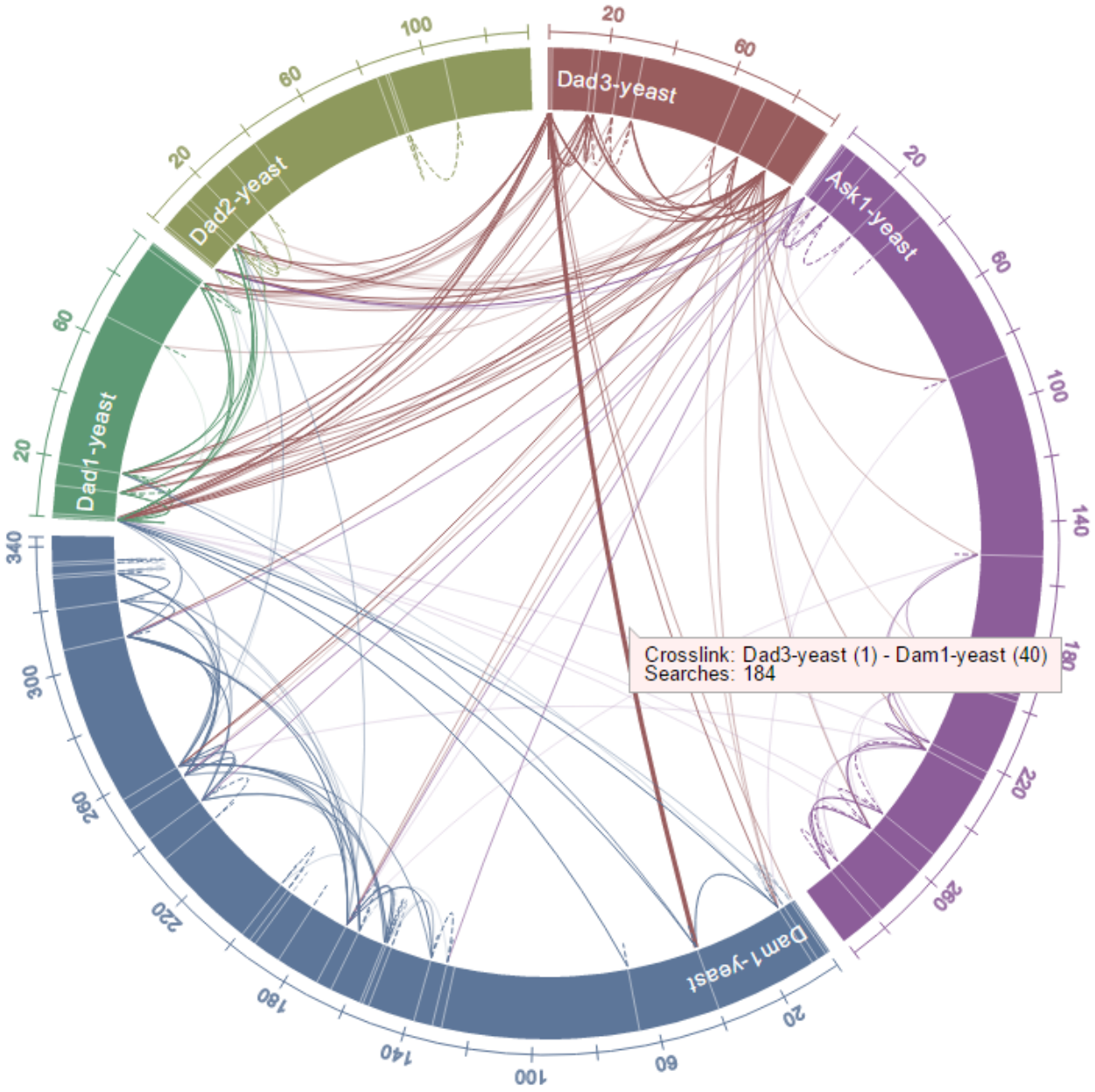
4.2.3 Rearrange Proteins

To rearrange proteins, simply drag and drop the protein in the protein list above the circle to the desired position in the order. The viewer will update automatically.



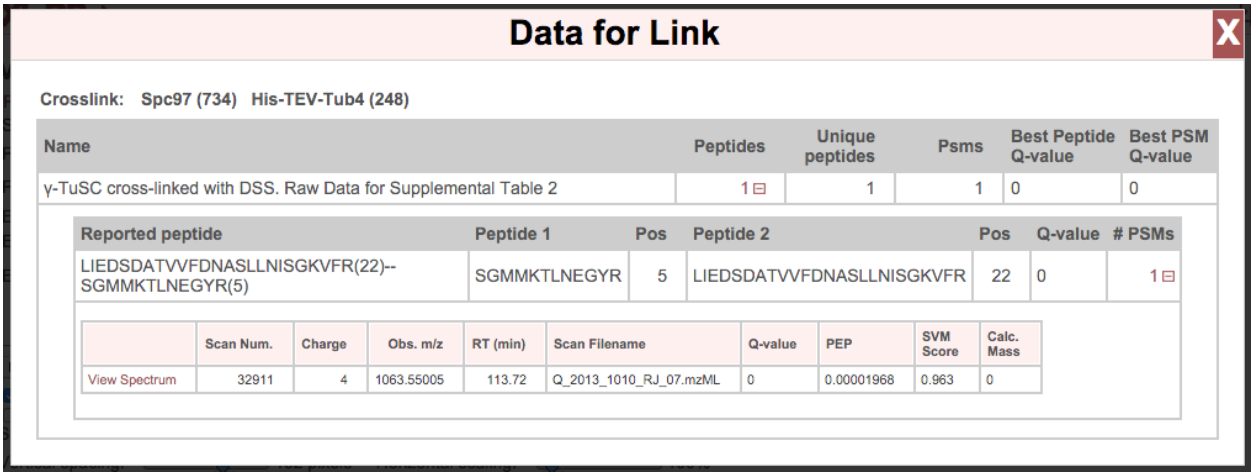
4.2.4 View Link Summary

To view summary information about a link, hover your mouse arrow over that link in the viewer. (Or tap, on touch devices.) This may be done for any link type. This will display the link type, protein(s), and position(s).



4.2.5 View Peptides, PSMs, and Spectra

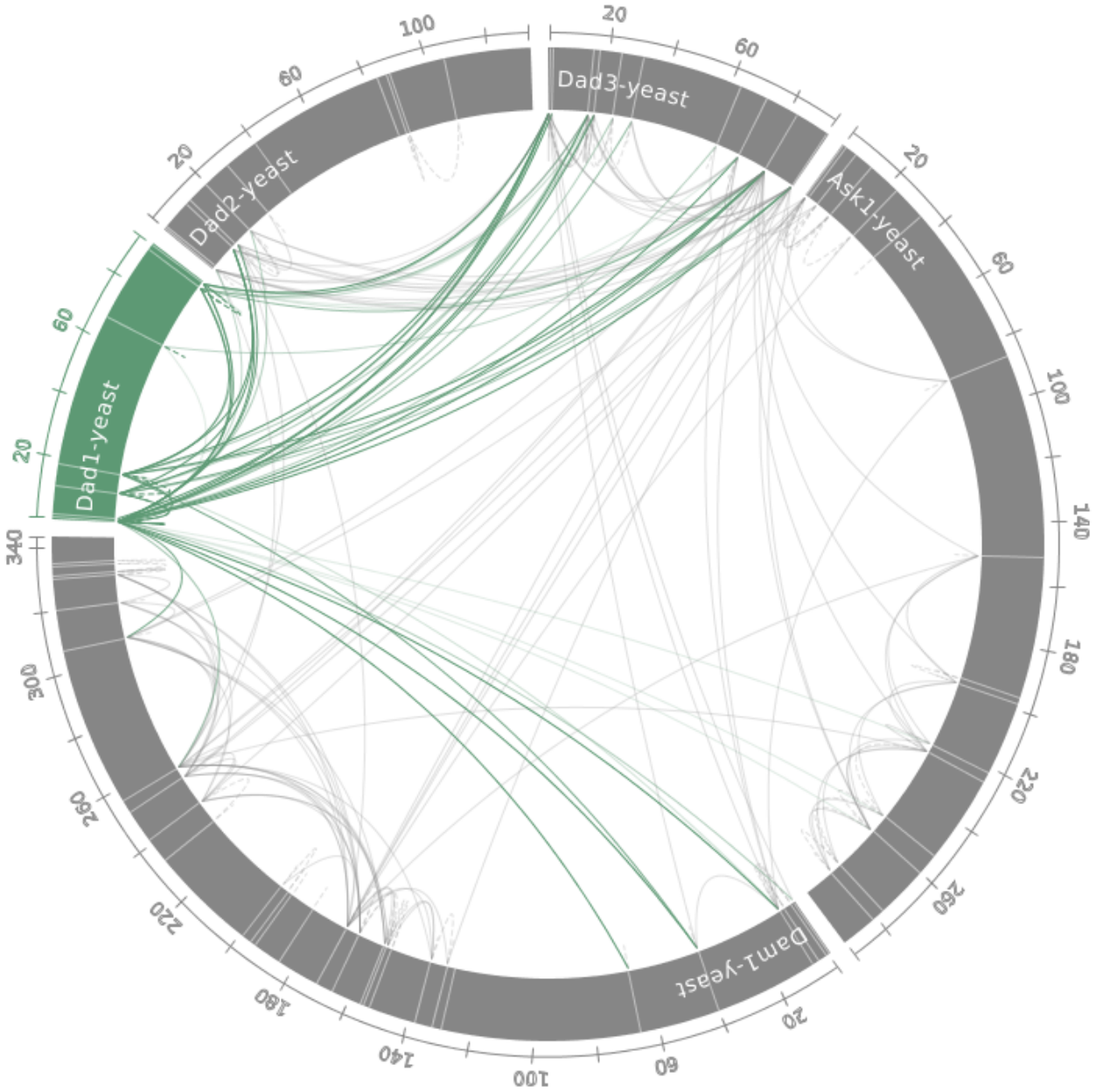
To view listings of the peptides and corresponding PSMs for a link, click on that link in the viewer. This will open an overlay window displaying a table of peptides and PSMs:



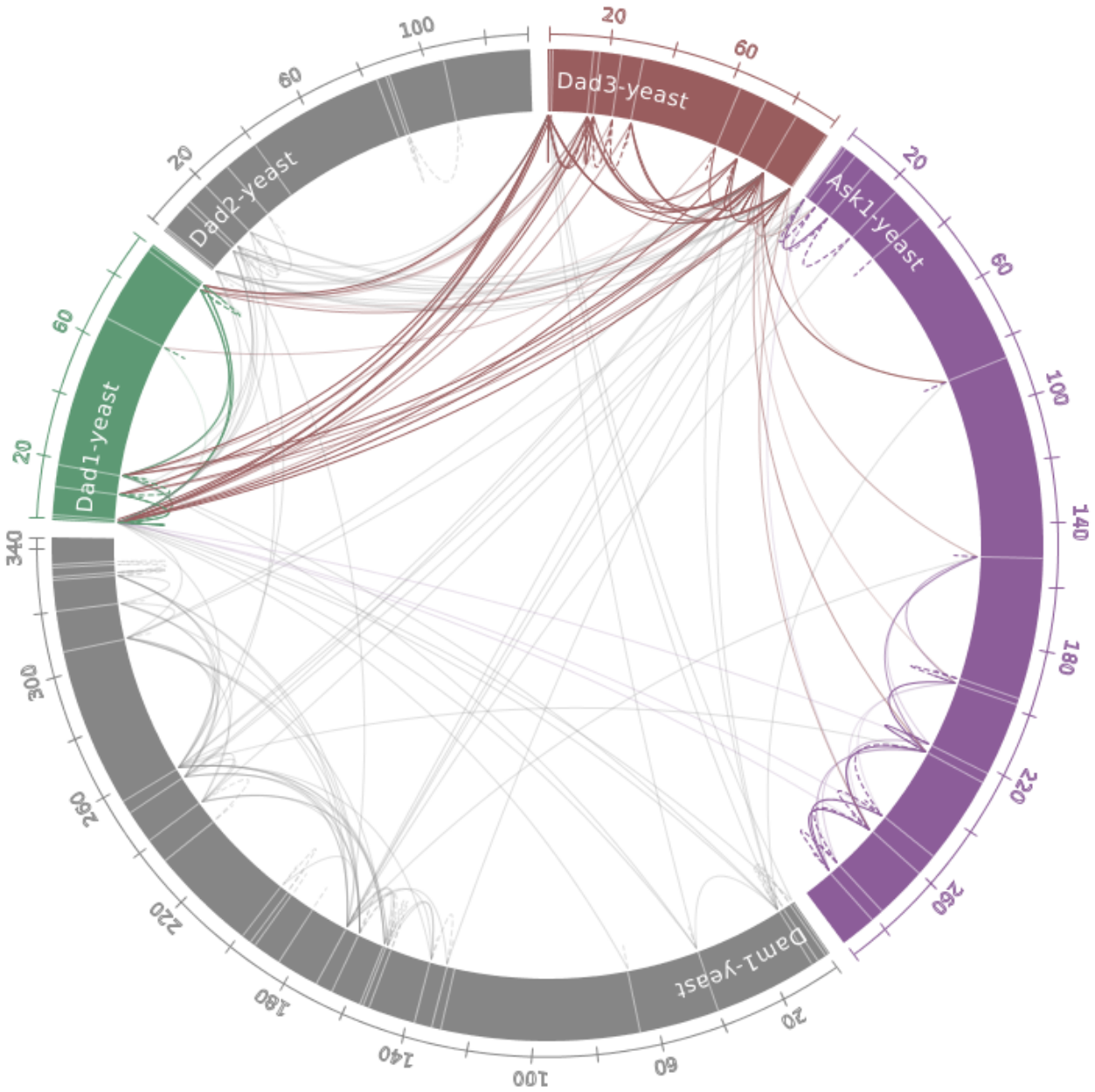
The top-level of this table are rows for each search in which this link was found. If multiple searches have been merged, each search will appear as a row in the table. Each search may be expanded by clicking on that row to view all peptides from that search that led to this link. Each peptide may be expanded by clicking on that row to view all PSMs for that peptide. Each PSM includes a "View Spectrum" link for viewing an annotated spectrum associated with that PSM. For help on our spectrum viewer, see the *Spectrum Viewer* page. Click the "X" in the top-right corner of the overlay (or click on the page anywhere outside of the overlay) to close the overlay window.

4.2.6 Highlight Proteins

Proteins may be highlighted by clicking on any of the protein bars. This will change the coloring scheme such that only links involving the highlighted protein(s) will be colored, all other links for all other proteins are shaded light gray:



Multiple proteins may be highlighted by holding shift and clicking protein bars:



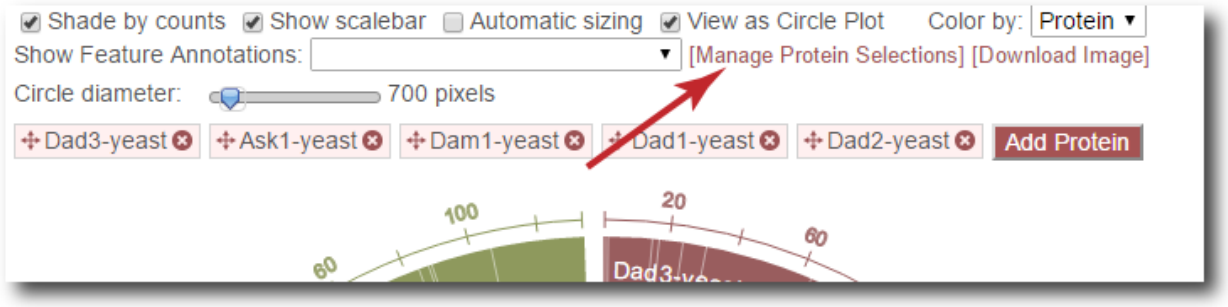
When multiple proteins are highlighted, all links within and between those proteins will be colored. Everything else will be greyed-out.

Remove Highlighting

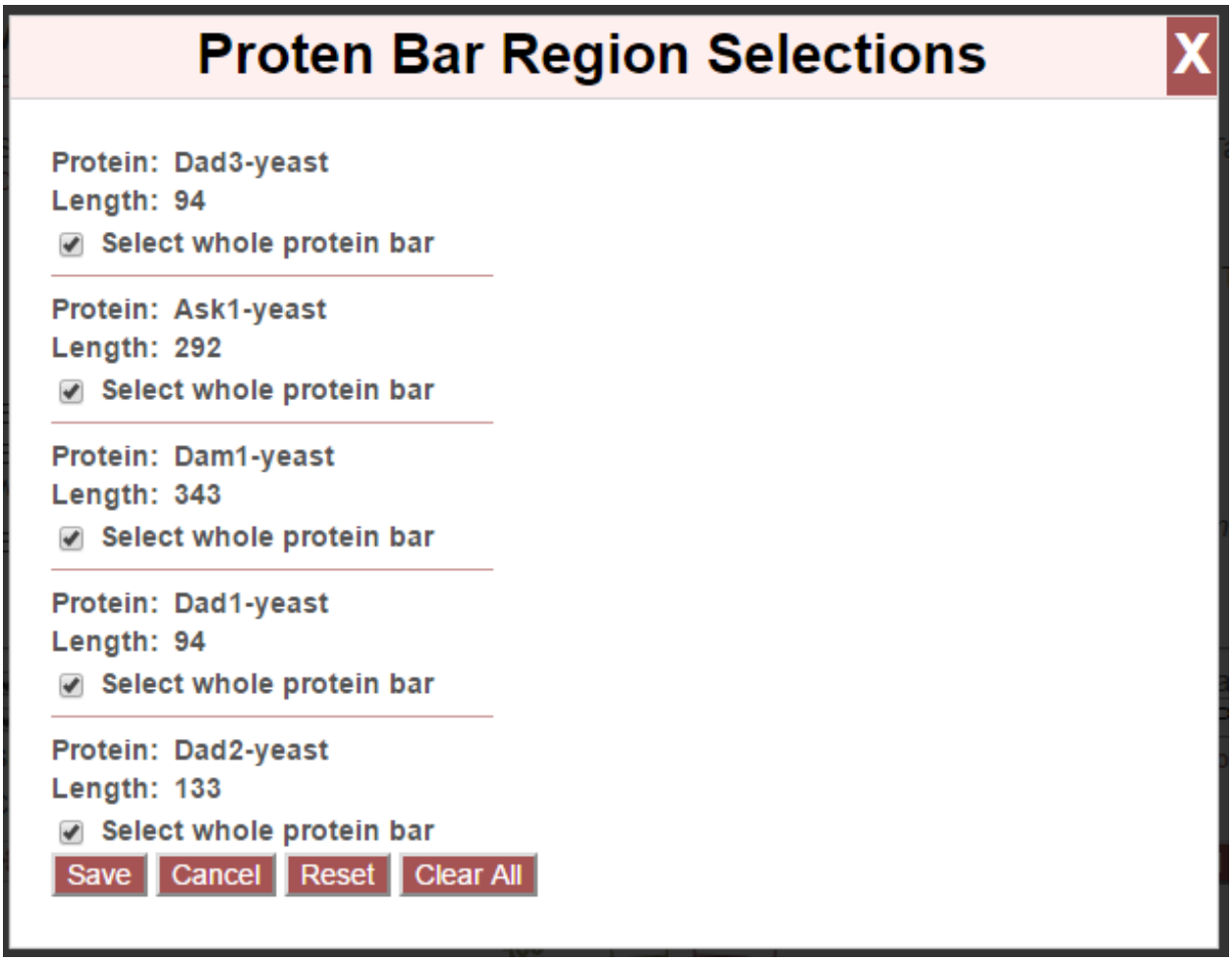
If a single protein is highlighted, click it to unhighlight it. If multiple proteins are highlighted, hold shift and click a highlighted protein to unhighlight it. If shift is not used, the viewer will highlight only the protein clicked.

4.2.7 Highlight Proteins Regions

Instead of highlighting entire proteins, it is possible highly only segments of proteins (protein regions). This is done by clicking the "[Manage Protein Selections]" link above the image:

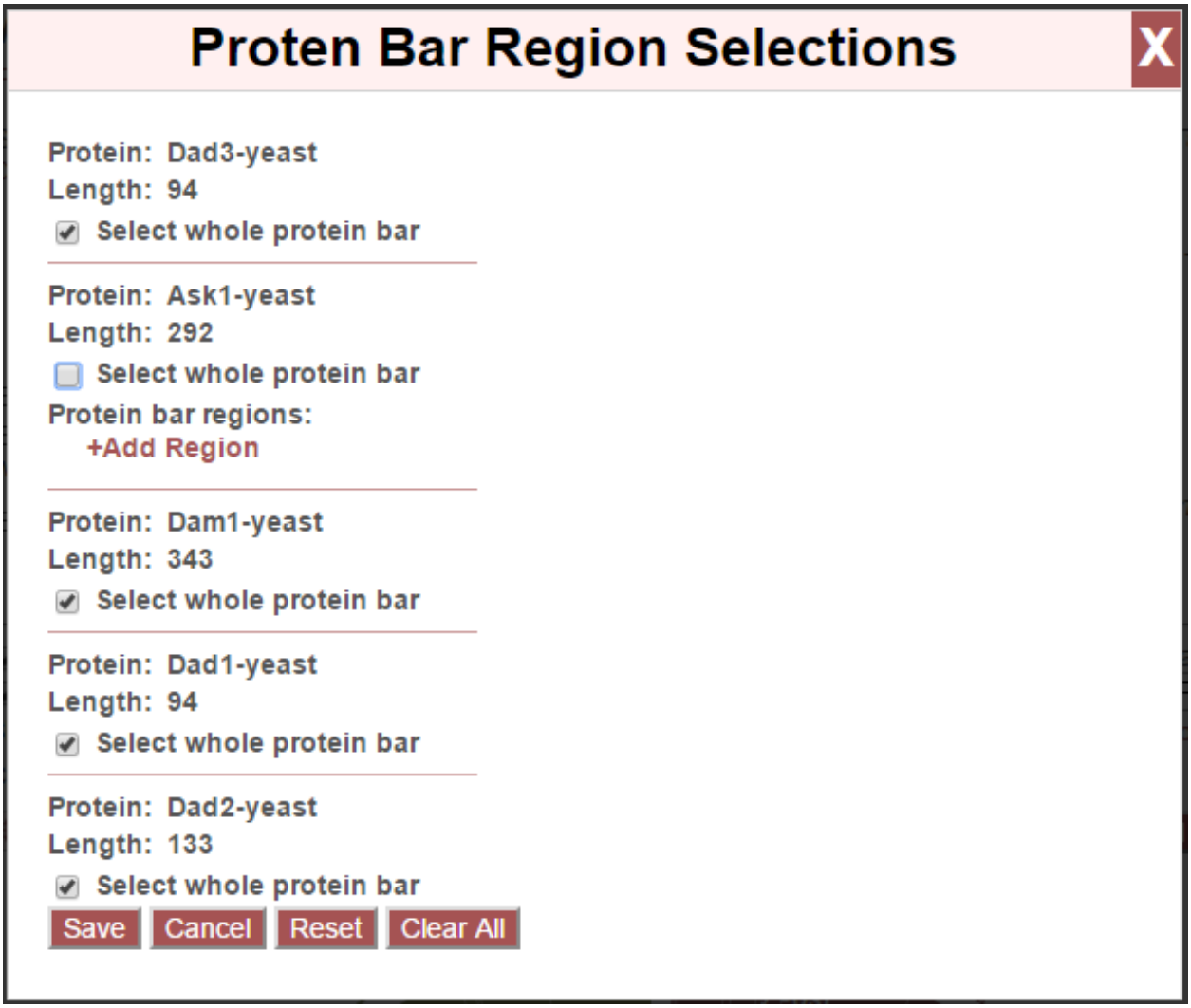


This opens an overlay with an interface for managing which regions of which proteins are highlighted:

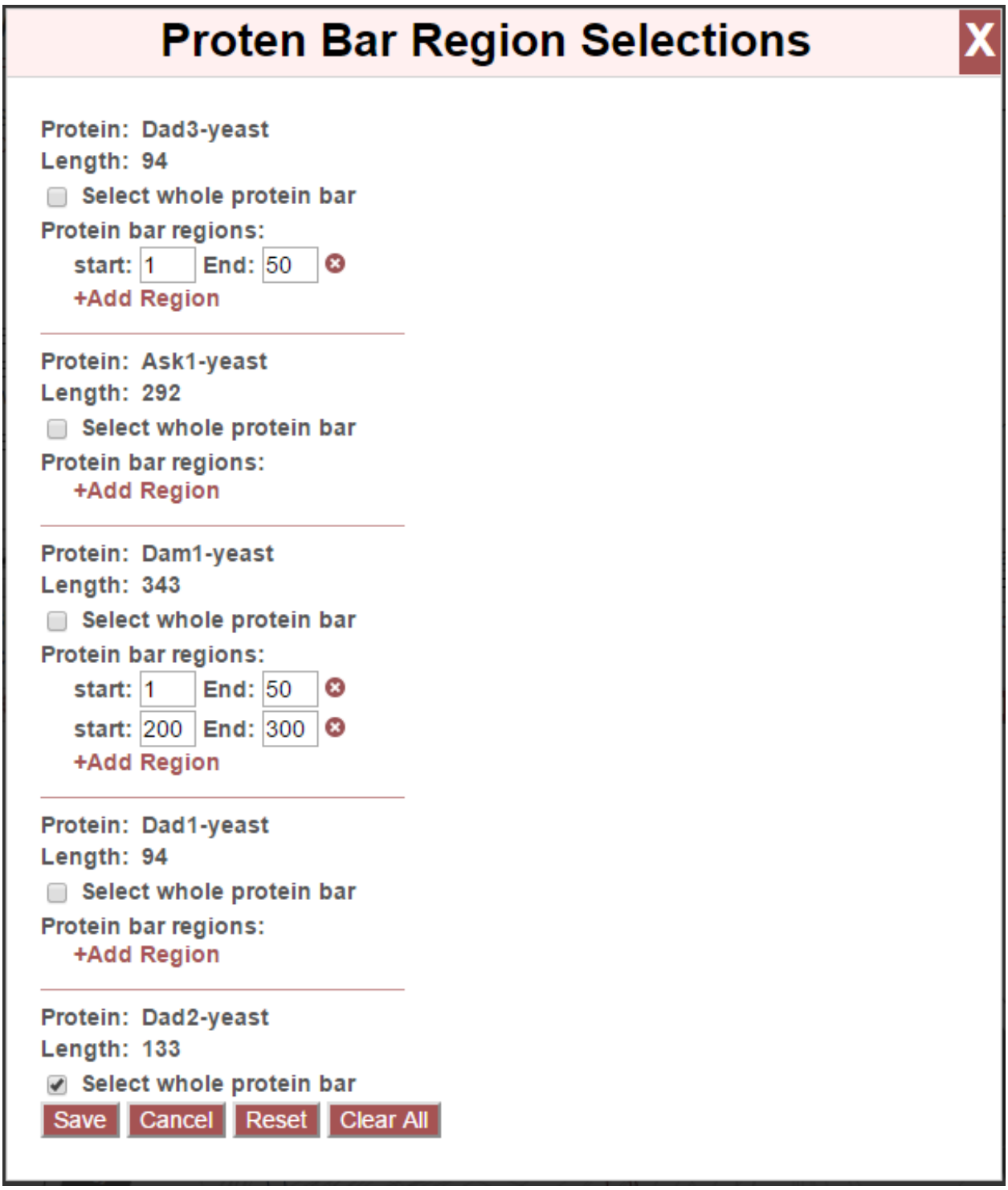


This overlay lists all proteins visible in the viewer and which regions are currently highlighted for each one. By default, all proteins are visible in their entirety, so the "Select whole protein bar" option is checked for all of them. When this checkbox is checked, that protein is highlighted in its entirety and no sub-regions may be selected.

To select sub-regions in a protein to highlight, uncheck the "Select whole protein bar" option:

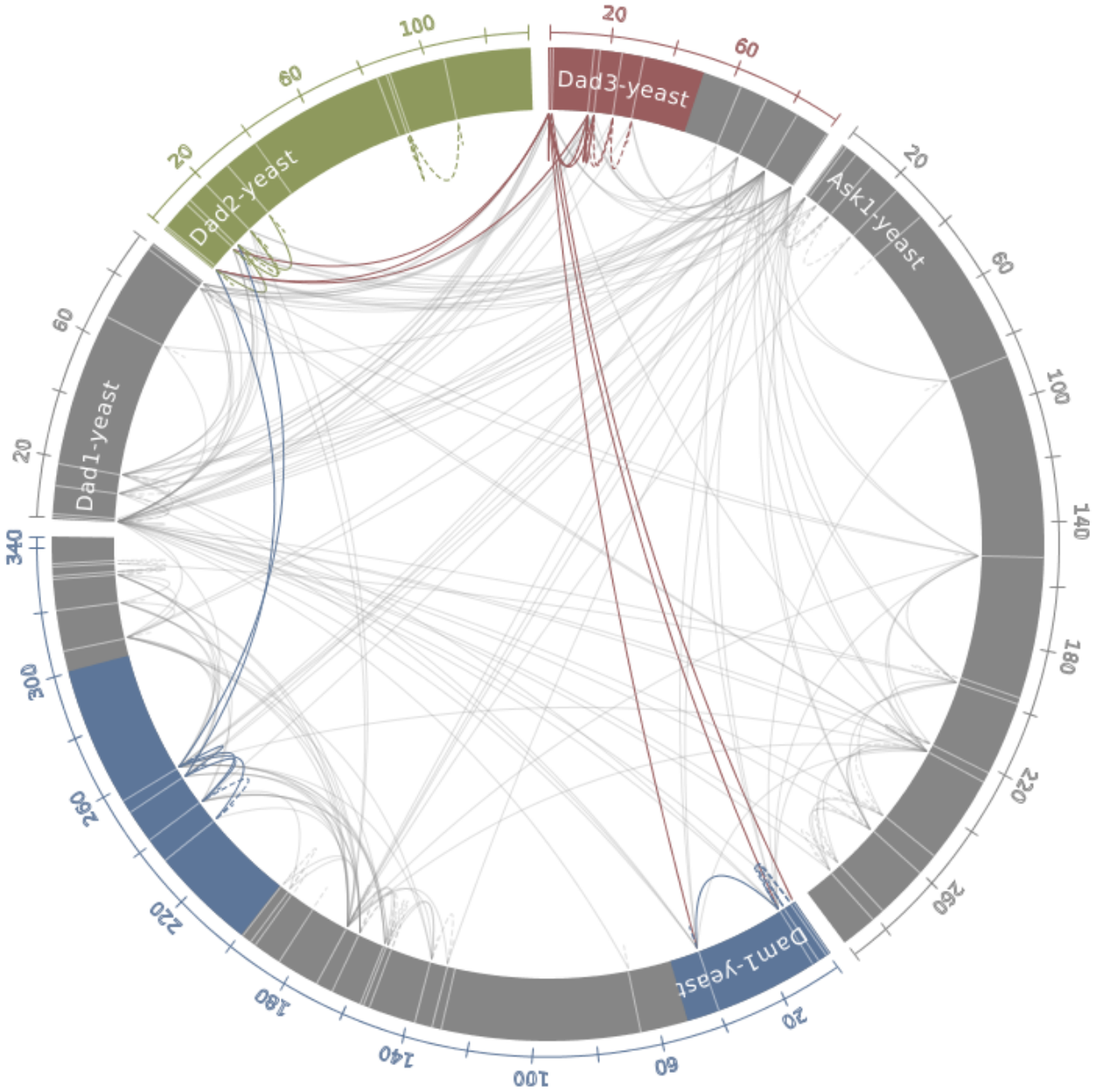


Click "+Add Region" to define the start and end positions for a region to highlight in that protein. This option may be used multiple times per protein to define multiple regions. Click the red "x" next to a region to remove it. Unchecking "Select whole protein bar" and not defining regions unhighlights that entire protein:



Click “Save” to save these settings and view the image with these defined regions. Click “Cancel” to make no changes and close overlay, “Reset” to reset regions to those visible in the image (without closing overlay), and “Clear All” to set to defaults (all proteins visible).

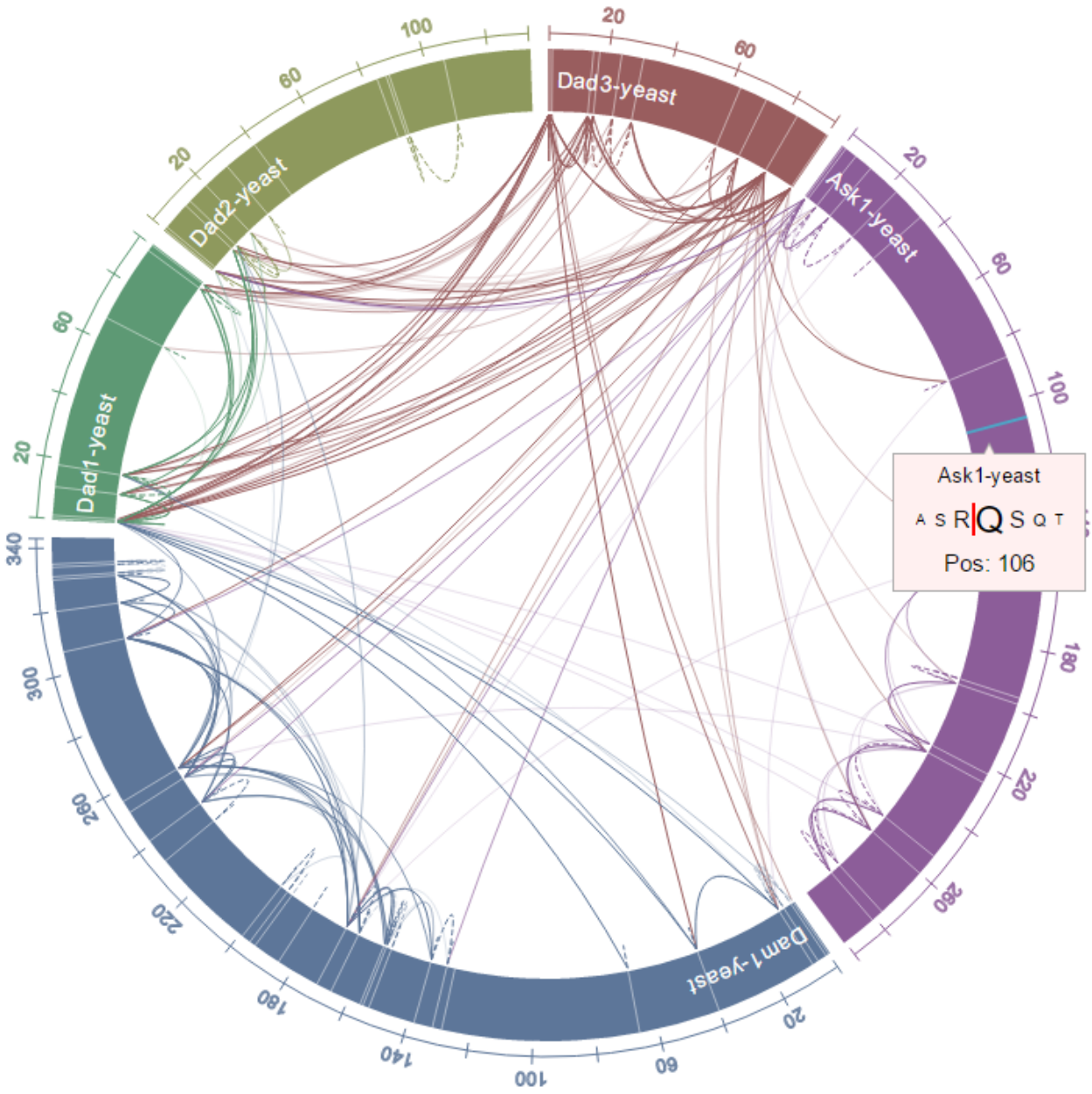
Clicking “Save” gives us:



Only links within and between the defined regions are colored. The reset are greyed-out.

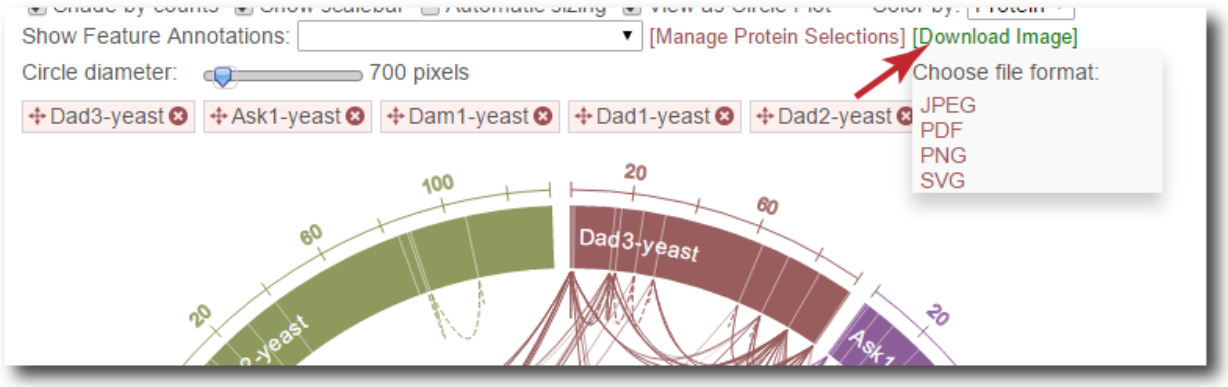
4.2.8 Local Sequence Information

Local sequence information in the protein bars may be viewed by hovering the mouse cursor over the protein bar. A tooltip will appear that shows the amino acid position number, the amino acid at that position, and neighboring amino acids. Amino acids that linkable with the cross-linker(s) used in the experiment(s) will be bolded and red. Vertical bars indicate sites that are cleavable by trypsin. This tooltip will slide and dynamically update along with the mouse cursor as it is moved along the protein bar.



4.2.9 Download Image

Mouse over the "[Download Image]" link to see image download options. Click on the format of choice to initiate a download of the image.



4.3 Viewer Options

4.3.1 Show crosslinks

Toggle the showing of inter-protein crosslinks.

4.3.2 Show self-crosslinks

Toggle the showing of intra-protein crosslinks.

4.3.3 Show looplinks

Toggle the showing of looplinks.

4.3.4 Show monolinks

Toggle the showing of monolinks.

4.3.5 Show linkable positions

Toggle the showing of which positions in the protein are linkable by the cross-linker(s) used in the experiment. The linkable positions are noted by white lines in the protein bar.

4.3.6 Show show tryptic positions

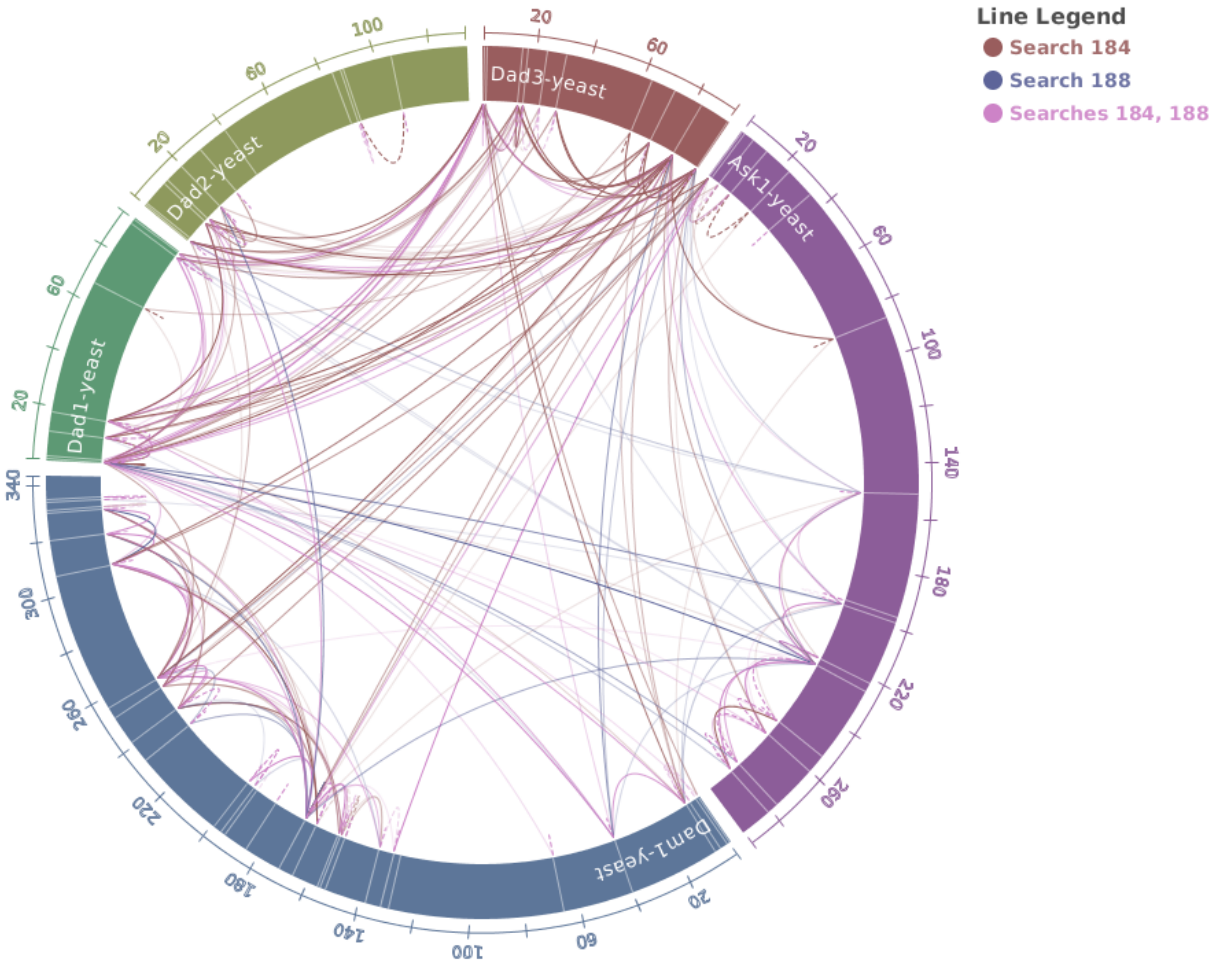
Toggle the showing of which positions in the protein are cleavable by trypsin, an enzyme commonly used to digest proteins in bottom-up proteomics experiments. The cleavable positions are noted by dashed white lines in the protein bar.

4.3.7 Shade by counts

If enabled, the opacity (transparency) of links reflects the number of PSMs found (or spectrum count) for the shown link. The shading scales from 1 PSM (minimum opacity, most transparent) to 10 PSMs (maximum opacity). Any link having 10 or more PSMs will have the maximum opacity.

4.3.8 Color by search

When merging multiple searches, this option changes the coloring scheme so that all links are colored by which search (or searches) they were found in at the given cutoffs. Each search is assigned a color, and each combination of searches are assigned other, distinct colors. It is possible to ascertain from the color in which search, or combination of searches, the individual link was found. A legend is provided with the graphic. This functionality is limited to a maximum of three searches.

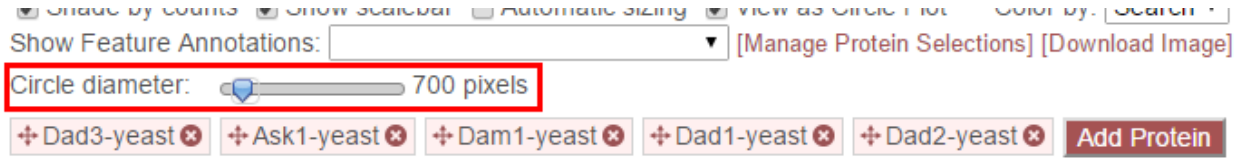


4.3.9 Show scalebar

Toggle the display of the scale bar on and off.

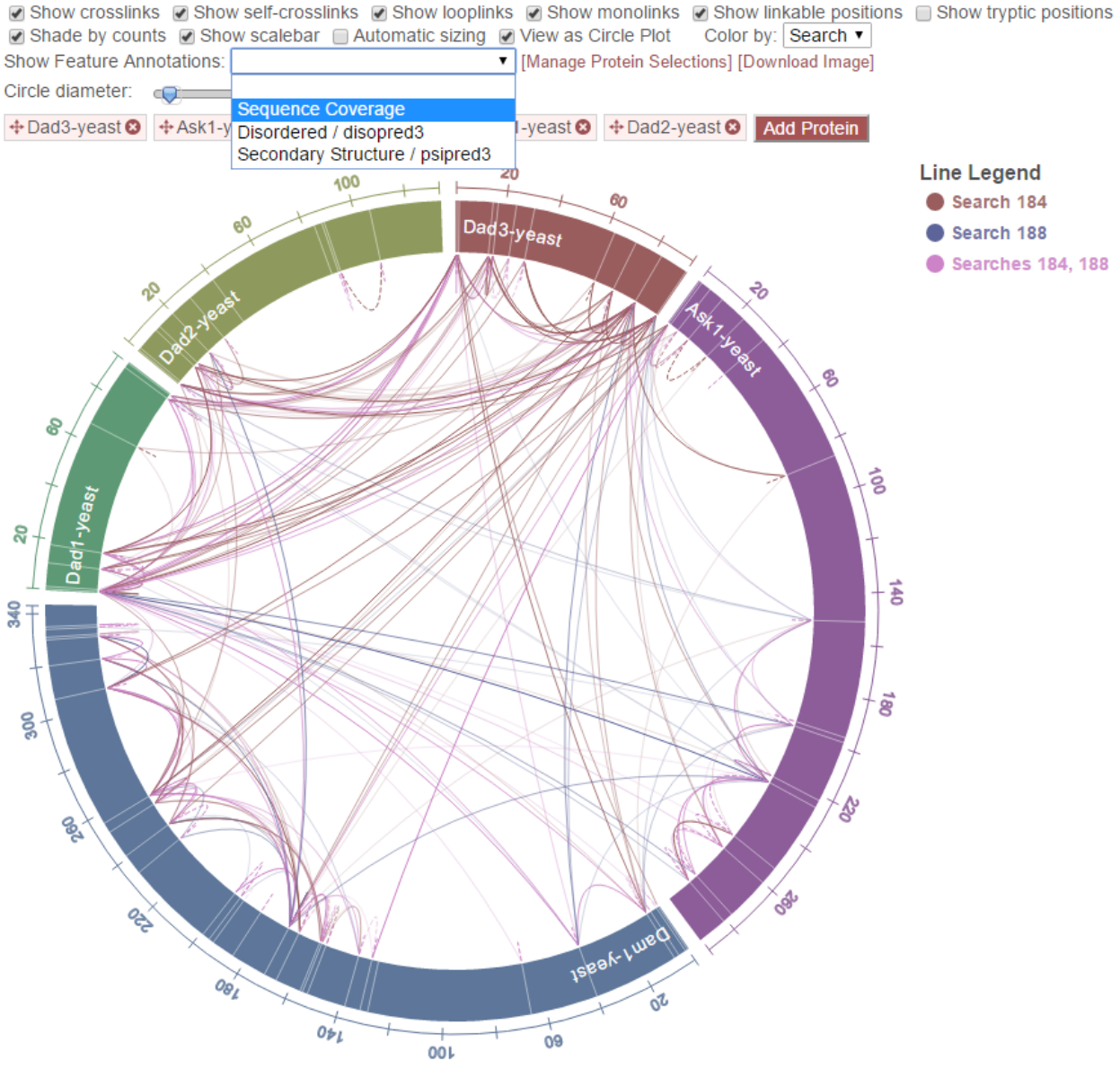
4.3.10 Automatic sizing

The viewer automatically sizes the circle plot with a diameter of 800 pixels. To change this, uncheck the "Automatic sizing" checkbox and use the slider to change the radius of the circle plot.



4.3.11 Show Feature Annotations

This option allows for the display of protein sequence feature annotations of various on the protein bars. To select a type of feature annotation, click the pull-down menu next to "Show Feature Annotations" and select a type:



This will retrieve the necessary data from the server and display the respective annotation as bars aligned outside of the protein bars on the circle:

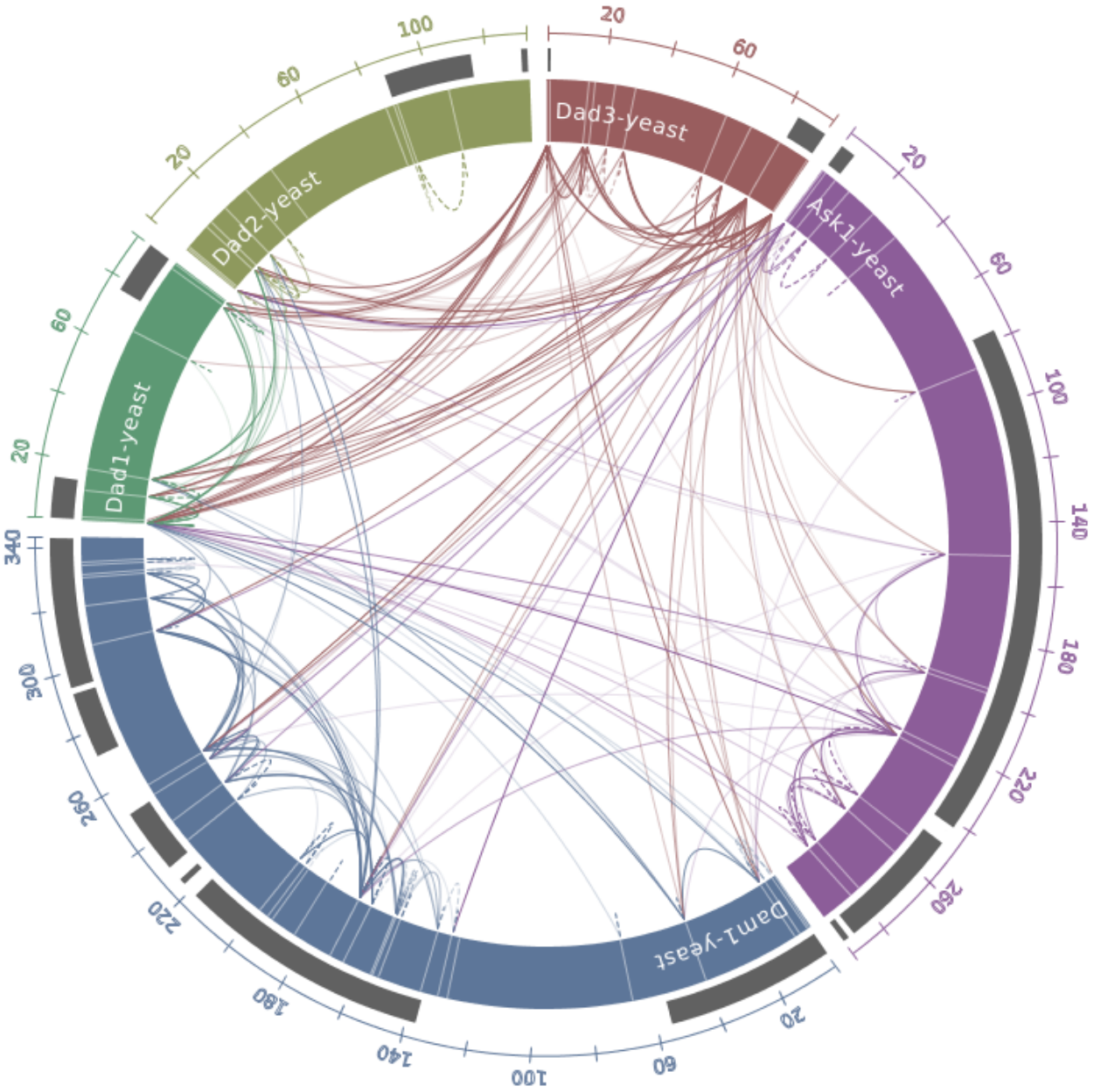
The types of feature annotations currently supported are:

Sequence Coverage

Sequence coverage shows which regions of the protein's sequence are covered by peptides of any type from the search(es) that meet the filtering criteria. An example of viewing the sequence coverage is shown above. The regions may be moused over to view exact start and stop residues.

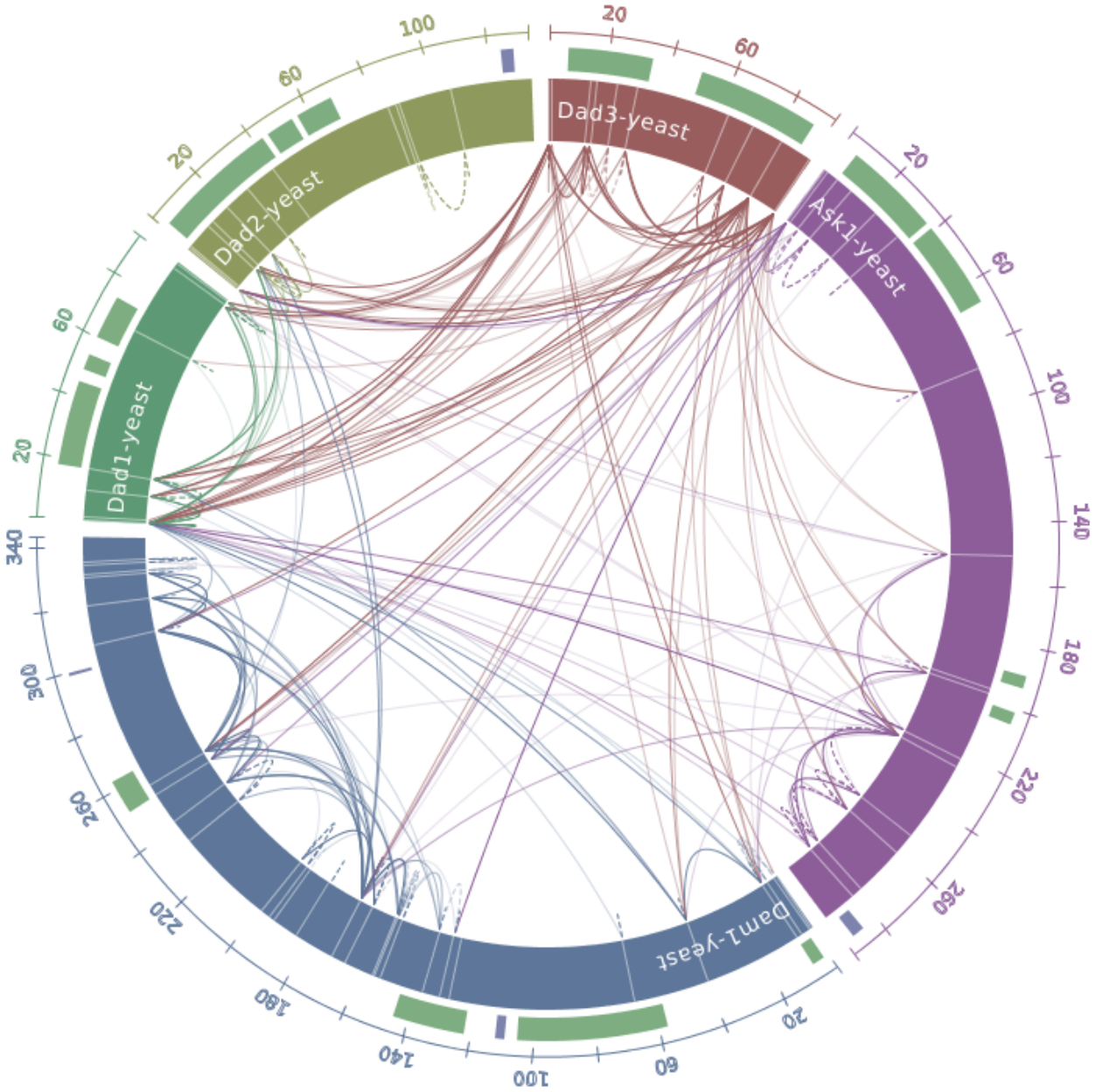
Predicted Disordered Regions

Selecting this option annotates the protein bars to show predicted disordered regions according to the DISOPRED3 algorithm. The regions may be moused over to view exact start and stop residues. This feature requires that PAWS be available, see: *Feature Annotations and PAWS*.



Predicted Secondary Structure

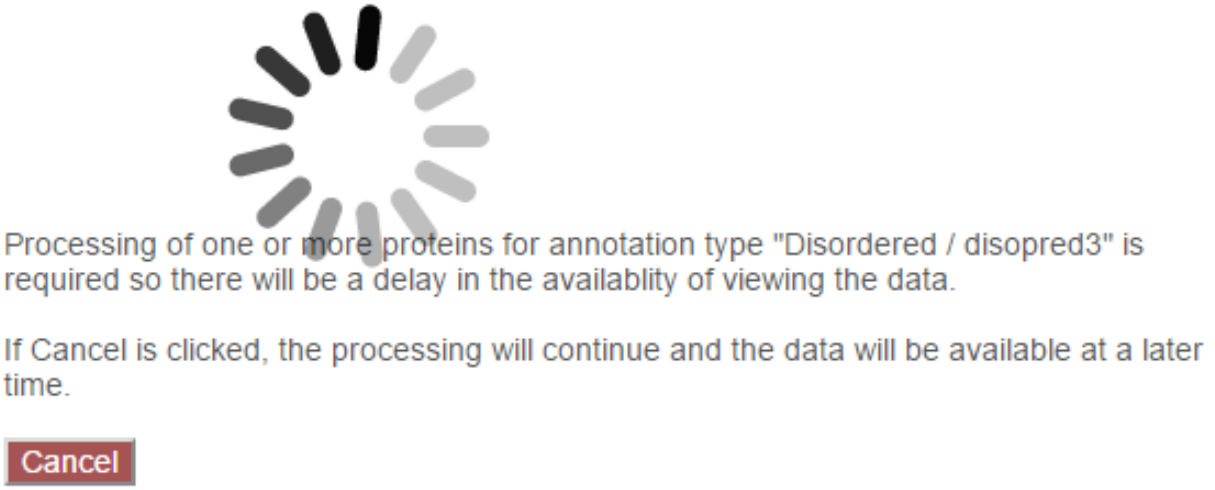
Selecting this option annotates the protein bars to show predictions for secondary structure according to the PSIPRED 3 algorithm. The regions may be moused over to view exact start and stop residues. This feature requires that PAWS be available, see: *Feature Annotations and PAWS*.



4.3.12 Feature Annotations and PAWS

Disordered regions and secondary structure require a separate, optional web application be installed by the site administrator that we have called PAWS, or Protein Annotation Web Services. Requests for these types of sequence annotations make a request to the PAWS service for information about the respective sequence. If available, that information is sent by PAWS to proxl and that information is shown. If not available, PAWS will initiate the running of DISOPRED3 or PSIPRED3 on the sequence, store the results in a database (for future use), and respond to proxl with the data.

As a consequence, if the sequence annotations for the requested sequence has not yet been run, there may be a delay before the data are shown in proxl. The user as the option of waiting for the data to be returned, or canceling and performing other actions. (Note: if canceled, the data are still processed and will be available on a subsequent request.)



For more information about PAWS, please see *PAWS Guide*.

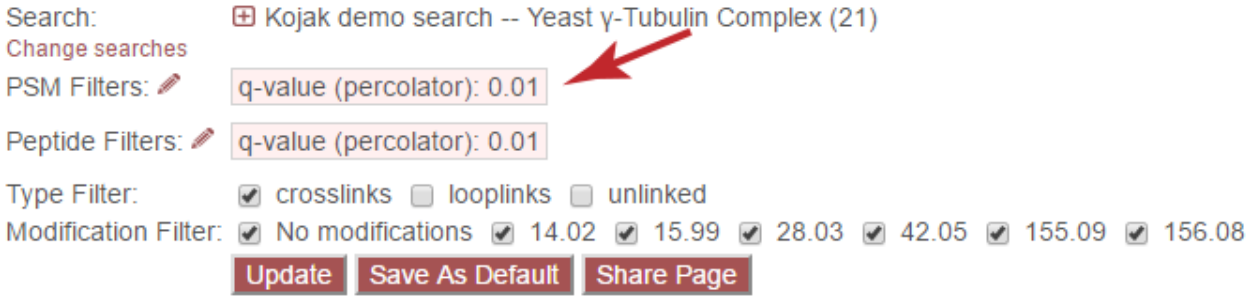
4.4 Filter Data

The data presented in the viewer may be filtered using the form at the top of the page. The filtering options are:

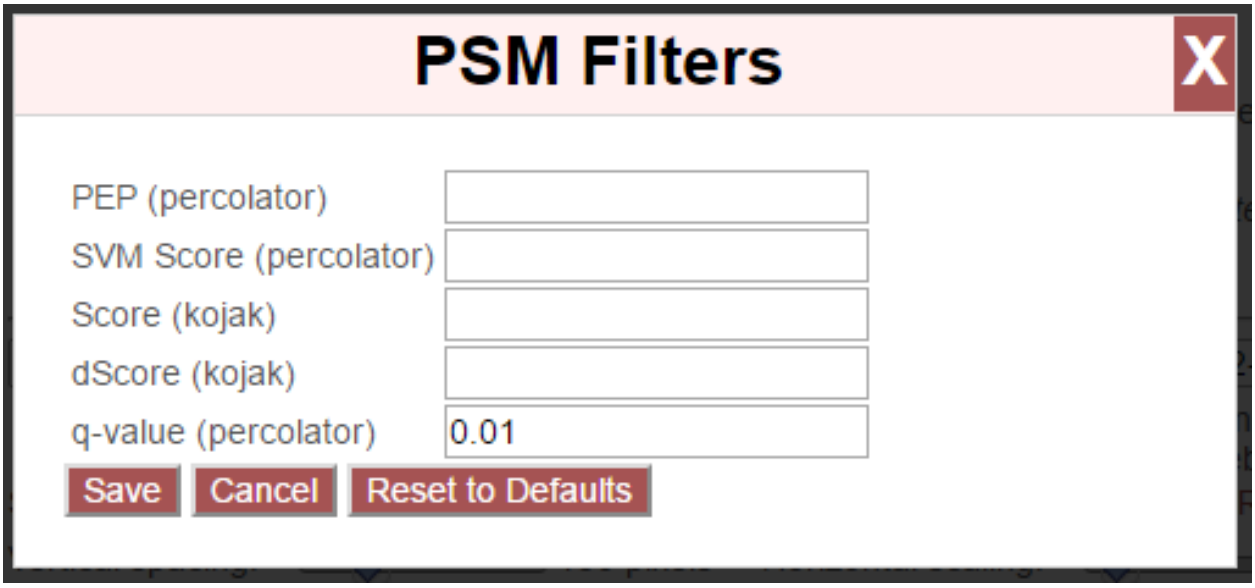
4.4.1 PSM Filters

The filters to apply at the PSM level. Only results which have at least one PSM that meets all of the selected critiera will be listed. When listing PSMs associated with peptides, only PSMs that meet all of the selected critiera will be listed.

To change the PSM-level filters, first click the pencil icon next to "PSM Filters":



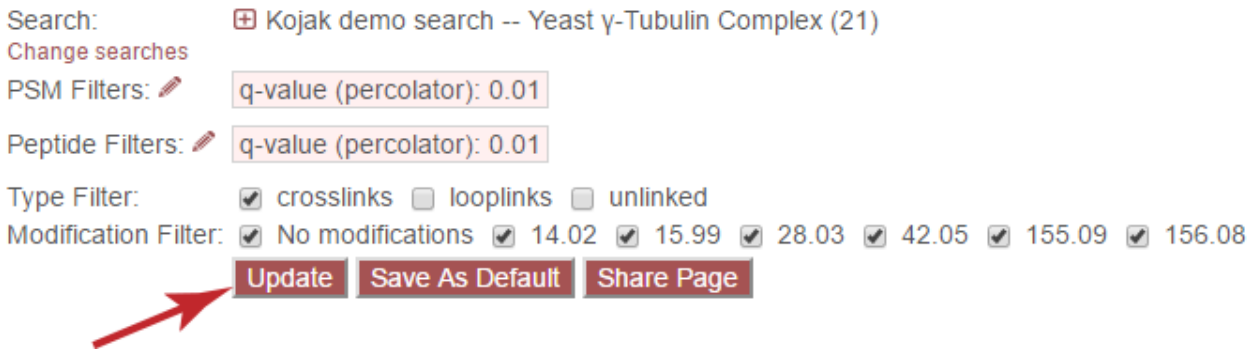
This opens an overlay with the containing the possible score types to use as PSM filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

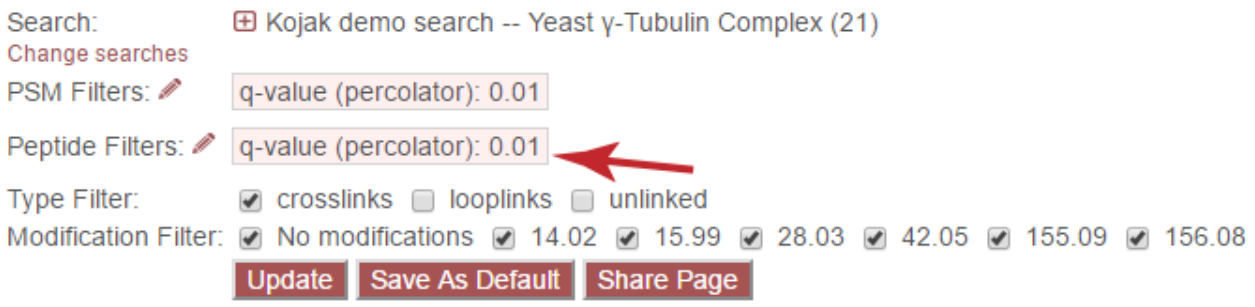
Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update From Database" button on the page to apply any new PSM- or peptide-level filters.



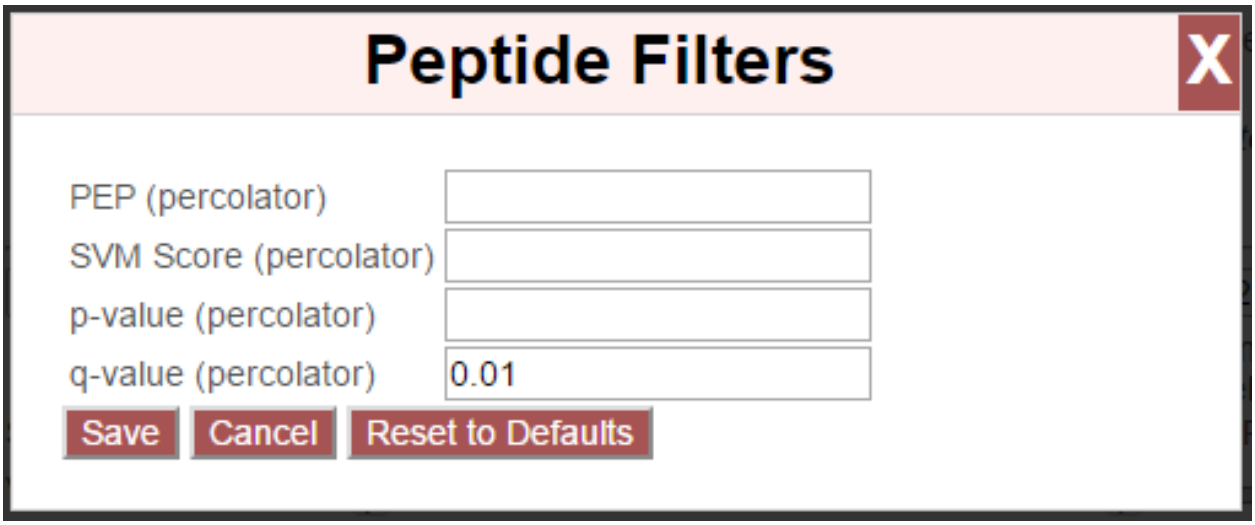
4.4.2 Peptide Filters

The filters to apply at the peptide level. Only results which have at least one peptide that meets all of the selected critiera will be listed.

To change the peptide-level filters, first click the pencil icon next to "Peptide Filters":



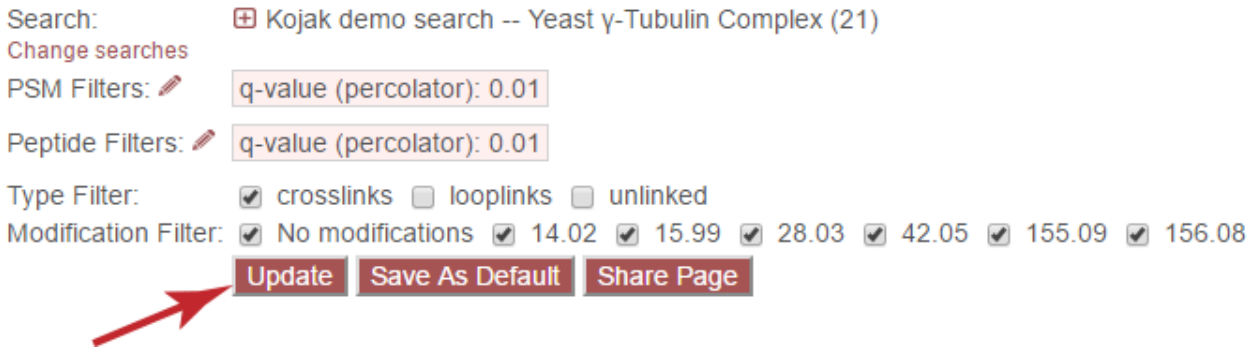
This opens an overlay with the containing the possible score types to use as peptide-level filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update From Database" button on the page to apply any new PSM- or peptide-level filters.



4.4.3 Exclude links with

Peptides with any of the checked attributes will not be shown. The attributes are:

- no unique peptides - If the link (crosslink, looplink, or monolink) was exclusively identified by peptides that also map to othe proteins
- only one PSM - If a given link was identified by a single PSM
- only one peptide - If a given link was identifed by a single peptide, where a peptide is the combination of sequence, linked positions, and modifications

4.4.4 Exclude proteins with

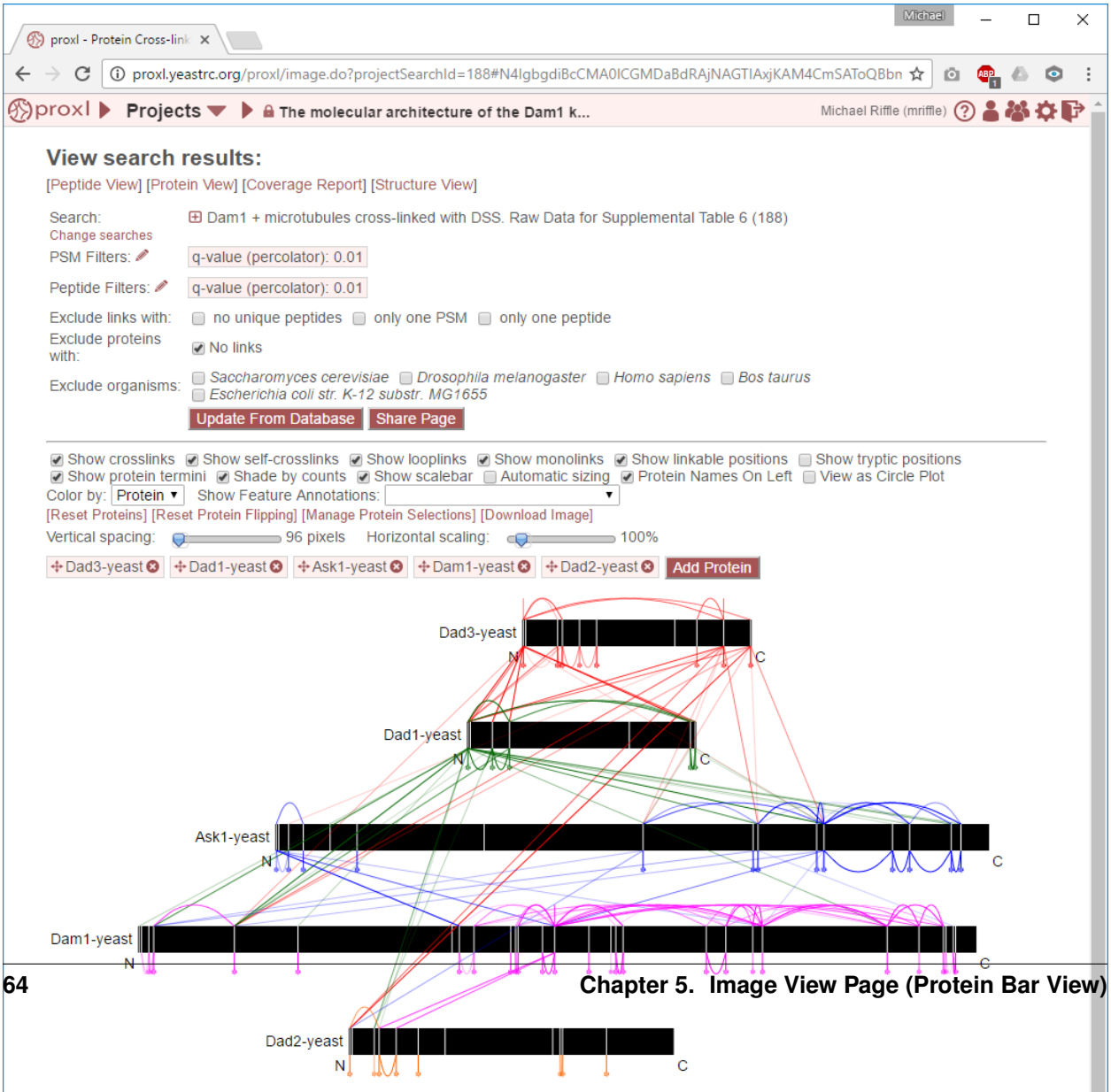
This option limits which proteins will appear in the pull-down menu for adding proteins to the viewer (see below). Proteins that contain any of the checked options will not appear. For example, checking 'No links' prevents proteins that do not contain crosslinks, looplinks, or monolinks from appearing.

4.4.5 Exclude organisms

This options limits which proteins will appear in the pull-down menu for adding proteins. No proteins from any of the checked organisms will appear.

CHAPTER 5

Image View Page (Protein Bar View)

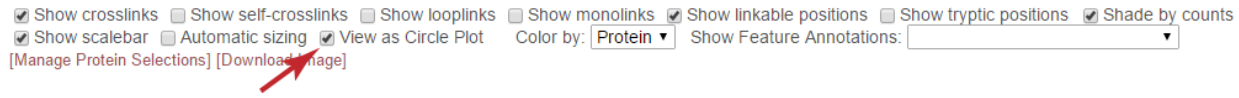


The image viewer displays a SVG rendering of select proteins as bars, the lengths of which are proportional to protein sequence length. The bars are annotated with positions of monolinks, looplinks, crosslinks, and other biological sequence annotations. The viewer is highly interactive and contains many options for customization (see below).

5.1 Viewer Basics

5.1.1 Switch to "Circle Plot" View

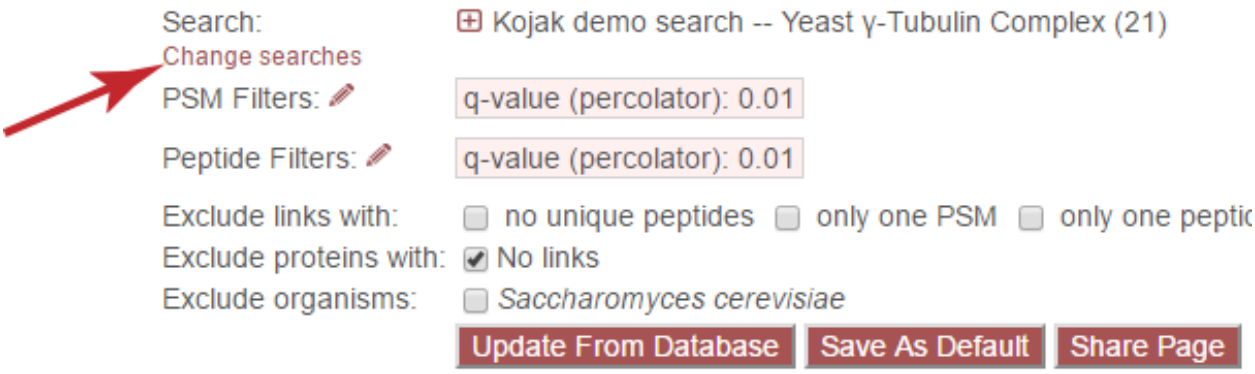
To switch to the "Circle Plot" view, click the checkbox next to "View as Circle Plot" in the viewer options. See *Image View Page (Circle Plot View)* for more information.



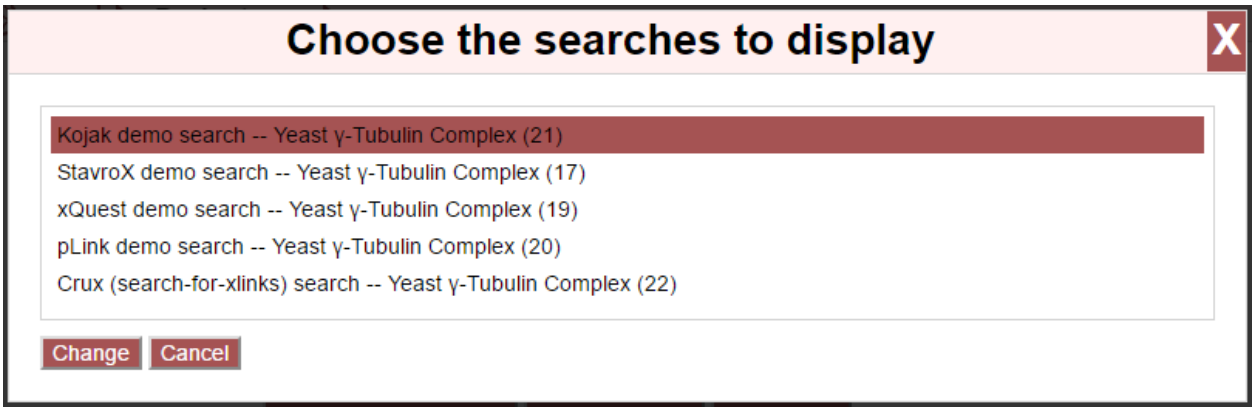
5.1.2 URL Captures State of Page

The URL of the page is dynamically updated at all times to reflect the complete state of the viewer–including filter parameters, protein bar positions, and all viewer options. As such, the URL may be bookmarked or shared to link to a viewer with the same content and appearance as the current viewer–just copy and paste from the address bar. For complicated visualizations, this is a convenient way to save and share you work. (Note, this link only works for other users who are listed on this project unless public access is enabled.)

5.1.3 Change Searches

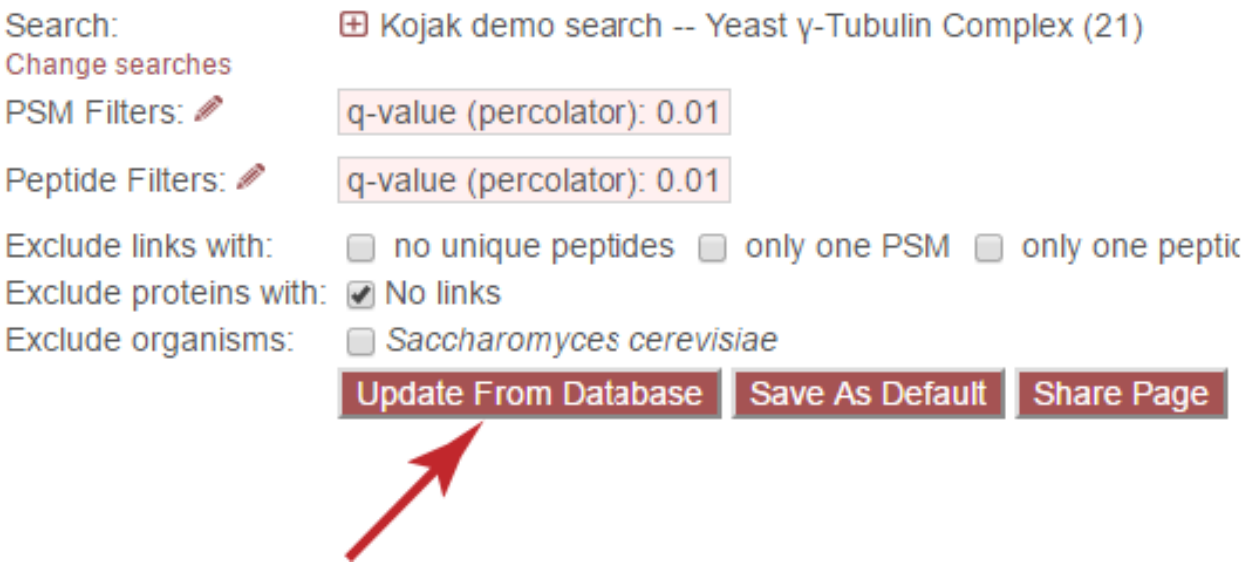


The "Change searches" link allows the user to change which searches are currently being displayed. None of the other options on the page (such as viewer settings) will be changed, only the searches from which data are displayed. Clicking the link causes the following overlay to be displayed:



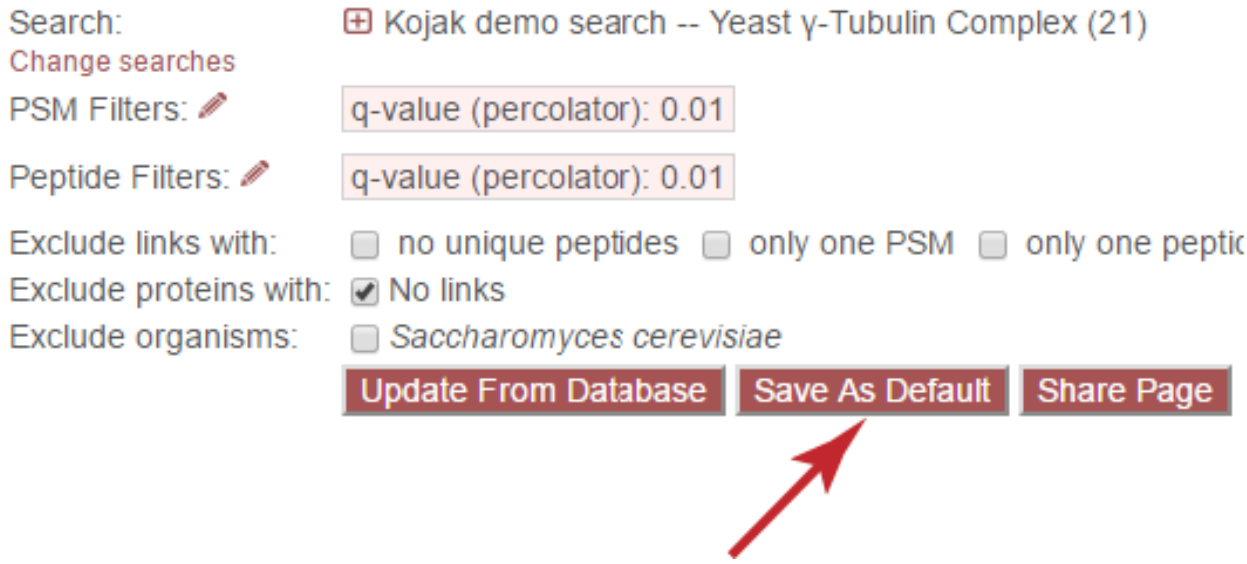
Select or de-select searches by clicking on them in the list. Once done, click "Change" to update the page with the new data or "Cancel" to close the overlay.

5.1.4 Update From Database



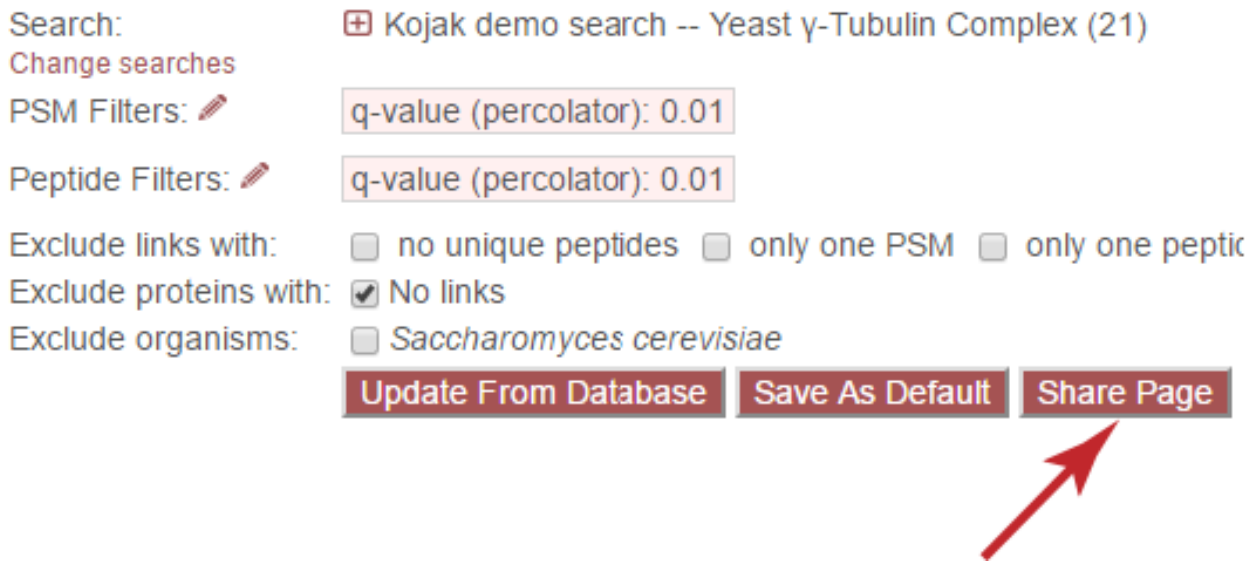
If the user changes any filter parameters–such as PSM/peptide score cutoffs, link type exclusions, protein exclusions, or taxonomic exclusions–this button must be clicked to reflect the new filter choices.

5.1.5 Save as Default

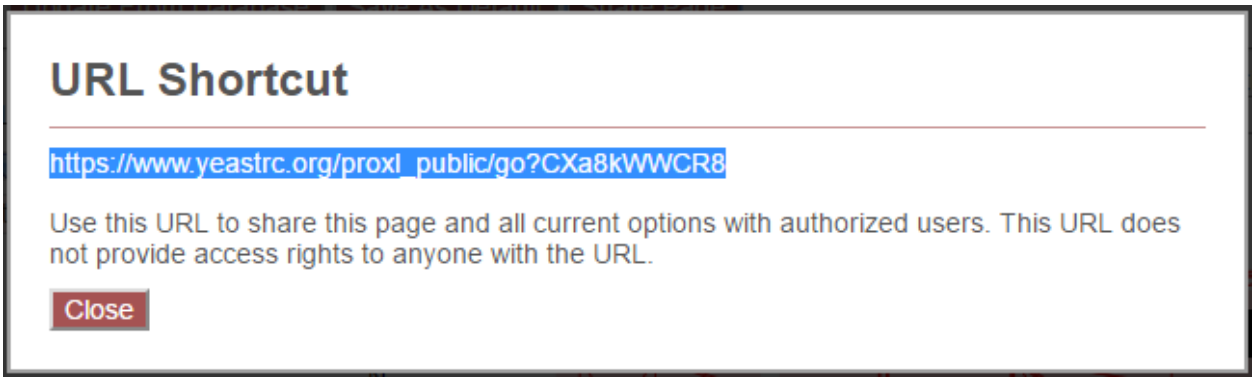


Project owners may click "Save as Default" to save the current URL (see above) as the default view of the "Image View" for this project. This default view will be populated with the same proteins, positions, and viewer options as when the button is clicked. This is a convenient way to share data with collaborators or the public that does not require that they manipulate the image viewer to see the data.

5.1.6 Share Page



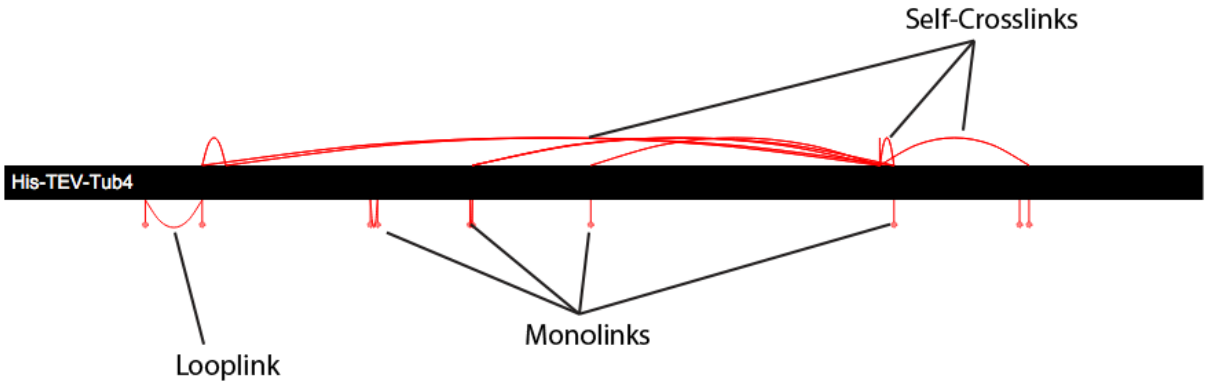
Clicking the "Share Page" button will generate a shortcut URL for viewing the current page, including all viewer settings, filter options, and so forth. The shortened URL will appear in an overlay as:



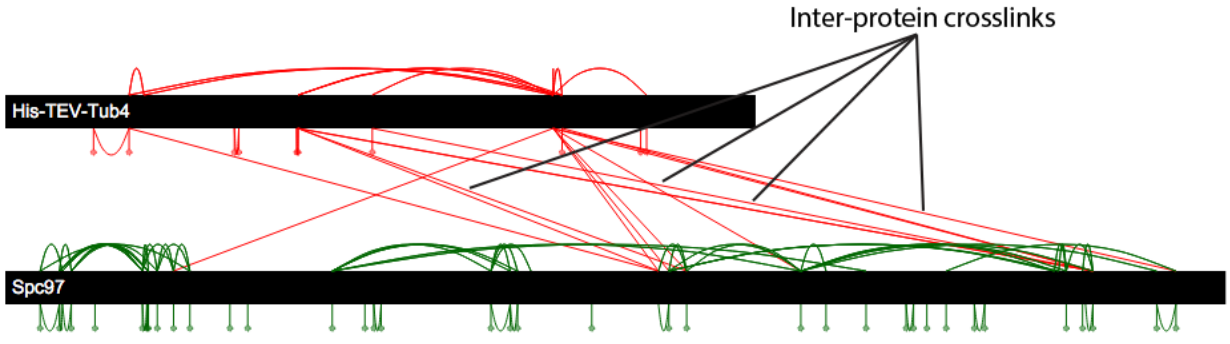
Copying and sharing the highlighted URL will direct users to the view of the page when the URL was generated. Note that this URL does not grant access to the page to any user that would not otherwise have access.

5.1.7 Viewer Display

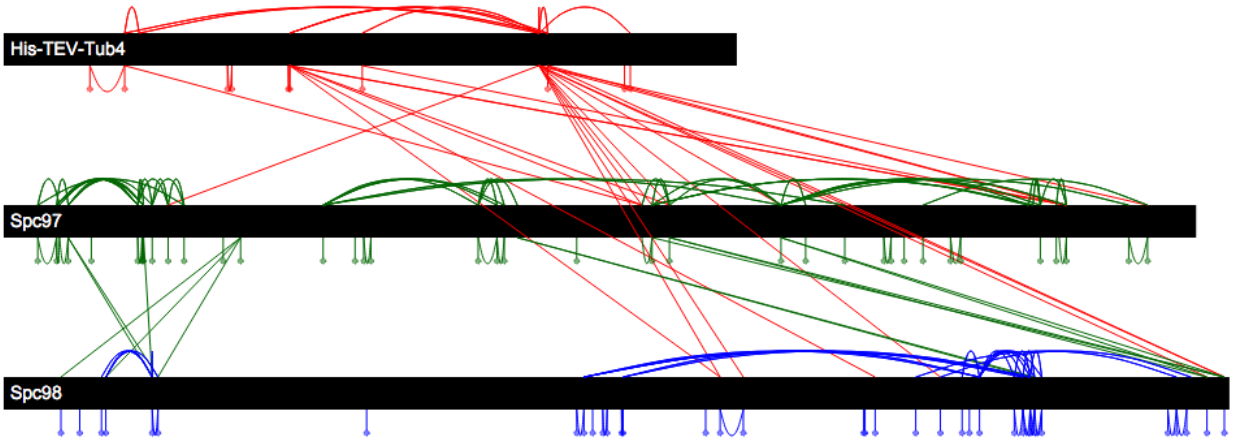
Below is a labeled example of a protein bar in the viewer, with most options disabled. Self-crosslinks, that is crosslinks where both linked peptides map to the same protein, appear as arcs on the top of the protein bar. These are contrasted with looplinks, where a single peptide contains two linked residues, which appear as arcs on the bottom of the protein bar. Monolinks appear as inverted lollipops (a line segment with a ball on the end) on the bottom of protein bars.



When a second protein is added to the viewer, crosslinks between the two proteins will appear as line segments connecting the two proteins, with the end points of the segments are the respective link positions in the proteins:



And a third protein:



5.1.8 Coloring

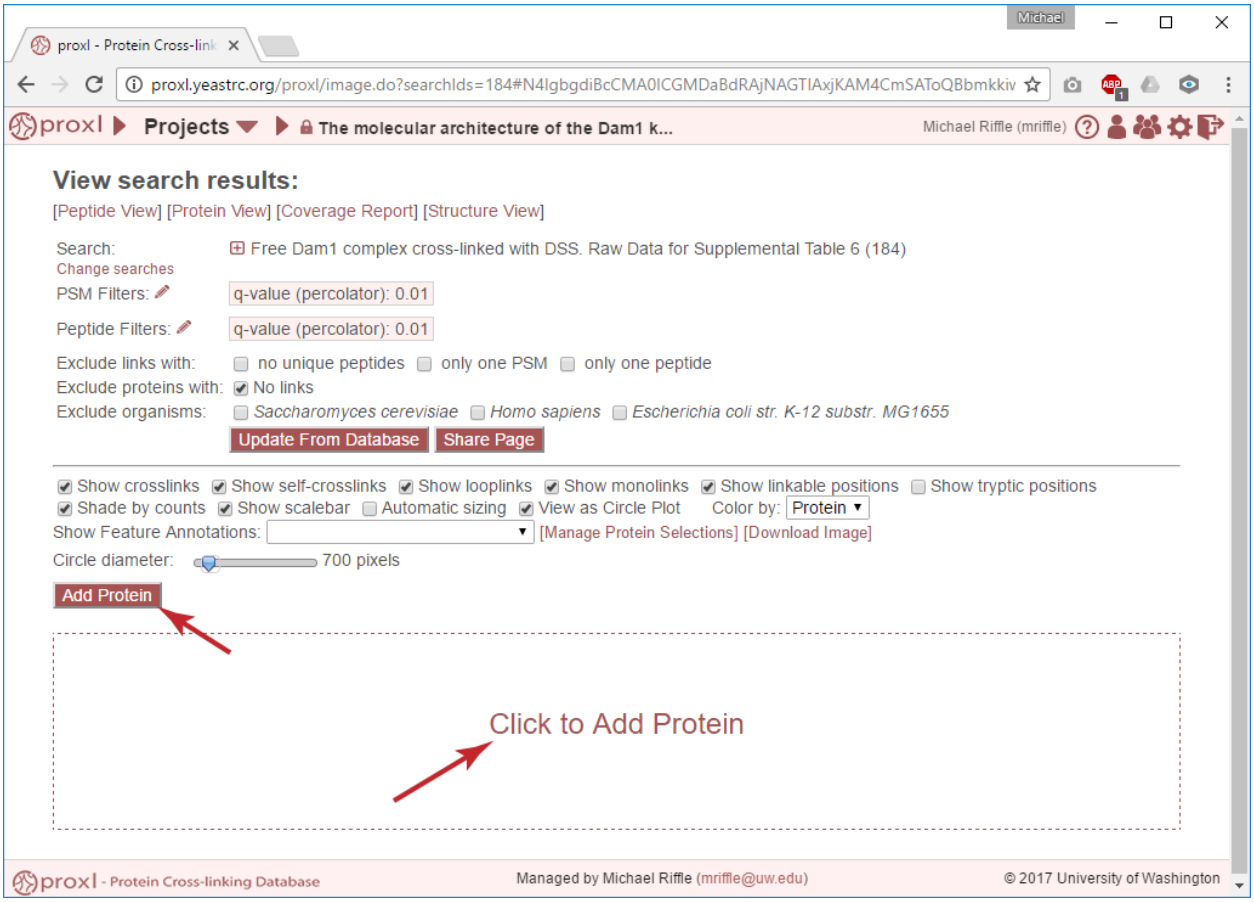
By default, the coloring of the links corresponds to the "originating" protein in the viewer. For example, the first protein is colored red. All of its self-crosslinks, looplinks, and monolinks will be red. And all inter-protein crosslinks originating from this protein will be red. The originating protein of a crosslink will be the protein which appears first in the viewer (from top-down). Similarly, the second protein is colored green. All of its links and originating inter-protein crosslinks will be colored green. This coloring scheme is intended to ease confusion about which links originated at which proteins, farther down the list of proteins.

This coloring scheme can be changed to show in which search(es) the link appears (See *Color by search*). And, specific proteins in the viewer can be highlighted to only color links involving those proteins (See *Highlight Proteins*).

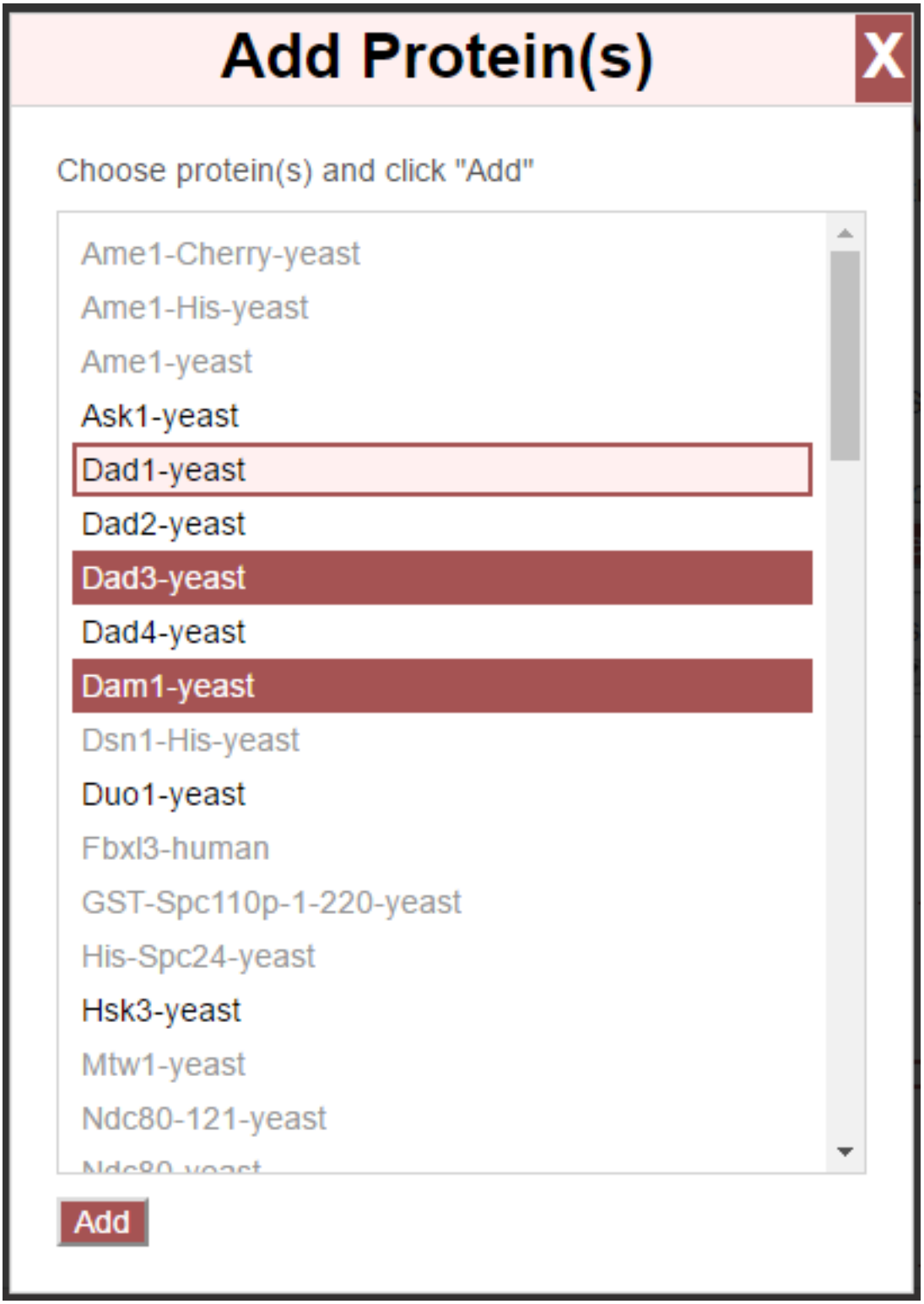
5.2 Viewer Interaction

5.2.1 Add a Protein

To add a protein to an empty viewer (no proteins visible), click either the "Add Protein" button or the "Click to Add Protein" text in the viewer area:



This will open the "Add Protein(s)" overlay, where one or more proteins may be selected:



Cick “Add” to add the selected proteins to the viewer. Click anywhere outside the overlay, or on the “X” in the top-right corner, to close the overlay.

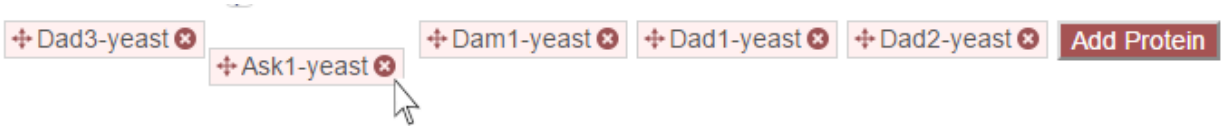
To add more proteins to the viewer, click the “Add Protein” button above the viewer.

5.2.2 Remove a Protein

To remove a protein from the viewer, click the small red (X) next to the name of the protein above the image area:

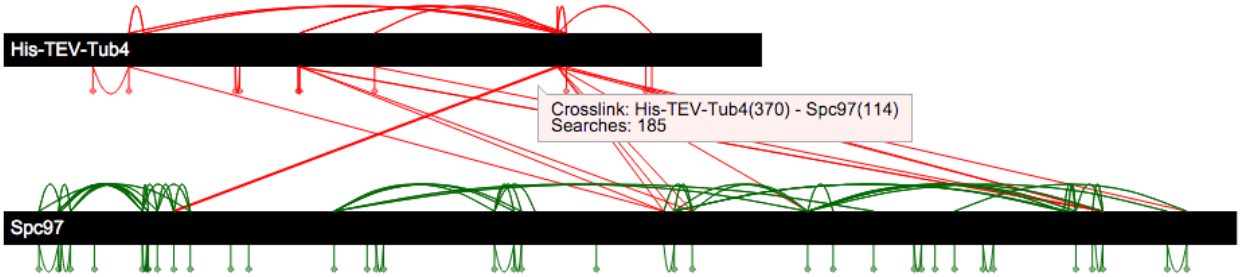
5.2.3 Rearrange Proteins

To rearrange proteins, simply drag and drop the protein in the protein list above the circle to the desired position in the order. The viewer will update automatically.



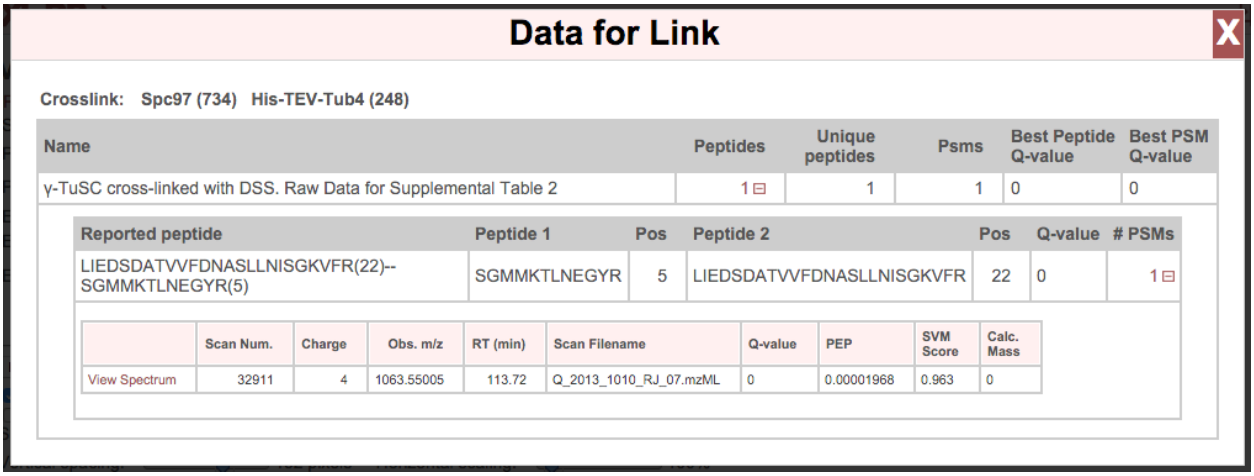
5.2.4 View Link Summary

To view summary information about a link, hover your mouse arrow over that link in the viewer. (Or tap, on touch devices.) This may be done for any link type. This will display the link type, protein(s), and position(s).



5.2.5 View Peptides, PSMs, and Spectra

To view listings of the peptides and corresponding PSMs for a link, click on that link in the viewer. This will open an overlay window displaying a table of peptides and PSMs:



The top-level of this table are rows for each search in which this link was found. If multiple searches have been merged, each search will appear as a row in the table. Each search may be expanded by clicking on that row to view all peptides from that search that led to this link. Each peptide may be expanded by clicking on that row to view all PSMs for that peptide. Each PSM includes a "View Spectrum" link for viewing an annotated spectrum associated with that PSM. For help on our spectrum viewer, see the *Spectrum Viewer* page. Click the "X" in the top-right corner of the overlay (or click on the page anywhere outside of the overlay) to close the overlay window.

5.2.6 Move Protein Bars

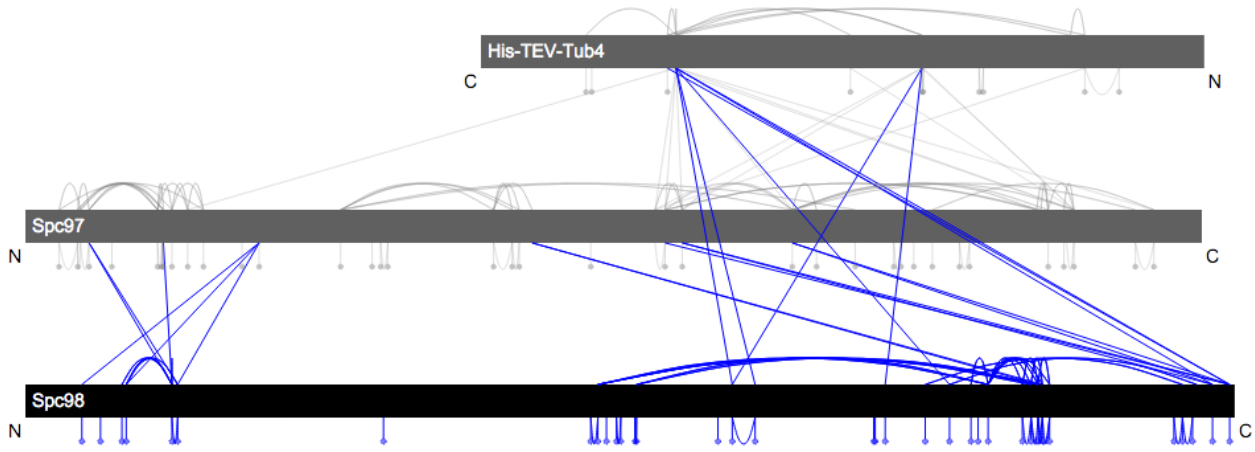
The protein bars may be moved to the left or right by clicking and dragging the bars in the desired direction.

5.2.7 Flip Protein Bars

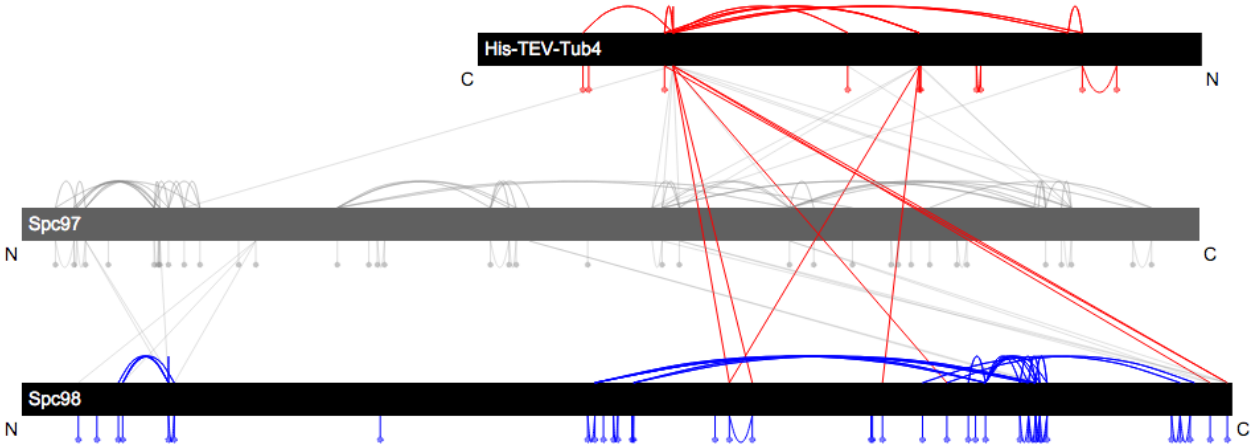
By default, the protein bars are laid out left-to-right from N-to-C terminus. This orientation may be flipped by double-clicking on the protein bar.

5.2.8 Highlight Proteins

Proteins may be highlighted by clicking on any of the protein bars. This will change the coloring scheme such that only links involving the highlighted protein(s) will be colored, all other links for all other proteins are shaded light gray:



Multiple proteins may be highlighted by holding shift and clicking protein bars:

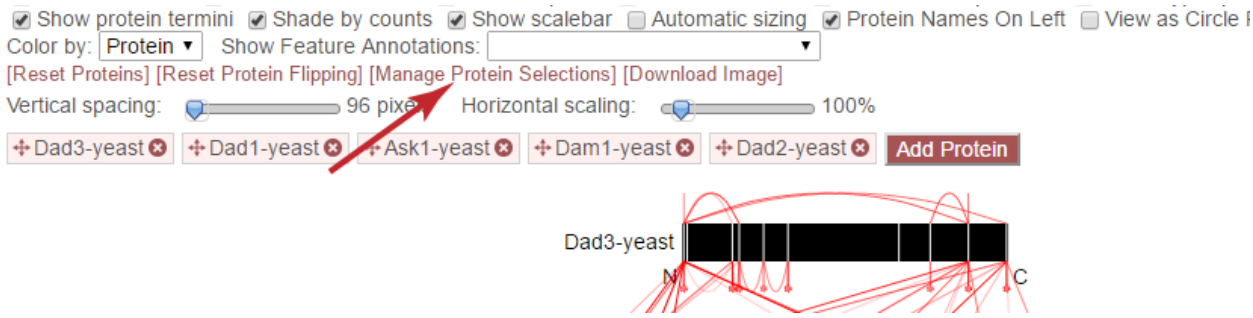


Remove Highlighting

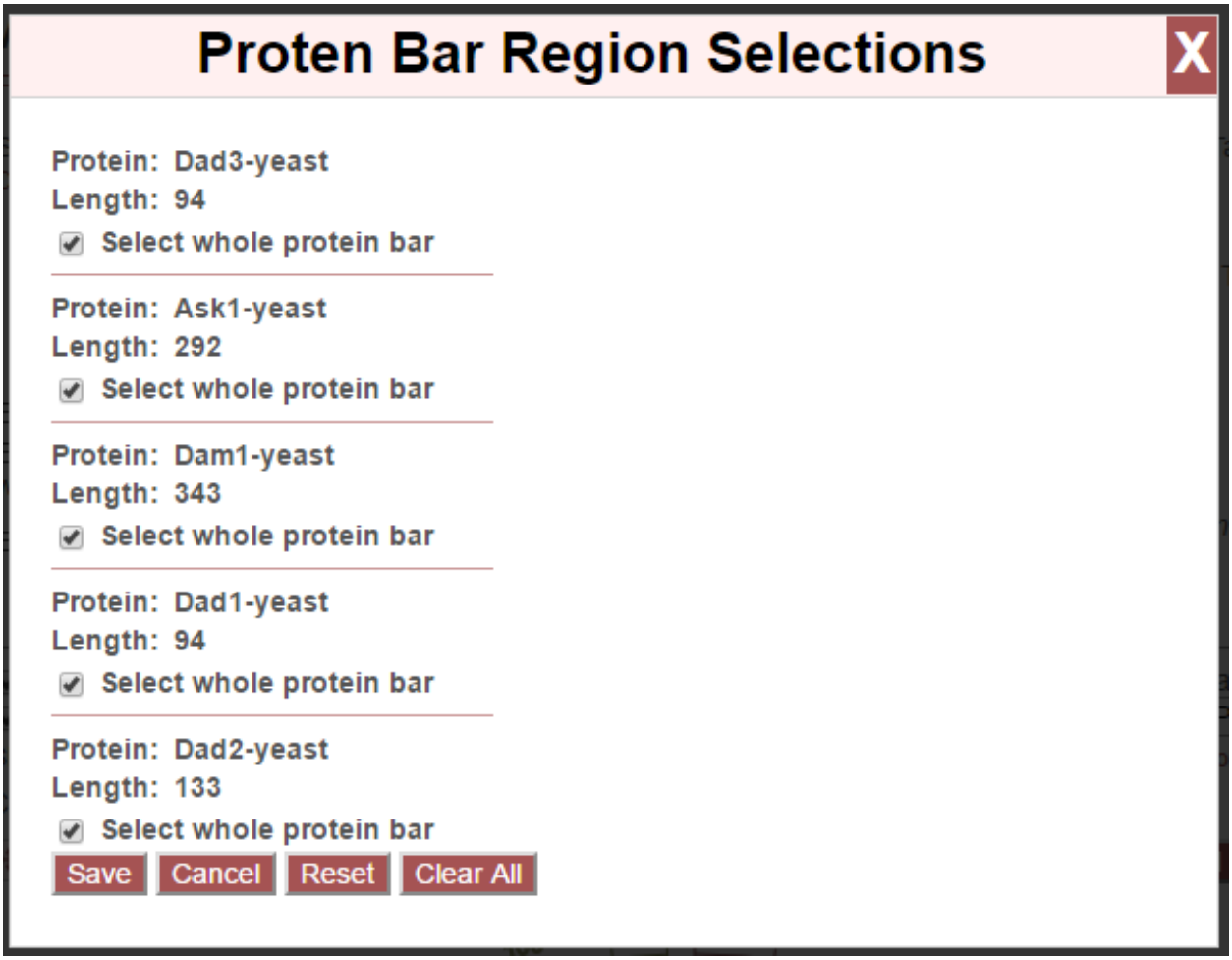
If a single protein is highlighted, click it to unhighlight it. If multiple proteins are highlighted, hold shift and click a highlighted protein to unhighlight it. If shift is not used, the viewer will highlight only the protein clicked.

5.2.9 Highlight Proteins Regions

Instead of highlighting entire proteins, it is possible highly only segments of proteins (protein regions). This is done by clicking the "[Manage Protein Selections]" link above the image:

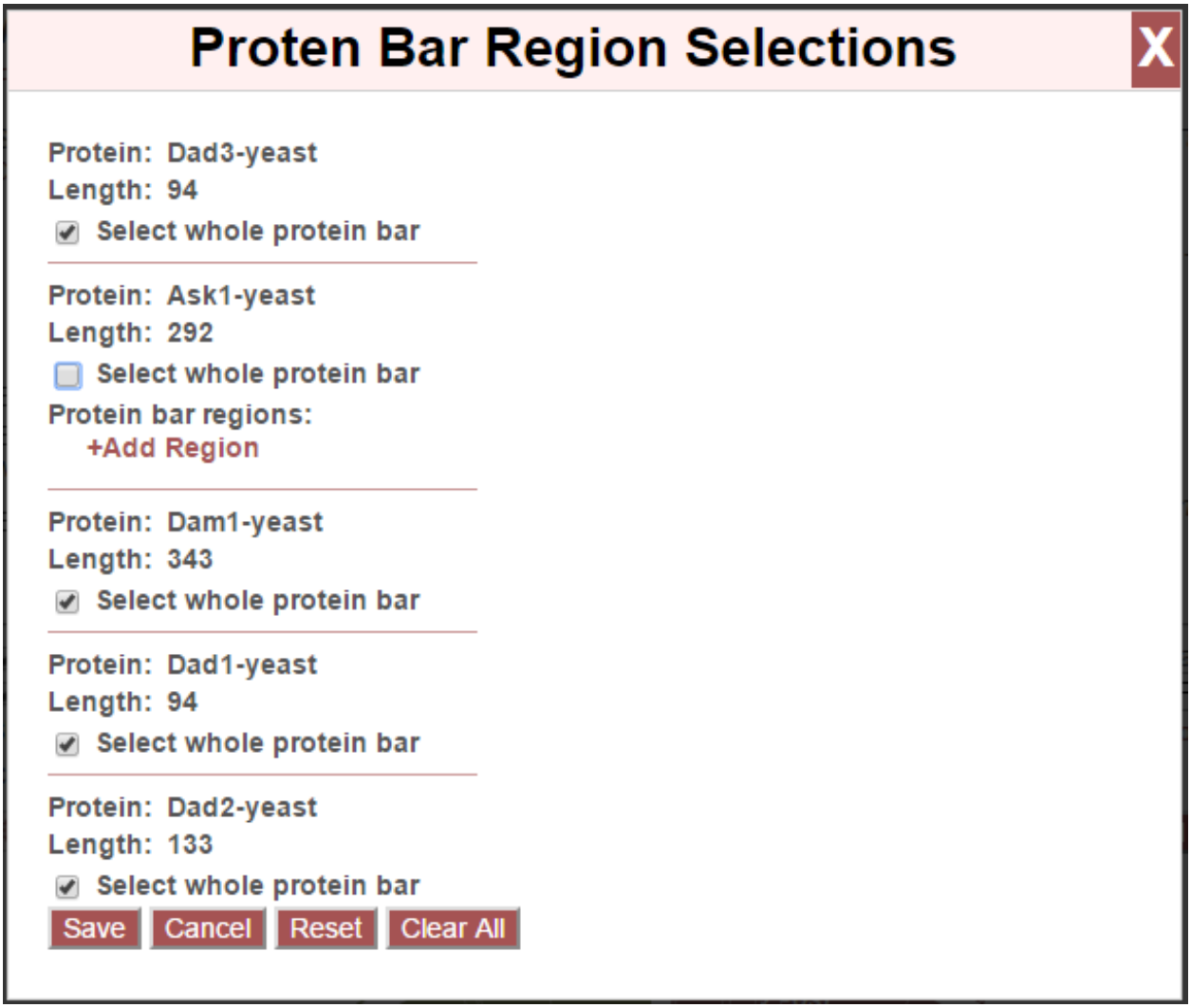


This opens an overlay with an interface for managing which regions of which proteins are highlighted:

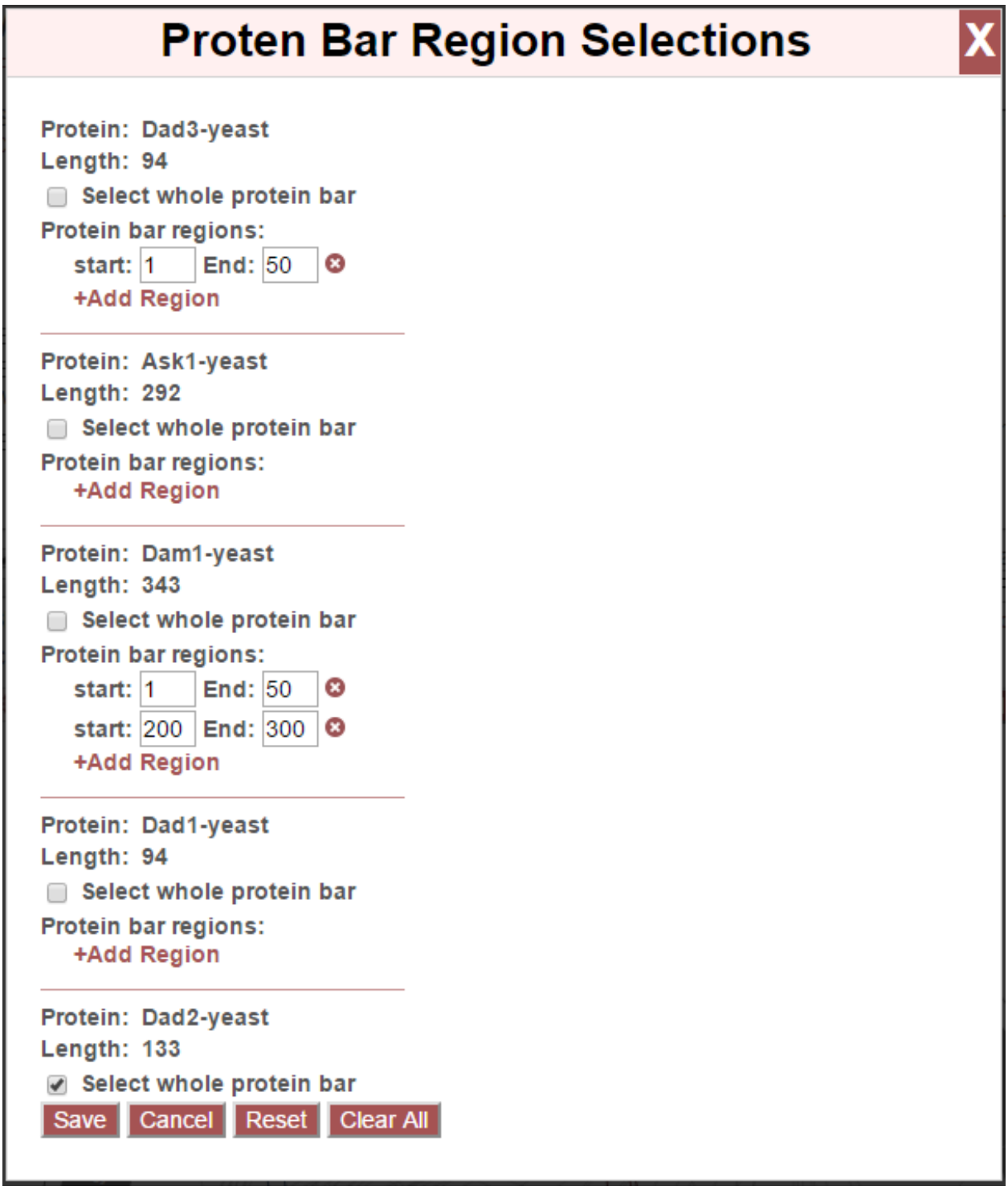


This overlay lists all proteins visible in the viewer and which regions are currently highlighted for each one. By default, all proteins are visible in their entirety, so the "Select whole protein bar" option is checked for all of them. When this checkbox is checked, that protein is highlighted in its entirety and no sub-regions may be selected.

To select sub-regions in a protein to highlight, uncheck the "Select whole protein bar" option:

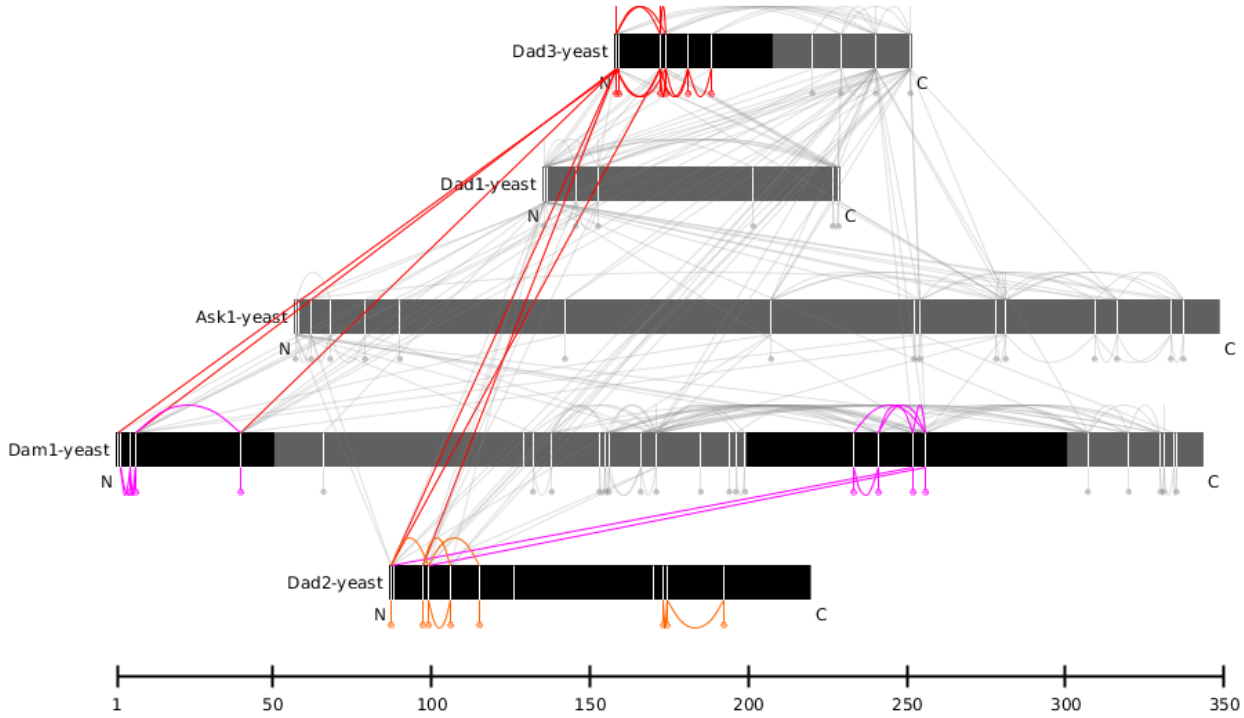


Click "+Add Region" to define the start and end positions for a region to highlight in that protein. This option may be used multiple times per protein to define multiple regions. Click the red "x" next to a region to remove it. Unchecking "Select whole protein bar" and not defining regions unhighlights that entire protein:



Click “Save” to save these settings and view the image with these defined regions. Click “Cancel” to make no changes and close overlay, “Reset” to reset regions to those visible in the image (without closing overlay), and “Clear All” to set to defaults (all proteins visible).

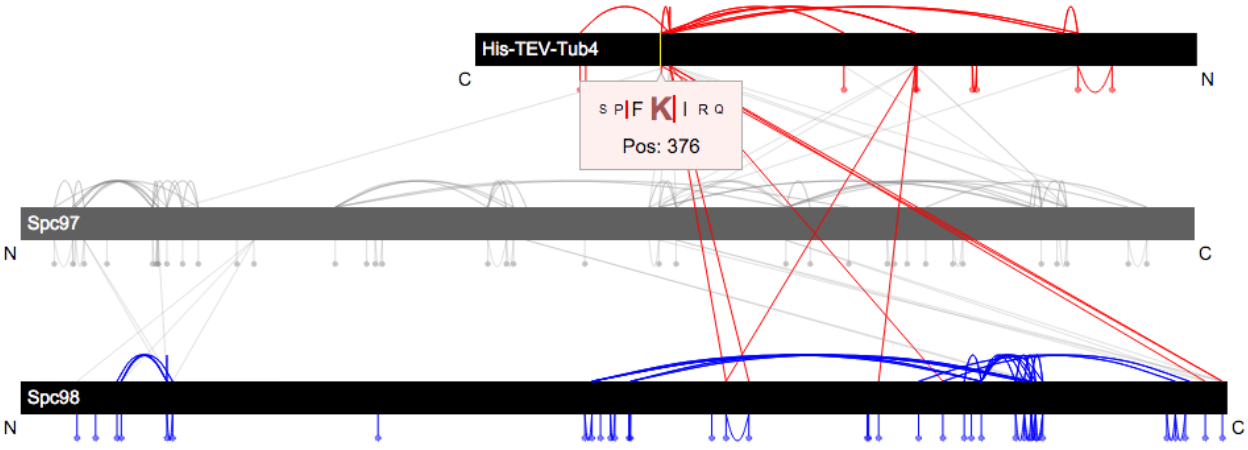
Clicking “Save” gives us:



Only links within and between the defined regions are colored. The reset are greyed-out.

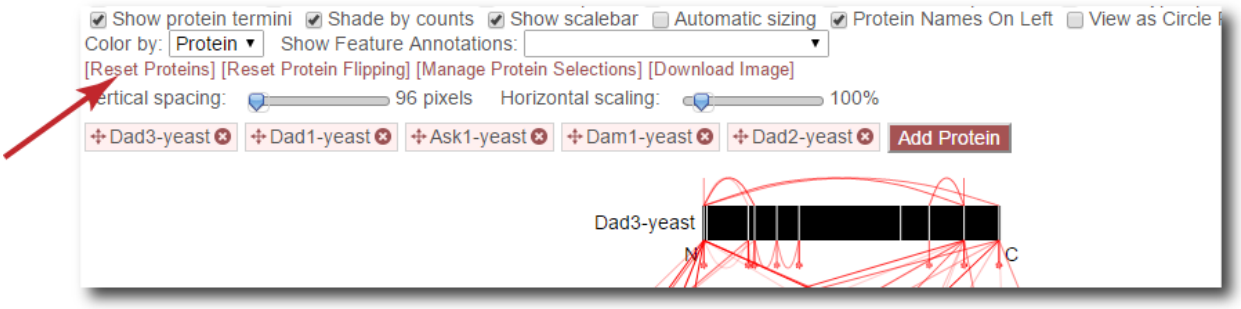
5.2.10 Local Sequence Information

Local sequence information in the protein bars may be viewed by hovering the mouse cursor over the protein bar. A tooltip will appear that shows the amino acid position number, the amino acid at that position, and neighboring amino acids. Amino acids that linkable with the crosslinker(s) used in the experiment(s) will be bolded and red. Vertical bars indicate sites that are cleavable by trypsin. This tooltip will slide and dynamically update along with the mouse cursor as it is moved along the protein bar.



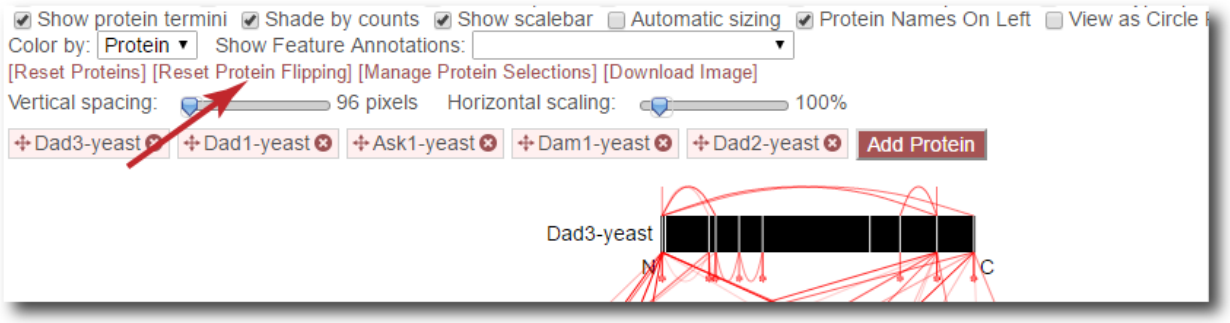
5.2.11 Reset Proteins

Reset the positioning of all protein bars so that left edges are aligned to left of viewer.



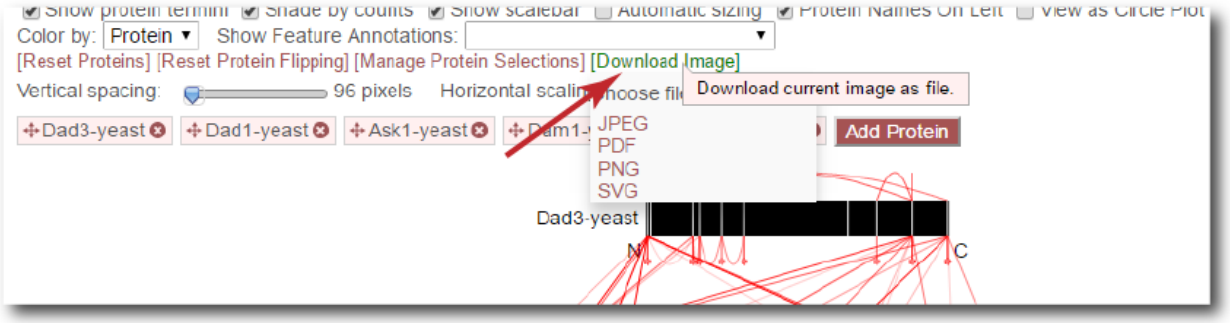
5.2.12 Reset Protein Flipping

Sets the left side of all protein bars to be the N-termini.



5.2.13 Download Image

Mouse over the "[Download Image]" link to see image download options. Click on the format of choice to initiate a download of the image.



5.3 Viewer Options

5.3.1 Show crosslinks

Toggle the showing of inter-protein crosslinks.

5.3.2 Show self-crosslinks

Toggle the showing of intra-protein crosslinks.

5.3.3 Show looplinks

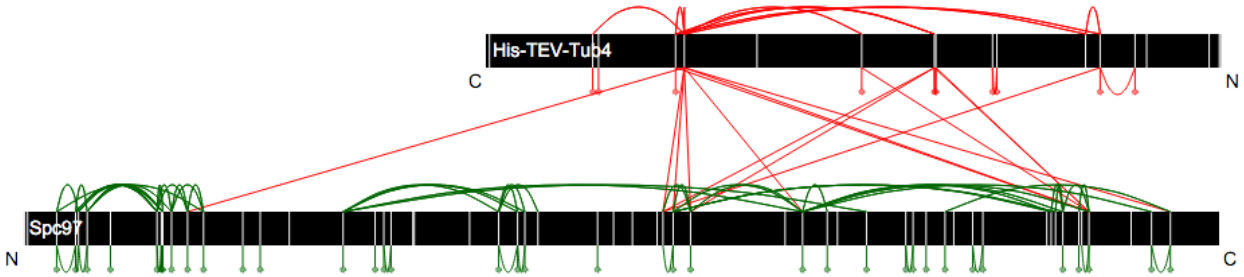
Toggle the showing of looplinks.

5.3.4 Show monolinks

Toggle the showing of monolinks.

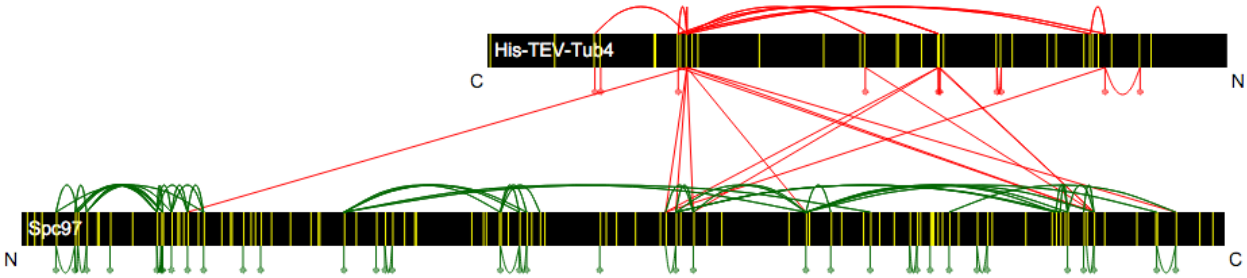
5.3.5 Show linkable positions

Toggle the showing of which positions in the protein are linkable by the crosslinker(s) used in the experiment. The linkable positions are noted by white lines in the protein bar.

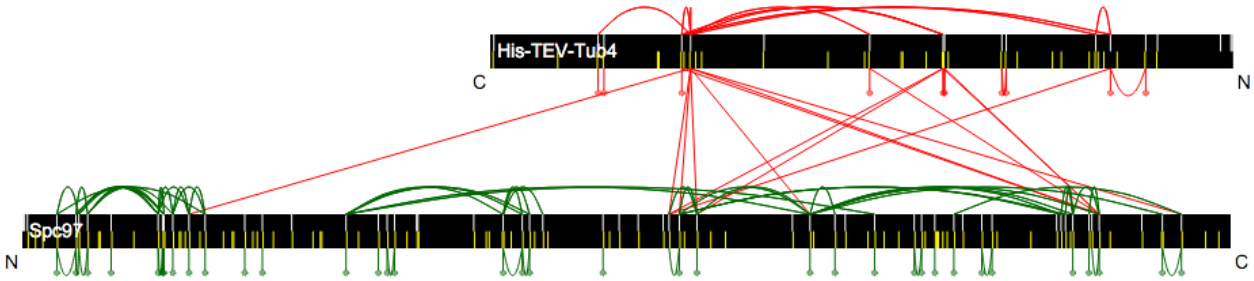


5.3.6 Show show tryptic positions

Toggle the showing of which positions in the protein are cleavable by trypsin, an enzyme commonly used to digest proteins in bottom-up proteomics experiments. The cleavable positions are noted by yellow lines in the protein bar.



If both linkable and tryptic positions are being displayed, each type is displayed by a half-height line to remove ambiguity caused by overlapping linkable and tryptic positions. Linkable sites are shown in white on the top-half of the protein bar, and tryptic positions in yellow on the bottom half.

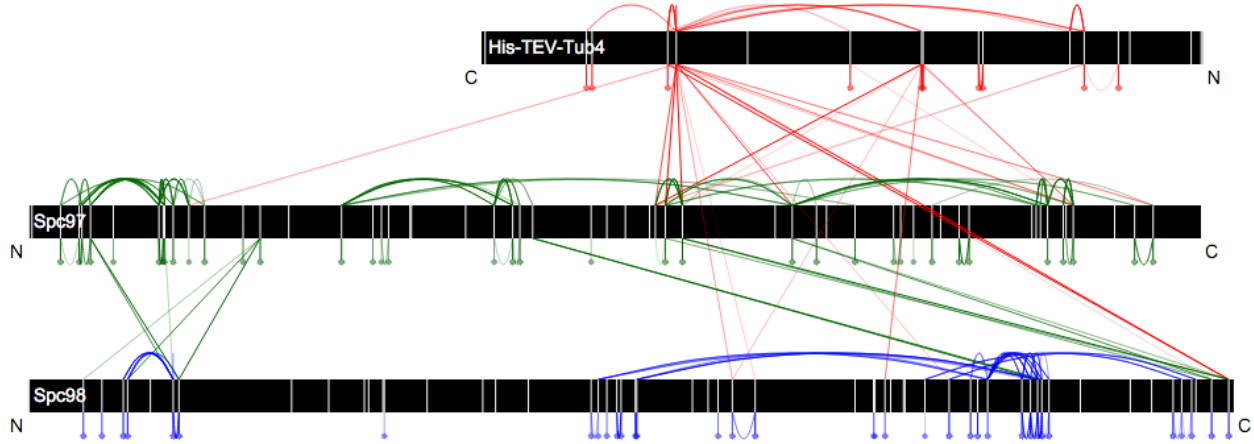


5.3.7 Show protein termini

Toggles the labelling of the N and C termini to the lower left and right of the protein bars.

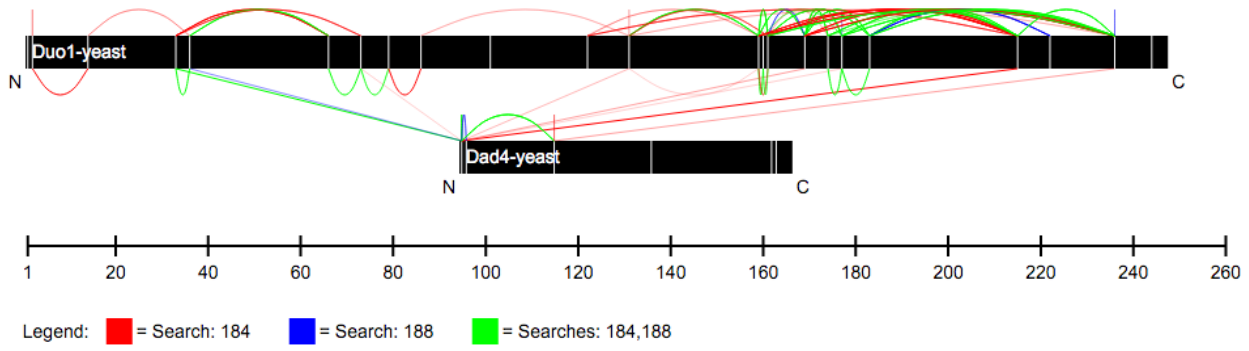
5.3.8 Shade by counts

If enabled, the opacity (transparency) of links reflects the number of PSMs found (or spectrum count) for the shown link. The shading scales from 1 PSM (minimum opacity, most transparent) to 10 PSMs (maximum opacity). Any link having 10 or more PSMs will have the maximum opacity.



5.3.9 Color by search

When merging multiple searches, this option changes the coloring scheme so that all links are colored by which search (or searches) they were found in at the given cutoffs. Each search is assigned a color, and each combination of searches are assigned other, distinct colors. It is possible to ascertain from the color in which search, or combination of searches, the individual link was found. A legend is provided beaneath the graphic. This functionality is limited to a maximum of three searches.



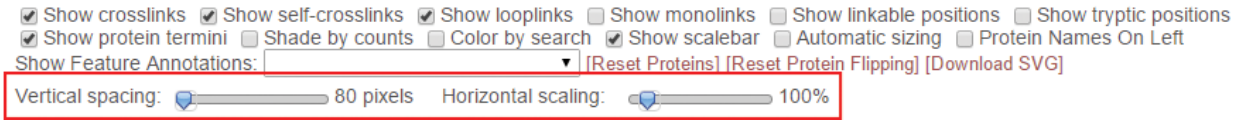
5.3.10 Show scalebar

Toggle the display of the scale bar on and off.

5.3.11 Automatic sizing

The viewer automatically determines a single horizontal scale for pixels/residue for all protein bars based on the length of the longest protein and the width of the browser window–such that the longest protein stretches the entire width of the window. This scaling is dynamically recalculated and redrawn as the width of the browser window is changed or as longer proteins are added to the viewer. Additionally, the viewer employs a default vertical distance between the protein bars.

These defaults may be disabled and manually altered by disabling this option. Disabling this option presents the two sliders below:



Vertical spacing

This slider adjusts the distance between the vertical bars, slide right to increase the distance.

Horizontal scaling

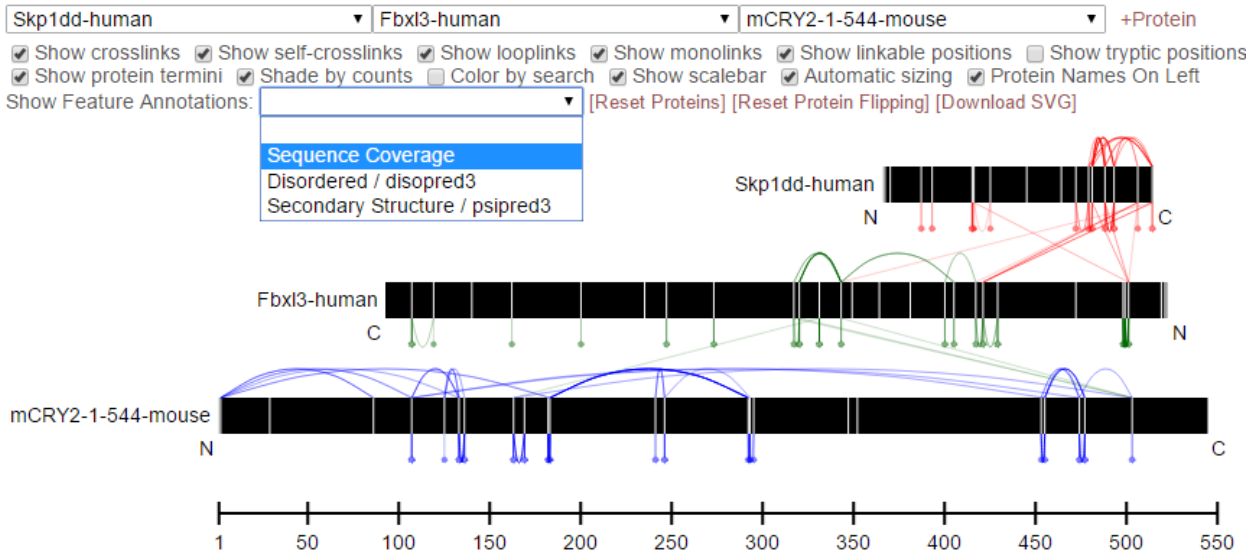
This slider adjusts the the number of pixels per residue, as a percentage of the default. 50% means the bars are scaled to be one-half as wide as they are by default. 400% means the bars are 4 times as wide. Slide left to decrease the width, slide right to increase the width.

5.3.12 Protein Names On Left

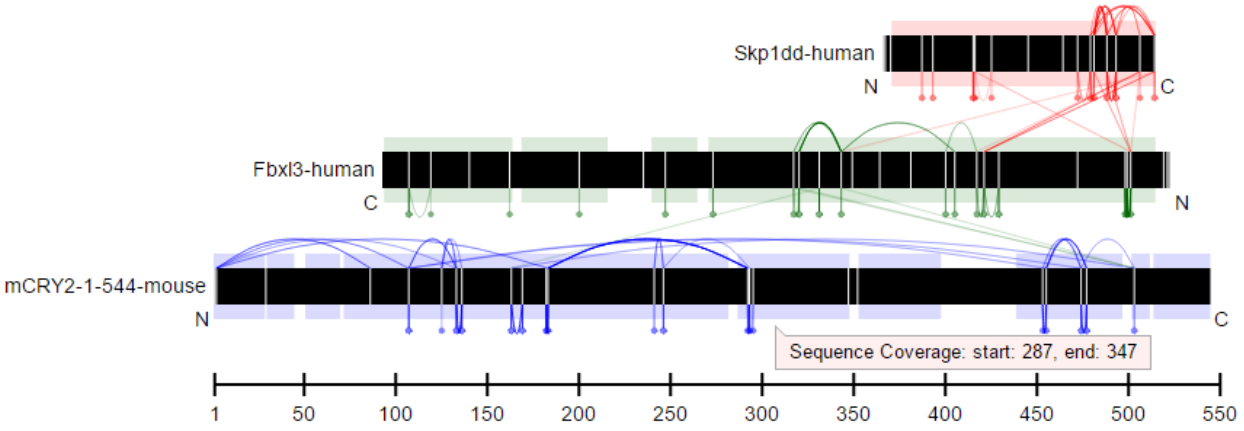
By default, protein names are placed within the protein bar, on the left side. This option will place the protein names outside and to the left of the protein bars.

5.3.13 Show Feature Annotations

This option allows for the display of protein sequence feature annotations of various on the protein bars. To select a type of feature annotation, click the pull-down menu next to "Show Feature Annotations" and select a type:



This will retrieve the necessary data from the server and display the respective annotation as a shaded region on the protein bars:



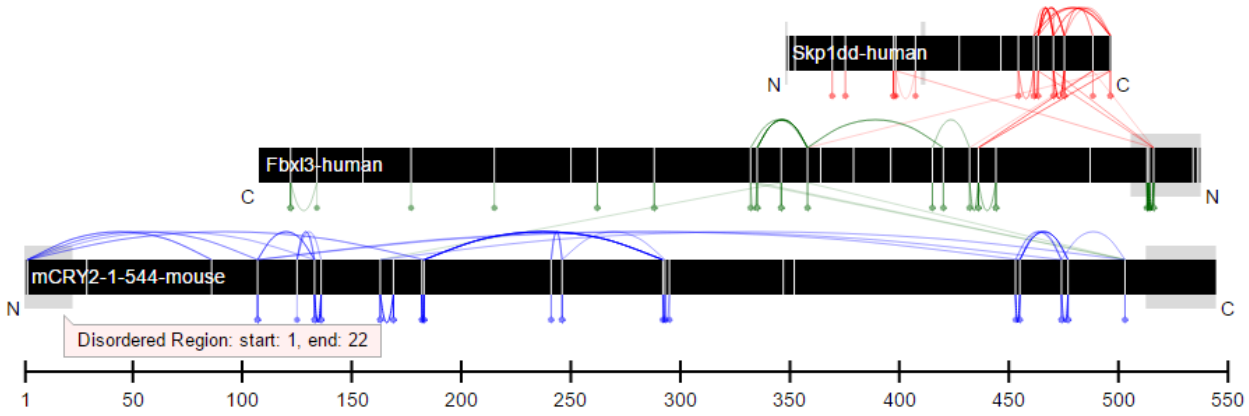
The types of feature annotations currently supported are:

Sequence Coverage

Sequence coverage shows which regions of the protein's sequence are covered by peptides of any type from the search(es) that meet the filtering criteria. An example of viewing the sequence coverage is shown above. The regions may be moused over to view exact start and stop residues.

Predicted Disordered Regions

Selecting this option annotates the protein bars to show predicted disordered regions according to the DISOPRED3 algorithm. The regions may be moused over to view exact start and stop residues. This feature requires that PAWS be available, see: *Feature Annotations and PAWS*.



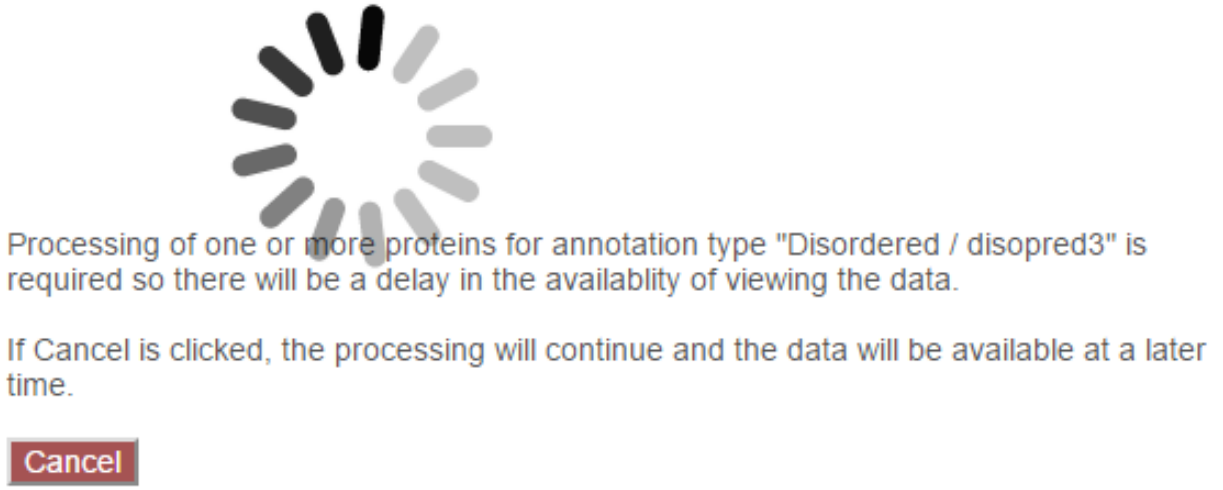
Predicted Secondary Structure

Selecting this option annotates the protein bars to show predictions for secondary structure according to the PSIPRED 3 algorithm. The regions may be moused over to view exact start and stop residues. This feature requires that PAWS be available, see: *Feature Annotations and PAWS*.

5.3.14 Feature Annotations and PAWS

Disordered regions and secondary structure require a separate, optional web application be installed by the site administrator that we have called PAWS, or Protein Annotation Web Services. Requests for these types of sequence annotations make a request to the PAWS service for information about the respective sequence. If available, that information is sent by PAWS to proxl and that information is shown. If not available, PAWS will initiate the running of DISOPRED3 or PSIPRED3 on the sequence, store the results in a database (for future use), and respond to proxl with the data.

As a consequence, if the sequence annotations for the requested sequence has not yet been run, there may be a delay before the data are shown in proxl. The user as the option of waiting for the data to be returned, or canceling and performing other actions. (Note: if canceled, the data are still processed and will be available on a subsequent request.)



For more information about PAWS, please see *PAWS Guide*.

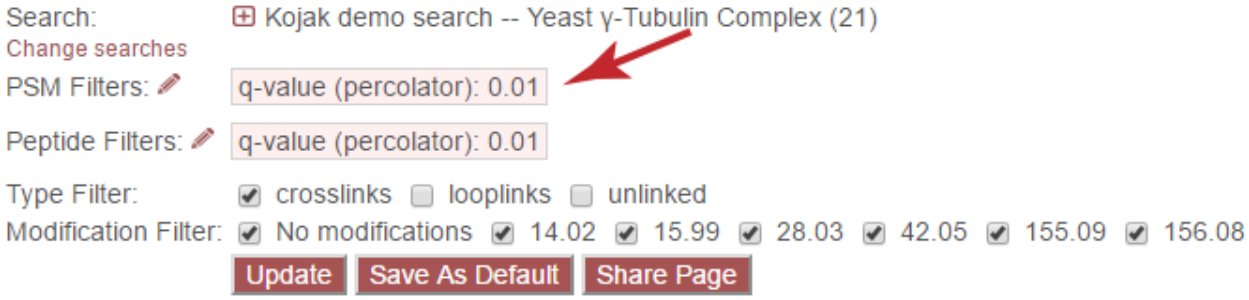
5.4 Filter Data

The data presented in the viewer may be filtered using the form at the top of the page. The filtering options are:

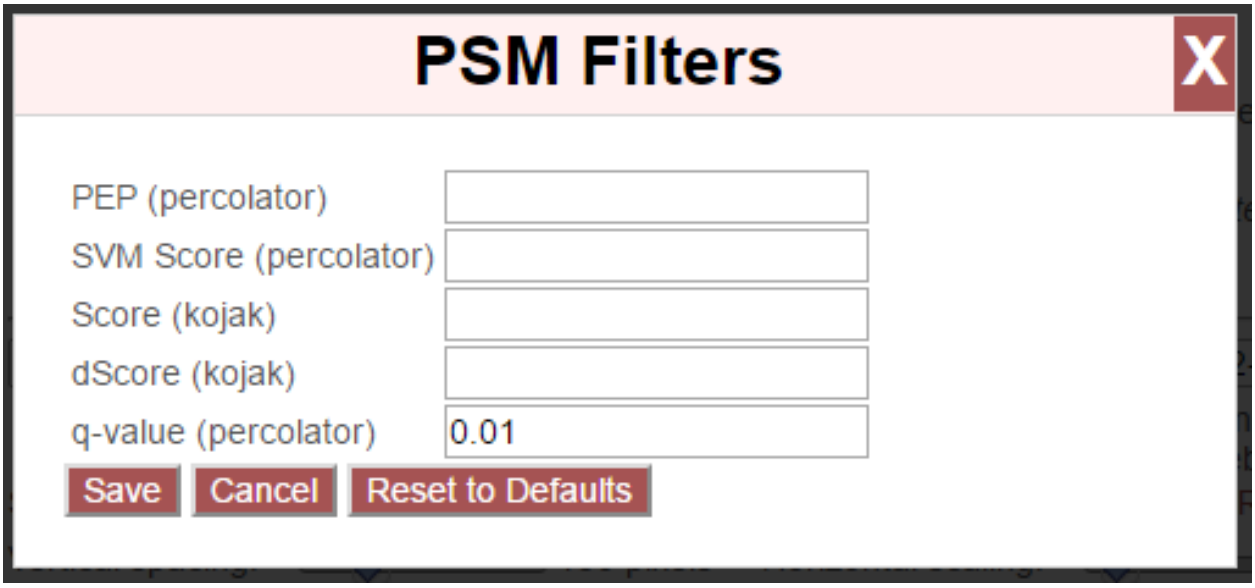
5.4.1 PSM Filters

The filters to apply at the PSM level. Only results which have at least one PSM that meets all of the selected critiera will be listed. When listing PSMs associated with peptides, only PSMs that meet all of the selected critiera will be listed.

To change the PSM-level filters, first click the pencil icon next to "PSM Filters":



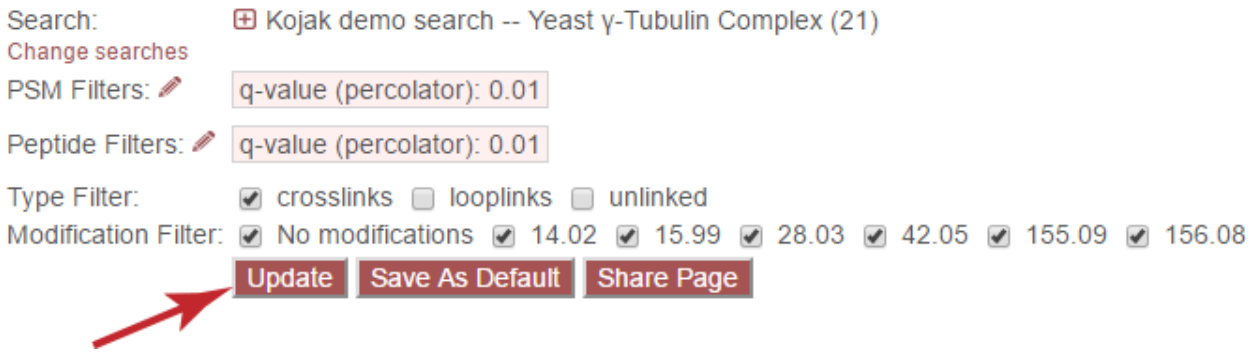
This opens an overlay with the containing the possible score types to use as PSM filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

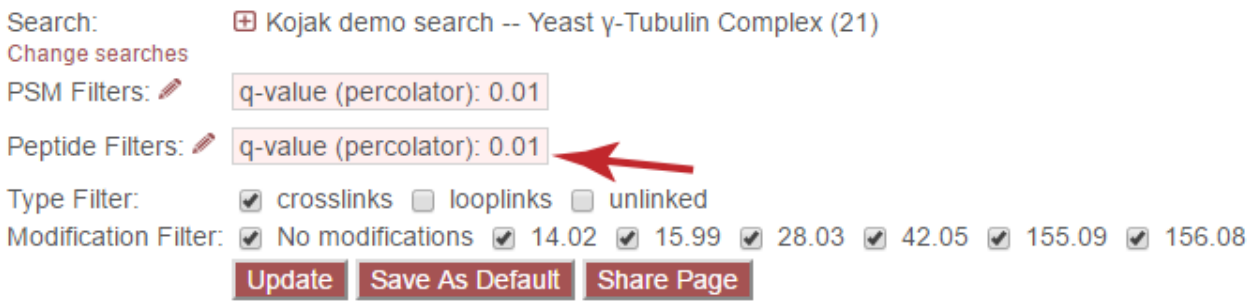
Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update From Database" button on the page to apply any new PSM- or peptide-level filters.



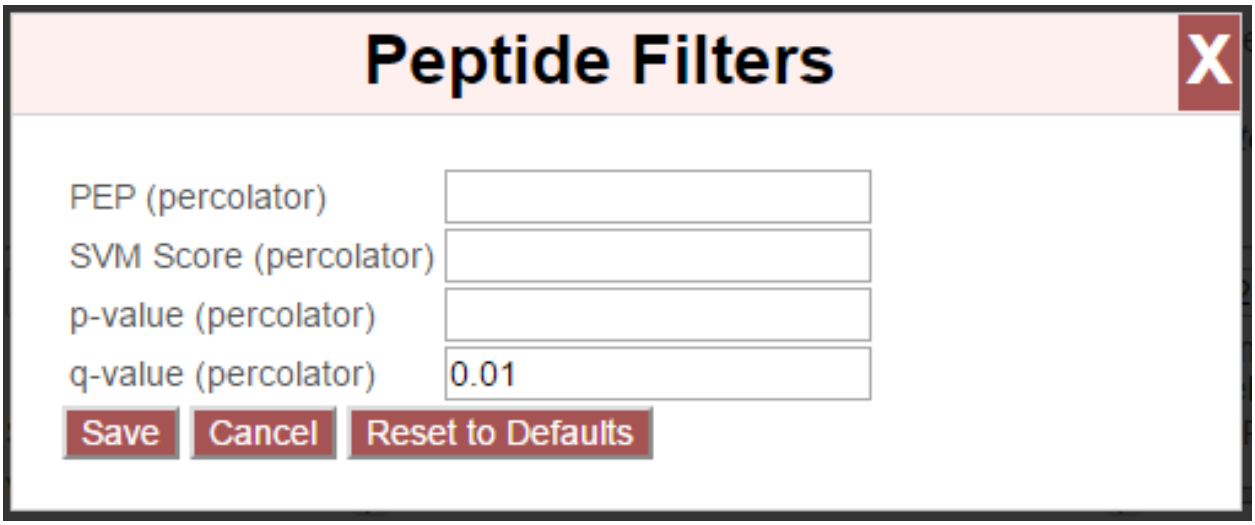
5.4.2 Peptide Filters

The filters to apply at the peptide level. Only results which have at least one peptide that meets all of the selected critiera will be listed.

To change the peptide-level filters, first click the pencil icon next to "Peptide Filters":



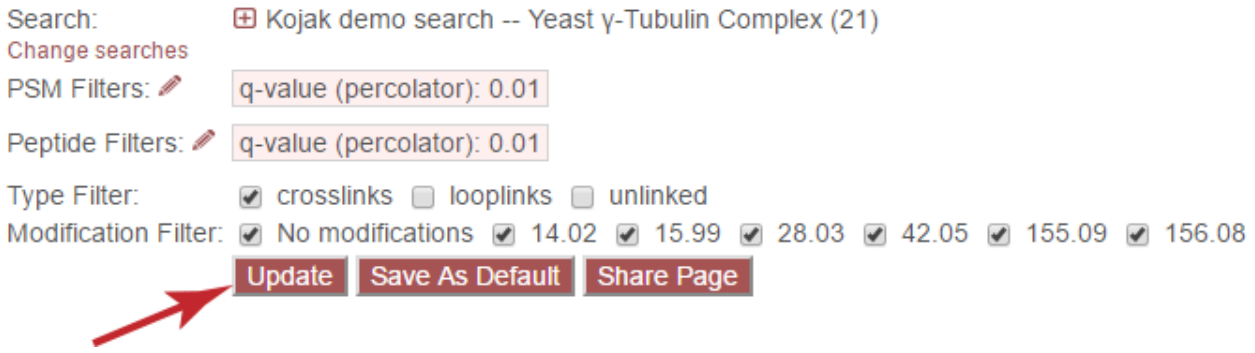
This opens an overlay with the containing the possible score types to use as peptide-level filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update From Database" button on the page to apply any new PSM- or peptide-level filters.



5.4.3 Exclude links with

Peptides with any of the checked attributes will not be shown. The attributes are:

- no unique peptides - If the link (crosslink, looplink, or monolink) was exclusively identified by peptides that also map to othe proteins
- only one PSM - If a given link was identified by a single PSM
- only one peptide - If a given link was identifed by a single peptide, where a peptide is the combination of sequence, linked positions, and modifications

5.4.4 Exclude proteins with

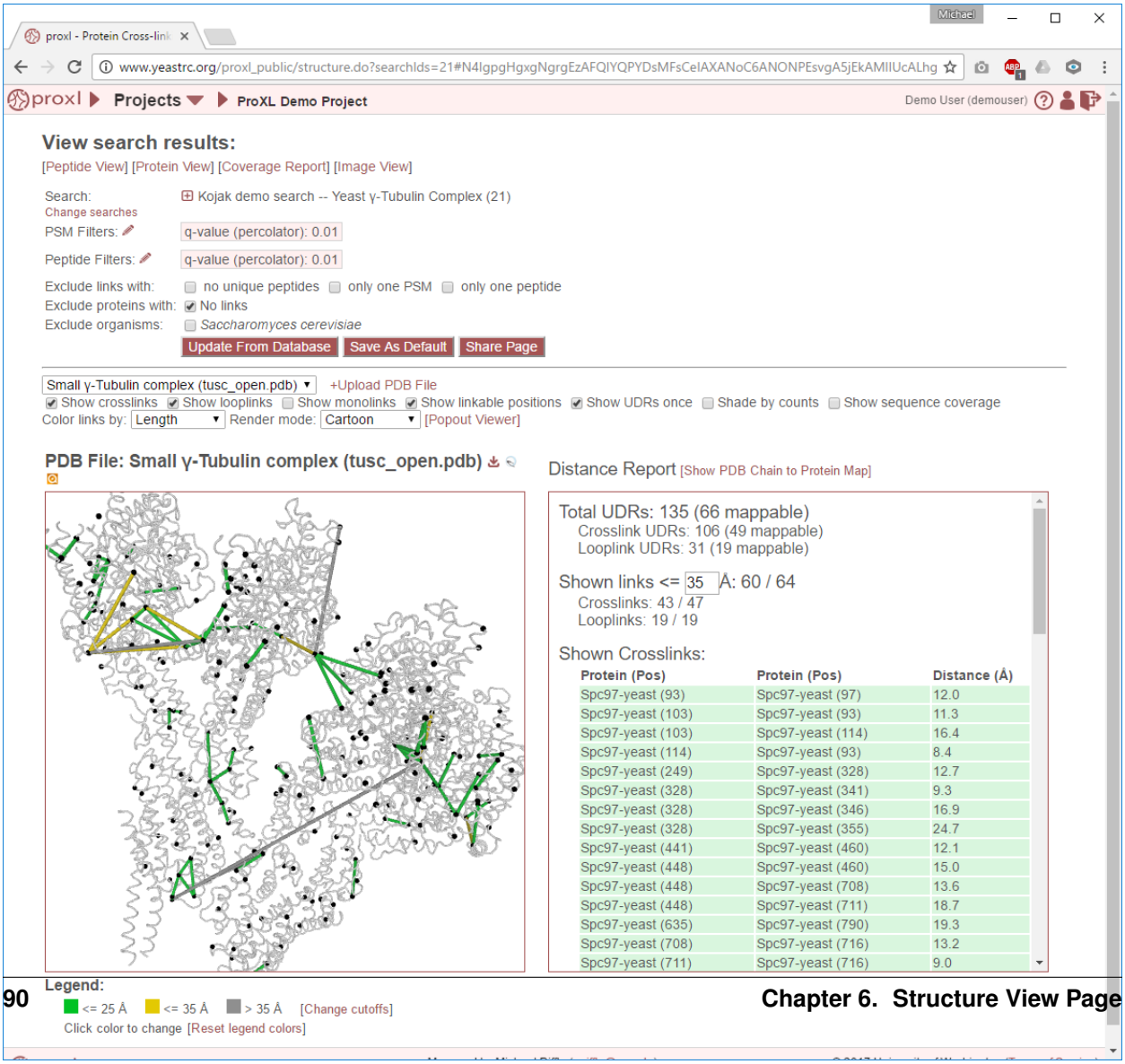
This option limits which proteins will appear in the pull-down menu for adding proteins to the viewer (see below). Proteins that contain any of the checked options will not appear. For example, checking 'No links' prevents proteins that do not contain crosslinks, looplinks, or monolinks from appearing.

5.4.5 Exclude organisms

This options limits which proteins will appear in the pull-down menu for adding proteins. No proteins from any of the checked organisms will appear.

CHAPTER 6

Structure View Page



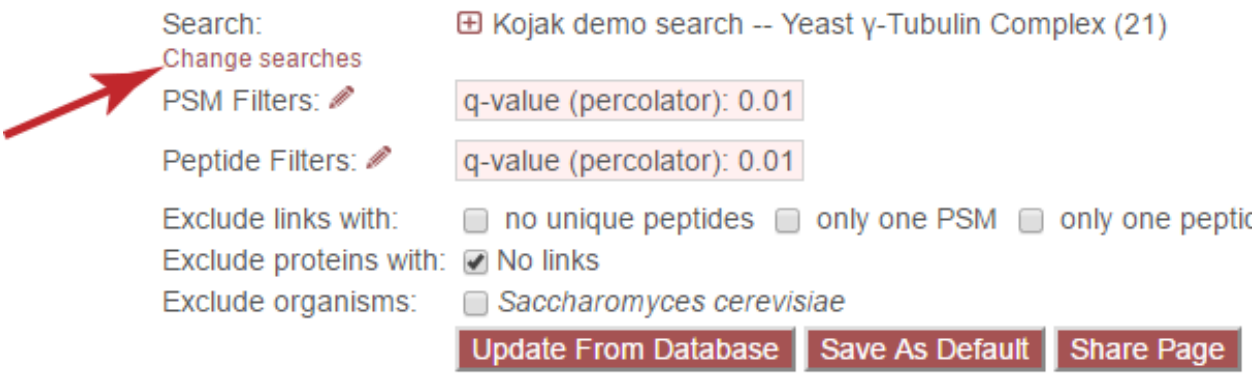
The structure viewer allows users to upload PDB files (including multi-protein complexes) from any source, align protein sequences from their experiment to sequences in those PDB files, and then visualize and interact with the crosslinking results on fully interactive 3D structures. The structure viewer includes multiple filtering and display options, distance reporting, and users may download the data for use in other structure visualization software or as high-quality raster images.

6.1 Viewer Basics

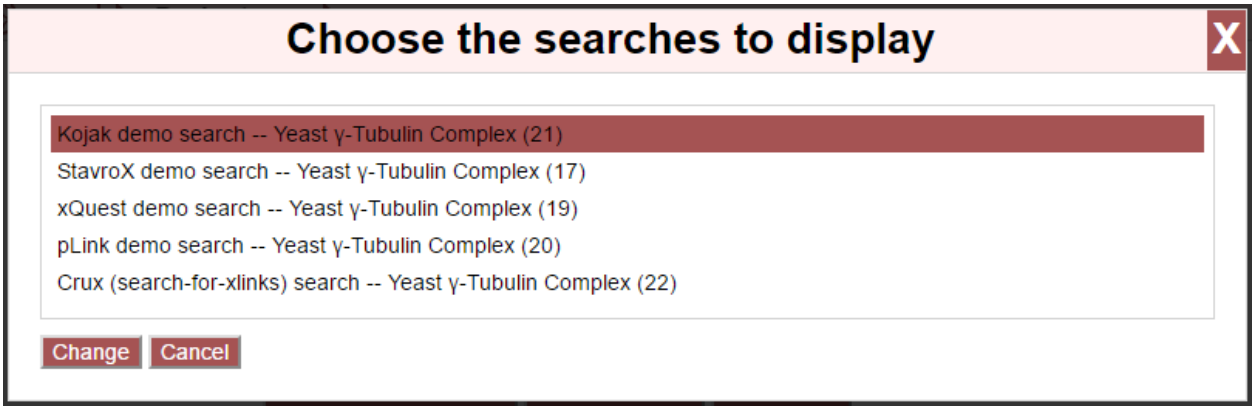
6.1.1 URL Captures State of Page

The URL of the page is dynamically updated at all times to reflect the complete state of the viewer–including filter parameters, protein bar positions, and all viewer options. As such, the URL may be bookmarked or shared to link to a viewer with the same content and appearance as the current viewer–just copy and paste from the address bar. For complicated visualizations, this is a convenient way to save and share you work. (Note, this link only works for other users who are listed on this project unless public access is enabled.)

6.1.2 Change Searches

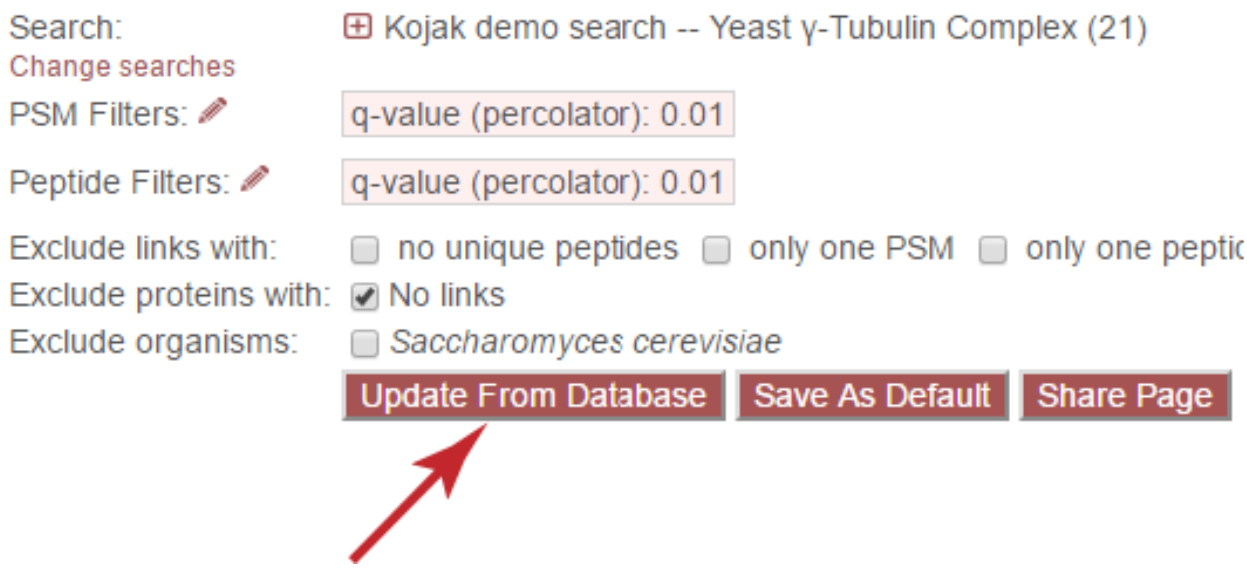


The "Change searches" link allows the user to change which searches are currently being displayed. None of the other options on the page (such as viewer settings) will be changed, only the searches from which data are displayed. Clicking the link causes the following overlay to be displayed:



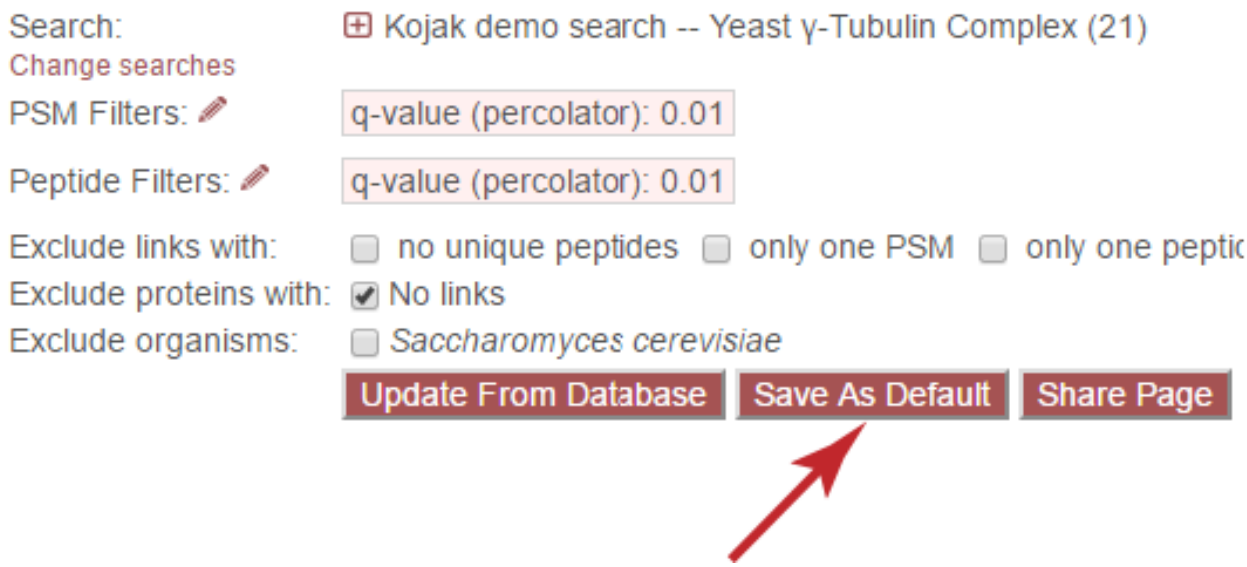
Select or de-select searches by clicking on them in the list. Once done, click "Change" to update the page with the new data or "Cancel" to close the overlay.

6.1.3 Update From Database



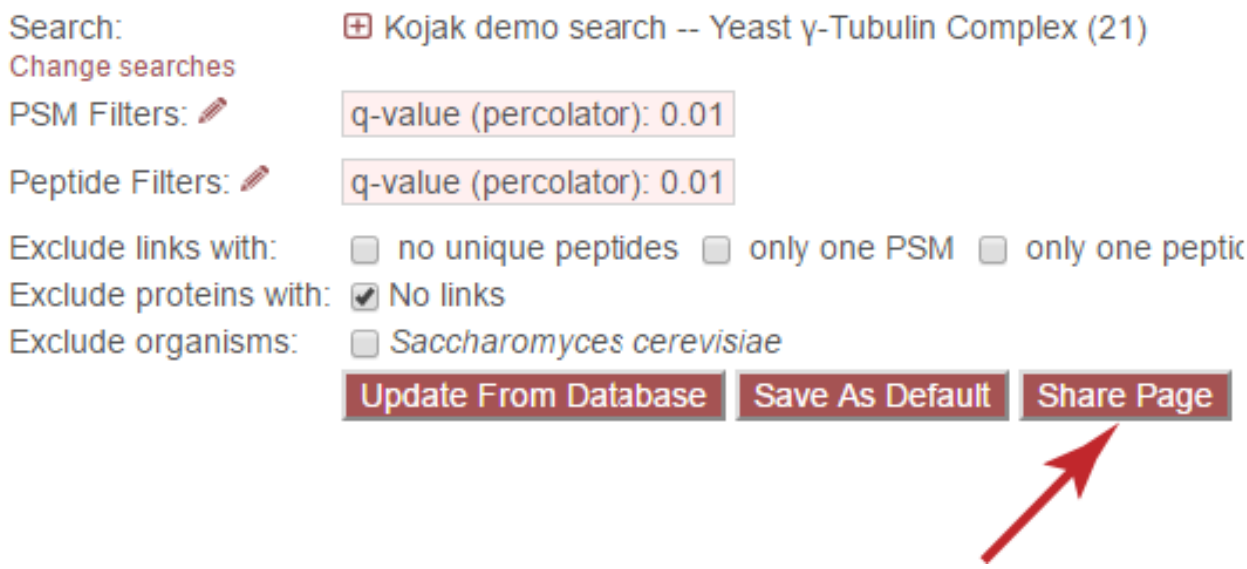
If the user changes any filter parameters–such as PSM/peptide score cutoffs, link type exclusions, protein exclusions, or taxonomic exclusions–this button must be clicked to reflect the new filter choices.

6.1.4 Save as Default



Project owners may click "Save as Default" to save the current URL (see above) as the default view of the "Structure View" for this search. This default view will be populated with the same proteins, positions, and viewer options as when the button is clicked. This is a convenient way to share data with collaborators or the public that does not require that they manipulate the image viewer to see the data.

6.1.5 Share Page

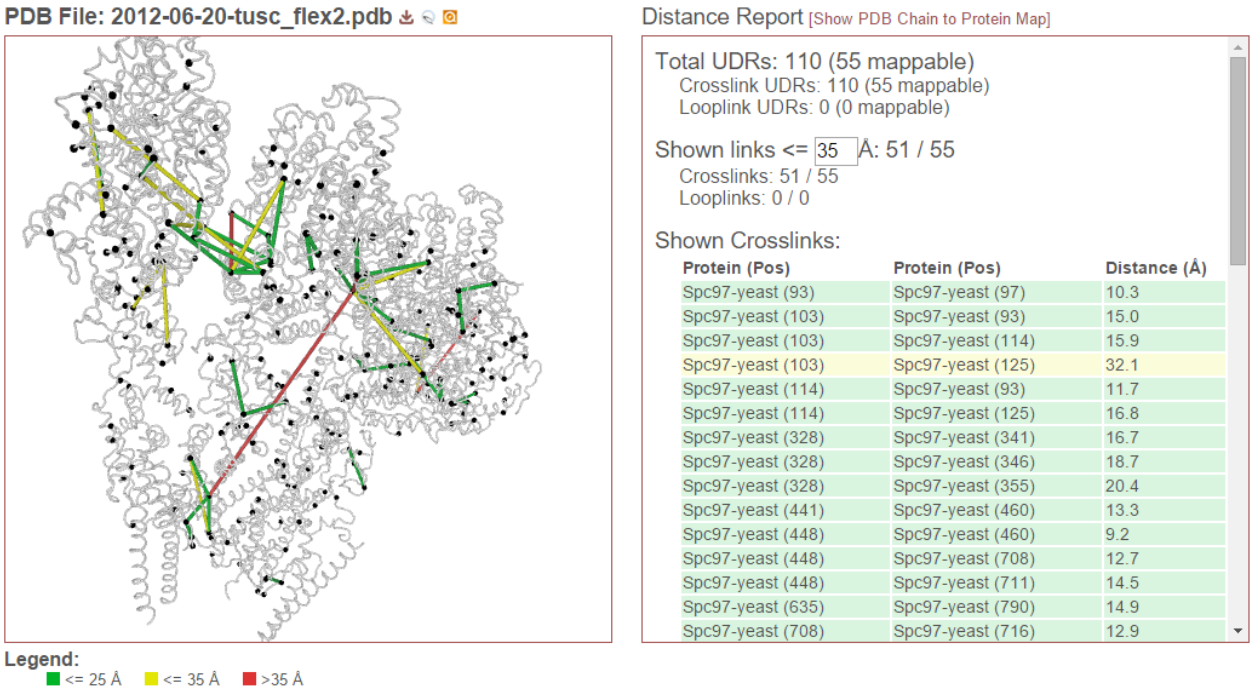


Clicking the "Share Page" button will generate a shortcut URL for viewing the current page, including all viewer settings, filter options, and so forth. The shortened URL will appear in an overlay as:

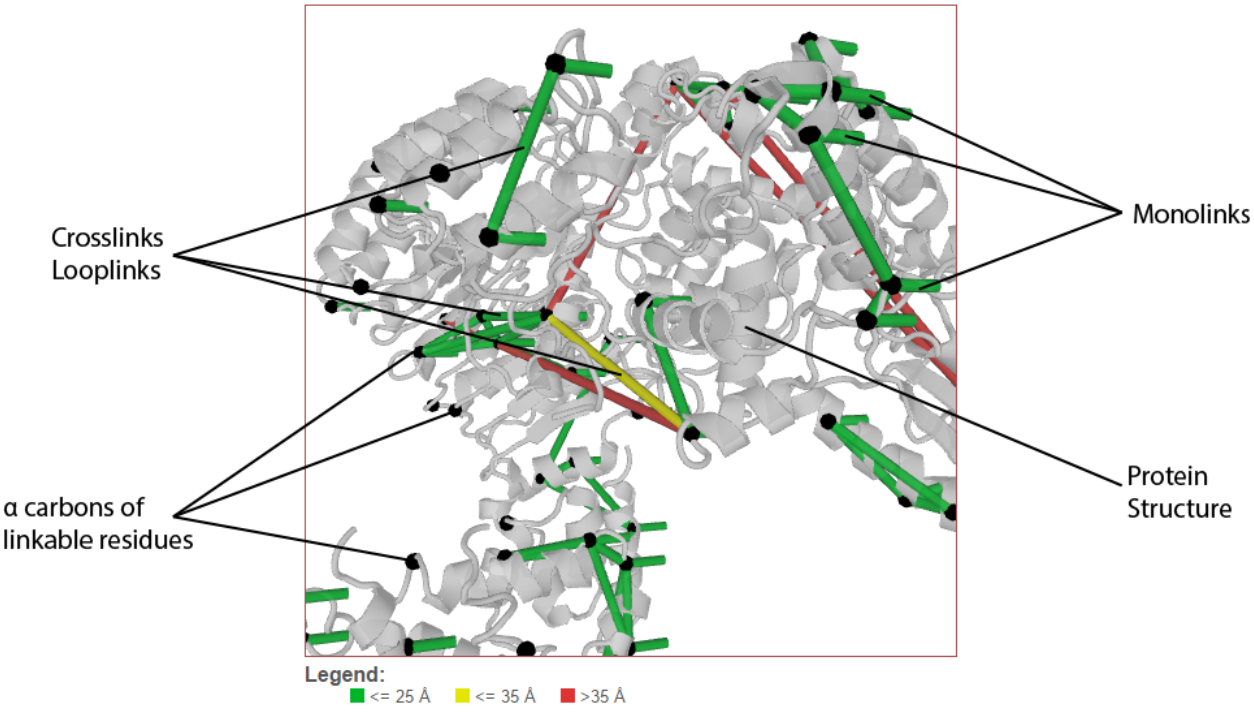
Copying and sharing the highlighted URL will direct users to the view of the page when the URL was generated. Note that this URL does not grant access to the page to any user that would not otherwise have access.

6.1.6 Viewer Display

The viewer is separated into two side-by-side panels. The left panel contains an interactive, 3D rendering of the PDB structure file, link locations from the experiment, and other sequence annotations for those proteins. The right panel contains either a report of the distances of the currently-rendered links (see *Distance Report*) or a form for choosing which protein-PDB chain alignments are currently being drawn (see *PDB Chain to Protein Map*).

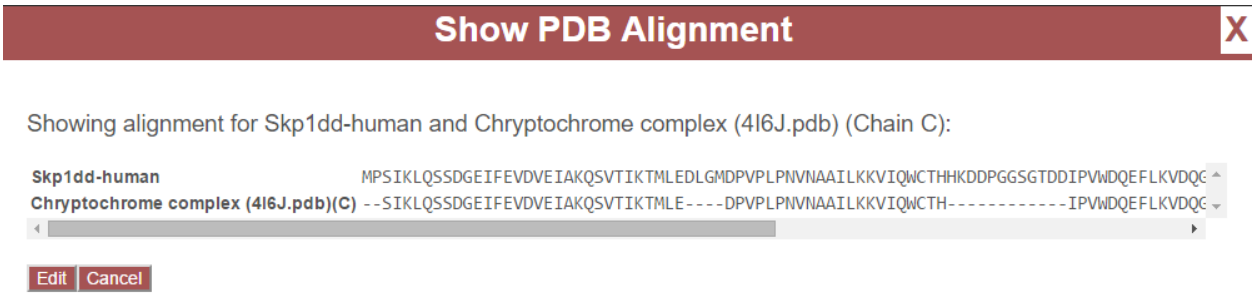


Crosslinks (and looplinks, if enabled) are displayed as rods connecting two points in the protein structure. By default, the color of these links is determined by the distance between the two points they connect. (This coloring scheme may be changed, see *Color links by*.) Monolinks are depicted as short rods connected to the structure on only one end. By default, "linkable" positions are shown as black spheres placed on the alpha carbons of residues in proteins that may react with the crosslinker used in the experiment.



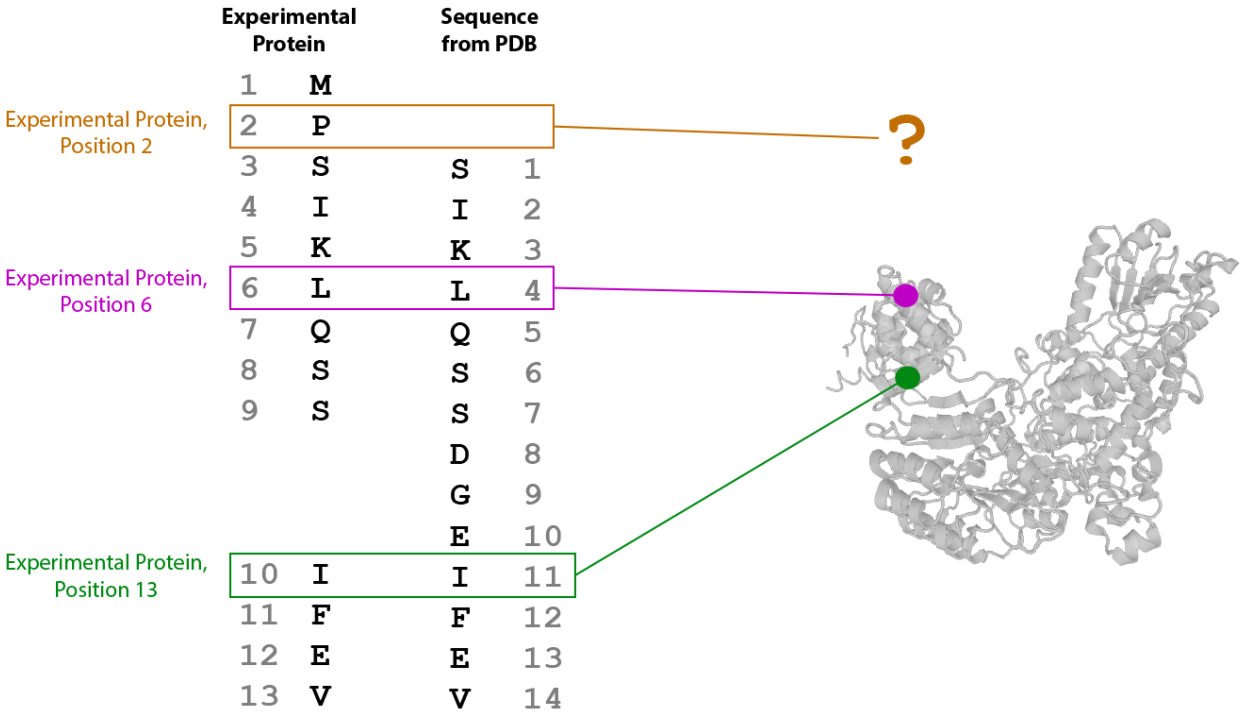
6.1.7 Mapping Residue Position to 3D Space

After a PDB file is uploaded, proteins found in the experiment may be mapped to chains from the PDB file based on sequence (see *Map PDB Chains to Proteins*). This process creates a mapping of residue positions in proteins found in the experiment to residue positions in the PDB structure. This mapping is very rarely 1 to 1, that is, the sequences of the protein and the PDB chain may be different–containing variation, insertions, or deletions. For example, view the following alignment between Skp1dd and chain C from 4I6J.pdb:



Note that position 1 in Skp1dd does not map to any position in chain C of the PDB file. Position 3 in Skp1dd maps to position 1 in chain C. Position 4 in Skip1dd maps to position 2, Position 5 maps to position 3, and so on. All insertions and deletions are taken into account to create a mapping for every position between the two sequences.

So, then, when drawing (or measuring distances between) positions that correspond to specific residues in experimental proteins, this mapping is used to lookup the corresponding positions in the PDB chain sequences. If the positions map to the PDB, the 3D coordinate position of the alpha carbon in the corresponding PDB residue is used as the location of the residue in the rendered structure. If the positions do not map to the PDB, they are never drawn or measured.



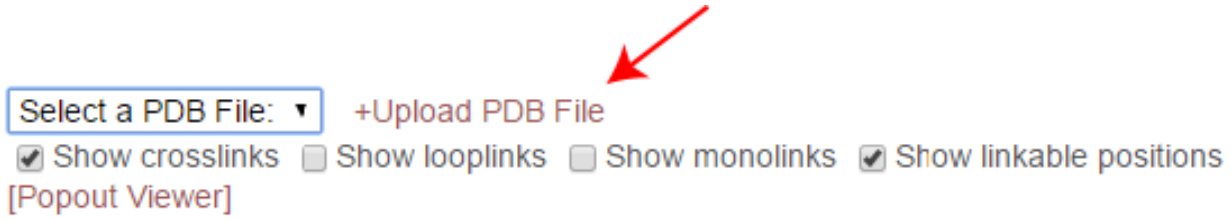
In proxl, links that map to the structure on both ends are said to be "mappable". The distance report displays the number of mappable links from the experiment out of the total number of observed links. See *Distance Report* for more information.

6.2 Uploading PDB Files and Mapping Proteins

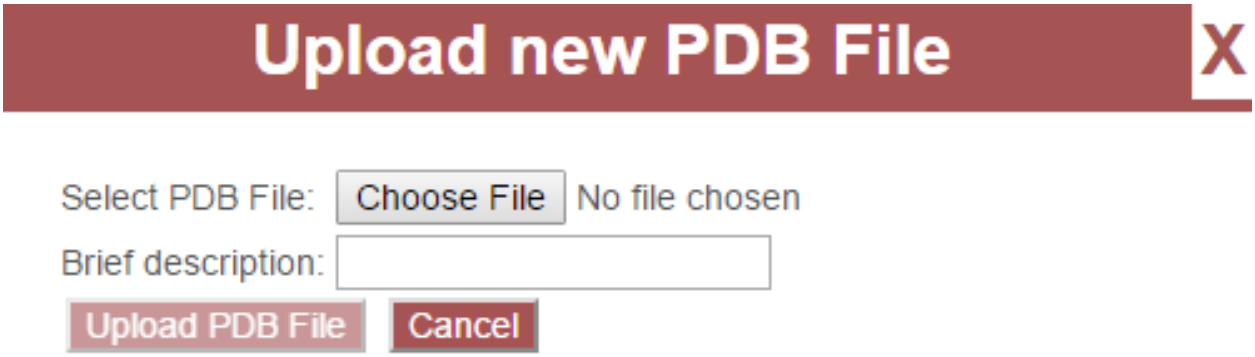
6.2.1 Upload PDB File

A PDB file in proxl is any file adhering to the PDB file format. These include public PDB files downloaded from the PDB database, or structures you have generated yourself using any number of programs– as long as they adhere to the PDB file format. The PDB file may contain structures for multiple proteins (such as for a protein complex). PDB files you upload to proxl are only visible to members of the project with which the data are associated. (If you enable public access on the project, the PDB file will be visible to public access users as well.)

To upload a PDB file, click the "+Upload PDB File" link next to the PDB file pull-down menu above the viewer:



This will open a dialog for uploading a PDB file. Click the button next to "Select PDB File" to select a PDB file on your computer:

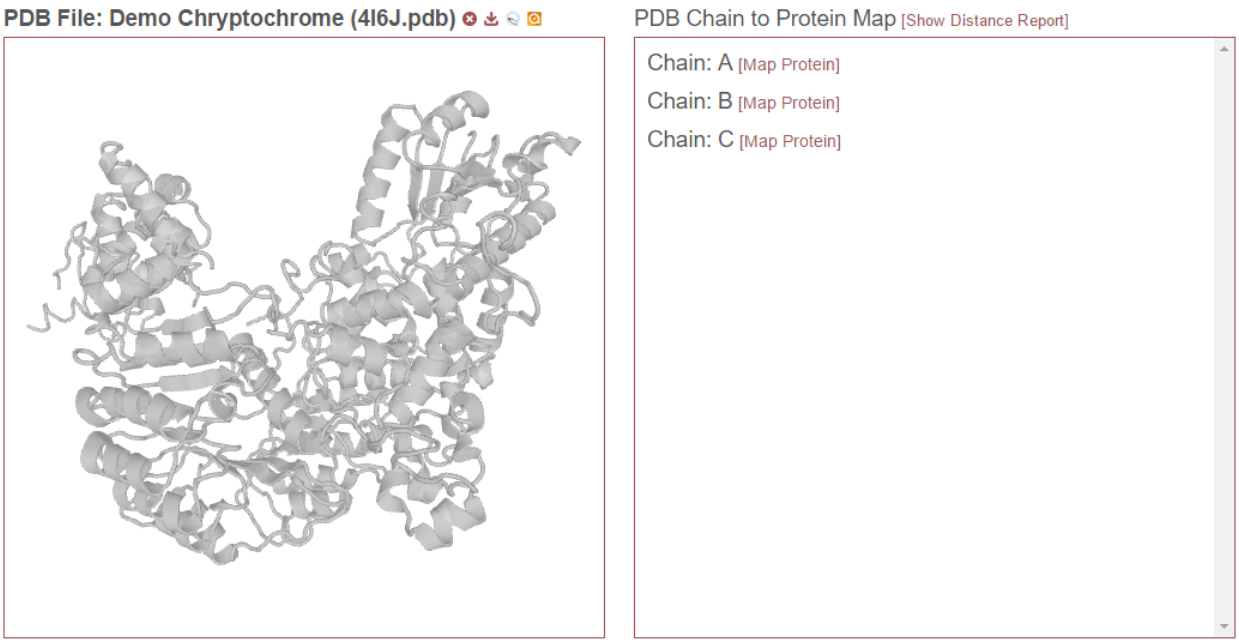


Enter a brief description for your PDB file and click "Upload PDB File." The PDB file will then be visible in the PDB File pull-down menu.

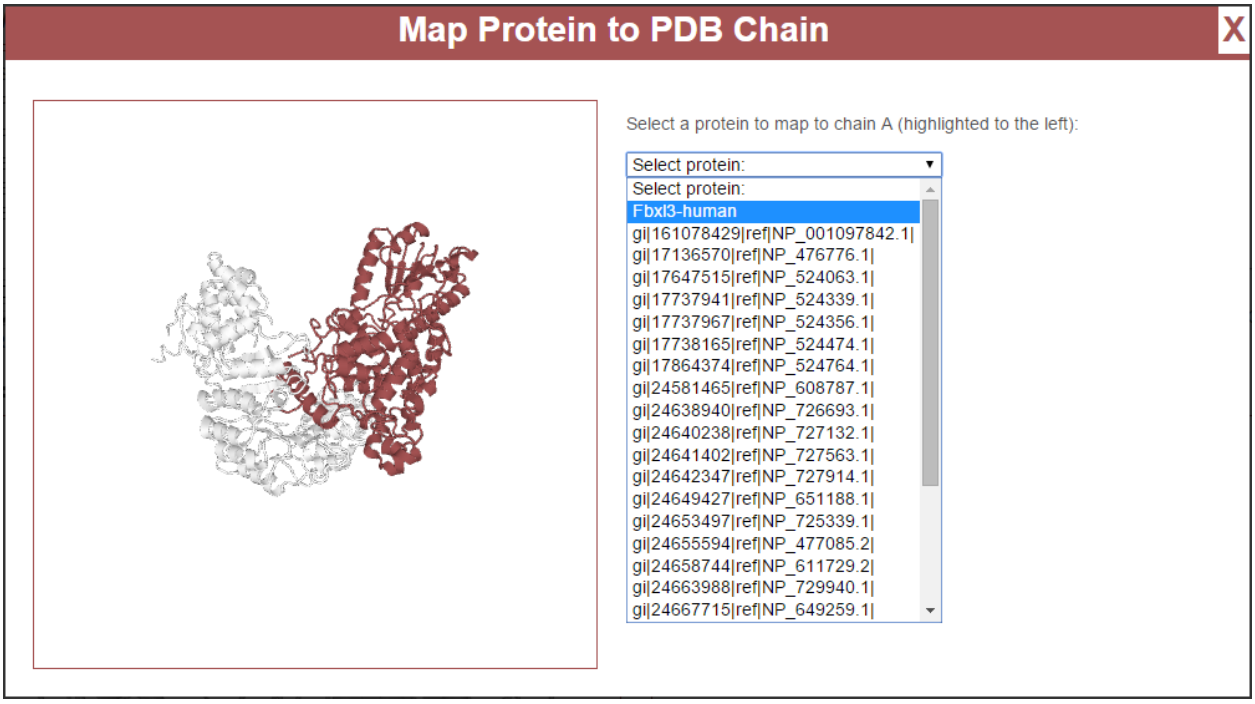
6.2.2 Map PDB Chains to Proteins

In order to calculate distances or view crosslinking data on a 3D structure, it is necessary to first map proteins from the experiment to sequences present in a PDB file. (To learn more about how proxl uses this mapping to find 3D positions, see *Mapping Residue Position to 3D Space*.) The sequences in the PDB file do not need to be 100% matches to the protein sequences in your experiment, and may contain insertions or deletions. However, the quality of the reported distances and visualized links depends on the matches being close. It is also not necessary to map proteins to all chains in the PDB file.

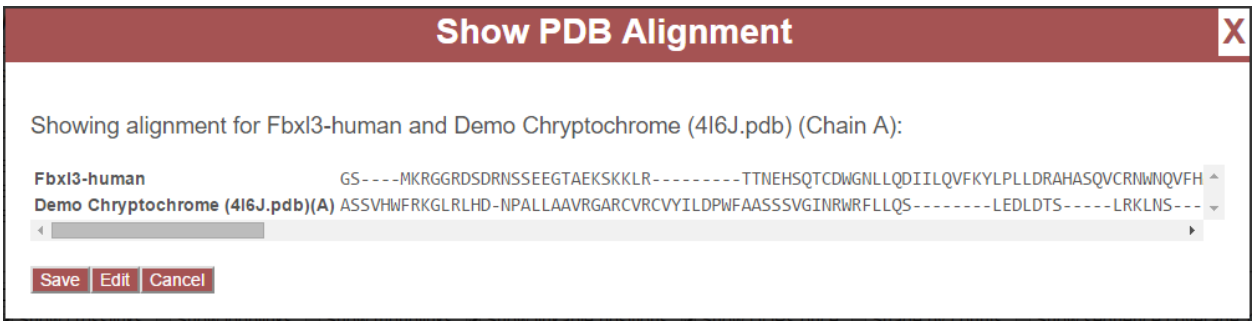
To perform this mapping, first select a PDB file in the PDB file pull-down menu. This will display the 3D structure from the PDB file in the left panel and, for proteins with no mapping, a "PDB Chain to Protein Map" in the right panel with no proteins listed for any of the chains.



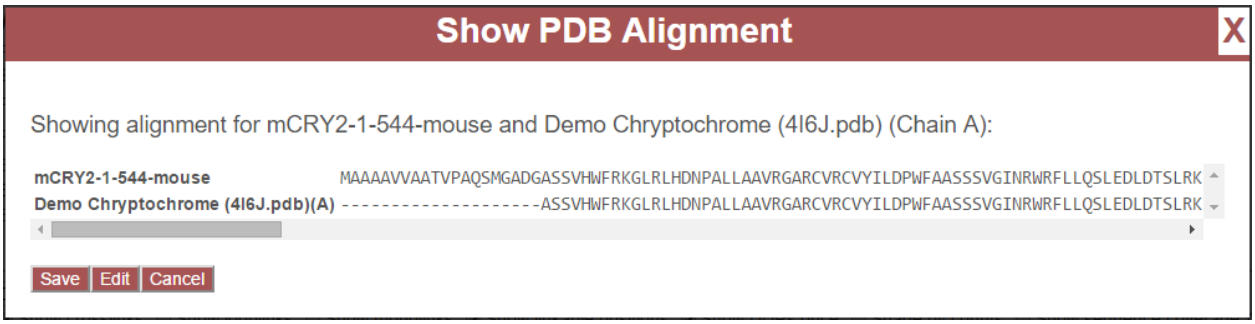
To begin, click the "[Map Protein]" link next to the desired PDB chain. This will open the following window:



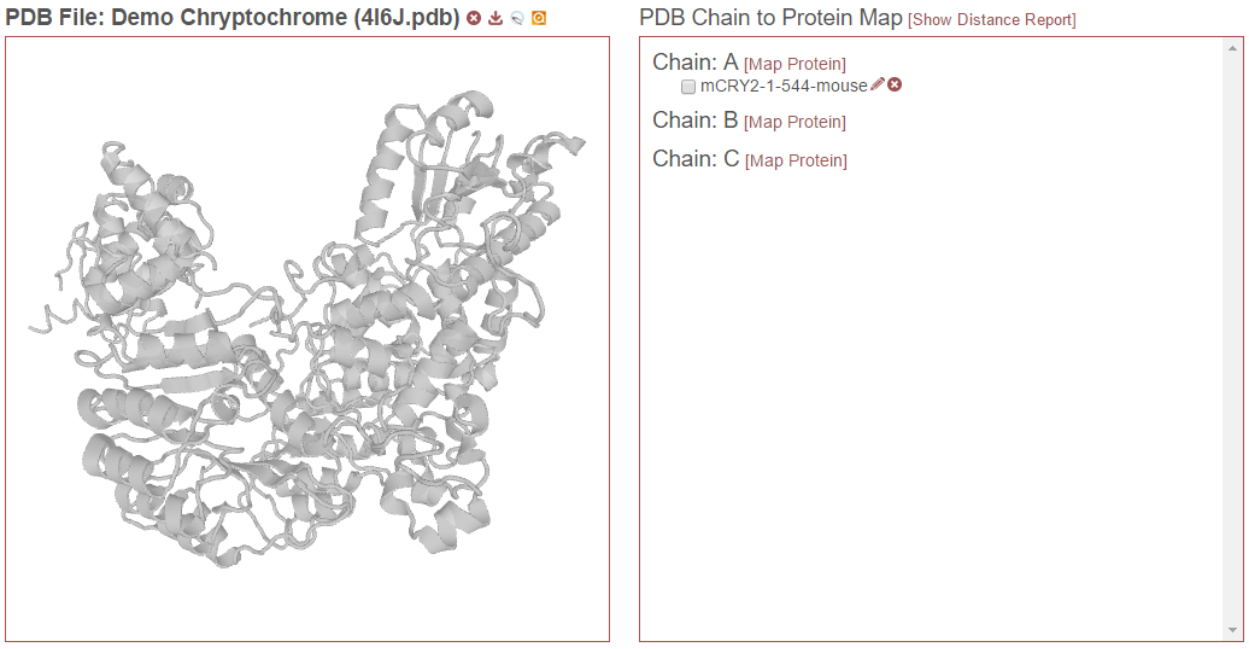
The panel to the left shows the structure from the PDB file, with the currently-selected chain highlighted in red. This rendering may be rotated and zoomed just as the main visualization, see *Structure Manipulation*. To the right is a pull-down list of all proteins found in the experiment. Click the one to be mapped to the currently-selected chain and click the "Map Protein to Structure" button. This will perform a local pairwise sequence alignment on the experimental protein's sequence and the sequence for the selected chain from the PDB file and present the results in the window:



This is an example of a bad match. The pairwise sequence alignment will always be successful, even in the case of poor matches. In the example above, "Fbxl3-human" was chosen and a sequence alignment was performed against chain A from 4I6J.pdb. Note that very few residues match between the alignments, and there are many insertions and deletions present. The horizontal scroll bar present beneath the alignment may be used to view the entire alignment. To reject this alignment, click "Cancel" to map a different protein.



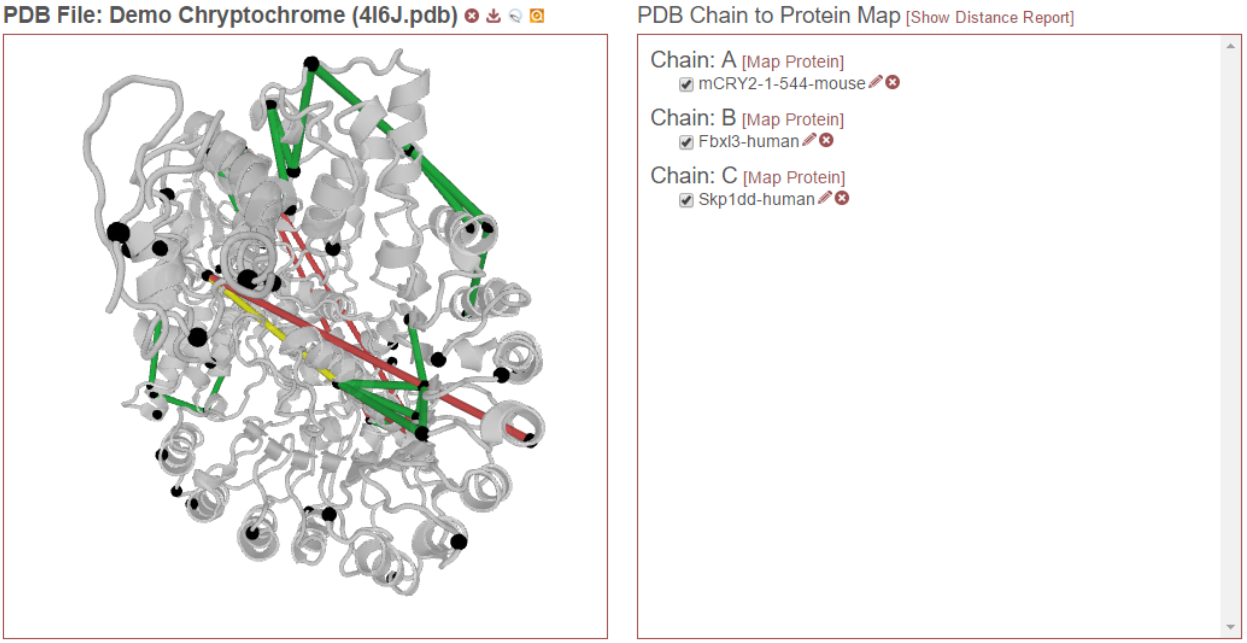
This is an example of a good match. In the example above, "mCRY2-1-544-mouse" was chosen as the experimental protein. All the matched residues are identical. There is a segment of sequence at the N-terminus of the experimental protein that is not present in the PDB file. To accept this match, click "Save." The mapped protein will now appear as associated with Chain A in the "PDB Chain to Protein Map":



This process can be repeated for as many other chains as desired. Additionally, each chain may be associated with multiple proteins from the experiment–such as in the case that homologs or proteoforms are present in the search

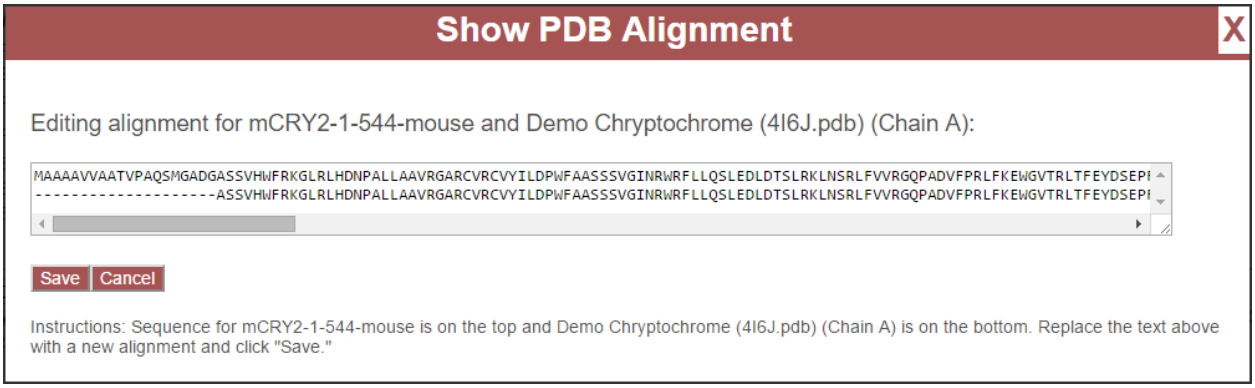
results. However, only up to one protein listed under a given chain may be checked at any given time.

Then, to view the crosslinking results on the structure, check the desired protein chain alignments (check box next to a given protein listed under a given chain). To learn more about the effects and implications of checking the alignments, see *PDB Chain to Protein Map*.



6.2.3 Edit PDB Chain to Protein Mapping

Although not generally advised, there are two methods for manually editing the alignment between an experimental protein and a PDB chain. First, during the initial mapping process described above, instead of clicking "Save" or "Cancel" on the final step, click "Edit." Or, after the mapping is complete, click the pencil icon next to the protein name associated with a chain in the PDB and click "Edit" in the resulting window. Both methods will open the alignment edit window:



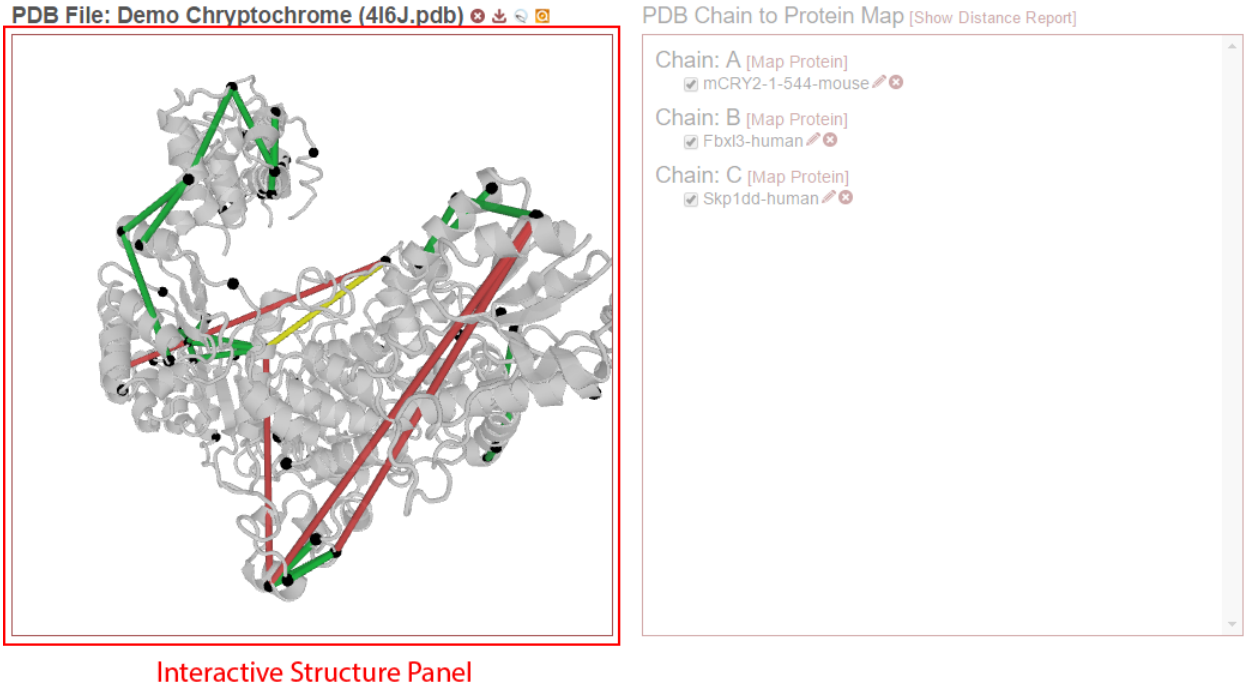
This interface consists of a simple text field containing the current pairwise sequence alignment, with the experimental protein on the top and the PDB sequence on the bottom. This alignment may be manually edited, either in this window or in an external text editor and pasted back into this window. The length of both aligned sequences, including the dashes (-) must be identical, and the sequences present for the experimental protein and PDB chain (without the dashes) must match the sequences on record. To cancel the process, click "Cancel". To save the manual alignment, click "Save."

6.2.4 Delete PDB Chain to Protein Mapping

Click the red circle (X) icon next to the protein listed for a chain in the PDB on the "PDB Chain to Protein Map" panel to remove that alignment from the database.

6.3 Interactive Structure Panel

This section describes the functionality and features of the left panel, where the 3D protein structures are presented:



6.3.1 Structure Manipulation

Rotation

Using a mouse, the structure may be rotated on a central axis by clicking and dragging in the panel–that is, click, hold down mouse button, and move the mouse cursor in the desired direction of rotation. On touch devices, simply tap, hold, and slide finger in desired direction of rotation.

Zooming

To zoom in or out using a mouse, use the scroll wheel–slide the scroll wheel towards you to zoom in and away from you to zoom out. On touch devices, pinch and zoom with two fingers to zoom in or to zoom out.

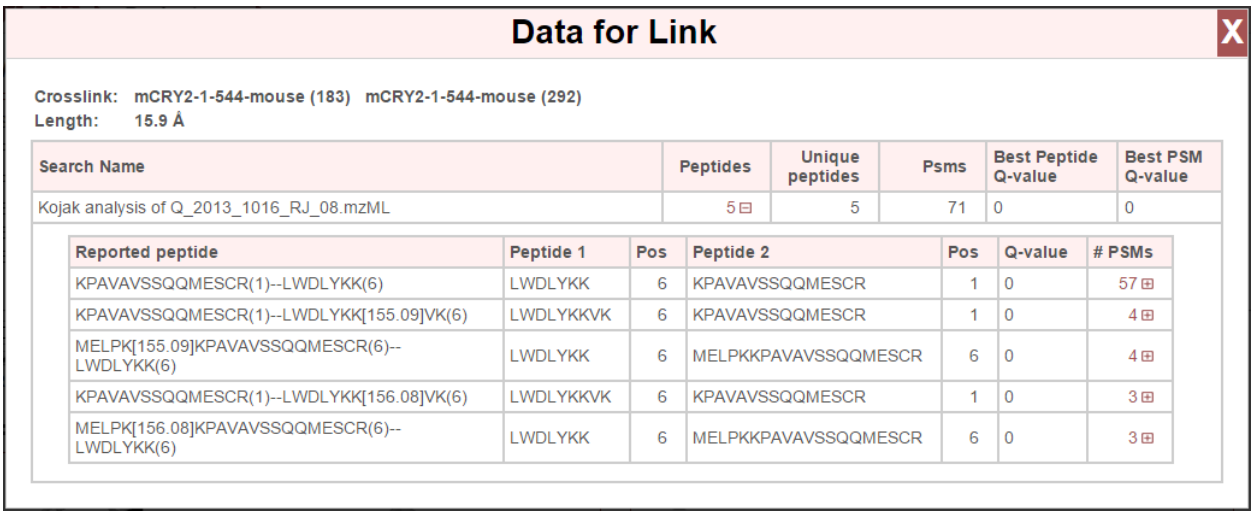
Re-centering

Double clicking on the structure will move the center of rotation to that point and re-center the view of the structure to that point.

6.3.2 Click on Links

Any link (crosslink, looplinks or monolink) may be clicked on to view underlying link information (such as which positions in which proteins are connected), a peptide list for peptides found to support those links, a list of corresponding PSMs, and view associated spectra.

Clicking a link will open a window that lists the link information and underlying peptide list:



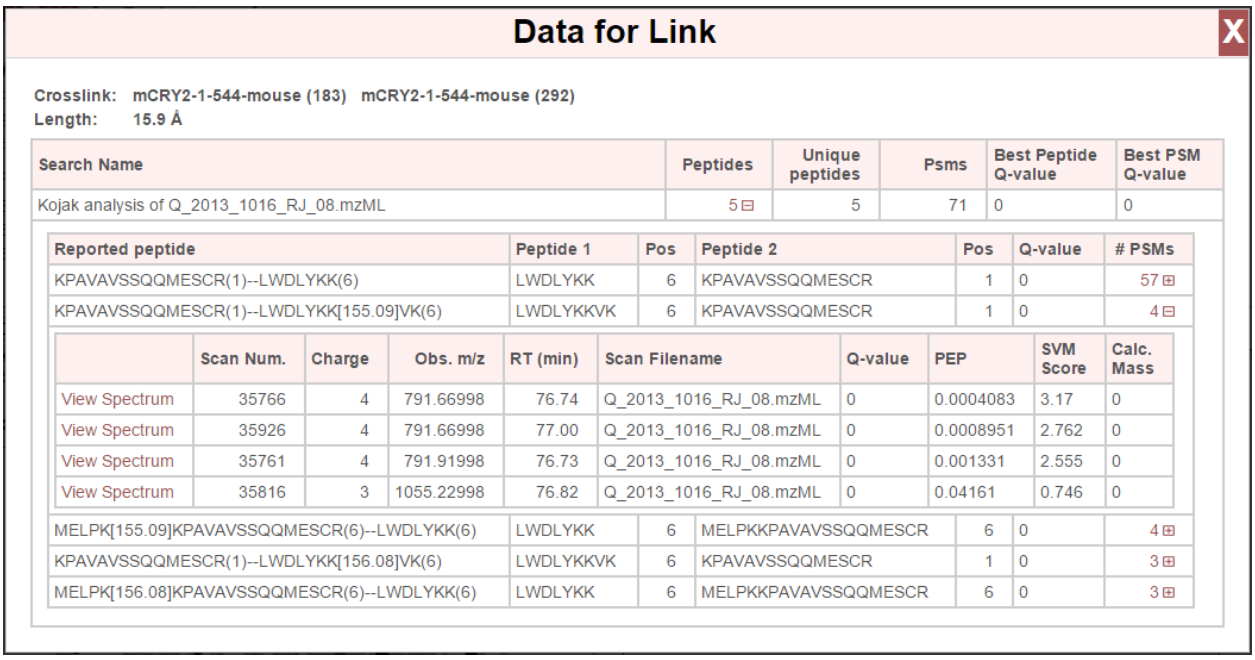
Data for Link

Crosslink: mCRY2-1-544-mouse (183) mCRY2-1-544-mouse (292)
Length: 15.9 Å

Search Name	Peptides	Unique peptides	Psms	Best Peptide Q-value	Best PSM Q-value
Kojak analysis of Q_2013_1016_RJ_08.mzML	5 ⊟	5	71	0	0

Reported peptide	Peptide 1	Pos	Peptide 2	Pos	Q-value	# PSMs
KPAVAVSSQQMESCR(1)--LWDLYKK(6)	LWDLYKK	6	KPAVAVSSQQMESCR	1	0	57 ⊞
KPAVAVSSQQMESCR(1)--LWDLYKK[155.09]VK(6)	LWDLYKKVK	6	KPAVAVSSQQMESCR	1	0	4 ⊞
MELPK[155.09]KPAVAVSSQQMESCR(6)--LWDLYKK(6)	LWDLYKK	6	MELPKKPAVAVSSQQMESCR	6	0	4 ⊞
KPAVAVSSQQMESCR(1)--LWDLYKK[156.08]VK(6)	LWDLYKKVK	6	KPAVAVSSQQMESCR	1	0	3 ⊞
MELPK[156.08]KPAVAVSSQQMESCR(6)--LWDLYKK(6)	LWDLYKK	6	MELPKKPAVAVSSQQMESCR	6	0	3 ⊞

At the top are the linked proteins, their respective positions, and the distance of that link. A list of peptides is then presented for each search (if viewing data from multiple searches, each search is listed) that support this link. Any of the rows listing peptides may be clicked and expanded to view the underlying PSMs:



Data for Link

Crosslink: mCRY2-1-544-mouse (183) mCRY2-1-544-mouse (292)
Length: 15.9 Å

Search Name	Peptides	Unique peptides	Psms	Best Peptide Q-value	Best PSM Q-value
Kojak analysis of Q_2013_1016_RJ_08.mzML	5 ⊟	5	71	0	0

Reported peptide	Peptide 1	Pos	Peptide 2	Pos	Q-value	# PSMs
KPAVAVSSQQMESCR(1)--LWDLYKK(6)	LWDLYKK	6	KPAVAVSSQQMESCR	1	0	57 ⊞
KPAVAVSSQQMESCR(1)--LWDLYKK[155.09]VK(6)	LWDLYKKVK	6	KPAVAVSSQQMESCR	1	0	4 ⊟

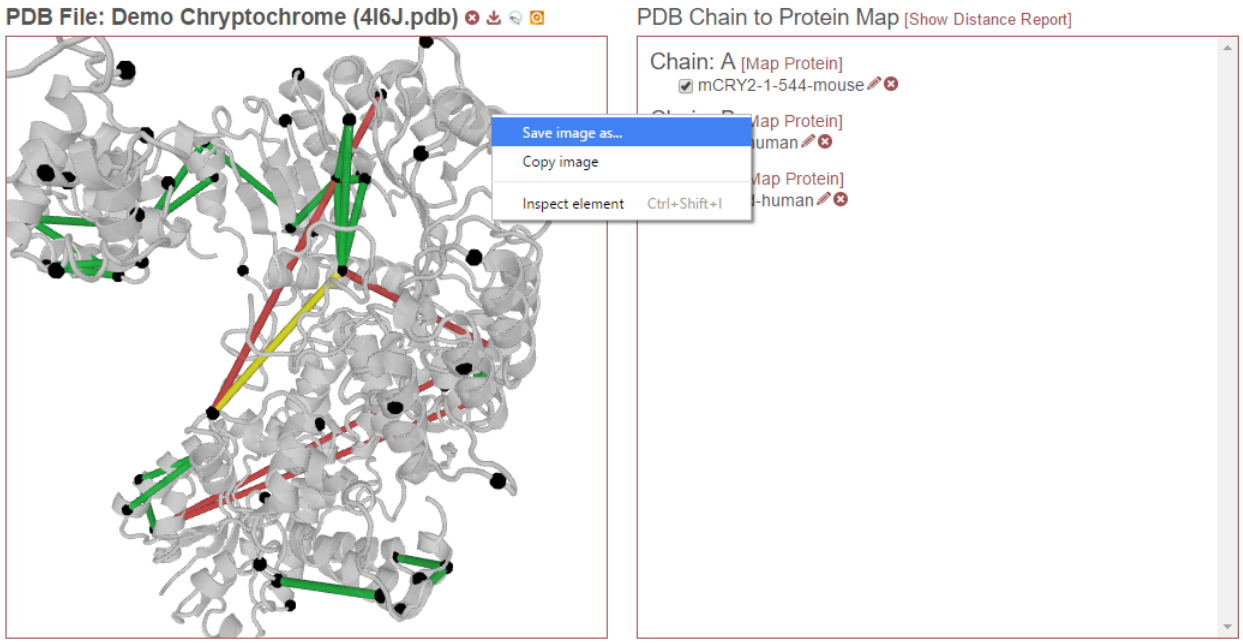
	Scan Num.	Charge	Obs. m/z	RT (min)	Scan Filename	Q-value	PEP	SVM Score	Calc. Mass
View Spectrum	35766	4	791.66998	76.74	Q_2013_1016_RJ_08.mzML	0	0.0004083	3.17	0
View Spectrum	35926	4	791.66998	77.00	Q_2013_1016_RJ_08.mzML	0	0.0008951	2.762	0
View Spectrum	35761	4	791.91998	76.73	Q_2013_1016_RJ_08.mzML	0	0.001331	2.555	0
View Spectrum	35816	3	1055.22998	76.82	Q_2013_1016_RJ_08.mzML	0	0.04161	0.746	0

MELPK[155.09]KPAVAVSSQQMESCR(6)--LWDLYKK(6)	LWDLYKK	6	MELPKKPAVAVSSQQMESCR	6	0	4 ⊞
KPAVAVSSQQMESCR(1)--LWDLYKK[156.08]VK(6)	LWDLYKKVK	6	KPAVAVSSQQMESCR	1	0	3 ⊞
MELPK[156.08]KPAVAVSSQQMESCR(6)--LWDLYKK(6)	LWDLYKK	6	MELPKKPAVAVSSQQMESCR	6	0	3 ⊞

The "View Spectrum" link for each PSM will show the respective spectrum using the built-in spectrum viewer. See *Spectrum Viewer* for more information.

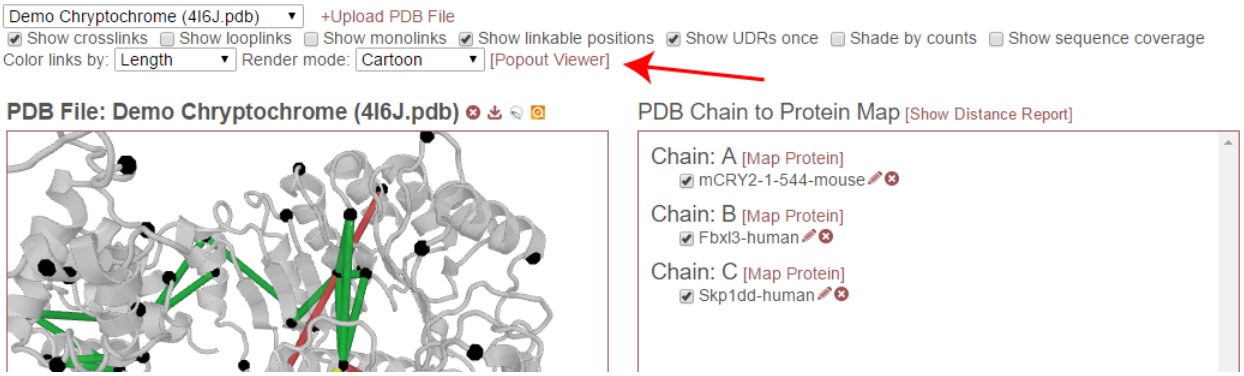
6.3.3 Save Image of Structure

Right-click the structure to save the current view as an image. This image has the same dimensions as the view window. To capture a high resolution image of the structure, use the *Popout Viewer* function to view the structure in a separate window, make that window as large as possible and then save the view as an image.

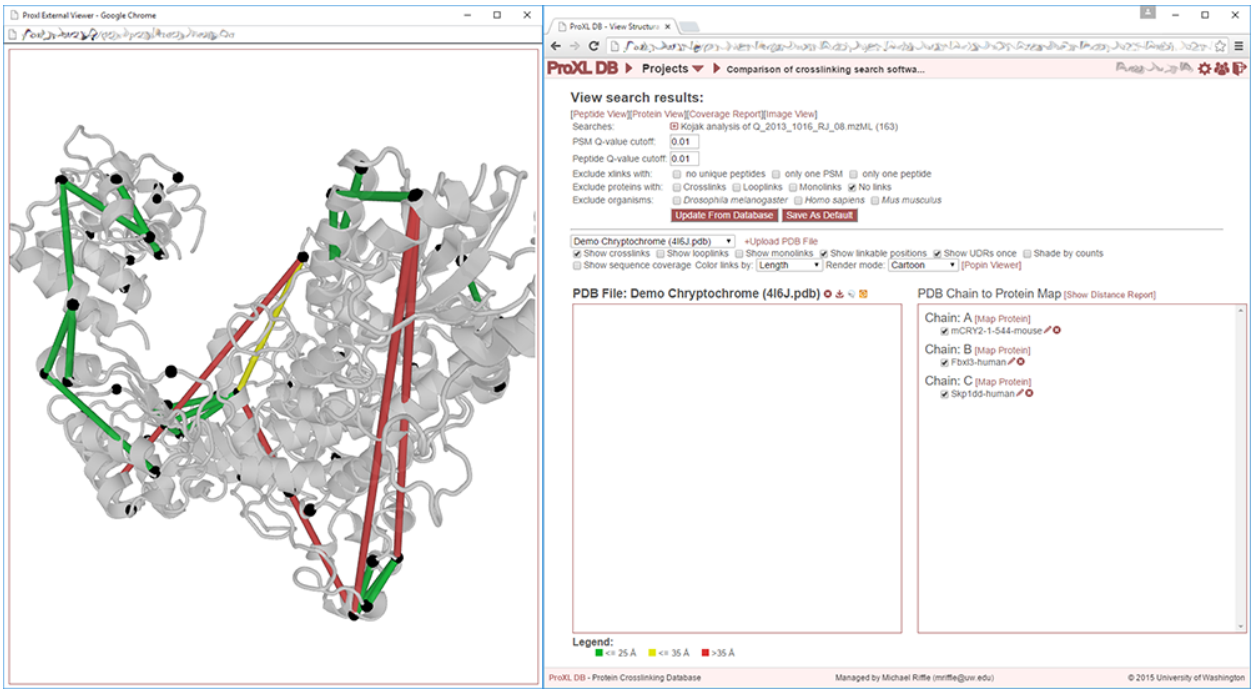


6.3.4 Popout Viewer

The structure viewer may be "popped out" into a separate window by clicking the "[Popout Viewer]" link next to the viewer options:



Clicking this link creates a new browser window with a structure viewer that fills that window:



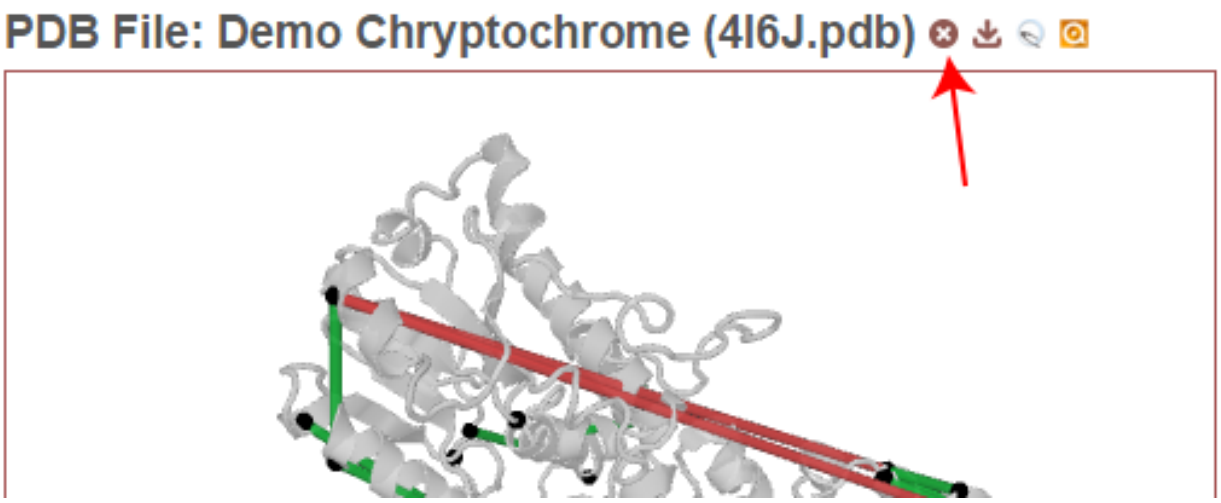
Resizing the new window dynamically changes the size of the structure viewer so that it always fills the window. The separate window allows for enhanced exploration of the structure. Not only is the structure larger and easier to see, but clicking on links in the viewer will open windows for those links in the original window without obscuring the structure. This is further enhanced in a multi-monitor environment, where it is possible to run the structure viewer full-screen on its own monitor.

The separate window also allows for the download of much higher resolution images of the structure, since the downloaded image has the same resolution as the 3D rendering of the structure. To download the image, right-click on the structure in the new window and choose to save the image to your computer or device.

6.3.5 Popin Viewer

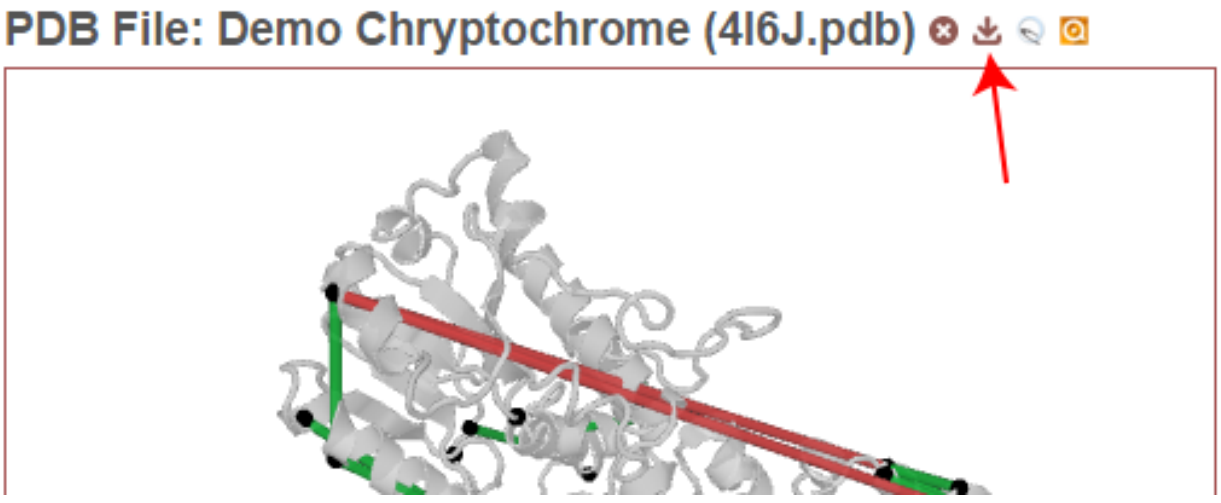
To pop the structure viewer back into the original panel, simply close the new window or click the "[Popin Viewer]" link where the "[Popout Viewer]" link was originally.

6.3.6 Delete PDB File



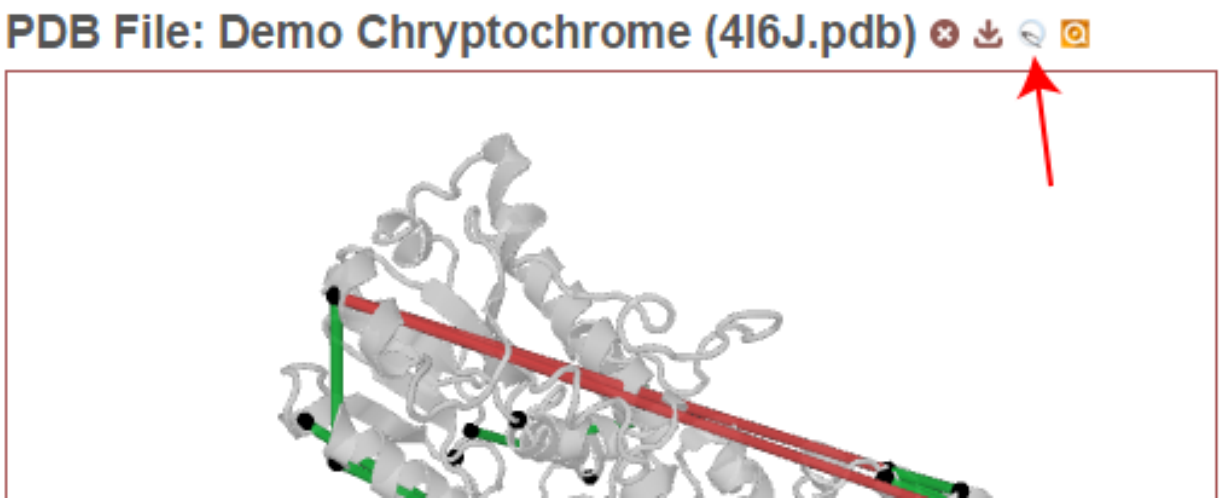
Project owners may delete a PDB file from by clicking the red (X) icon above the structure view panel. This will remove the PDB file from this project and make it unavailable to users of the project. Note, if the same PDB file has been uploaded to other projects, those will be unaffected.

6.3.7 Download PDB File

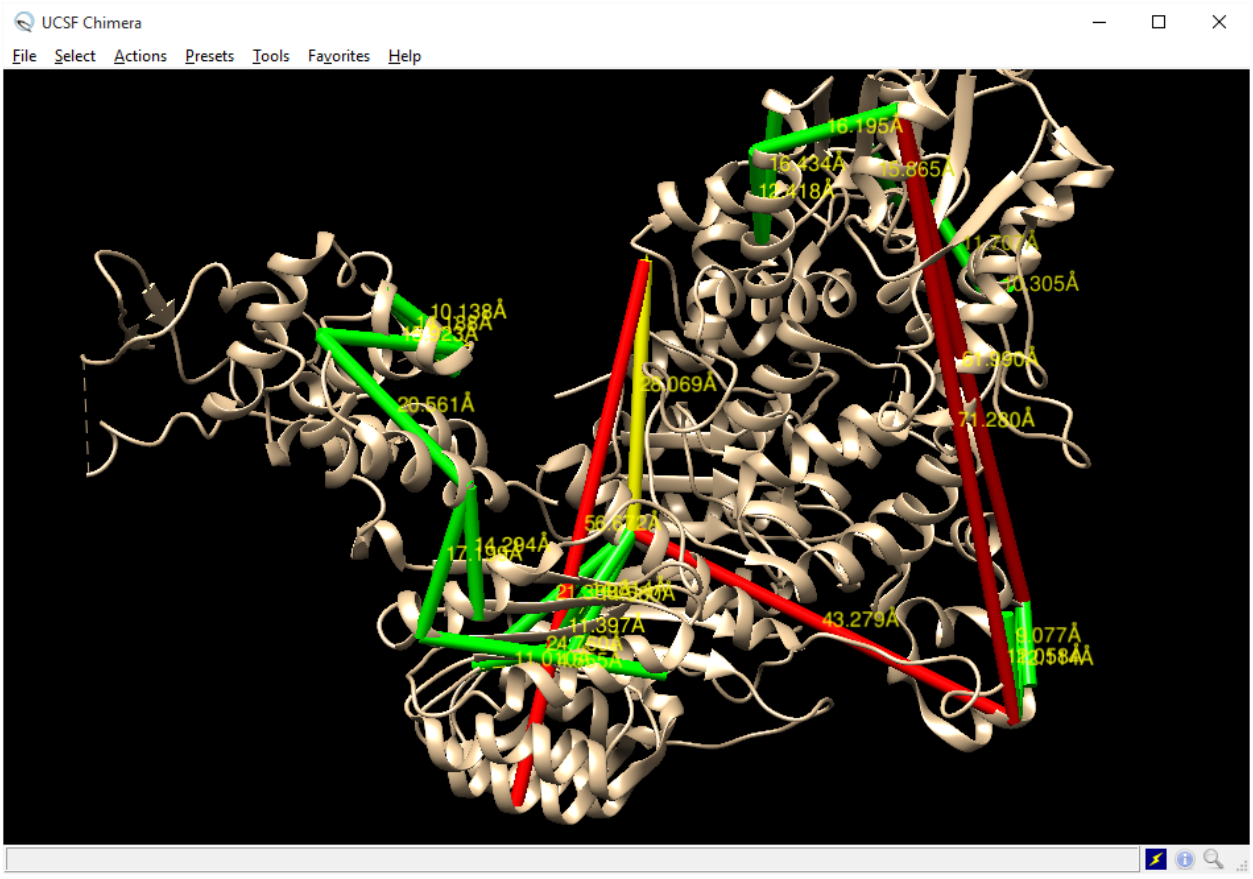


Project members (or public access users, if enabled) may download the PDB file by clicking the download icon above the structure view panel. This file is suitable for viewing in structure viewing software such as Chimera or Pymol.

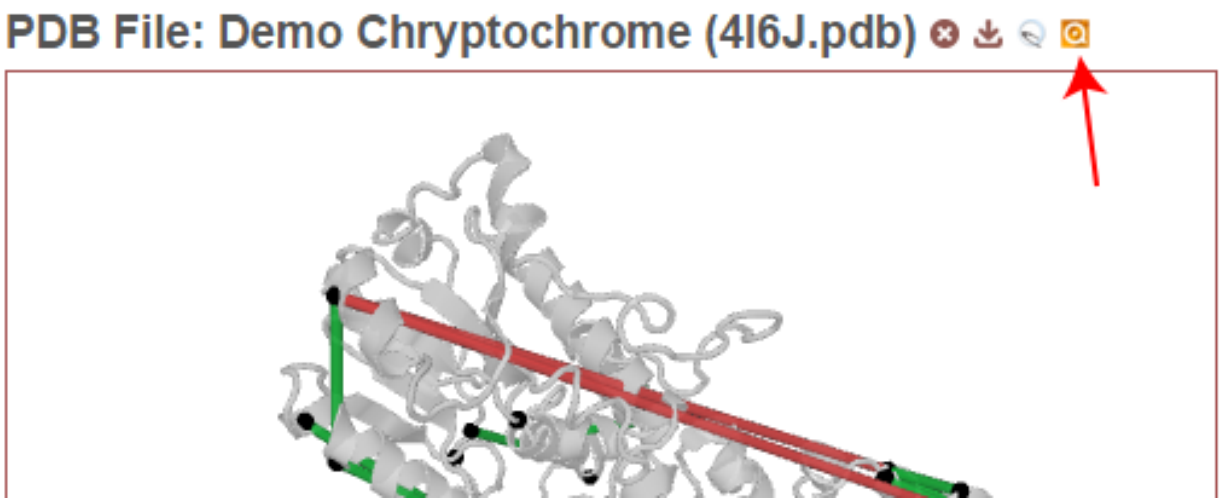
6.3.8 Download Chimera Script



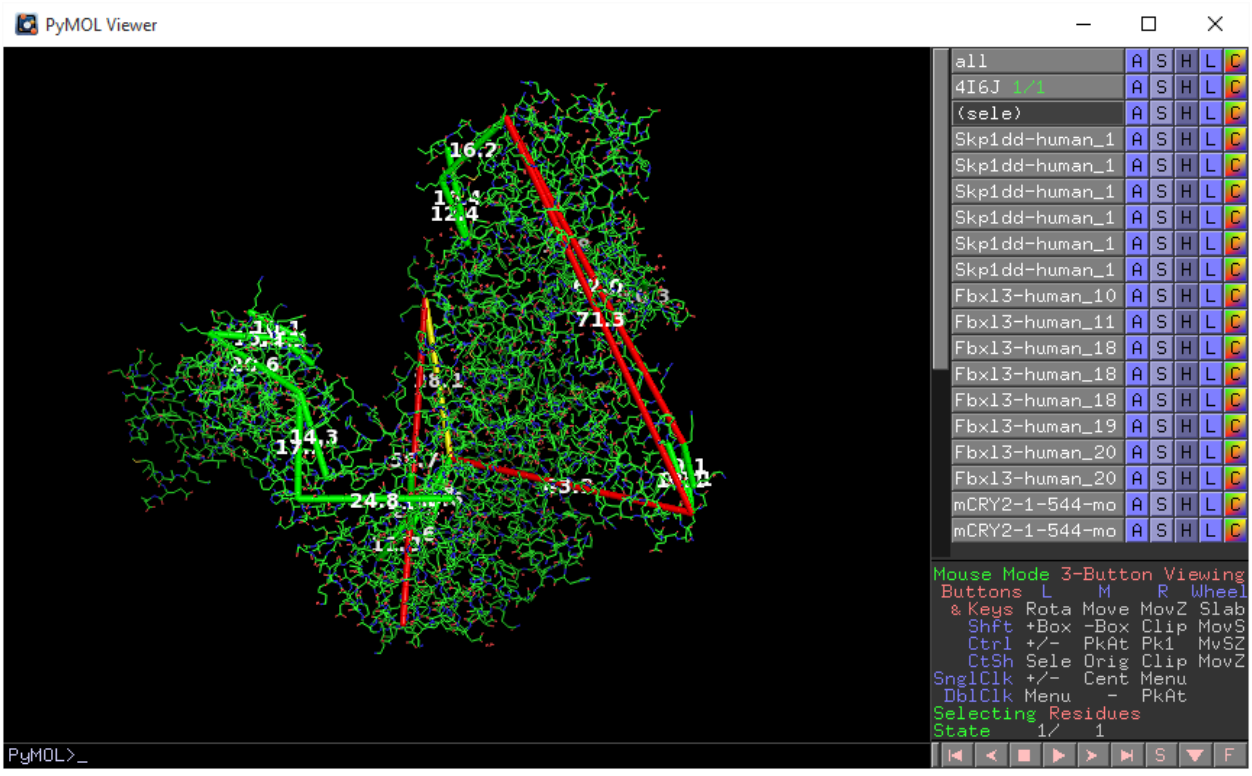
Project members (or public access users, if enabled) may download a Chimera script that will draw the currently-visible links onto the current PDB file by clicking the Chimera icon above the stucture view panel. This script may be run in Chimera after loading the PDB file, by choosing "File->Open", choosing the script that was downloaded, and choosing "Chimera Commands" as the file type.



6.3.9 Download Pymol Script



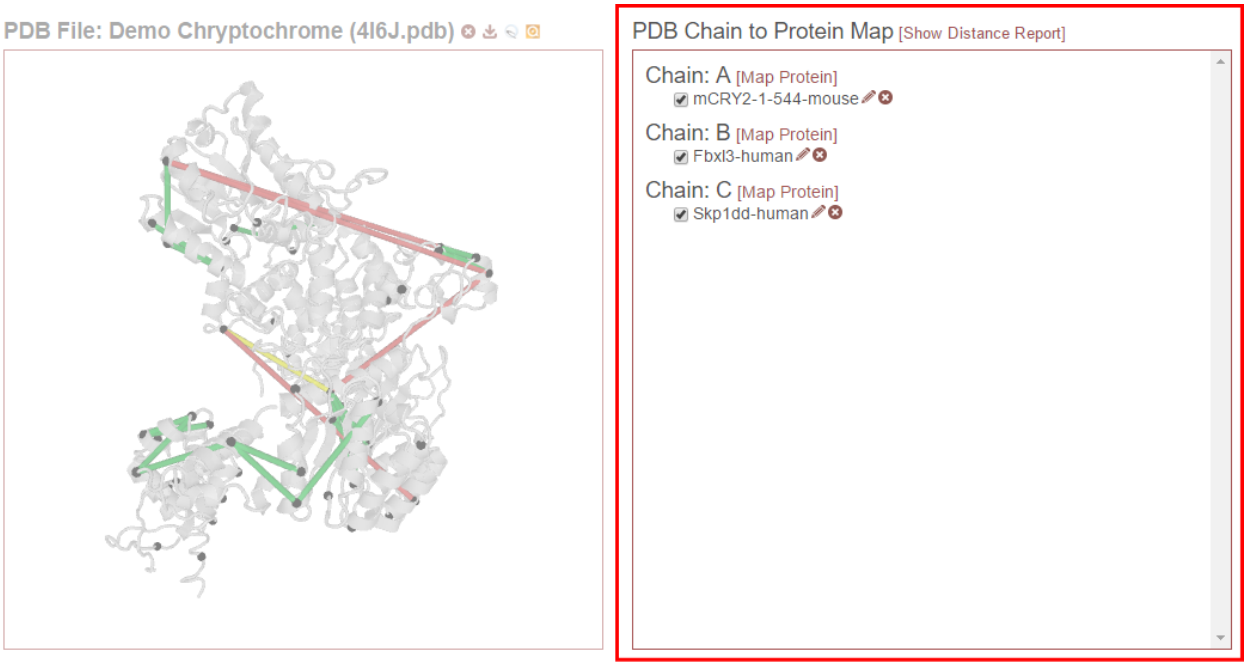
Project members (or public access users, if enabled) may download a Pymol script that will draw the currently-visible links onto the current PDB file by clicking the Pymol icon above the stucture view panel. This script may be run in Pymol after loading the PDB file by typing "@C:locationtoscript.txt" in the Pymol command line (where C:locationtoscript.txt is the actual location of the downloaded file).



6.4 PDB Chain to Protein Map

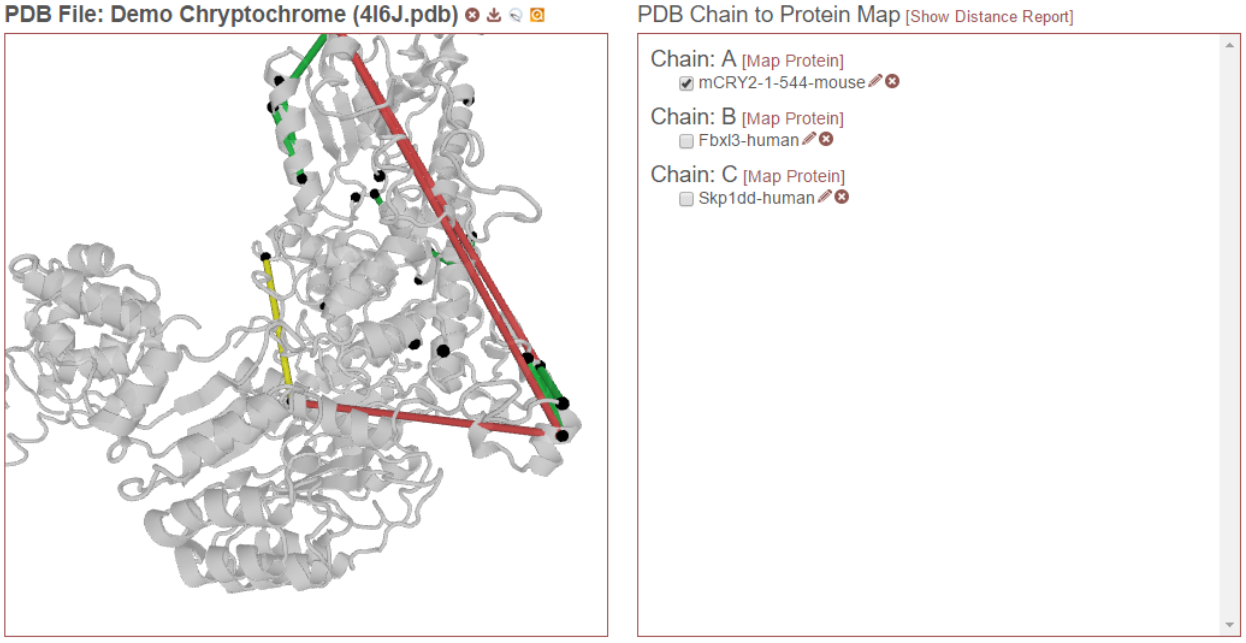
This section describes the functionality and features of the default right panel, where the proteins from the experiment that have been aligned to chains in the PDB file are displayed. (If the Distance Report is currently visible in place of

this panel, click the "[Show PDB Chain to Protein Map]" link above the report.)



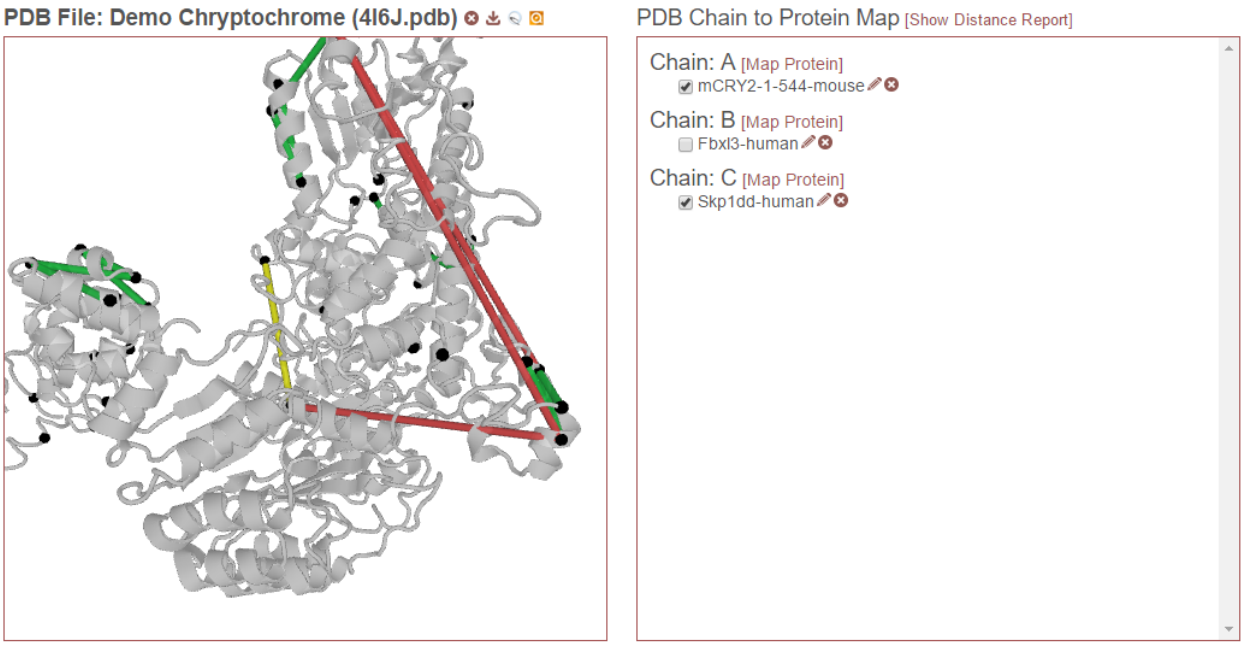
PDB Chain to Protein Map

The PDB Chain to Protein Map shows which proteins from the experiment have been mapped to which chains present in the currently-selected PDB file. The check box next to each protein indicates if that specific alignment is currently being considered when drawing links on the structure or generating the distance report. Alignments (that is a specific protein mapped to a PDB chain) that are not checked are not being considered. For example, in the figure below, only "mCRY2-1-544-mouse" is selected in Chain A. The only links drawn on the structure will only involve that protein and will only be drawn on Chain A. (If that protein were to other chains, and they were not checked, only chain A would be considered for drawing links or reporting distances.)

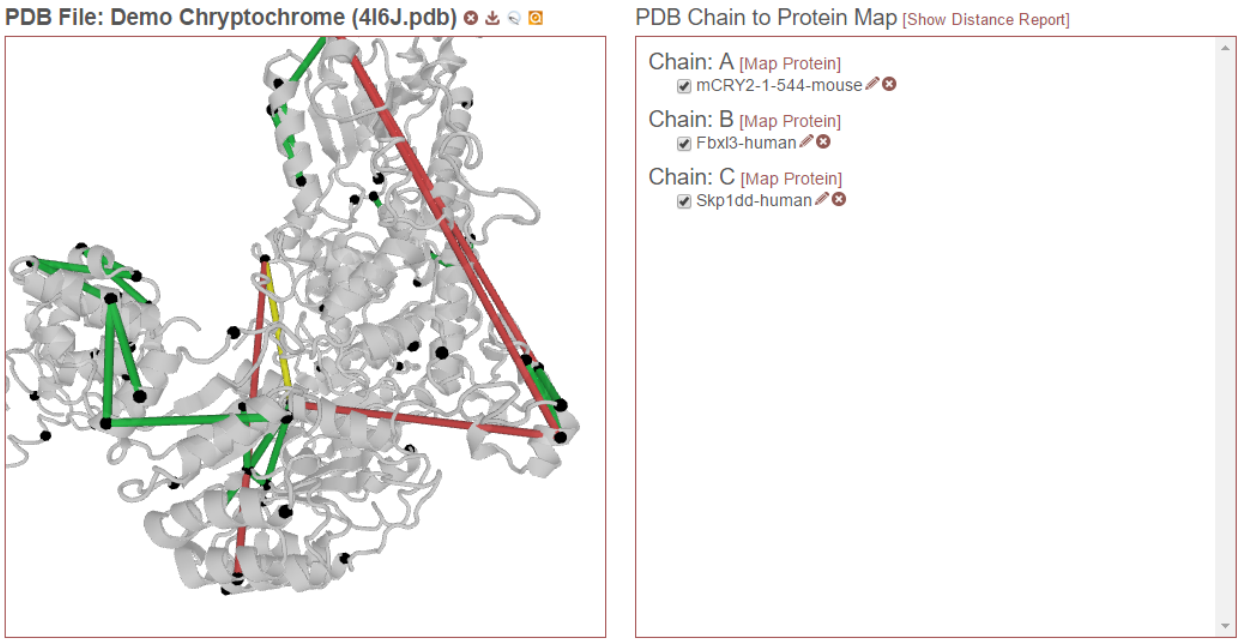


Then, in the example below, "Skp1dd-human" on chain C is also checked. Now this protein aligned to this chain will

also be considered for drawing links and reporting distances:



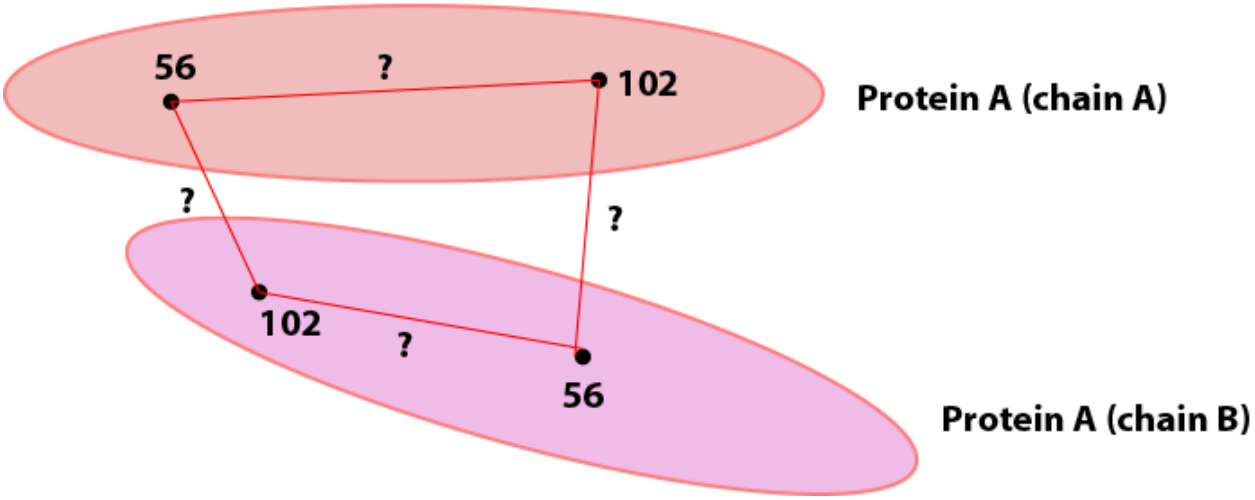
Finally, in the example below, "Fbxl3-human" on chain B is also selected. Now this alignment will also be considered:



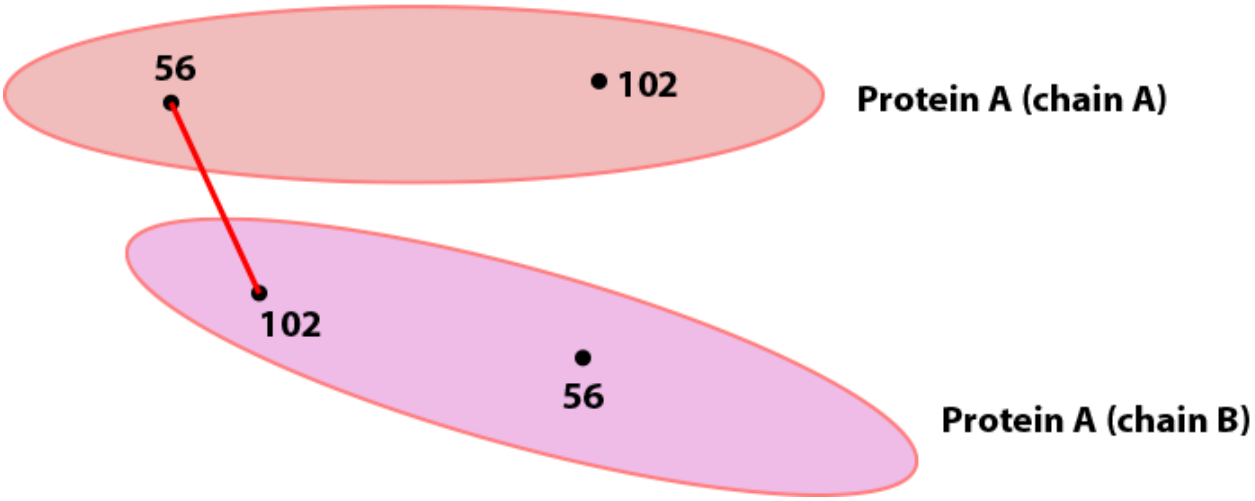
Note that it is possible for a given chain in a PDB to be aligned to multiple proteins in the experiment, such as when homologs or proteoforms were present in the proteomics search database. In this case, only one alignment for a given chain may be checked at a time.

Further note that it is also possible for the same protein to be mapped to multiple chains in the PDB, such as if the PDB depicts an oligomer. In this case, it is certainly possible to check that same protein multiple times for the different chains; however, this creates multiple possibilities for where links may be drawn on the PDB. As an example, the figure below depicts a cartoon of the same protein mapped two chains, A and B. A crosslink between positions 56 and 102 was observed. If the alignment for Protein A is checked for both chains A and B, there are four possibilities for

where to draw the crosslink (each corresponding to a different distance):



By default, proxl will pick the shortest possibility and only draw that one:



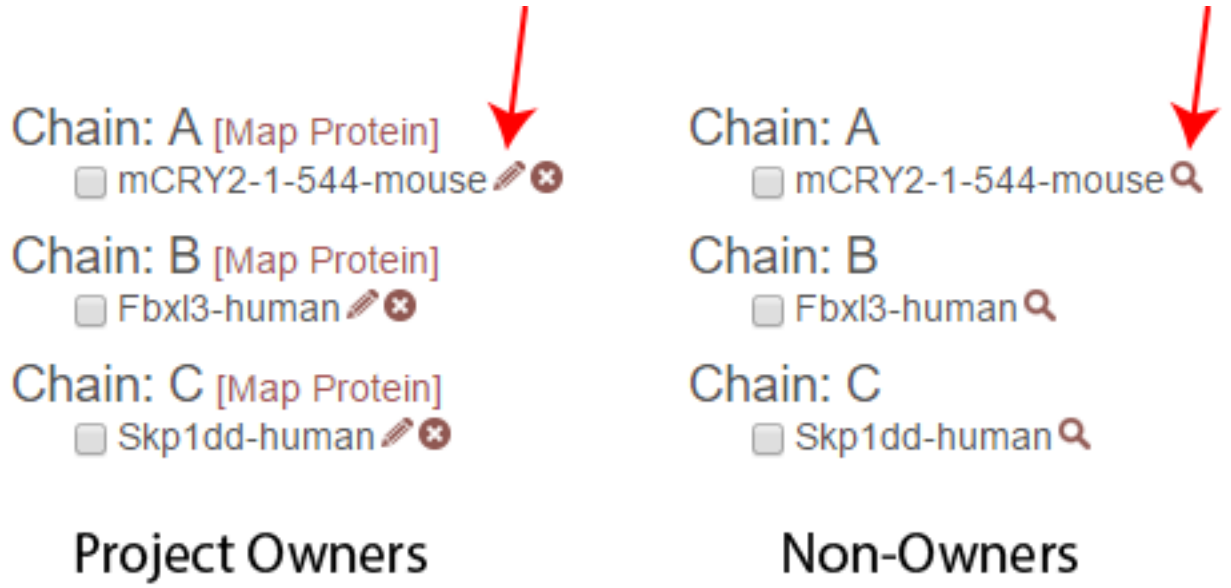
This behavior may be changed, see *Show UDRs once* for more information.

6.4.1 Add New Alignment

For a description of how to map proteins to chains in the PDB file, see *Map PDB Chains to Proteins*.

6.4.2 View Alignment

As shown below, project owners and non-owners have a different set of options associated with protein/PDB alignments:



Owners may click on the edit icon (pencil icon) to either view or manually edit the alignment. Non owners may click on the examine icon (magnifying glass icon) to view the alignment.

6.4.3 Edit Alignment

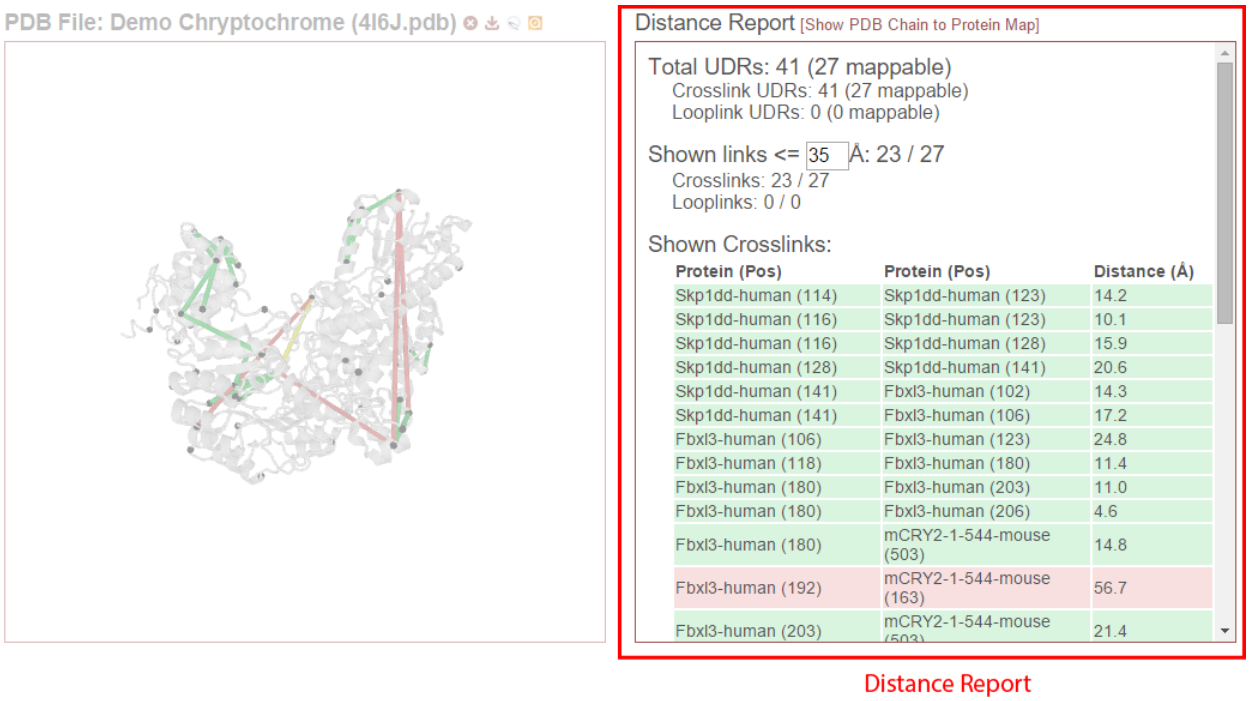
Project owners may click on the edit icon (pencil icon) next to the protein/PDB alignment they wish to view or edit.

6.4.4 Delete Alignment

Project owners may click the delete icon ((X) icon) next to the protein/PDB alignment they wish to delete. Deleting the alignment removes it from the database and makes it unavailable for all users.

6.5 Distance Report

This section describes the functionality and features of the right panel when the distance report is being displayed. To display the distance report, click the "[Show Distance Report]" link above the PDB Chain to Protein Map.



The distance report shows three primary types of data: 1) Total UDRs, or the number of observed links of each type that were mappable to the structure, 2) Shown Links, or the number and proportion of the currently-visible links that have distances less than or equal to a user-supplied distance cutoff, and 3) Shown Crosslinks/Shown Looplinks, which is a table of shown links and their distances

Links are drawn (and distances measured) between alpha carbons from PDB residues mapped to the respective residues from the experimental protein. To learn more about this mapping, see *Mapping Residue Position to 3D Space*.

6.5.1 Total UDRs

UDR stands for "unique distance restraint." In proxl, this is the name for a distinct position in a protein linked to a distinct position in the same or another protein. This section of the distance report reports the total number of such UDRs present in the data, and how many of those are mappable to the structure. To learn more about what "mappable" means or about how proxl maps protein positions to PDB structure, see *Mapping Residue Position to 3D Space*.

6.5.2 Shown links

This section shows the number of UDRs (that were mappable to the structure) have distances less than or equal to the supplied distance cutoff. (This defaults to 35 Angstroms.) This cutoff may be changed simply by changing its value in the text field–the report will dynamic update as the value is changed.

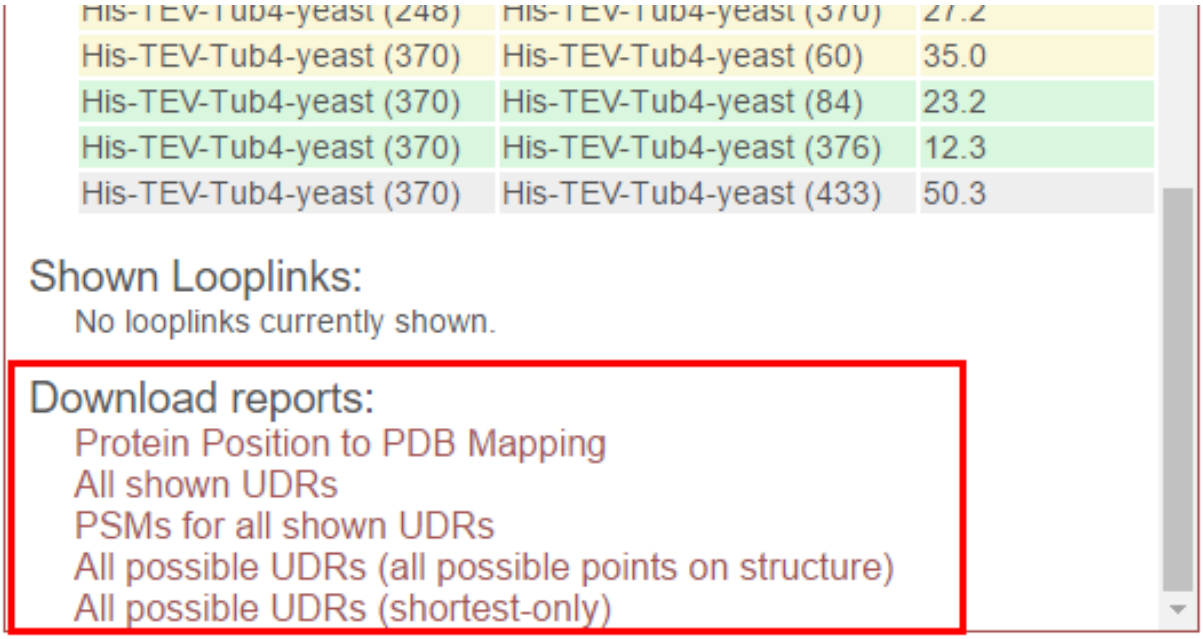
6.5.3 Shown Crosslinks / Shown Looplinks

These tables list the crosslinks and/or looplinks currently being shown on the structure. The coloring of the rows matches the coloring of those links on the structure–even if an alternate coloring scheme is chosen (see *Color links by*). Each of the rows shows the positions in respective proteins that are linked, as well as the distance of that link in Angstroms. All of the currently drawn crosslinks and/or looplinks will be listed in the table.

Each row may be clicked on (or tapped on touch devices) to view the underlying peptides, PSMs, and spectra. See *Click on Links* for more information.

6.5.4 Download Reports

At the bottom of the "Distance Report" panel is "Download reports". Click on the name of the report to initiate a download.



Available reports are:

- Protein Position to PDB Mapping - A tab-delimited file showing the pair-wise mapping of the protein sequences in the experiment to the PDB file.
- All shown UDRs - A tab-delimited report of all currently-displayed UDRs, showing protein positions, distances, and PSM/peptide scoring statistics.
- PSMs for all shown UDRs - A tab-delimited report of all PSMs for all currently-displayed UDRs, including scoring statistics for those PSMs.
- All possible UDRs (all possible points on structure) - A tab-delimited report of all theoretically possible cross-links on the structure, given the cross-linker that was used.
- All possible UDRs (shortest-only) - A tab-delimited report of all possible cross-links on the structure, given the cross-linker that was used. Only the shortest cross-link for each unique protein pair and position is reported. This is useful when proteins appear more than once in the structure.

6.6 Viewer Options

6.6.1 Show crosslinks

Toggle the showing of crosslinks on and off.

6.6.2 Show looplinks

Toggle the showing of looplinks on and off.

6.6.3 Show monolinks

Toggle the showing of monolinks on and off.

6.6.4 Show linkable positions

Toggle the showing of linkable positions. Linkable positions appear as black spheres on the alpha carbons of residues in the PDB that mapped to linkable positions in the experimental protein. A linkable position is a residue expected to be theoretically linkable by the crosslinker used in the experiment.

6.6.5 Show UDRs once

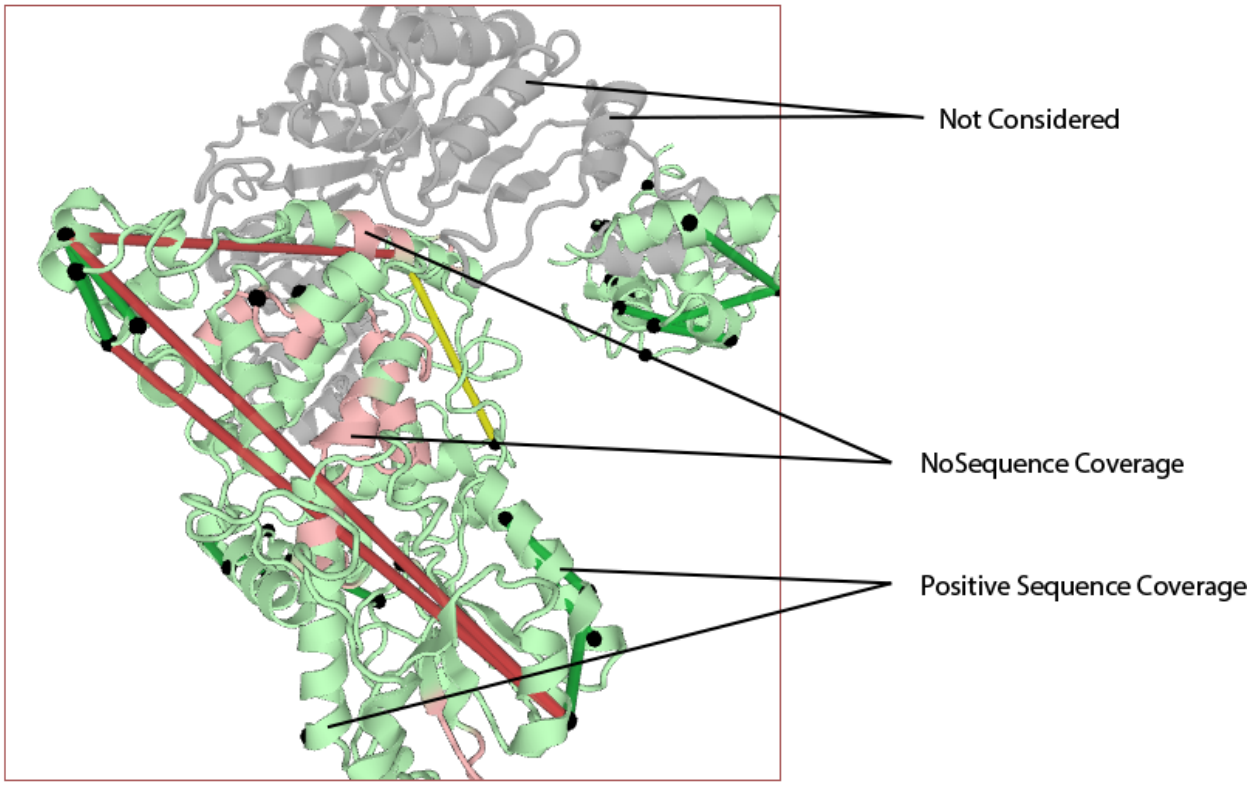
Maybe I should delete this feature.

6.6.6 Shade by counts

If enabled, the opacity (transparency) of links reflects the number of PSMs found (or spectrum count) for the shown link. The shading scales from 1 PSM (minimum opacity, most transparent) to 10 PSMs (maximum opacity). Any link having 10 or more PSMs will have the maximum opacity.

6.6.7 Show sequence coverage

If enabled, the color of the structure, itself, will reflect the sequence coverage for the protein mapped to the respective chain. Segments of the structure will be shaded green where those residues map to experimental protein residues that were contained in any observed peptide (at the given cutoffs). Segments will be shaded red where no peptides were observed that contained those residues. Areas of the structure not being considered (either residues that do not map to experimental protein residues orno protein/PDB aligments are enabled for that chain) will appear as white.



6.6.8 Color links by

This pull-down menu alters the coloring applied to the links. The three options are 1) Length (default), 2) Type, and 3) Search.

Length (default)

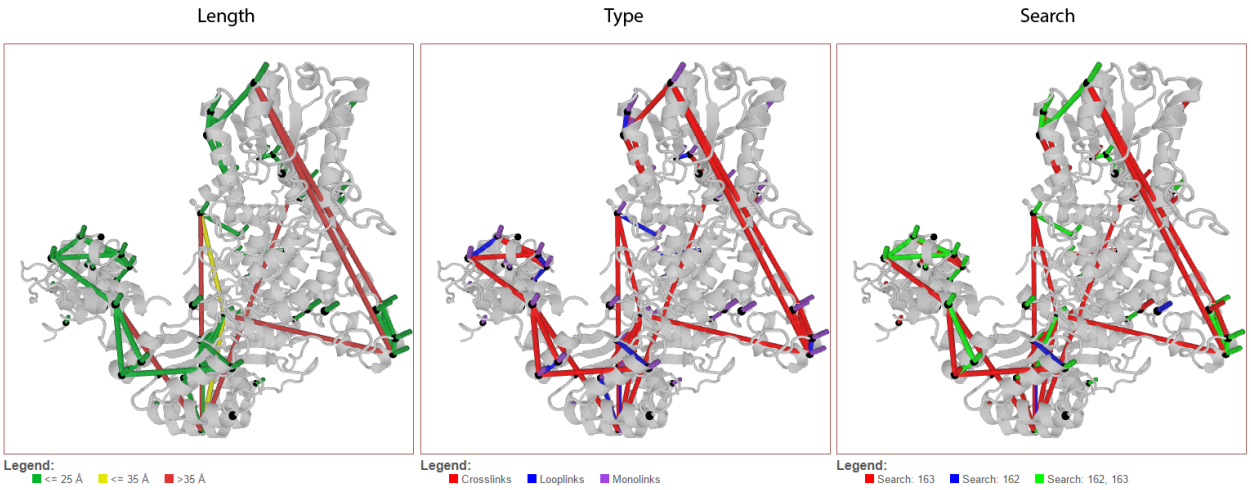
This option colors links based on their length.

Type

This option colors links based on their type–crosslink, looplink, or monolink.

Search

When merging multiple searches (maximum of 3), this option colors links based on which search(es) they were found in. This allows for a quick, structure-based comparison of results between searches.



6.6.9 Render mode

Provides access to different rendering modes for the structure, itself. The options are:

Cartoon (default)

Provides a helix, strand coil cartoon view of the structure.

Smooth Line

Draws the structure as a smoothed line tracing consecutive alpha carbons..

Trace

Draws the structure as straight lines connecting consecutive alpha carbons.

Lines

Draws the struture at atomic resolution using lines for bonds.

Points

Draws the structure as point cloud of all atoms.

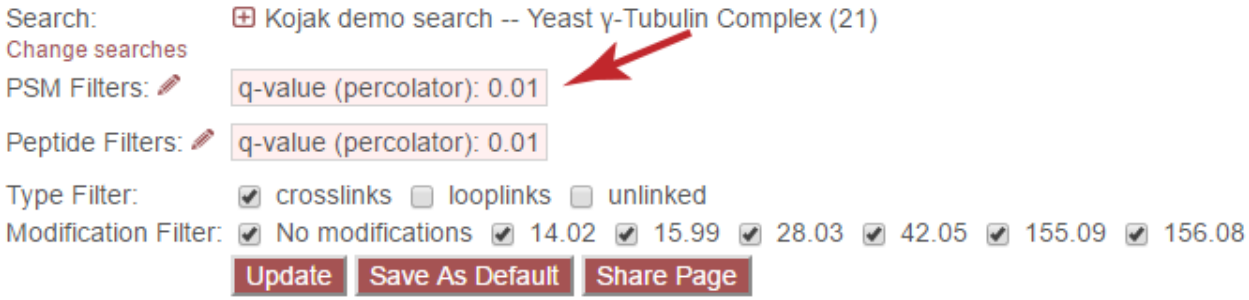
6.7 Filter Data

The data presented in the viewer may be filtered using the form at the top of the page and clicking “Update From Database”. The filtering options are:

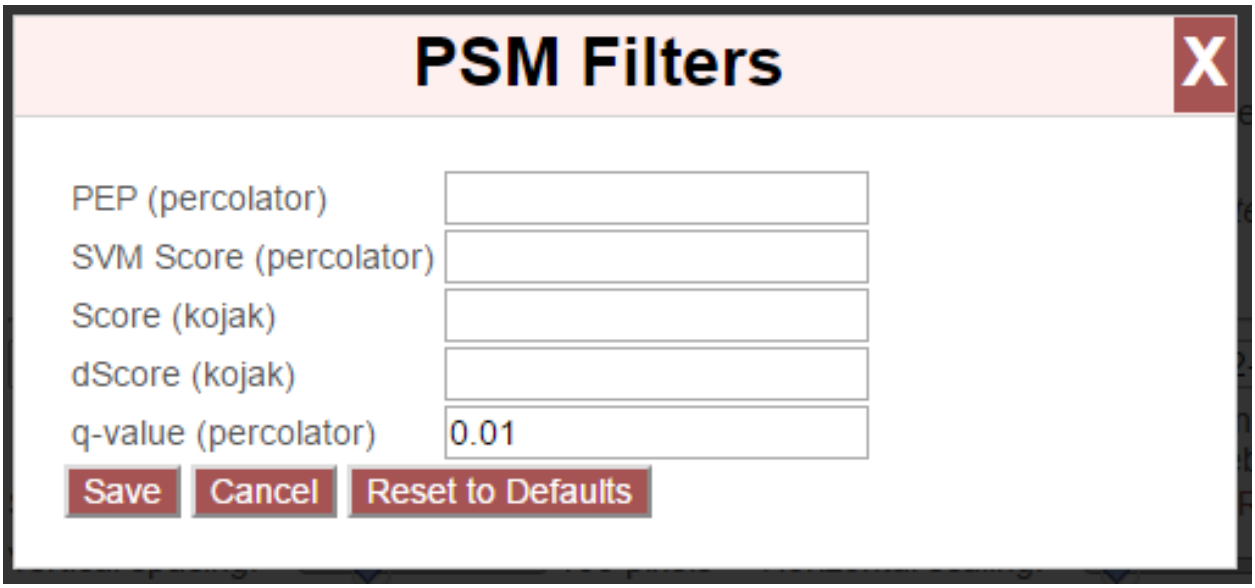
6.7.1 PSM Filters

The filters to apply at the PSM level. Only results which have at least one PSM that meets all of the selected critiera will be listed. When listing PSMs associated with peptides, only PSMs that meet all of the selected critiera will be listed.

To change the PSM-level filters, first click the pencil icon next to "PSM Filters":



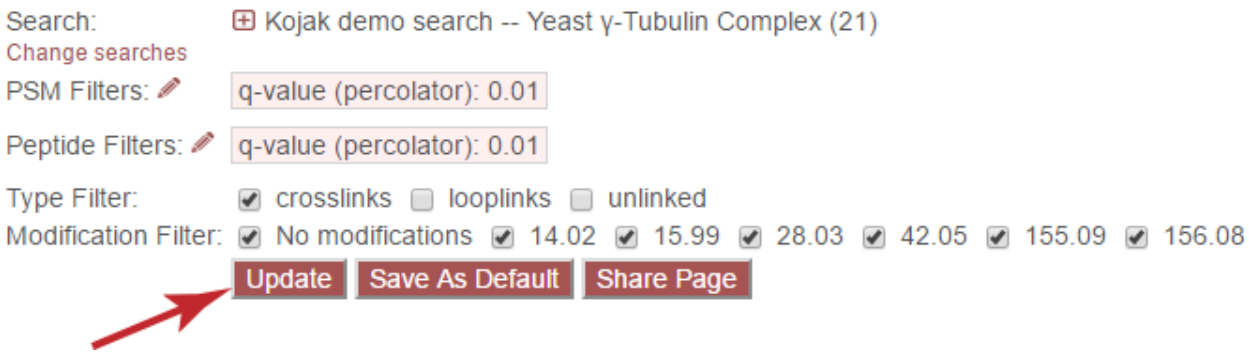
This opens an overlay with the containing the possible score types to use as PSM filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

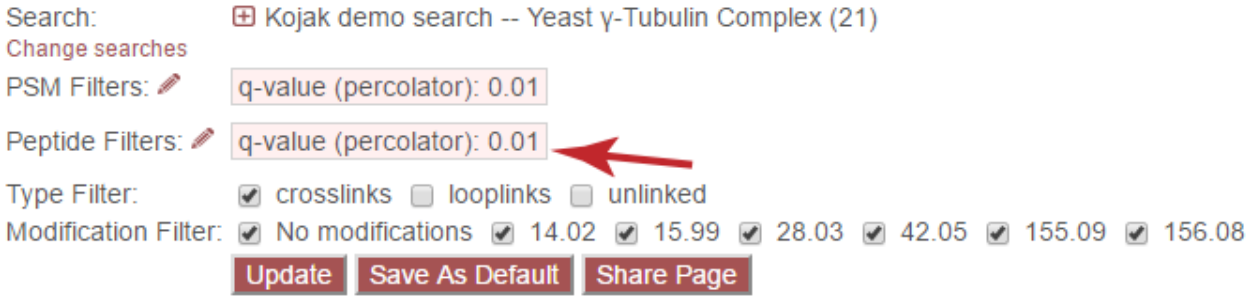
Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update From Database" button on the page to apply any new PSM- or peptide-level filters.



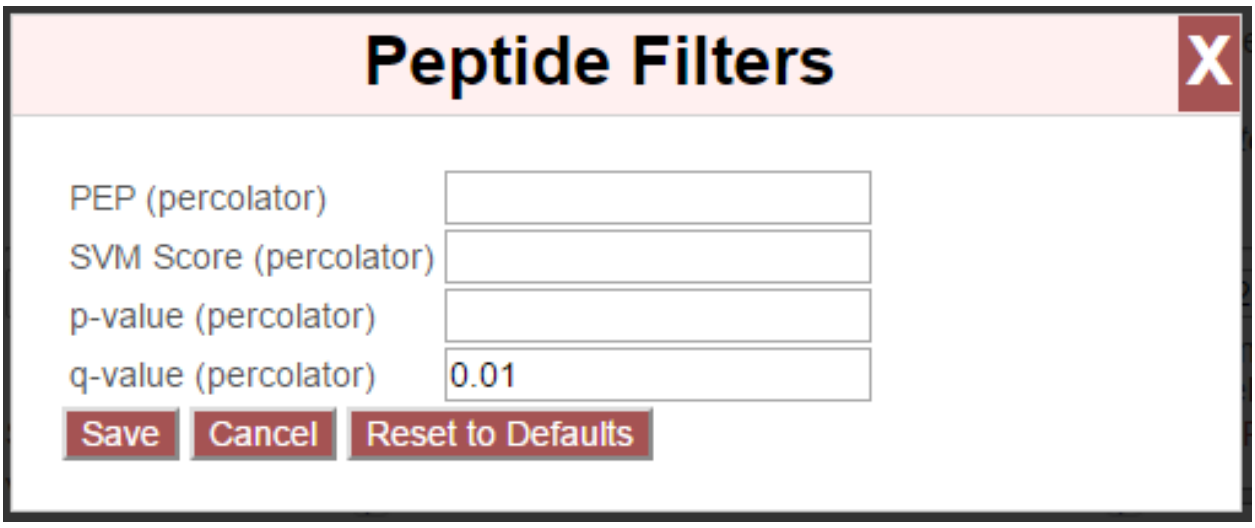
6.7.2 Peptide Filters

The filters to apply at the peptide level. Only results which have at least one peptide that meets all of the selected critiera will be listed.

To change the peptide-level filters, first click the pencil icon next to "Peptide Filters":



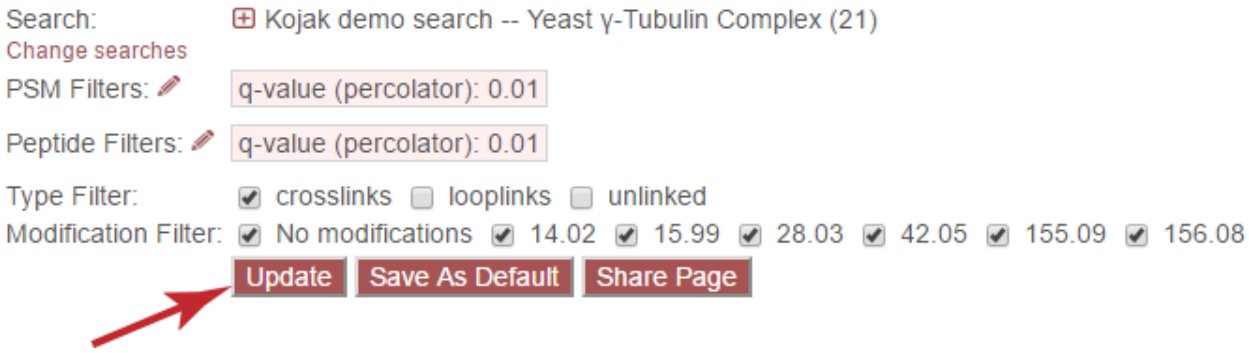
This opens an overlay with the containing the possible score types to use as peptide-level filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update From Database" button on the page to apply any new PSM- or peptide-level filters.



6.7.3 Exclude links with

Links that exclusively have peptides with any of the checked attributes will not be shown. The attributes are:

- no unique peptides - If the link (crosslink, looplink, or monolink) was exclusively identified by peptides that also map to othe proteins
- only one PSM - If a given link was identified by a single PSM
- only one peptide - If a given link was identifed by a single peptide, where a peptide is the combination of sequence, linked positions, and modifications

6.7.4 Exclude proteins with

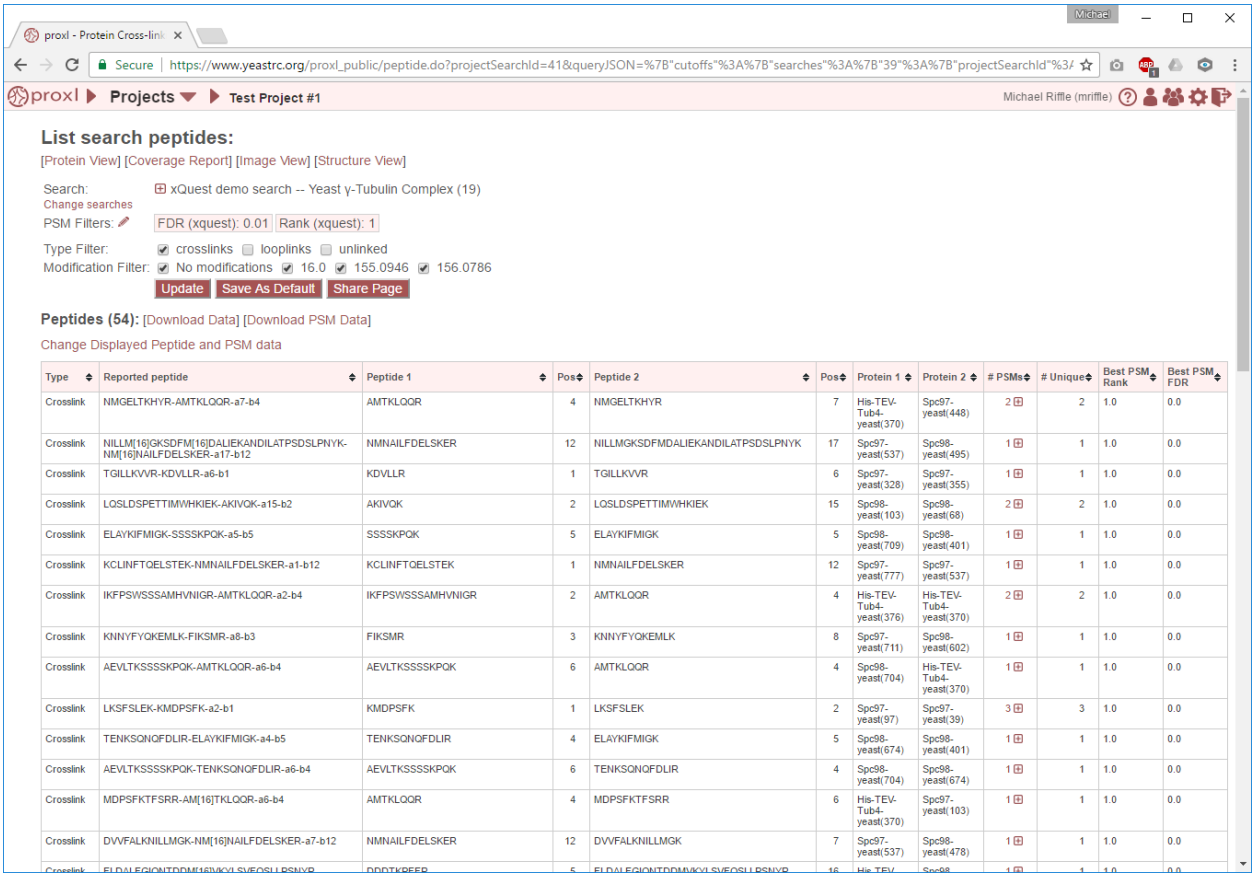
This option limits which proteins will appear in the pull-down menu for mapping proteins to PDB chains. Proteins that contain any of the checked options will not appear. For example, checking 'No links' prevents proteins that do not contain crosslinks, looplinks, or monolinks from appearing. This will also prevent any links involving excluded proteins from being drawn on the structure.

6.7.5 Exclude organisms

This options limits which proteins will appear in the pull-down menu for mapping proteins to PDB chains. No proteins from any of the checked organisms will appear. This will also prevent any links involving excluded proteins from being drawn on the structure.

CHAPTER 7

Peptides View Page



The peptide view page provides a table view of the data at the peptide level. That is, all peptides (and crosslinked peptide pairs) identified by the search software may be viewed on this page–along with accompanying peptide spectrum matches (PSMs) and tandem mass spectra. The data presented may be filtered according to confidence, type of peptide, and which modifications are present on the peptide. Note, this document covers the peptide view page for a

single search. For the view page seen when merging multiple searches, see *Merged Peptides View Page*.

7.1 Search Information

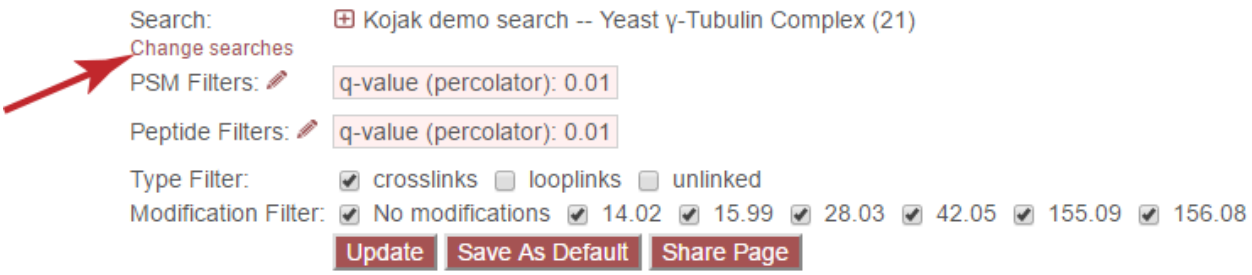
The name of the search (and internal search ID reference number) from which these data were obtained is shown first. The red [+] icon may be clicked to reveal details about the search.

7.1.1 Search Details

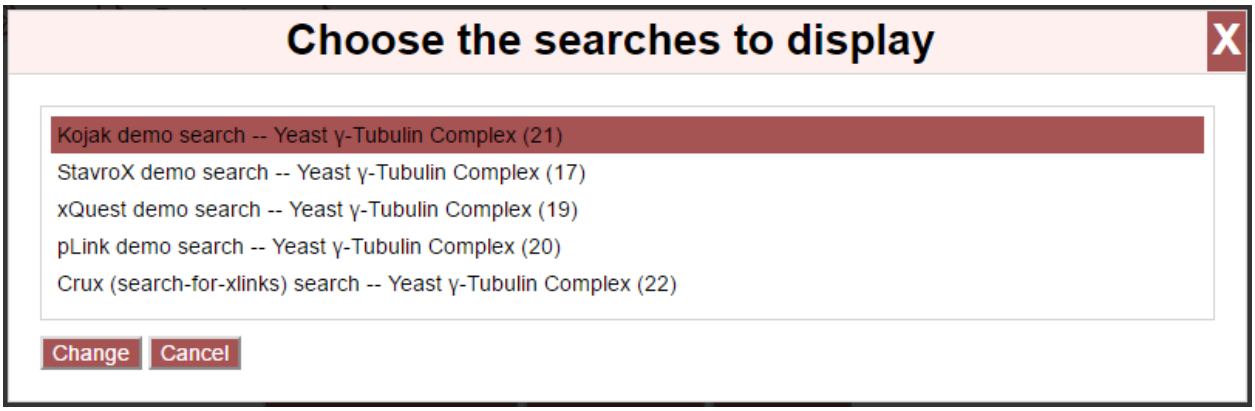
The "Path" is the location on disk from which the data were imported. The "Linker" is the name of the crossinker used in the experiment. "Search Program(s)" is the name and version number of the PSM search software used. "Upload date" is the date the data were uploaded into proxl. "FASTA file" is the name of the FASTA file used to perform the PSM search.

7.2 General Options

7.2.1 Change Searches

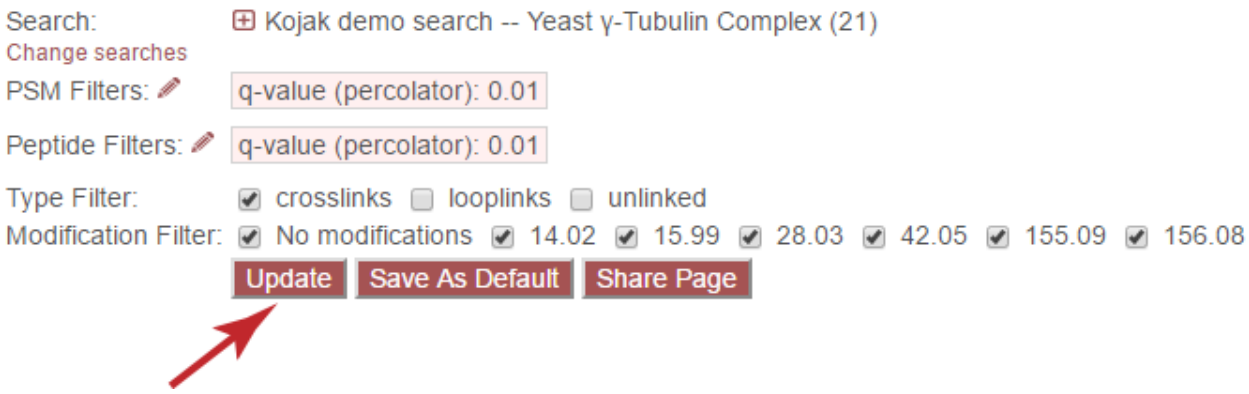


The "Change searches" link allows the user to change which searches are currently being displayed. Clicking the link causes the following overlay to be displayed:



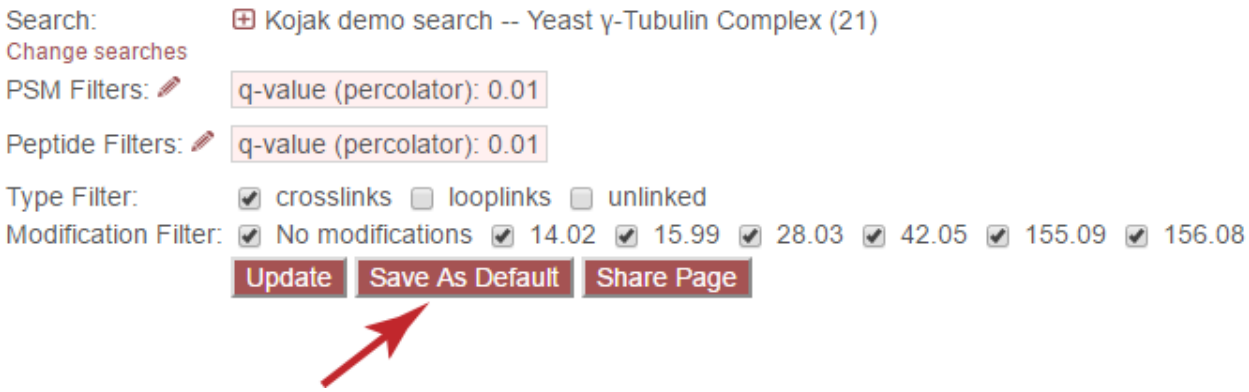
Select or de-select searches by clicking on them in the list. Once done, click "Change" to update the page with the new data or "Cancel" to close the overlay.

7.2.2 Update From Database



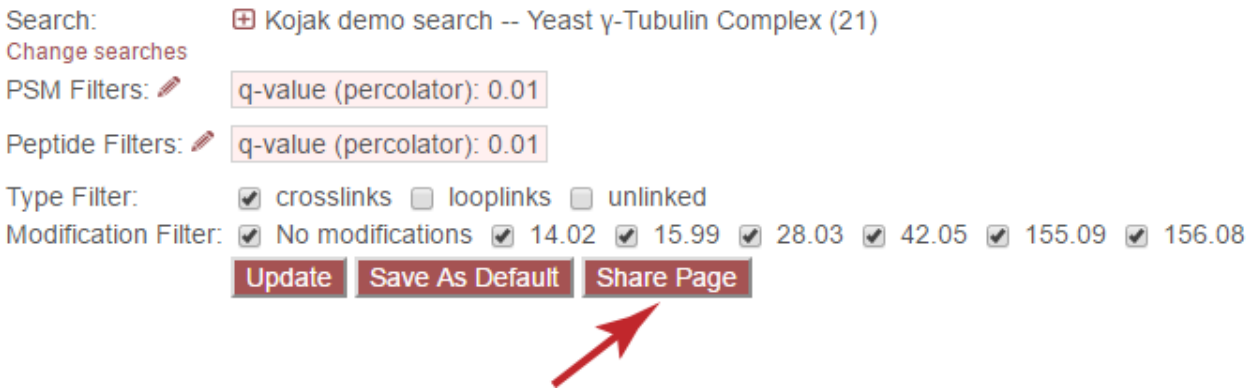
If the user changes any filter parameters–such as PSM/peptide score cutoffs–this button must be clicked to reflect the new filter choices.

7.2.3 Save as Default



Project owners may click "Save as Default" to save the current URL as the default view of the "Peptide View" for this search. This default view will be populated with the same options as when the button is clicked. This is a convenient way to share data with collaborators or the public that does not require that they manipulate the image viewer to see the data.

7.2.4 Share Page



Clicking the "Share Page" button will generate a shortcut URL for viewing the current page. The shortened URL will appear in an overlay as:

Copying and sharing the highlighted URL will direct users to the view of the page when the URL was generated. Note that this URL does not grant access to the page to any user that would not otherwise have access.

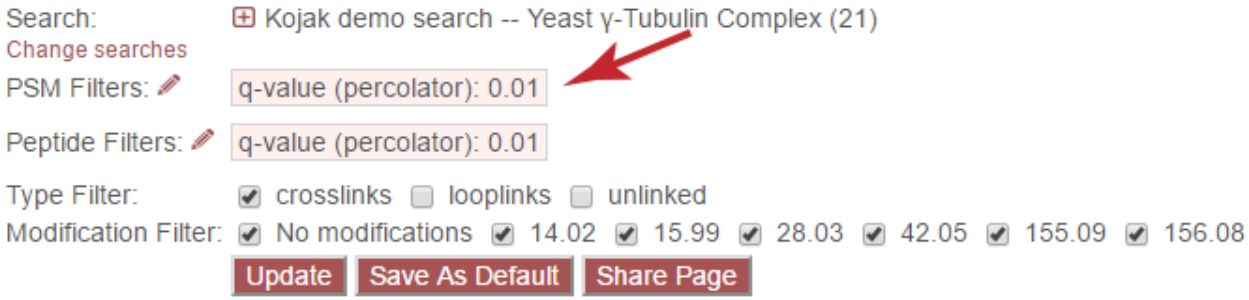
7.3 Filter Data

The data presented may be filtered according to the following criteria. Note: Only peptides that meet ALL of the specified criteria are returned.

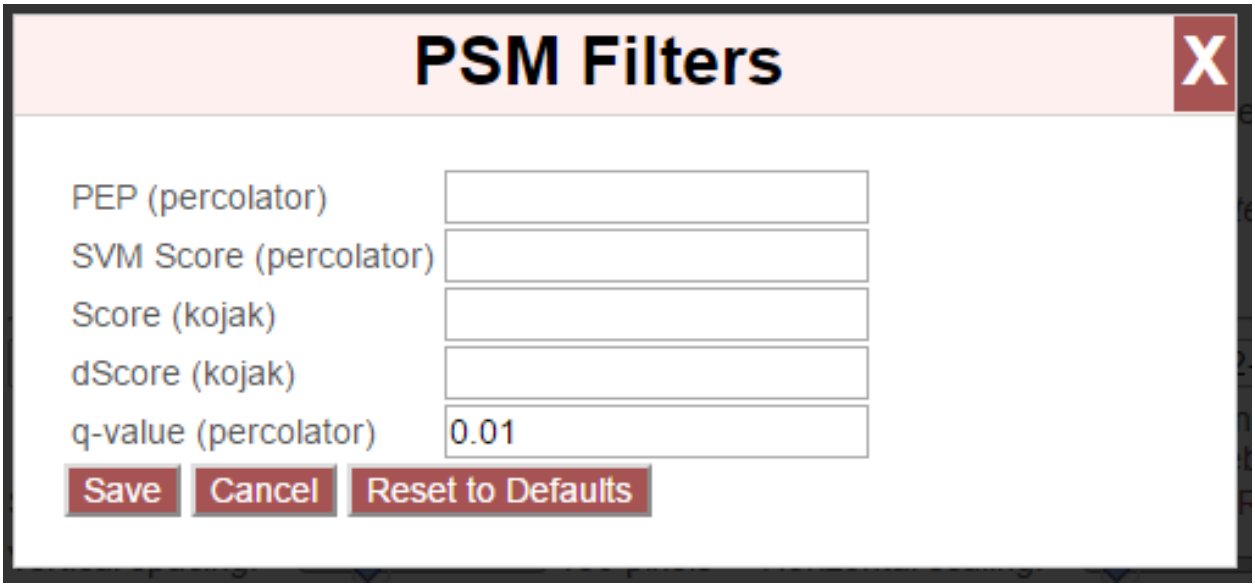
7.3.1 PSM Filters

The filters to apply at the PSM level. Only results which have at least one PSM that meets all of the selected critiera will be listed. When listing PSMs associated with peptides, only PSMs that meet all of the selected critiera will be listed.

To change the PSM-level filters, first click the pencil icon next to "PSM Filters":



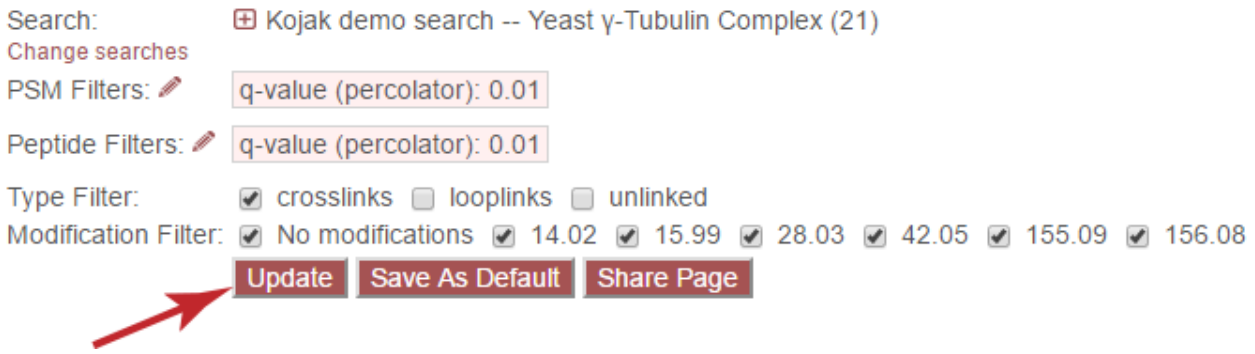
This opens an overlay with the containing the possible score types to use as PSM filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

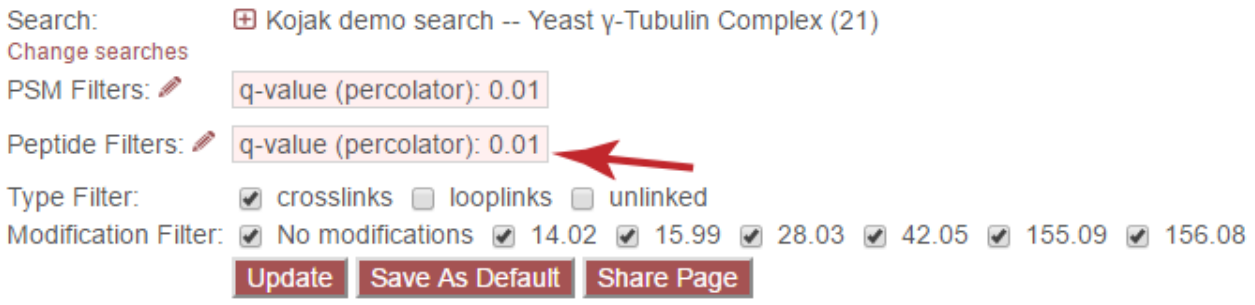
Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update" button on the page to apply any new PSM- or peptide-level filters.



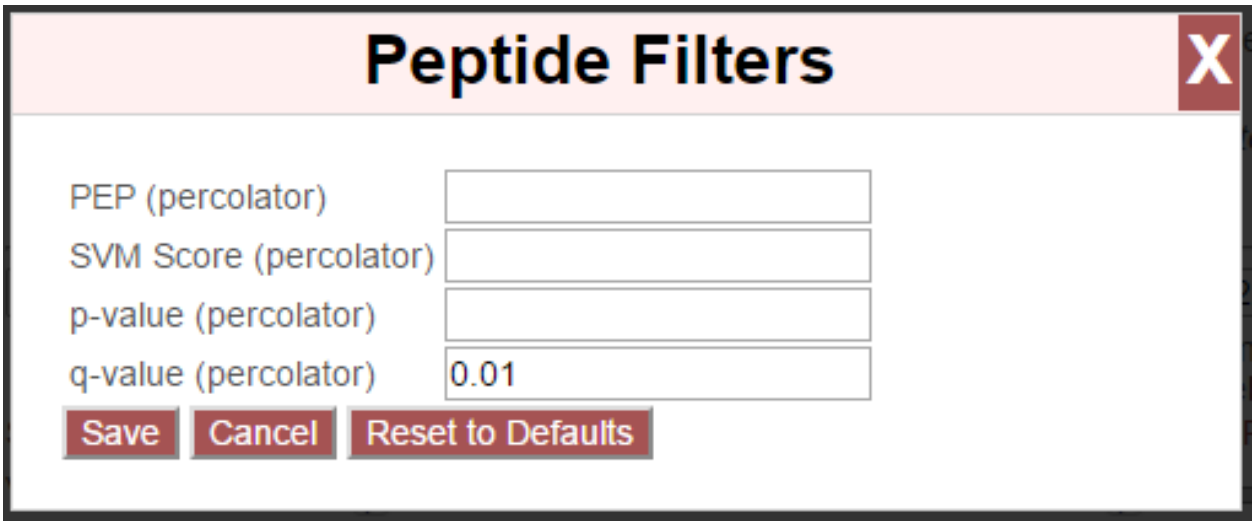
7.3.2 Peptide Filters

The filters to apply at the peptide level. Only results which have at least one peptide that meets all of the selected critiera will be listed.

To change the peptide-level filters, first click the pencil icon next to "Peptide Filters":



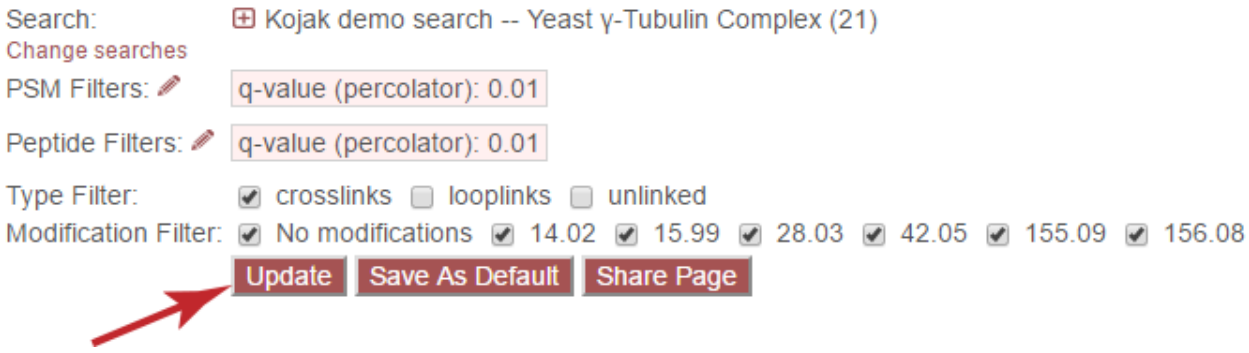
This opens an overlay with the containing the possible score types to use as peptide-level filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update" button on the page to apply any new PSM- or peptide-level filters.



7.3.3 Type filter

Only peptides of the checked type(s) will be returned. Proxl defines the types as:

- crosslink - A pair of peptides linked by a crosslinker.
- looplink - A single peptide with two residues linked by a crosslinker.
- unlinked - The peptide without a crosslinker on any residue.

Checking multiple boxes will include any peptide that has at least one of the checked types. I.e., checking 'crosslinks' and 'looplinks' will only include peptides that are either crosslinks or looplinks. Only checking 'crosslinks' will only return crosslinked peptides.

7.3.4 Modification filter

Only peptides with at least one of the checked modifications will be included. Note that monolinks are considered modifications of residues in proxl, so the mass of the crosslinker when found on monolinks is included here.

7.3.5 Update

In order to apply new filter parameters to the shown data, the "Update" button must be clicked. This will fetch filtered data from the proxl server and display the data on the web page.

7.3.6 Save As Default

Project owners may save the current filter parameters as the default view of the data on this page by clicking this button. This default view will be shown when users follow links to the "Peptide View" for this search.

7.4 Table Description

Above the table is the text, "Peptides (#)", where # is the number of distinct reported peptides were found for this search. A distinct peptide is the combination of peptide sequence(s), linked positions in those peptides, and the location and type of post-translational modifications.

7.4.1 Columns

The columns are described below. Note that all column headers may be clicked to toggle between ascending and descending sorting of that column. Holding the shift key while clicking column headers allow sorting on multiple columns.

Type

The type of peptide (crosslink, looplink, or unlinked).

Reported peptide

The peptide as it was reported by the search program used.

Peptide 1

The parsed sequence of the peptide (or the first peptide in the case of crosslinks).

Pos

The position in that peptide containing the linker.

Peptide 2

The parse sequence of the second peptide in the crosslink.

Pos

The position in that peptide containing the linker.

Protein 1

The protein(s) to which the first peptide matches, and the position in that protein to which the linker position in that peptide matched. Mouse-over the protein name to get a description.

Protein 2

The protein(s) to which the second peptide matches, and the position in that protein to which the linker position in that peptide matched. Mouse-over the protein name to get a description.

Peptide-level Scores

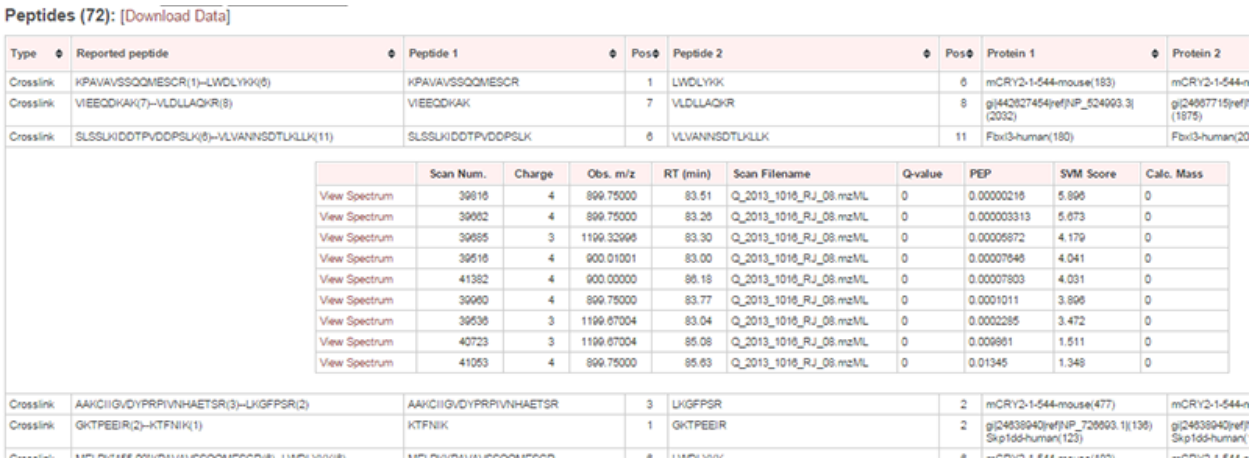
If peptide-level scores are available for this search, the scores will appear as separate columns.

Best PSM-level Scores

Columns will appear for each PSM-level score on which the results are currently being filtered. Each of these columns will show the best PSM-level score for each respective PSM-level filters. E.g., if p-value is being used as a PSM-level score, the best PSM p-value will be displayed for each peptide.

7.5 View PSMs

All PSMs meeting the current filtering criteria that map to a given peptide can by shown by clicking on the table row containing that peptide.



7.5.1 Columns

The PSMs appear in a table with the following columns:

Scan Num.

The scan number from the spectral file (e.g., mzML file)

Charge

The predicted charge state of the precursor ion.

Obs. m/z

The observed m/z of the precursor ion.

RT (min)

The retention time in minutes.

Scan Filename

The filename of the scan file.

PSM-level scores

Each PSM-level score will appear as a separate column.

7.5.2 View Spectra

The annotated mass spectrum may be viewed for any PSM by clicking the "View Spectrum" link. For help on our spectrum viewer, see the *Spectrum Viewer* page.

7.6 Sort Data

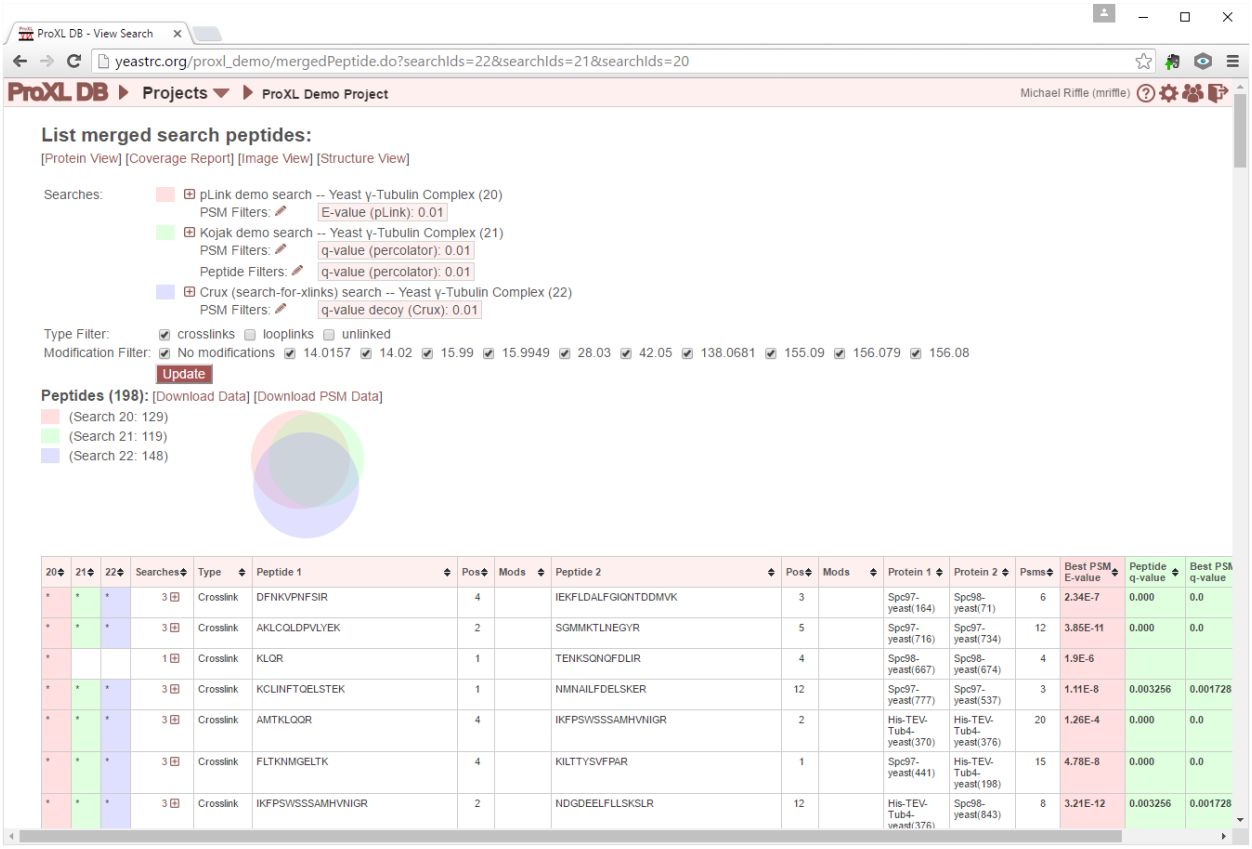
All column headers may be clicked to toggle between ascending and descending sorting of that column. Holding the shift key while clicking column headers allow sorting on multiple columns.

7.7 Download Data

Clicking the [Download Data] link will download the shown data as a tab-delimited text file.

CHAPTER 8

Merged Peptides View Page



To reach this page, select multiple searches on the project page and click "View Merged Peptides". (See *Project Overview Page*.) This page combines and collates the data from multiple searches and presents the results as an interactive table. The searches do not need to be from the same software pipeline. For example, different versions of the same program may be compared, or the results from entirely different programs (e.g., Kojak and XQuest) may be compared. Currently, the total number of merged searches must be 2 or 3 and must be from the same project. Note,

for the peptide view page seen when viewing a single search, see *Peptides View Page*.

8.1 Search List

The list of merged searches is presented below the top navigation. Each search is shown next to its assigned color for the page, and the color referencing this search is retained in the Euler diagram and in the peptide table. Clicking the [+] icon will expand that search to view details:

8.1.1 Search Details

The "Path" is the location on disk from which the data were imported. The "Linker" is the name of the crossinker used in the experiment. "Search Program(s)" is the name and version number of the PSM search software used. "Upload date" is the date the data were uploaded into proxl. "FASTA file" is the name of the FASTA file used to perform the PSM search.

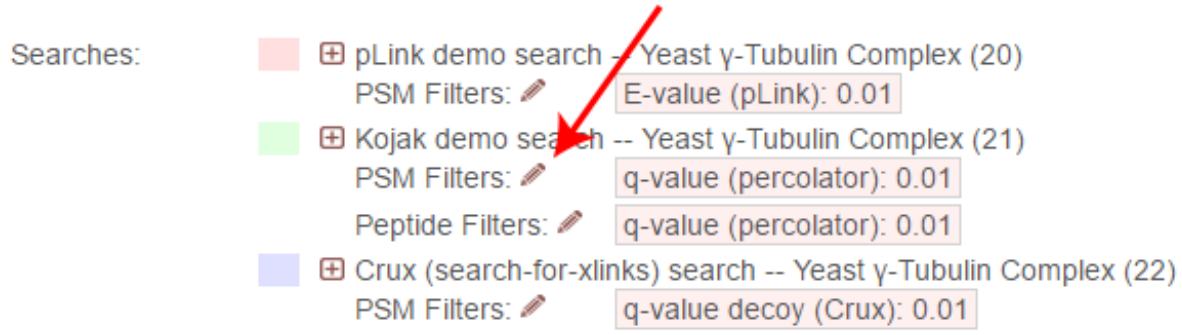
8.1.2 Search Filter

Each search is filtered separately, according to its own native score types. To change the filters for each search, click the pencil icon next to "PSM Filters:" or "Peptide Filters:" next to each search.

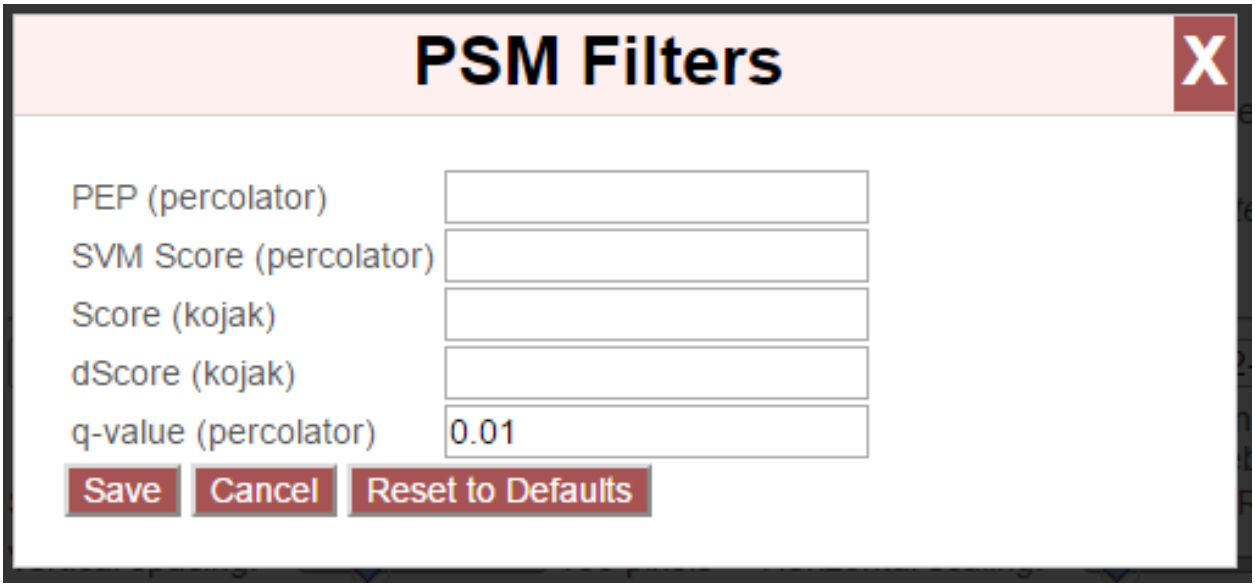
PSM Filters

The filters to apply at the PSM level. Only results which have at least one PSM that meets all of the selected critiera will be listed. When listing PSMs associated with peptides, only PSMs that meet all of the selected critiera will be listed.

To change the PSM-level filters, first click the pencil icon next to "PSM Filters":



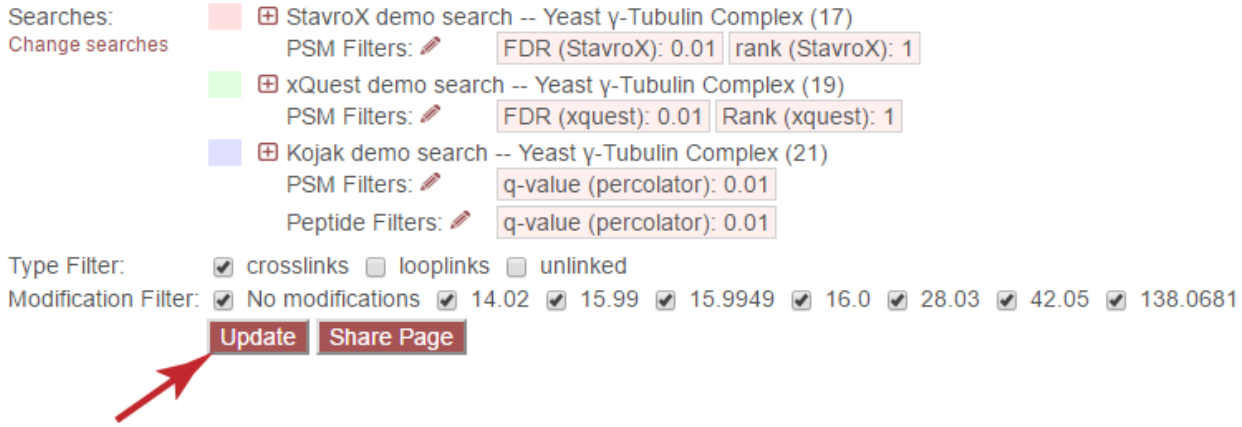
This opens an overlay with the containing the possible score types to use as PSM filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

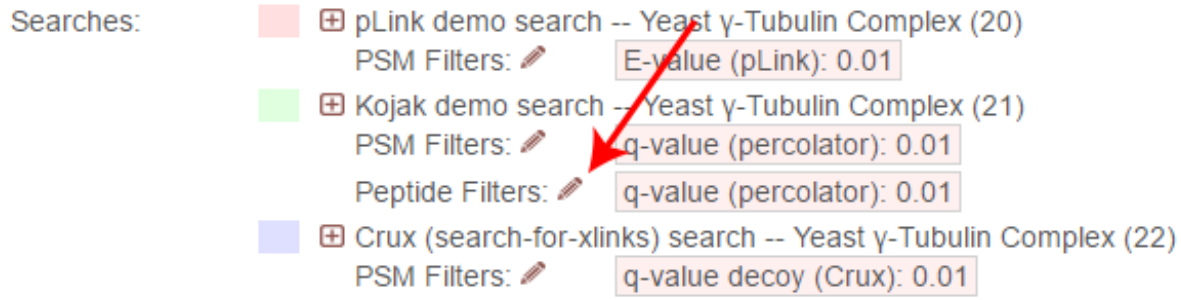
Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update" button on the page to apply any new PSM- or peptide-level filters.



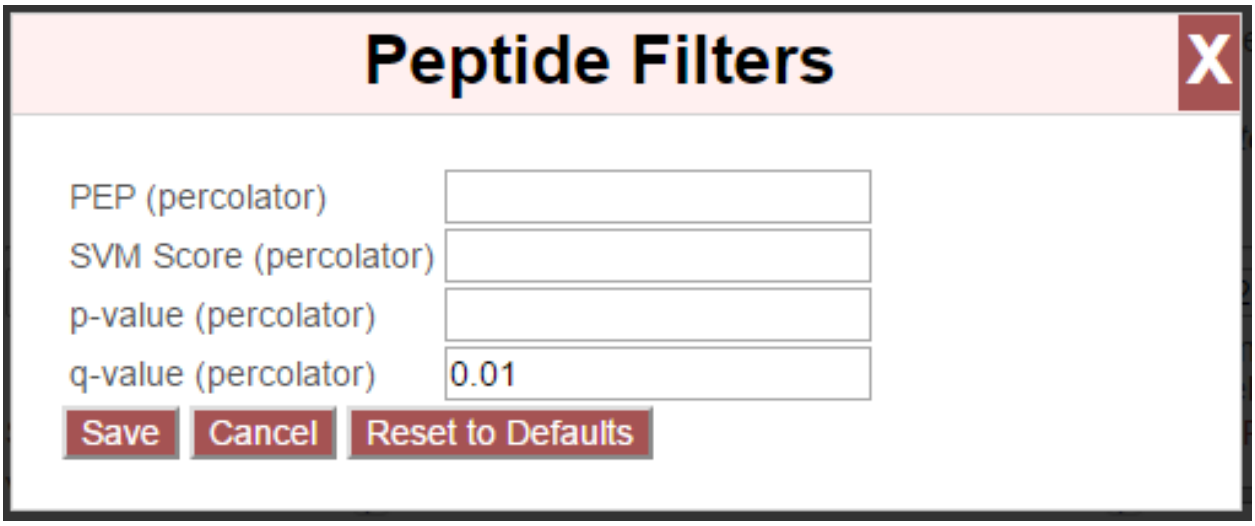
Peptide Filters

The filters to apply at the peptide level. Only results which have at least one peptide that meets all of the selected critiera will be listed.

To change the peptide-level filters, first click the pencil icon next to "Peptide Filters":



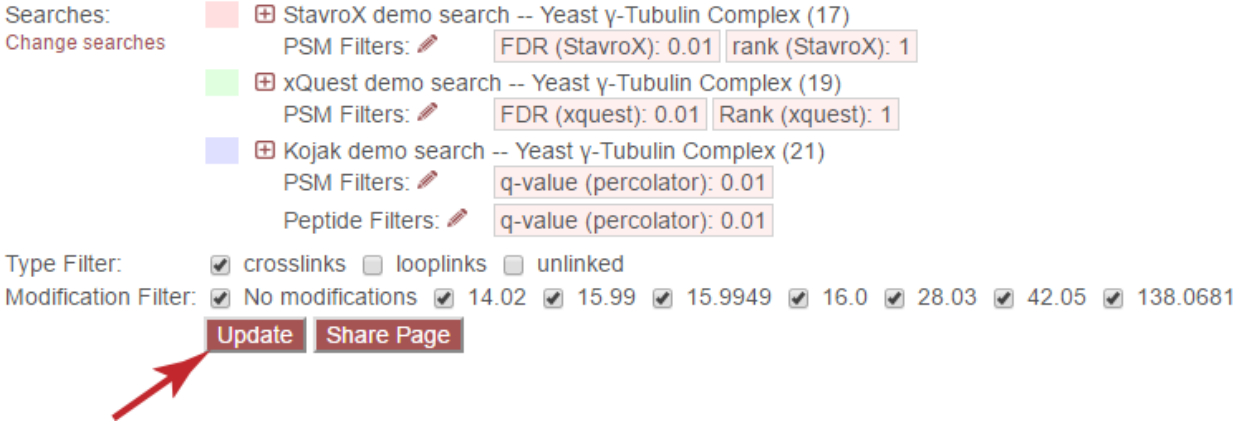
This opens an overlay with the containing the possible score types to use as peptide-level filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

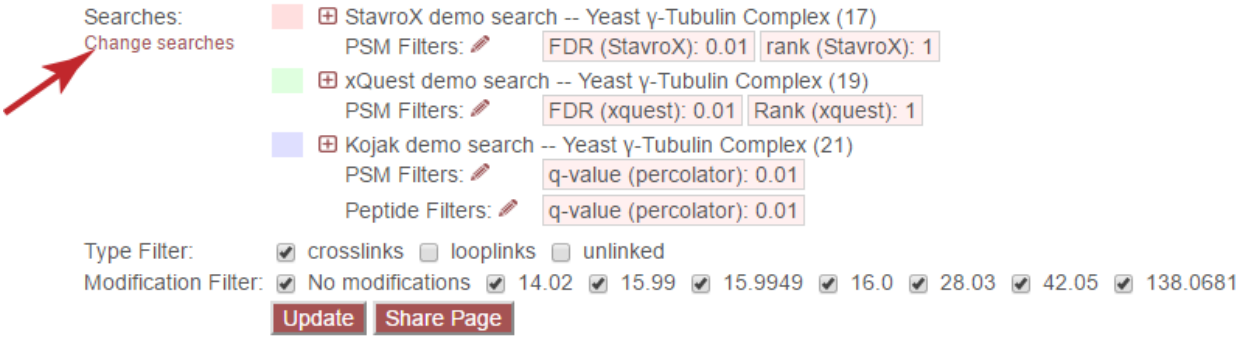
The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update" button on the page to apply any new PSM- or peptide-level filters.

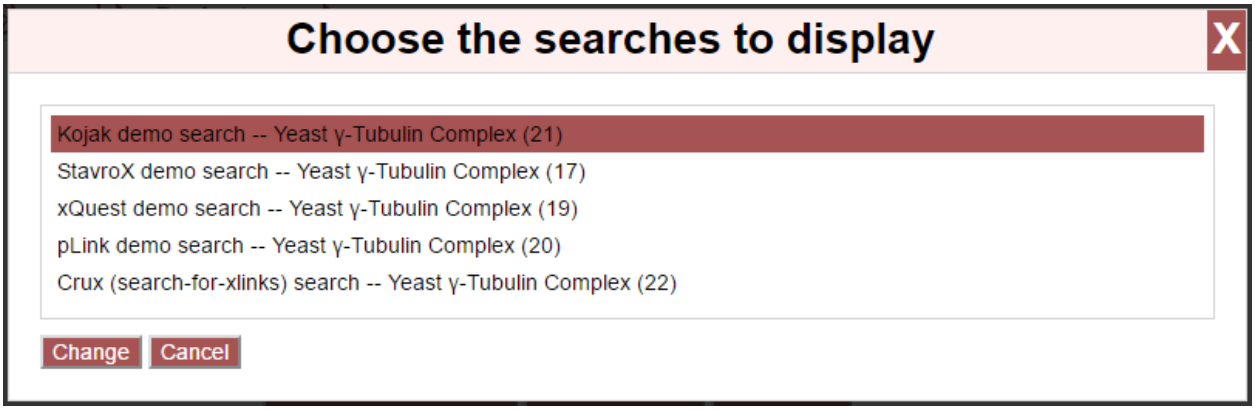


8.2 General Options

8.2.1 Change Searches

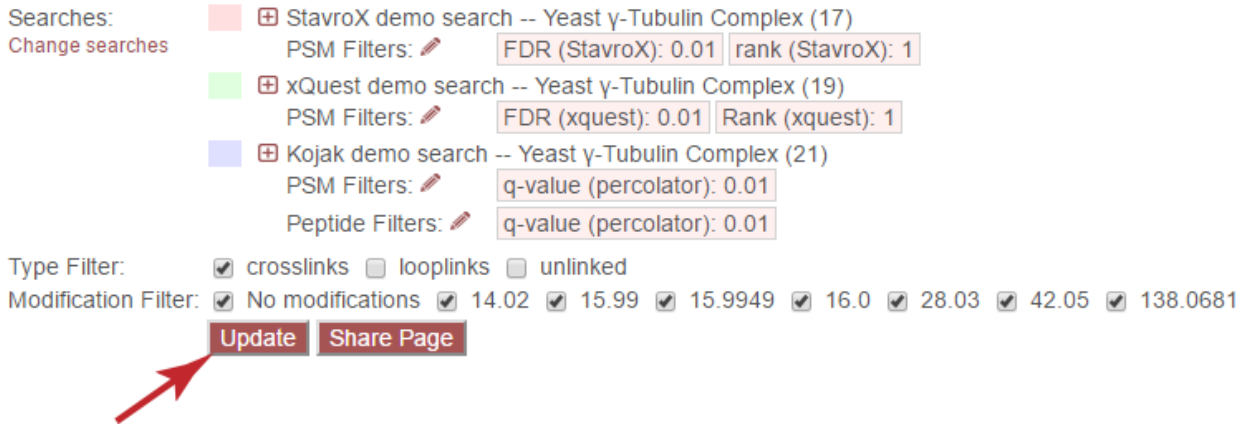


The "Change searches" link allows the user to change which searches are currently being displayed. Clicking the link causes the following overlay to be displayed:



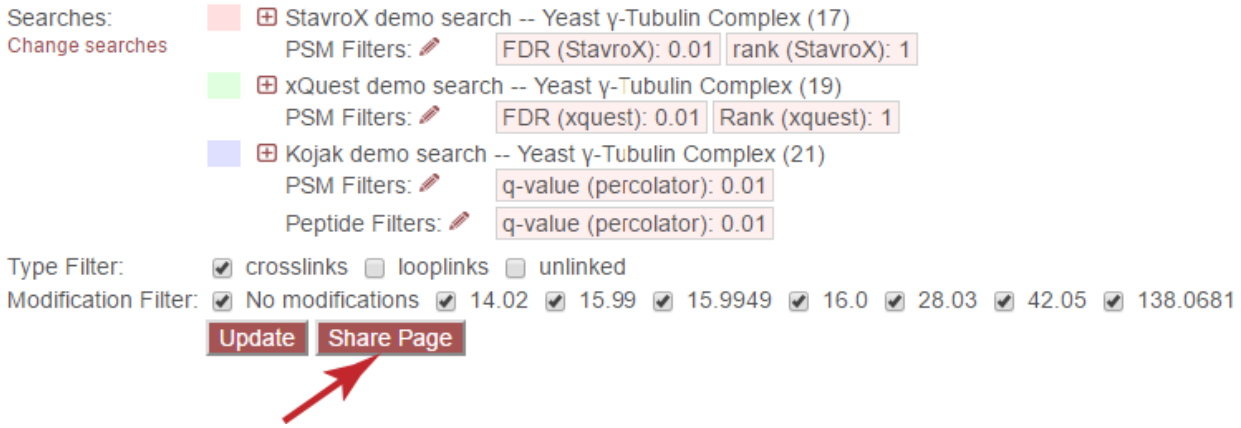
Select or de-select searches by clicking on them in the list. Once done, click "Change" to update the page with the new data or "Cancel" to close the overlay.

8.2.2 Update From Database

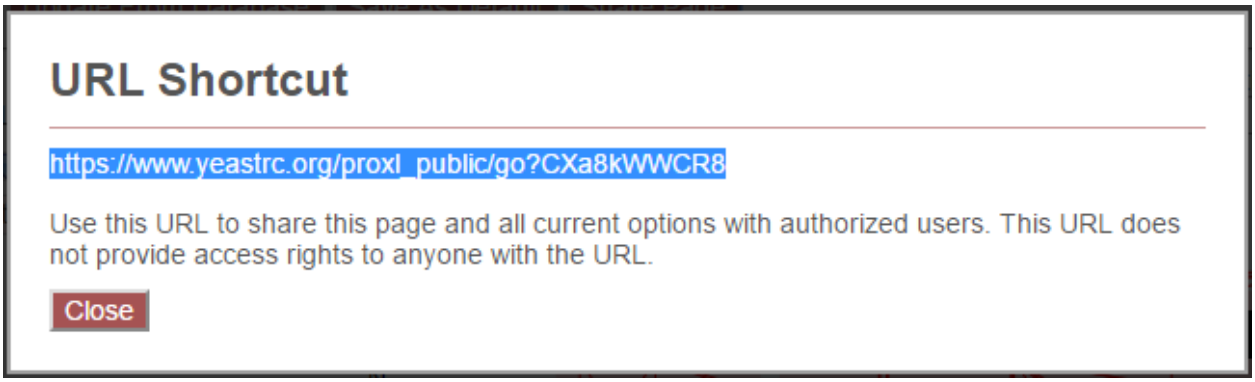


If the user changes any filter parameters–such as PSM/peptide score cutoffs–this button must be clicked to reflect the new filter choices.

8.2.3 Share Page



Clicking the "Share Page" button will generate a shortcut URL for viewing the current page. The shortened URL will appear in an overlay as:



Copying and sharing the highlighted URL will direct users to the view of the page when the URL was generated. Note that this URL does not grant access to the page to any user that would not otherwise have access.

8.3 Filter Data

8.3.1 Type filter

Only peptides of the checked type(s) will be returned. Proxl defines the types as:

- crosslink - A pair of peptides linked by a crosslinker.
- looplink - A single peptide with two residues linked by a crosslinker.
- unlinked - The peptide without a crosslinker on any residue.

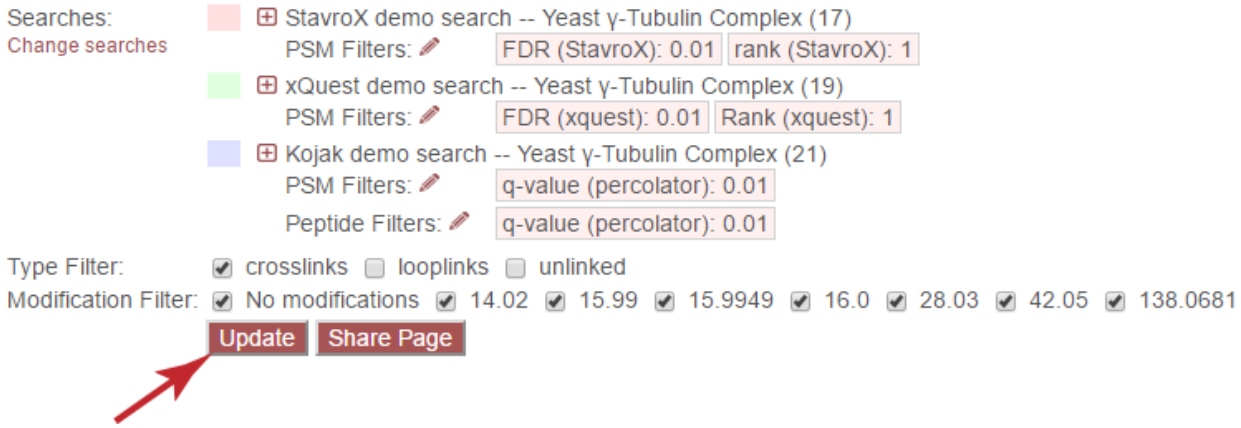
Checking multiple boxes will include any peptide that has at least one of the checked types. I.e., checking 'crosslinks' and 'looplinks' will only include peptides that are either crosslinks or looplinks. Only checking 'crosslinks' will only return crosslinked peptides.

8.3.2 Modification filter

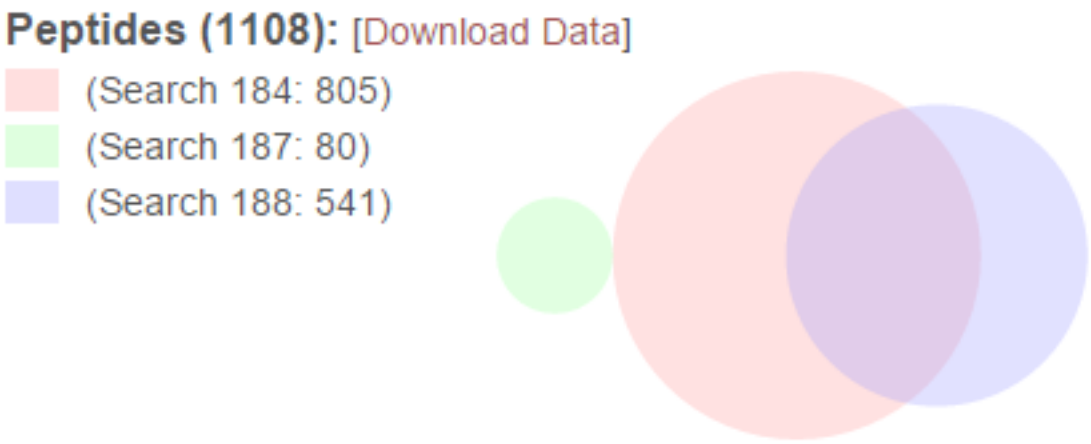
Only peptides with at least one of the checked modifications will be included. Note that monolinks are considered modifications of residues in proxl, so the mass of the crosslinker when found on monolinks is included here.

8.3.3 Update

Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update" button on the page to apply any new PSM- or peptide-level filters.



8.4 Euler diagram



The Euler diagram (similar to a Venn diagram) provides a graphical depiction of the relative sizes and overlap between the peptides found in the merged searches. The colors in the diagram match the colors used for the search list above. The search list is provided to the left of the diagram with their associated colors as a legend. The labels for each color include the search ID number and the number of distinct peptides found in each of the merged searches. The total number of peptides resulting from the merge is presented in the header above the legend next to "Peptides".

The "[Download Data]" link in the legend header will download the data in the table as a tab-delimited text file.

8.5 Table Description

The table presents columns describing the peptides and indicates in which of the merged searches the peptides were found. There is one row per peptide. A peptide on this page is defined as the unique combination of peptide sequence(s), link positions in those peptides, and dynamic modifications present on the peptides. So an unmodified peptide and a modified peptide with the same sequence will appear as separate rows in the table. Each row in the table may be clicked on to expand and view the peptide-level statistics for the given peptide from each search. Each of these searches may then be clicked on to view PSMs and spectra from those searches.

8.5.1 Columns

The columns are described below. Note that all column headers may be clicked to toggle between ascending and descending sorting of that column. Holding the shift key while clicking column headers allow sorting on multiple columns.

Search Columns

The first 1-3 columns will be labeled with search ID numbers as headers, and provide an indication for whether or not the peptide in that row was found in that search. If found in that search, the cell for this search in this row will be shaded the same color associated with that search in the Euler diagram and search list at the top of the page. The column will also contain an asterisk. If not found, this cell is empty.

Searches

The number of the merged searches that contain this peptide. The [+] icon indicates that the row may be clicked on to be expanded to show underlying searches in which this peptide as found, the stats for this peptide from each search, and the ability to view PSMs and associated spectra.

Type

The type of peptide (crosslink, looplink, or unlinked).

Peptide 1

The parsed sequence of the peptide (or the first peptide in the case of crosslinks).

Pos

The position in that peptide containing the linker.

Mods

A comma-delimited list of dynamic modifications found for peptide 1 in the form of position(mass). E.g., 17(15.99), 20(14.02)

Peptide 2

The parse sequence of the second peptide in the crosslink.

Pos

The position in that peptide containing the linker.

Mods

A comma-delimited list of dynamic modifications found for peptide 2 in the form of position(mass). E.g., 17(15.99), 20(14.02)

Protein 1

The protein(s) to which the first peptide matches, and the position in that protein to which the linker position in that peptide matched. Mouse-over the protein name to get a description.

Protein 2

The protein(s) to which the second peptide matches, and the position in that protein to which the linker position in that peptide matched. Mouse-over the protein name to get a description.

PSMs

The total number of combined PSMs from each search for this peptide that meet the filtering critera. Note: click the table row containing the peptide to see the PSMs.

Best PSM- and Peptide-level Scores

Separate columns, color-coded for each search, display the best PSM- and peptide-level scores currently being used to filter the data from each search.

8.6 View PSMs

To view PSMs for a given peptide, first click on a row in the table to expand and view the peptide-level statistics for a given peptide from each search in which it was found (at the given cutoffs). Each of these searches may be clicked to expand and view all PSMs that meet the current filtering criteria.

8.6.1 Columns

The PSMs appear in a table with the following columns:

Scan Num.

The scan number from the spectral file (e.g., mzML file)

Obs. m/z

The observed m/z of the precursor ion.

Charge

The predicted charge state of the precursor ion.

RT (min)

The retention time in minutes.

Scan Filename

The filename of the scan file.

Scores

A column for each PSM-level score from the respective search.

8.6.2 View Spectra

The annotated mass spectrum may be viewed for any PSM by clicking the "View Spectrum" link. For help on our spectrum viewer, see the *Spectrum Viewer* page.

8.7 Sort Data

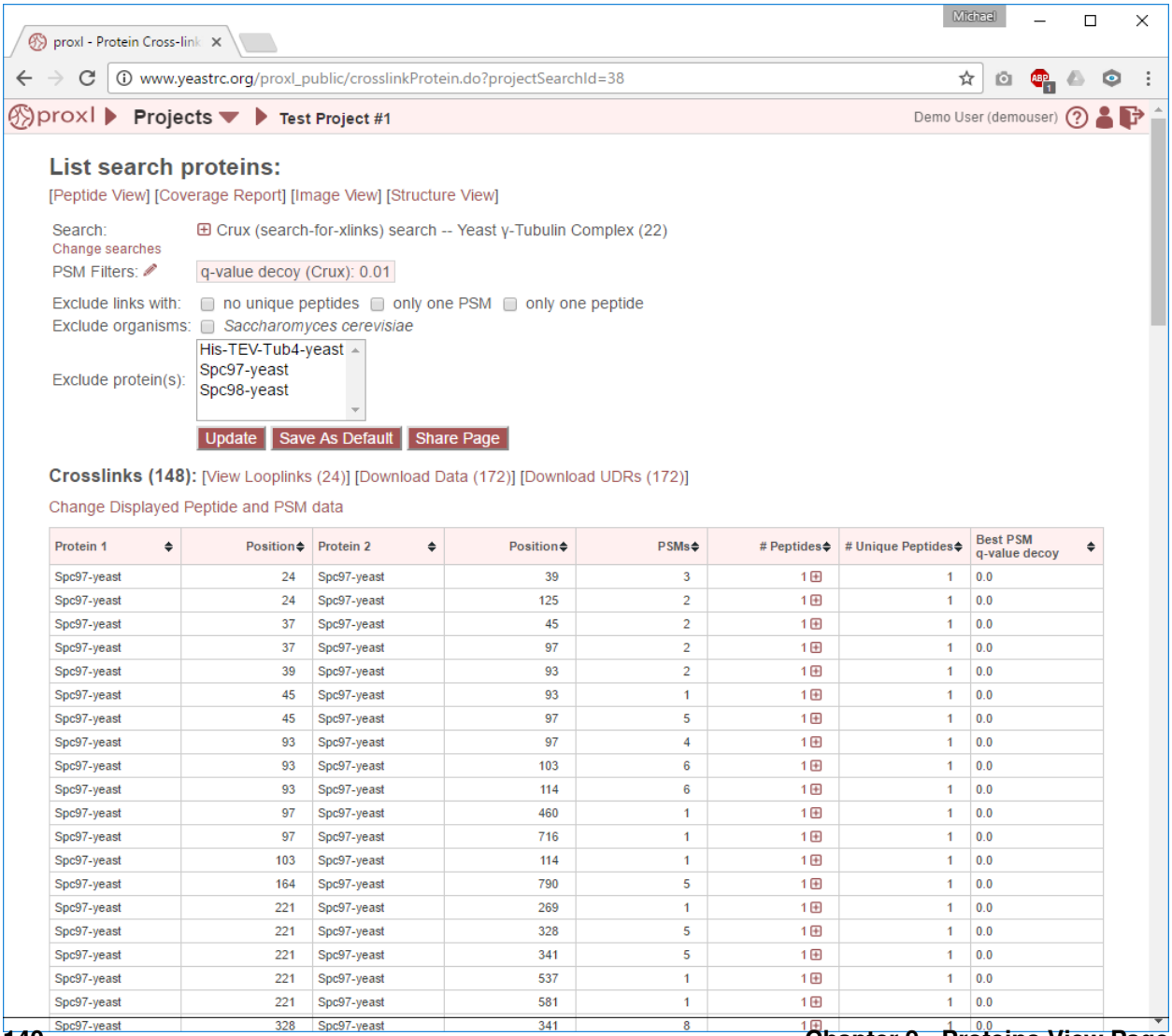
All column headers may be clicked to toggle between ascending and descending sorting of that column. Holding the shift key while clicking column headers allow sorting on multiple columns.

8.8 Download Data

Clicking the [Download Data] link in the header of the Euler diagram will download the shown data as a tab-delimited text file.

CHAPTER 9

Proteins View Page



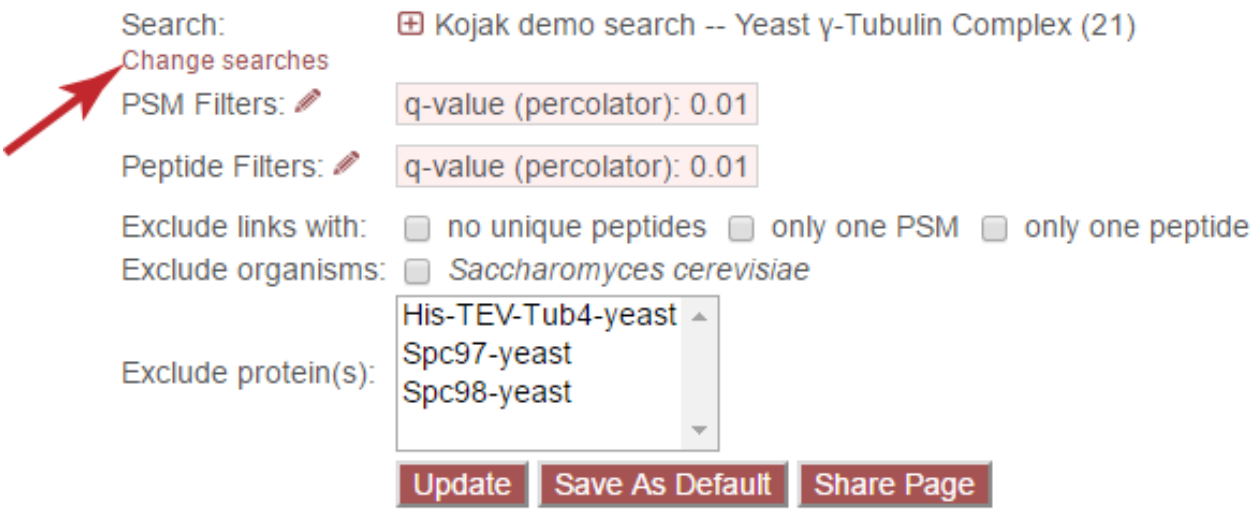
Protein 1	Position	Protein 2	Position	PSMs	# Peptides	# Unique Peptides	Best PSM q-value decoy
Spc97-yeast	24	Spc97-yeast	39	3	1	1	0.0
Spc97-yeast	24	Spc97-yeast	125	2	1	1	0.0
Spc97-yeast	37	Spc97-yeast	45	2	1	1	0.0
Spc97-yeast	37	Spc97-yeast	97	2	1	1	0.0
Spc97-yeast	39	Spc97-yeast	93	2	1	1	0.0
Spc97-yeast	45	Spc97-yeast	93	1	1	1	0.0
Spc97-yeast	45	Spc97-yeast	97	5	1	1	0.0
Spc97-yeast	93	Spc97-yeast	97	4	1	1	0.0
Spc97-yeast	93	Spc97-yeast	103	6	1	1	0.0
Spc97-yeast	93	Spc97-yeast	114	6	1	1	0.0
Spc97-yeast	97	Spc97-yeast	460	1	1	1	0.0
Spc97-yeast	97	Spc97-yeast	716	1	1	1	0.0
Spc97-yeast	103	Spc97-yeast	114	1	1	1	0.0
Spc97-yeast	164	Spc97-yeast	790	5	1	1	0.0
Spc97-yeast	221	Spc97-yeast	269	1	1	1	0.0
Spc97-yeast	221	Spc97-yeast	328	5	1	1	0.0
Spc97-yeast	221	Spc97-yeast	341	5	1	1	0.0
Spc97-yeast	221	Spc97-yeast	537	1	1	1	0.0
Spc97-yeast	221	Spc97-yeast	581	1	1	1	0.0
Spc97-yeast	328	Spc97-yeast	341	8	1	1	0.0

The protein view page provides a table view of crosslinks or looplinks at the protein level. Each row in the table corresponds to a unique crosslink (specific position in protein A linked to a specific position in protein B) or a unique looplink (specific pair of positions in a protein). The data may be filtered according to confidence, taxonomy, or individually by protein. For the view page seen when merging multiple searches, see *Merged Protein View Page*.

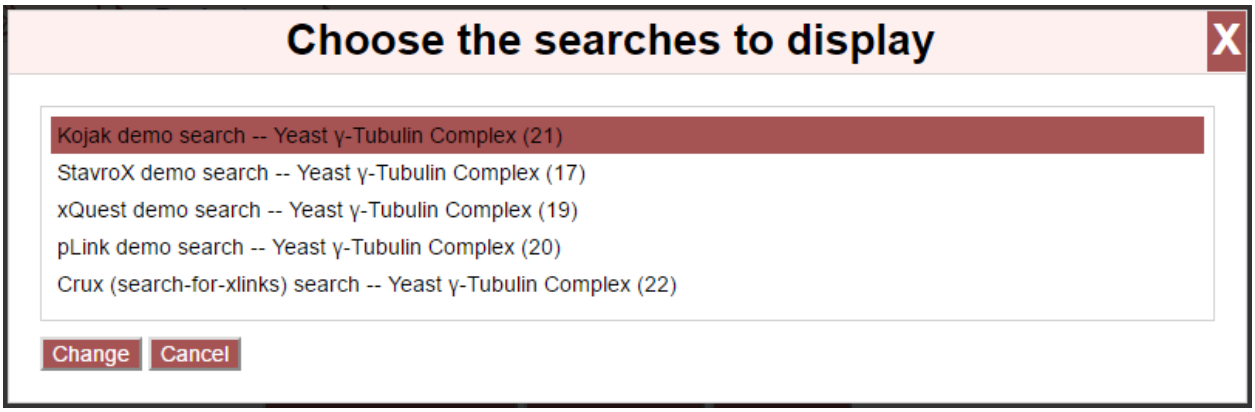
Note: If any identified peptides map to multiple proteins, those proteins are listed here as separate rows. For example if peptide 1 is linked to peptide 2, and peptide 1 maps to protein A and peptide 2 maps to proteins B and C, rows will be present for A-B and A-C. This may dramatically increase the number of reported crosslinks if your protein database is redundant in terms of homologous proteins or proteoforms or if small peptides are mapping to many proteins. The filtering options described below are meant to help eliminate this redundancy in reported proteins.

9.1 General Options

9.1.1 Change Searches

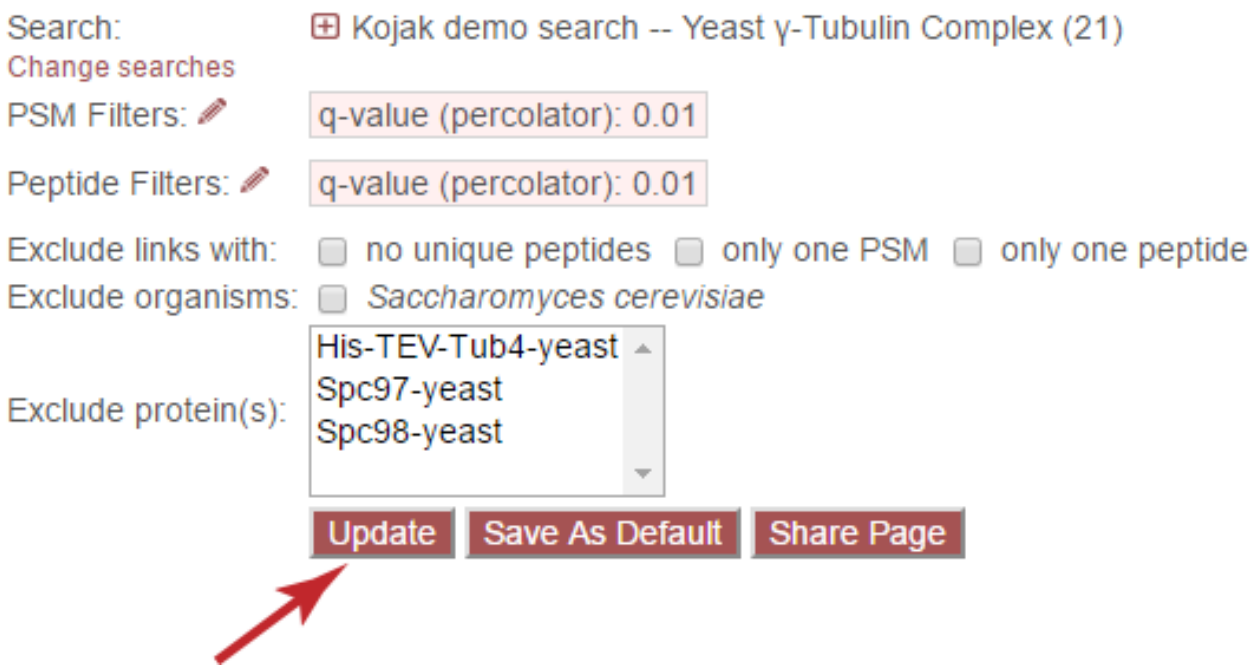


The "Change searches" link allows the user to change which searches are currently being displayed. Clicking the link causes the following overlay to be displayed:



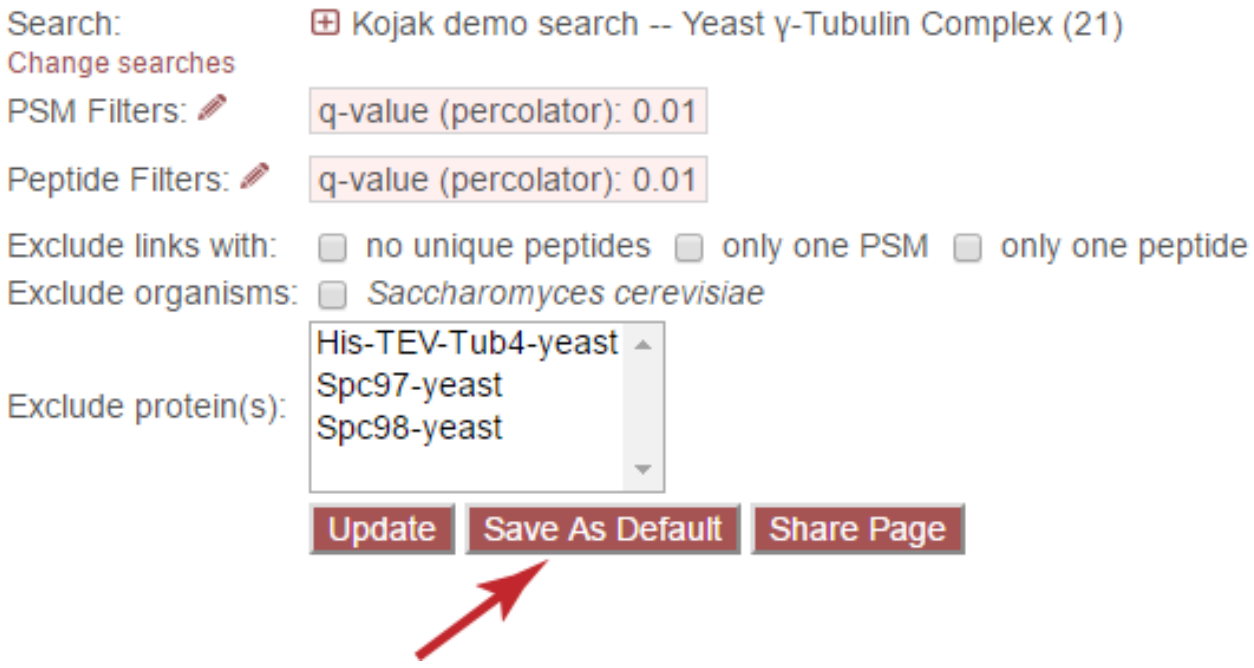
Select or de-select searches by clicking on them in the list. Once done, click "Change" to update the page with the new data or "Cancel" to close the overlay.

9.1.2 Update From Database



If the user changes any filter parameters–such as PSM/peptide score cutoffs–this button must be clicked to reflect the new filter choices.

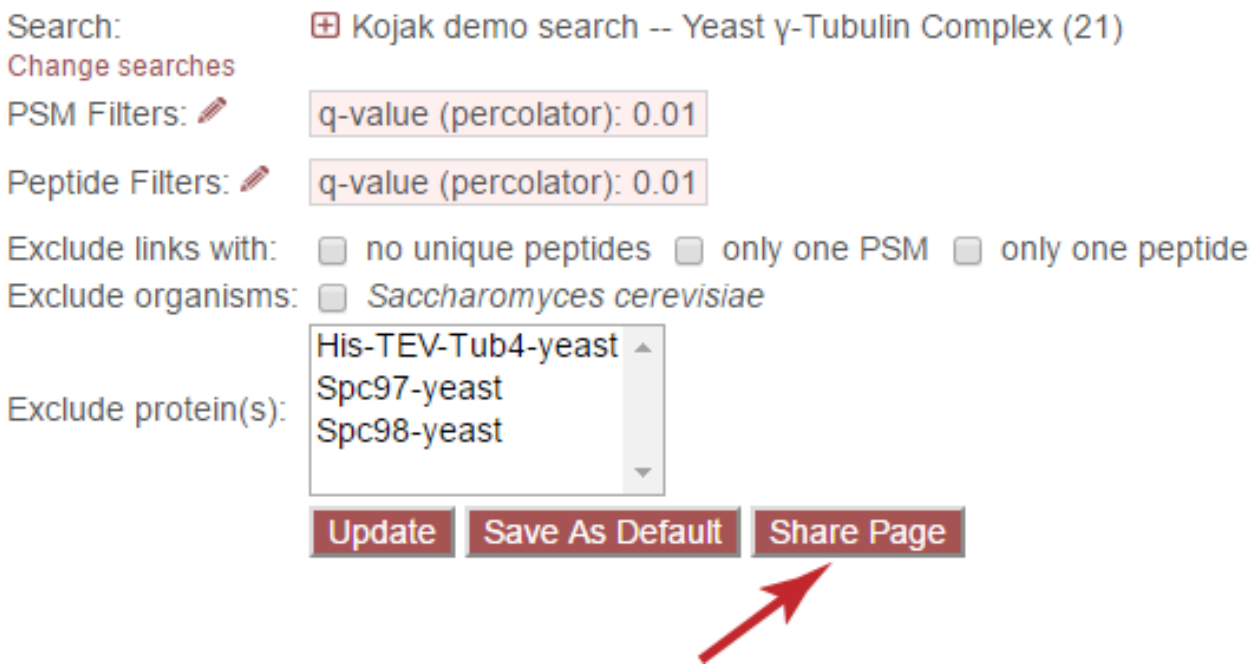
9.1.3 Save as Default



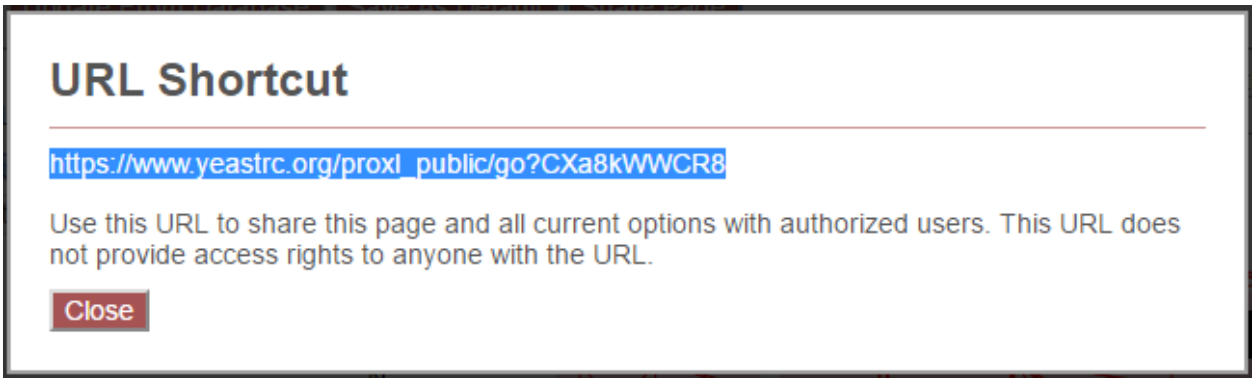
Project owners may click "Save as Default" to save the current URL as the default view of the "Protein View" for this search. This default view will be populated with the same options as when the button is clicked. This is a convenient

way to share data with collaborators or the public that does not require that they manipulate the image viewer to see the data.

9.1.4 Share Page



Clicking the "Share Page" button will generate a shortcut URL for viewing the current page. The shortened URL will appear in an overlay as:



Copying and sharing the highlighted URL will direct users to the view of the page when the URL was generated. Note that this URL does not grant access to the page to any user that would not otherwise have access.

9.2 View Looplinks

By default, the table shows crosslinks. To switch to looplinks, click the [View Looplinks] link at the top of the table. To view crosslinks again, click the [View Crosslinks] link that appears at the top of the table.

9.3 Download Data

All crosslinks and looplinks that meet the current filtering criteria may be downloaded as tab-delimited text by cliking the [Download Data (#)] link above the table. # indicates the number of rows in the table.

9.4 Download UDRs

UDR stands for "unique distance restraint", which takes its name from 3D modelling terminology. A UDR, in proxl, is any specific position in a protein linked to a specific position in another protein, whether it is a crosslink or a looplink. The [Download UDRs (#)] link downloads a non-redundant tab-delimited text table of these UDRs consolidated from the crosslinks and looplinks. The # is the number of UDRs.

9.5 Search Information

The name of the search (and internal search ID reference number) from which these data were obtained is shown first. The red [+] icon may be clicked to reveal more information about the search, including the path the data were imported from, the linker that was used, the upload date, and the FASTA file that was searched.

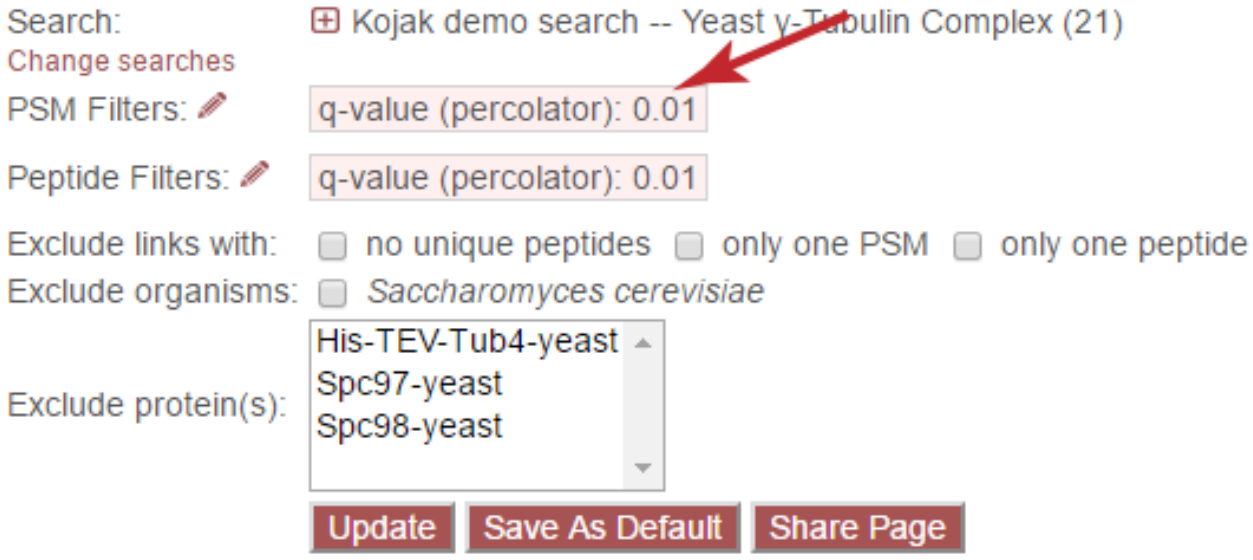
9.6 Filter Data

The data presented may be filtered according to the following criteria. Note: Only crosslinks or looplinks that meet ALL the filter criteria are shown.

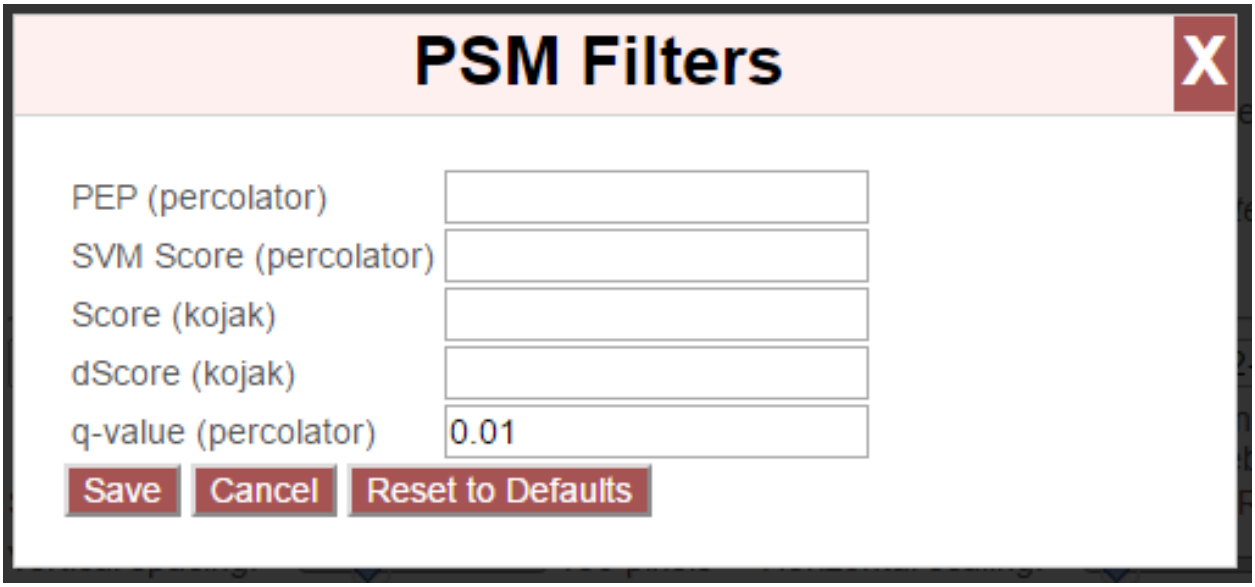
9.6.1 PSM Filters

The filters to apply at the PSM level. Only results which have at least one PSM that meets all of the selected critiera will be listed. When listing PSMs associated with peptides, only PSMs that meet all of the selected critiera will be listed.

To change the PSM-level filters, first click the pencil icon next to "PSM Filters":



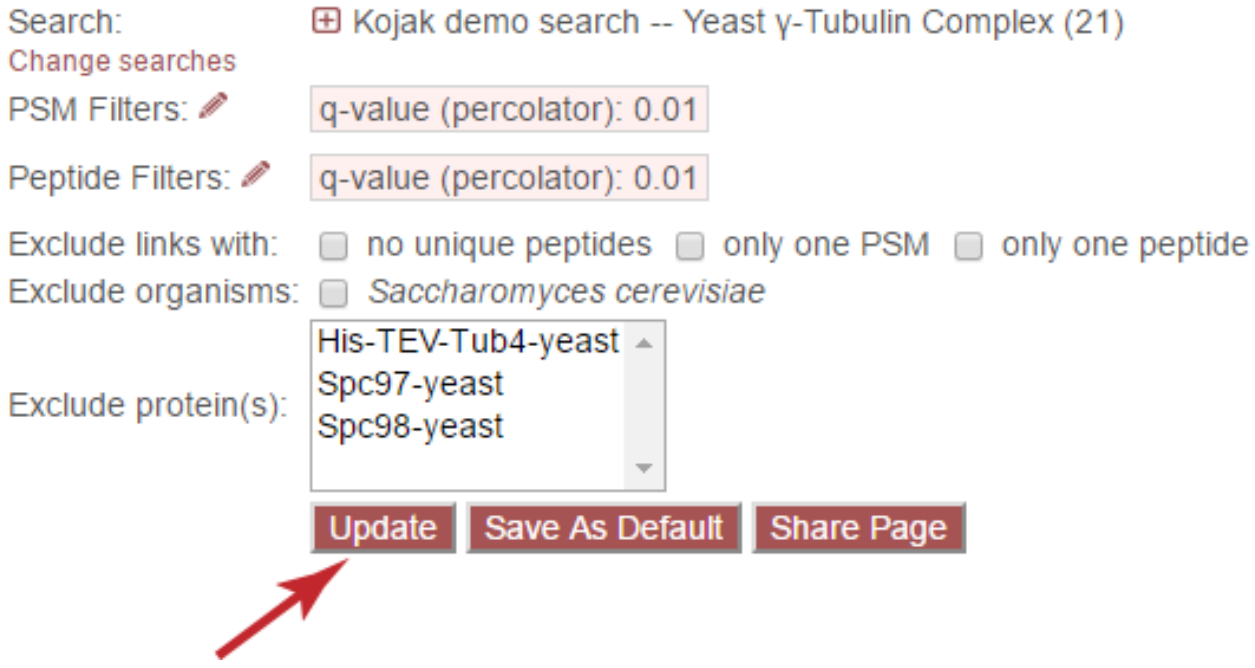
This opens an overlay with the containing the possible score types to use as PSM filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

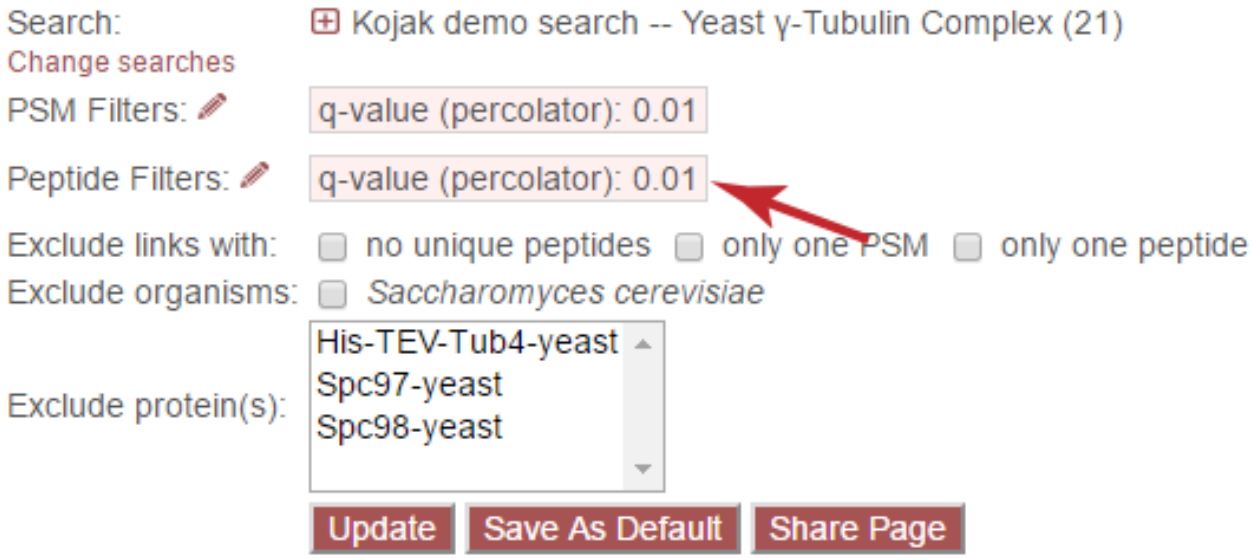
Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update" button on the page to apply any new PSM- or peptide-level filters.



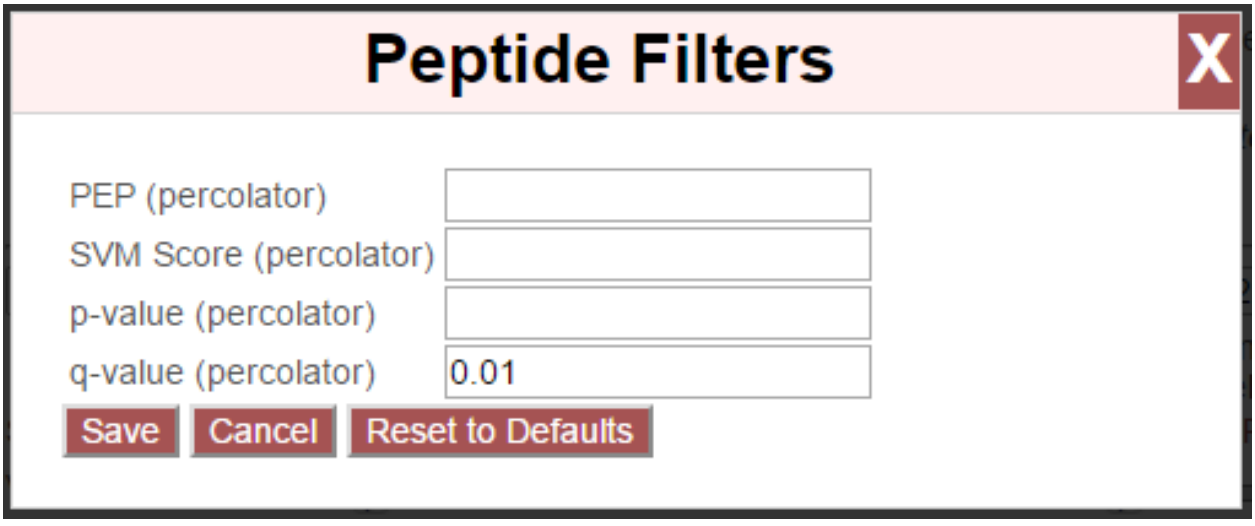
9.6.2 Peptide Filters

Some search algorithms, such as Percolator, provide statistics at the peptide level that may be used for filtering. If applicable, peptide-level filtering options may be set here. Only results which have at least one peptide that meets all of the selected critiera will be listed.

To change the peptide-level filters, first click the pencil icon next to "Peptide Filters":



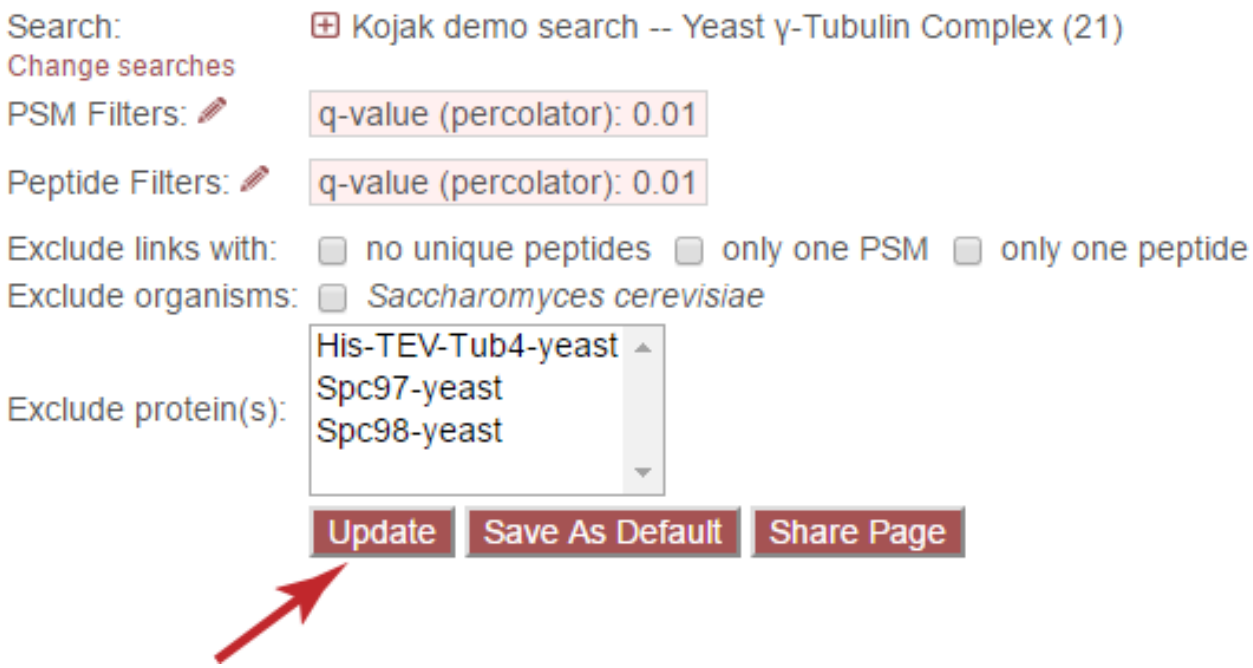
This opens an overlay with the containing the possible score types to use as peptide-level filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update" button on the page to apply any new PSM- or peptide-level filters.



9.6.3 Exclude links with

Crosslinks or looplinks that have any of the checked attributes will be excluded. The attributes are:

- no unique peptides - If all peptides that ID either one of the crosslinked proteins also map to another protein
- only one PSM - If a given crosslink or looplink was identified by a single PSM
- only one peptide - If a given crosslink or looplink was identifed by a single peptide, where a peptide is the combination of sequence, linked positions, and modifications

9.6.4 Exclude organisms

Any links containing a protein that maps to any of the checked organisms will be excluded. The list of organisms presented was gathered by the proteins found in the search. Useful for filtering out groups of contaminant proteins.

9.6.5 Exclude protein(s)

Any links containing a any of the selected proteins will be excluded. Multiple proteins may be selected or unselected using control-click (command-click on the mac) or shift-click. Useful for filtering out individual contaminant proteins.

9.6.6 Update

In order to apply new filter parameters to the shown data, the "Update" button must be clicked. This will fetch filtered data from the proxl server and display the data on the web page.

9.6.7 Save As Default

Project owners may save the current filter parameters as the default view of the data on this page by clicking this button. This default view will be shown when users follow links to the "Protein View" for this search.

9.7 Table Description

Above the table is the text "Crosslinks (#)" or "Looplinks (#)". # is the number of rows in the table. The [View Looplinks (#)] links will change to viewing looplinks, where # is the number of looplinks that will be shown. The [Download Data (#)] link downloads the data as tab delimited text (see *Download UDRs*) and [Download UDRs (#)] downloads the UDRs as tab delimited text (see *Download Data*).

9.7.1 Columns

The columns are described below. Note that all column headers may be clicked to toggle between ascending and descending sorting of that column. Holding the shift key while clicking column headers allow sorting on multiple columns.

Protein 1 and 2 (Crosslink-only)

In the case of crosslinks, these are the crosslinked proteins

Position (Crosslink-only)

This is the crosslinked position in the respective proteins, where the first residue is counted as position 1.

Protein (Looplink-only)

In the case of looplinks, this is the looplinked protein

Position 1 and 2 (Looplink-only)

In looplinked proteins, these are the positions in the protein that are linked.

PSMs

The total number of PSMs (peptide spectrum matches) meeting the cutoff that identified either crosslinked (crosslink view) or looplinked (looplink view) peptides that mapped to the reported proteins and positions.

Peptides

The total number of identified crosslinked (crosslink view) or looplinked (looplink view) peptides that mapped to the reported proteins and positions. Only peptides that meet the current filtering criteria are counted.

Note: The individual peptides may be viewed by clicking a row in the table to view a table of peptides. Rows in that peptide table may also be viewed to view the underlying PSMs and view spectra. See *View Peptides*.

Unique Peptides

Of the # of peptides, the total number that uniquely mapped to this protein pair (crosslink view) or protein (looplink view).

Best Peptide-level Scores

If peptide-level filters are being used, the best score from all peptides matching to the indicated proteins and positions will be displayed for each filter.

Best PSM-level Scores

If PSM-level filters are being used, the best score from all PSMs matching to the indicated proteins and positions will be displayed for each filter.

9.8 View Peptides

All peptides that meet the current filters that were mapped to a protein-level crosslink or looplink may be seen by clicking on the respective row in the table. Additionally, all rows of this peptide table may clicked to view all PSMs associated with that peptide identification. (See *View PSMs*.)

Protein 1	Position	Protein 2	Position	PSMs	# Peptides	# Unique Peptides	Best Peptide q-value	Best PSM q-value
Spc97-yeast	45	Spc97-yeast	103	3	2 ⊞	2	0.0	0.0
Spc97-yeast	221	Spc97-yeast	341	7	2 ⊞	2	0.0	0.0
Spc97-yeast	355	Spc98-yeast	709	4	2 ⊟	2	0.0	0.0

Reported peptide	Peptide 1	Pos	Peptide 2	Pos	q-value	PEP	p-value	# PSMs	# Unique	Best PSM q-value
TGILLKVVR(6)--SSSSKPQK(5)	SSSSKPQK	5	TGILLKVVR	6	0.000	0.007231	0.009524	3 ⊞	3	0.0
SSSSKPQKHAIEK(5)--TGILLKVVR(6)	SSSSKPQKHAIEK	5	TGILLKVVR	6	0.003256	0.02414	0.01905	1 ⊞	1	0.001728

Spc97-yeast	460	Spc98-yeast	843	4	2 ⊞	2	0.0	0.0
Spc97-yeast	537	Spc98-yeast	843	3	2 ⊞	2	0.0	0.0

9.8.1 Columns

The peptides appear in a table with the following columns:

Reported peptide

The peptide identificaton as it was reported by the respective search program.

Peptide 1 and 2 (Crosslink-only)

The sequences of the two crosslinked peptides.

Pos (Crosslink-only)

The positions in the respective peptides that were crosslinked (starting at 1).

Peptide (Looplink-only)

The sequence of the looplinked peptide.

Pos 1 and 2 (Looplink-only)

The positions in the peptide that were looplinked.

Scores

A column for each peptide-level score is shown.

PSMs

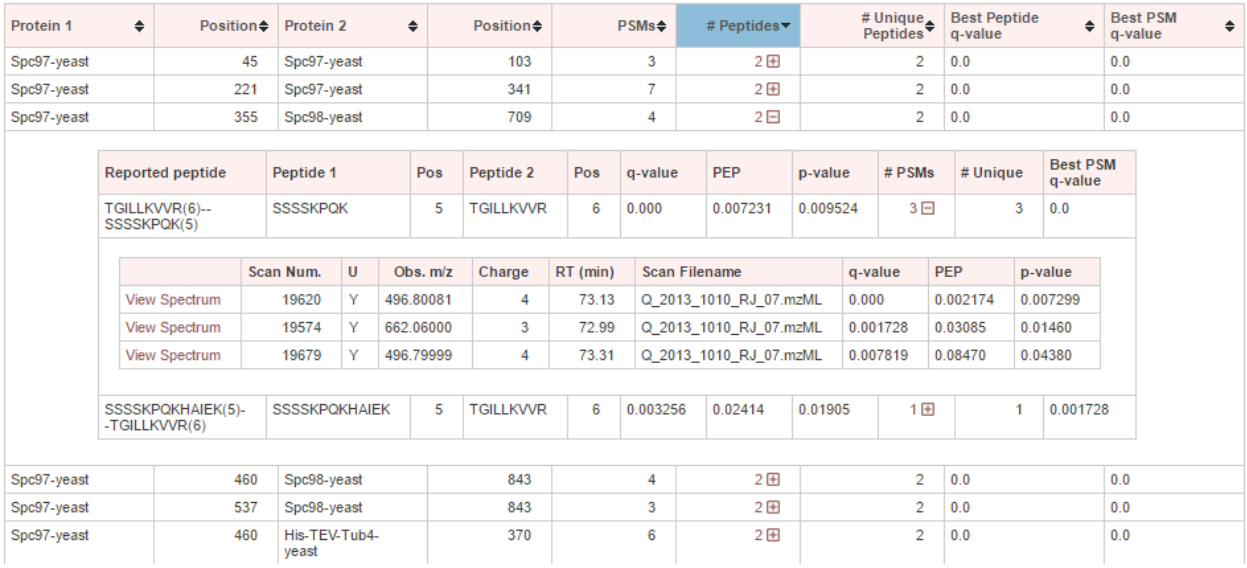
The number of PSMs that meet the cutoff criteria that identified this peptide.

Best PSM-level Scores

If PSM-level filters are being used, the best score from all PSMs matching to this peptide for each score on which PSMs are being filtered.

9.9 View PSMs

All PSMs meeting the current filtering criteria that map to a given peptide can by shown by clicking on the table row containing that peptide.



Protein 1	Position	Protein 2	Position	PSMs	# Peptides	# Unique Peptides	Best Peptide q-value	Best PSM q-value
Spc97-yeast	45	Spc97-yeast	103	3	2 ⊞	2	0.0	0.0
Spc97-yeast	221	Spc97-yeast	341	7	2 ⊞	2	0.0	0.0
Spc97-yeast	355	Spc98-yeast	709	4	2 ⊟	2	0.0	0.0

Reported peptide	Peptide 1	Pos	Peptide 2	Pos	q-value	PEP	p-value	# PSMs	# Unique	Best PSM q-value
TGILLKVVR(6)--SSSSKPQK(5)	SSSSKPQK	5	TGILLKVVR	6	0.000	0.007231	0.009524	3 ⊟	3	0.0

	Scan Num.	U	Obs. m/z	Charge	RT (min)	Scan Filename	q-value	PEP	p-value
View Spectrum	19620	Y	496.80081	4	73.13	Q_2013_1010_RJ_07.mzML	0.000	0.002174	0.007299
View Spectrum	19574	Y	662.06000	3	72.99	Q_2013_1010_RJ_07.mzML	0.001728	0.03085	0.01460
View Spectrum	19679	Y	496.79999	4	73.31	Q_2013_1010_RJ_07.mzML	0.007819	0.08470	0.04380

SSSSKPQKHAIEK(5)--TGILLKVVR(6)	SSSSKPQKHAIEK	5	TGILLKVVR	6	0.003256	0.02414	0.01905	1 ⊞	1	0.001728

Spc97-yeast	460	Spc98-yeast	843	4	2 ⊞	2	0.0	0.0
Spc97-yeast	537	Spc98-yeast	843	3	2 ⊞	2	0.0	0.0
Spc97-yeast	460	His-TEV-Tub4-yeast	370	6	2 ⊞	2	0.0	0.0

9.9.1 Columns

The PSMs appear in a table with the following columns:

Scan Num.

The scan number from the spectral file (e.g., mzML file)

Charge

The predicted charge state of the precursor ion.

Obs. m/z

The observed m/z of the precursor ion.

RT (min)

The retention time in minutes.

Scan Filename

The filename of the scan file.

Scores

A column for each PSM-level score or annotation.

9.9.2 View Spectra

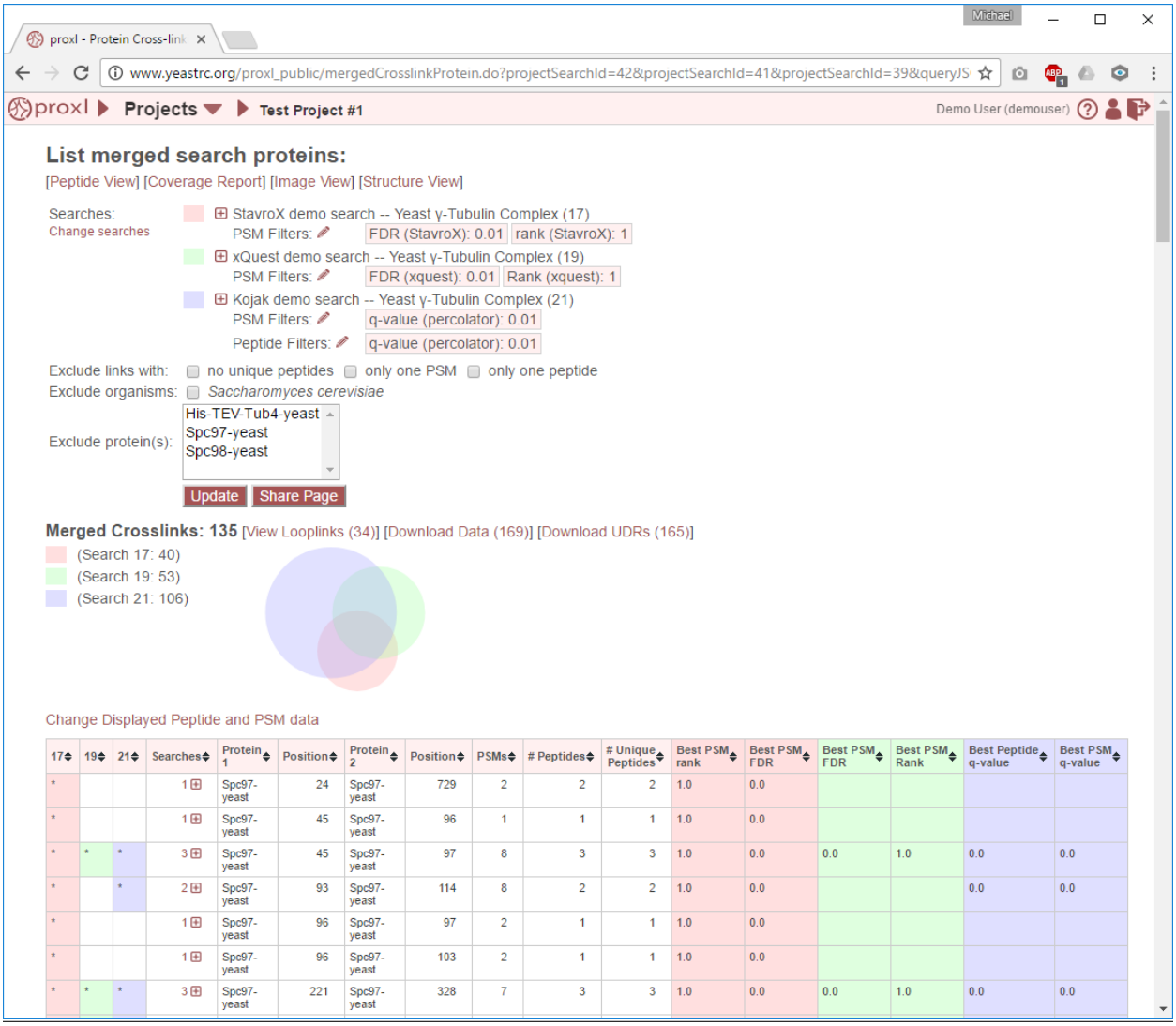
The annotated mass spectrum may be viewed for any PSM by clicking the "View Spectrum" link. For help on our spectrum viewer, see the *Spectrum Viewer* page.

9.10 Sort Data

All column headers may be clicked to toggle between ascending and descending sorting of that column. Holding the shift key while clicking column headers allow sorting on multiple columns.

CHAPTER 10

Merged Protein View Page



17	19	21	Searches	Protein 1	Position	Protein 2	Position	PSMs	# Peptides	# Unique Peptides	Best PSM rank	Best PSM FDR	Best PSM FDR	Best PSM Rank	Best Peptide q-value	Best PSM q-value
*			1	Spc97-yeast	24	Spc97-yeast	729	2	2	2	1.0	0.0				
*			1	Spc97-yeast	45	Spc97-yeast	96	1	1	1	1.0	0.0				
*	*	*	3	Spc97-yeast	45	Spc97-yeast	97	8	3	3	1.0	0.0	0.0	1.0	0.0	0.0
*		*	2	Spc97-yeast	93	Spc97-yeast	114	8	2	2	1.0	0.0			0.0	0.0
*			1	Spc97-yeast	96	Spc97-yeast	97	2	1	1	1.0	0.0				
*			1	Spc97-yeast	96	Spc97-yeast	103	2	1	1	1.0	0.0				
*	*	*	3	Spc97-yeast	221	Spc97-yeast	328	7	3	3	1.0	0.0	0.0	1.0	0.0	0.0

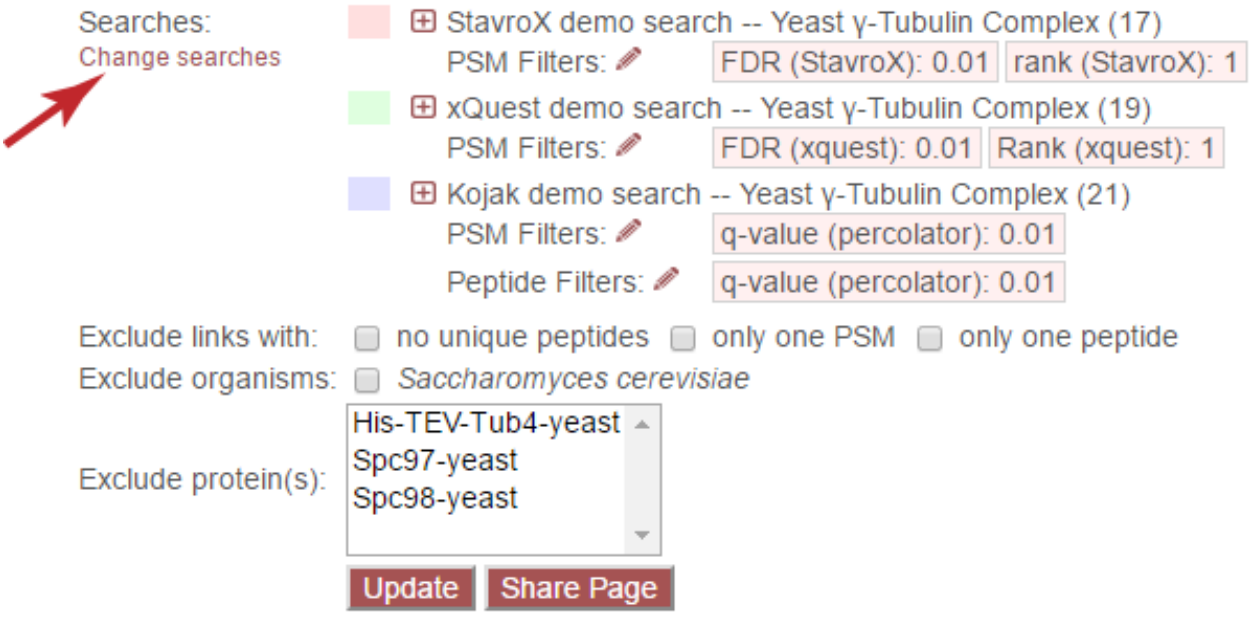
To reach this page, select multiple searches on the project page and click "View Merged Proteins". (See *Project Overview Page*.) This page combines and collates the data from multiple searches and presents the results as an interactive table. The searches do not need to be from the same software pipeline. For example, different versions of the same program may be compared, or the results from entirely different programs (e.g., Kojak and XQuest) may be compared. Currently, the total number of merged searches must be 2 or 3 and must be from the same project. For the protein view page seen when viewing a single search, see *Proteins View Page*.

This page is designed to present the data at the protein level, or the UDR level. That is to say that rows in the table represent distinct proteins linked at distinct positions. Rows in the table may be clicked on to view the data at the individual search level, and each of those searches may be expanded to view the underlying proteins, PSMs, and spectra.

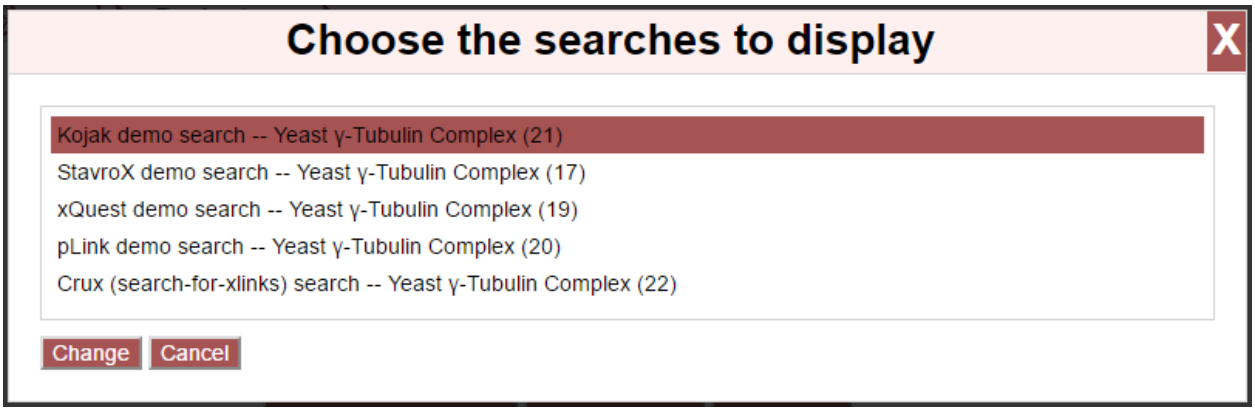
Note: If any identified peptides map to multiple proteins, those proteins are listed here as separate rows. For example if peptide 1 is linked to peptide 2, and peptide 1 maps to protein A and peptide 2 maps to proteins B and C, rows will be present for A-B and A-C. This may dramatically increase the number of reported crosslinks if your protein database is redundant in terms of homologous proteins or proteoforms or if small peptides are mapping to many proteins. The filtering options described below are meant to help eliminate this redundancy in reported proteins.

10.1 General Options

10.1.1 Change Searches

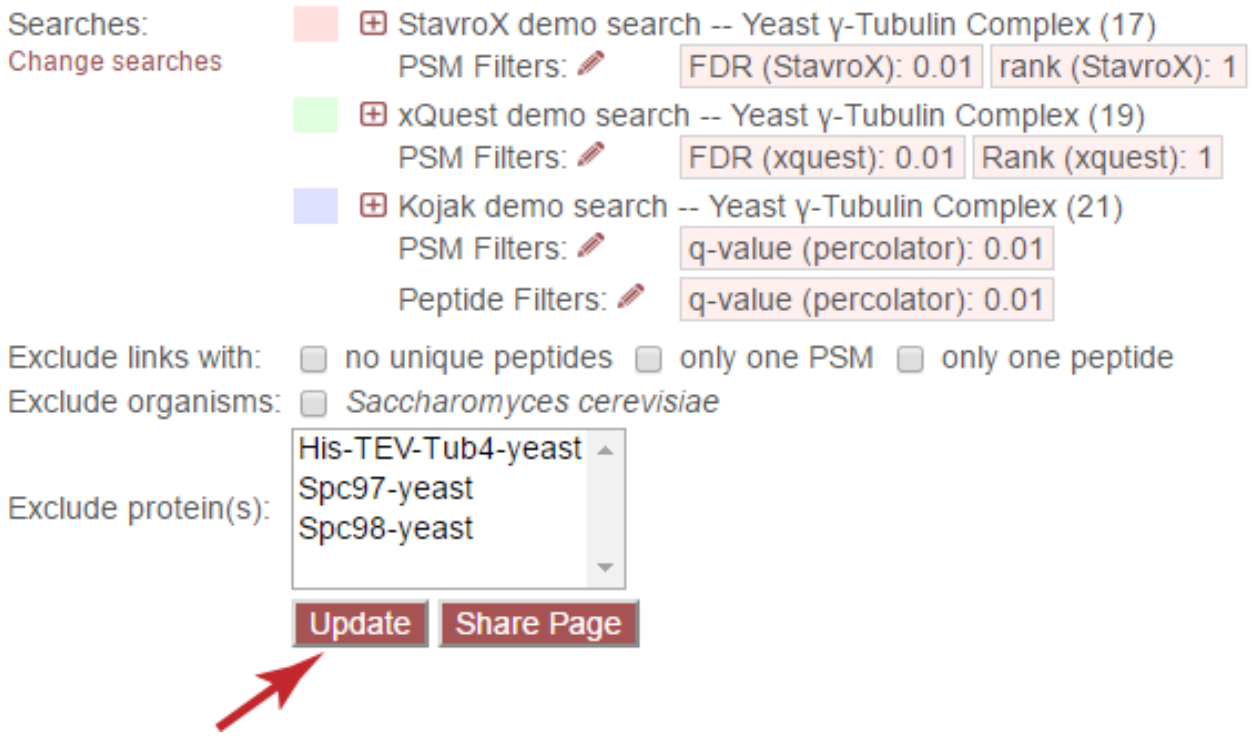


The "Change searches" link allows the user to change which searches are currently being displayed. Clicking the link causes the following overlay to be displayed:



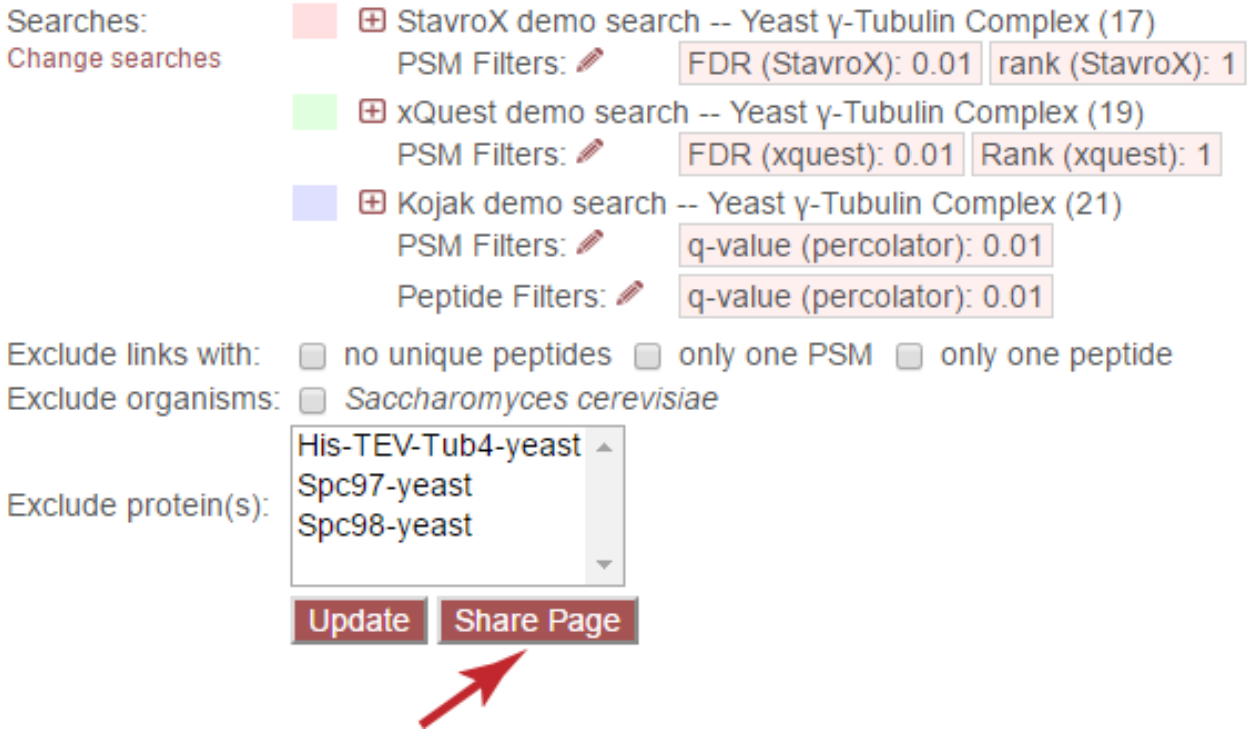
Select or de-select searches by clicking on them in the list. Once done, click "Change" to update the page with the new data or "Cancel" to close the overlay.

10.1.2 Update From Database

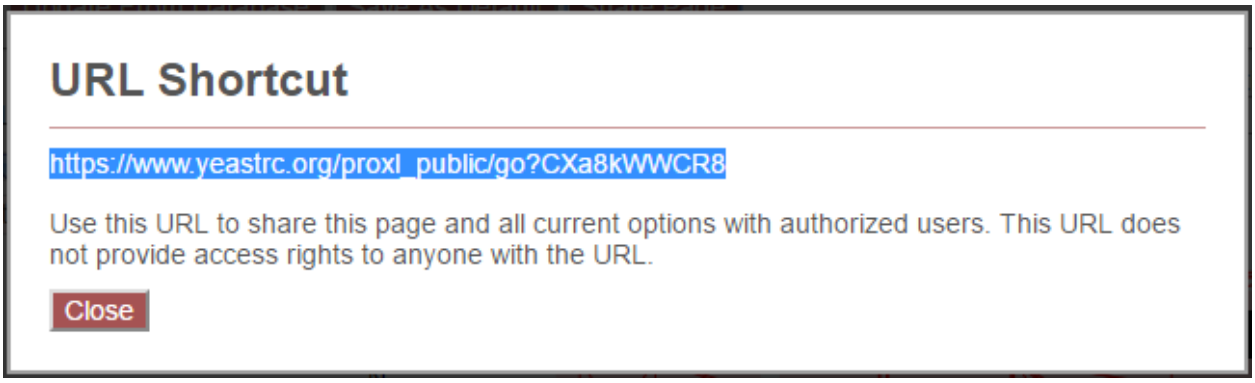


If the user changes any filter parameters–such as PSM/peptide score cutoffs–this button must be clicked to reflect the new filter choices.

10.1.3 Share Page



Clicking the "Share Page" button will generate a shortcut URL for viewing the current page. The shortened URL will appear in an overlay as:



Copying and sharing the highlighted URL will direct users to the view of the page when the URL was generated. Note that this URL does not grant access to the page to any user that would not otherwise have access.

10.2 Search List

The list of merged searches is presented below the top navigation. Each search is shown next to its assigned color for the page, and the color referencing this search is retained in the Euler diagram and in the peptide table. Clicking the [+] icon will expand that search to view details:

10.2.1 Search Details

The "Path" is the location on disk from which the data were imported. The "Linker" is the name of the crossinker used in the experiment. "Search Program(s)" is the name and version number of the PSM search software used. "Upload date" is the date the data were uploaded into proxl. "FASTA file" is the name of the FASTA file used to perform the PSM search.

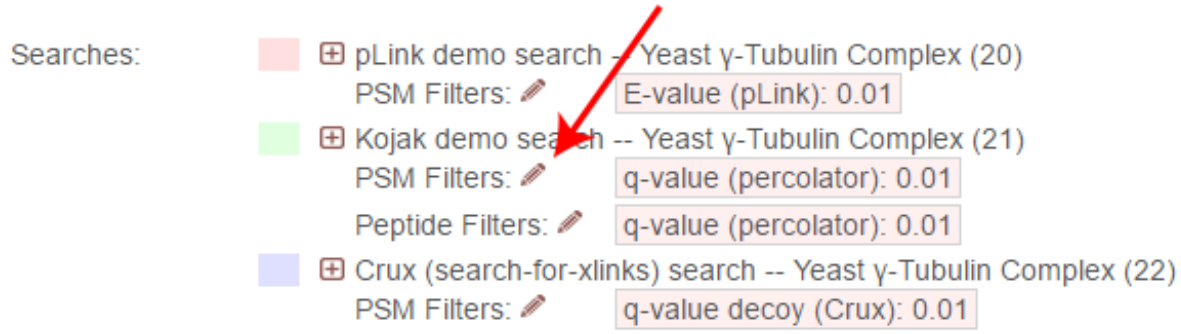
10.2.2 Search Filter

Each search is filtered separately, according to its own native score types. To change the filters for each search, click the pencil icon next to "PSM Filters:" or "Peptide Filters:" next to each search.

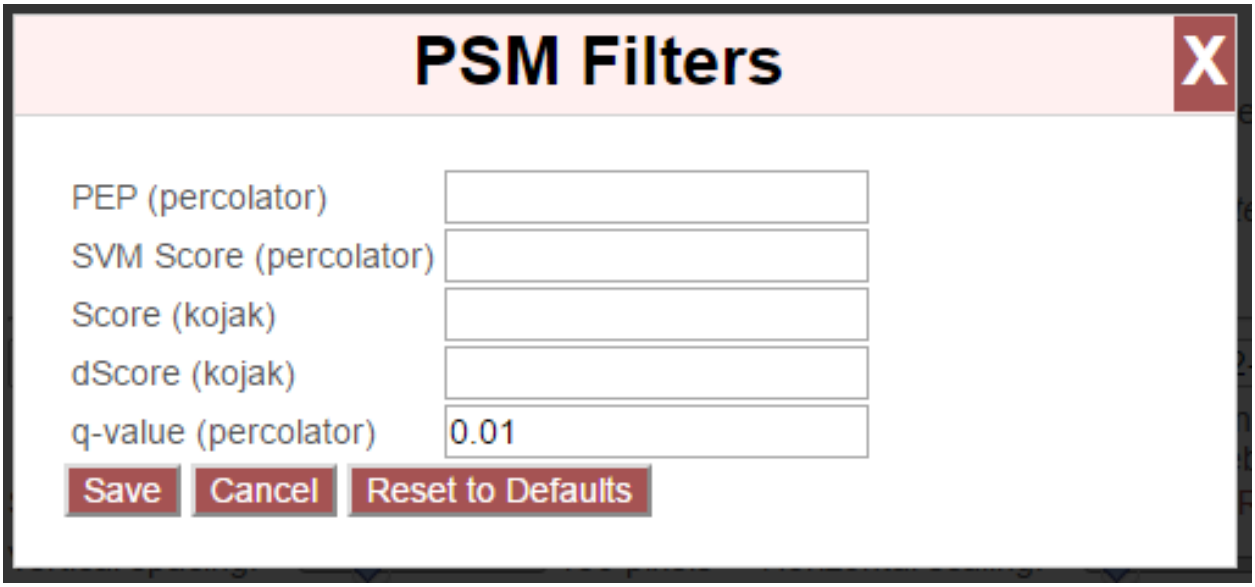
PSM Filters

The filters to apply at the PSM level. Only results which have at least one PSM that meets all of the selected critiera will be listed. When listing PSMs associated with peptides, only PSMs that meet all of the selected critiera will be listed.

To change the PSM-level filters, first click the pencil icon next to "PSM Filters" or the filter tag:



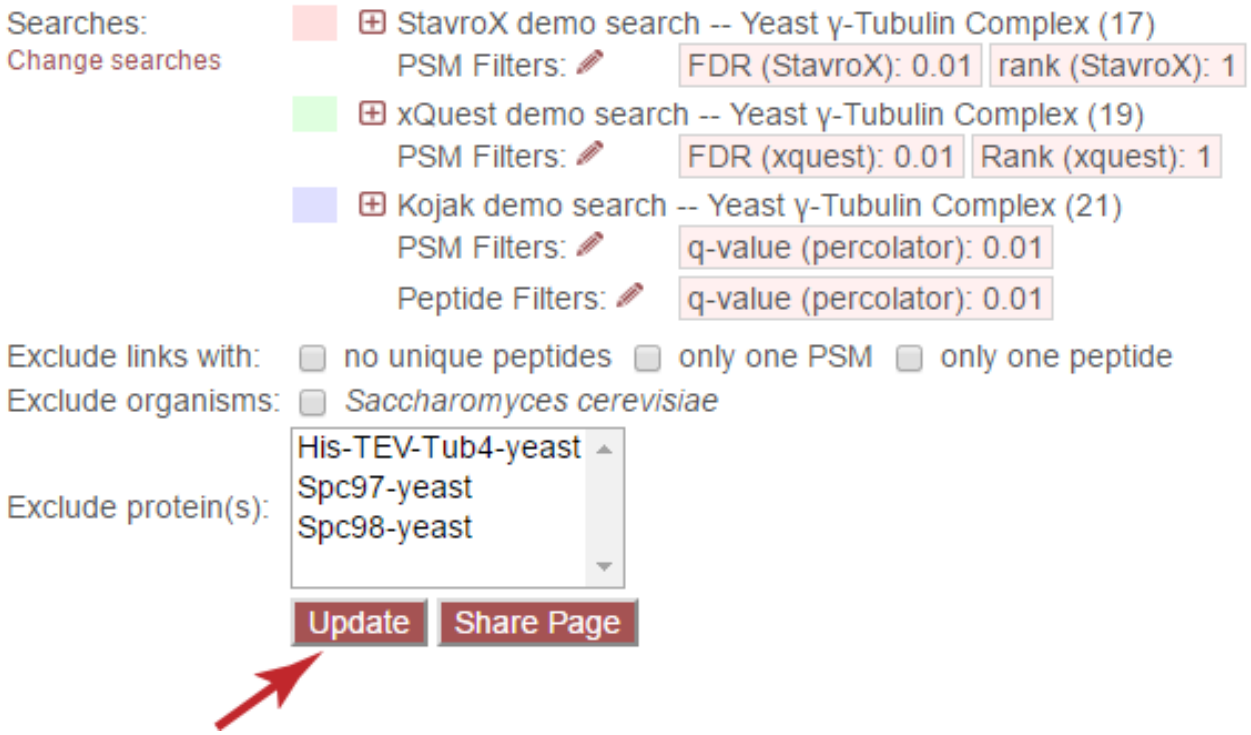
This opens an overlay with the containing the possible score types to use as PSM filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

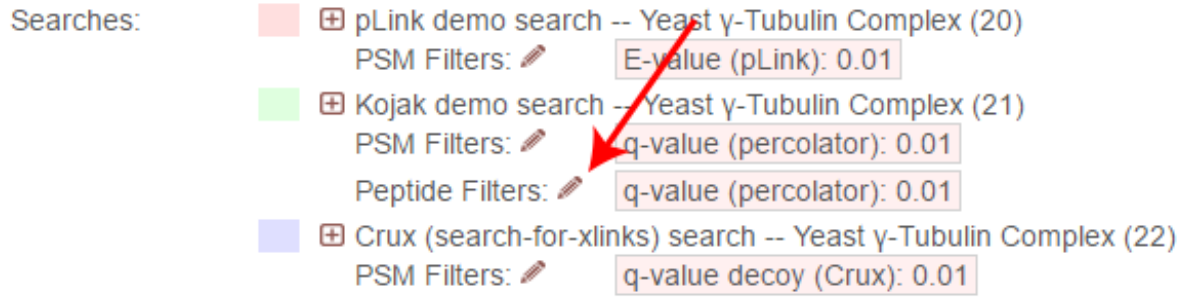
Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update" button on the page to apply any new PSM- or peptide-level filters.



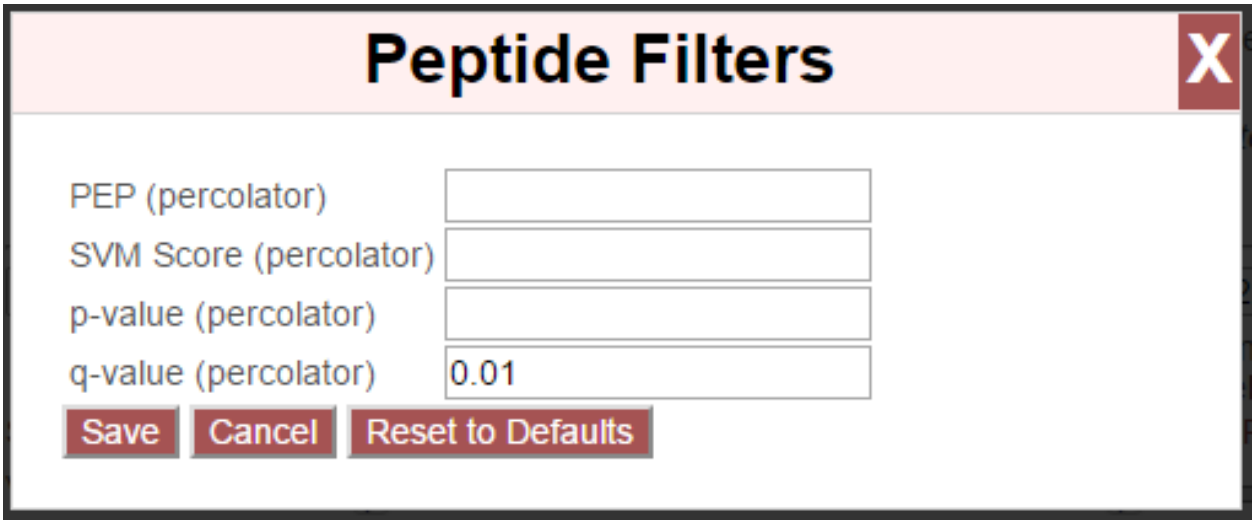
Peptide Filters

Some search algorithms, such as Percolator, provide statistics at the peptide level that may be used for filtering. If applicable, peptide-level filtering options may be set here. Only results which have at least one peptide that meets all of the selected critiera will be listed.

To change the peptide-level filters, first click the pencil icon next to "Peptide Filters":



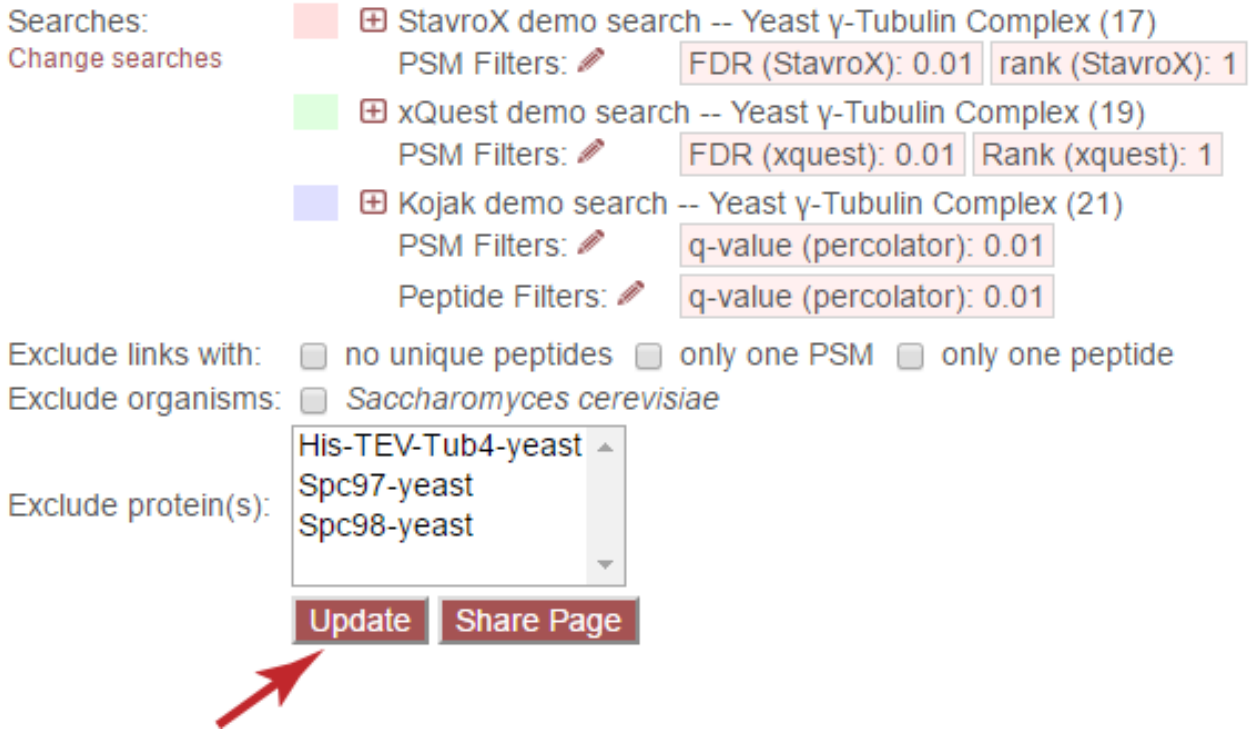
This opens an overlay with the containing the possible score types to use as peptide-level filters for this search. To change the cutoff values to be used for any of these score types, enter the value next to the score type. proxl will correctly handle scores for which larger values are more significant or scores for which smaller values are more signiciant.



To save the new values to the page, click the "Save" button. To cancel, click "Cancel".

The "Reset to Defaults" button will reset the cutoff values to the defaults specified by the proxl XML file uploaded to the database. This typically represents the suggested cutoffs by the author of the respective search program.

Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update" button on the page to apply any new PSM- or peptide-level filters.



10.3 Filter Data

The data presented may be filtered according to the following criteria. Note: Only crosslinks or looplinks that meet ALL the filter criteria are shown.

10.3.1 Exclude links with

Crosslinks or looplinks that have any of the checked attributes will be excluded. The attributes are:

- no unique peptides - If all peptides that ID either one of the crosslinked proteins also map to another protein
- only one PSM - If a given crosslink or looplink was identified by a single PSM
- only one peptide - If a given crosslink or looplink was identifed by a single peptide, where a peptide is the combination of sequence, linked positions, and modifications

10.3.2 Exclude organisms

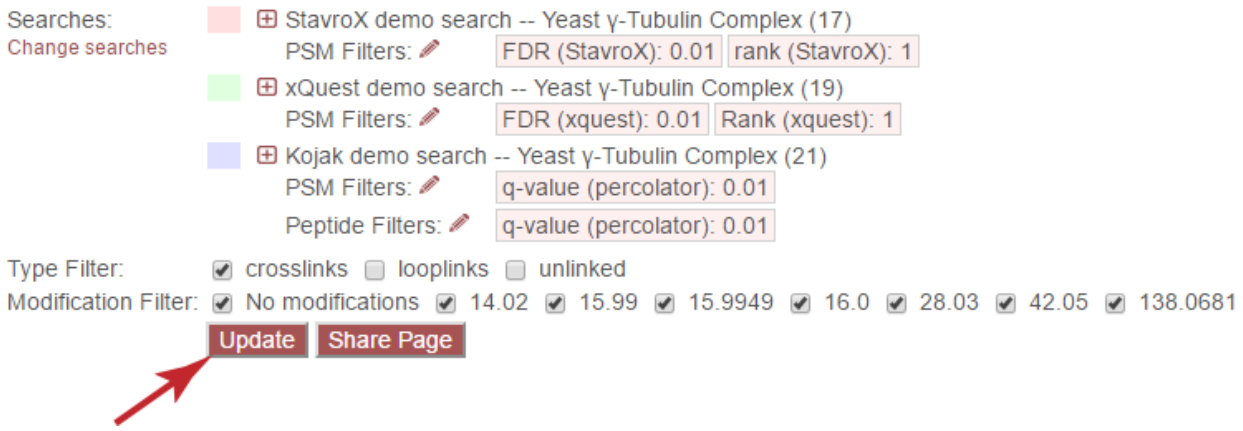
Any links containing a protein that maps to any of the checked organisms will be excluded. The list of organisms presented was gathered by the proteins found in the search. Useful for filtering out groups of contaminant proteins.

10.3.3 Exclude protein(s)

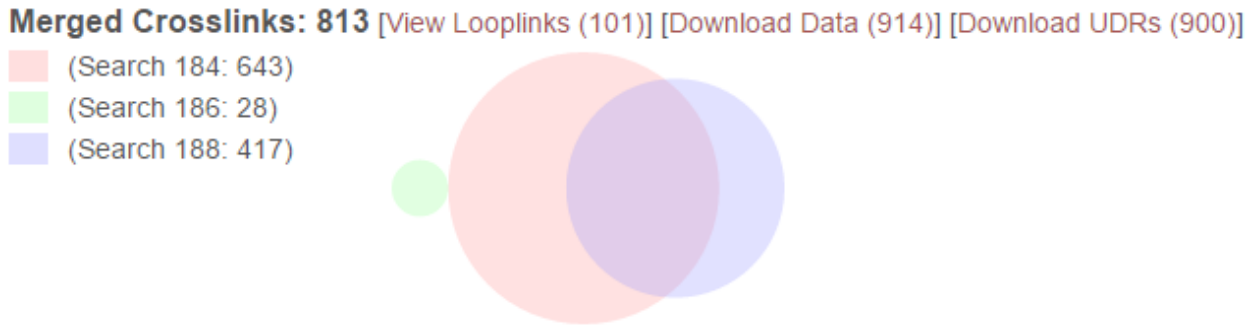
Any links containing a any of the selected proteins will be excluded. Multiple proteins may be selected or unselected using control-click (command-click on the mac) or shift-click. Useful for filtering out individual contaminant proteins.

10.3.4 Update

Important: It is necessary to update the data on the page after changing filter cutoff values. After clicking the "Save" button, you must click the "Update" button on the page to apply any new PSM- or peptide-level filters.



10.4 Euler diagram



The Euler diagram (similar to a Venn diagram) provides a graphical depiction of the relative sizes and overlap between the proteins/UDRs found in the merged searches. The colors in the diagram match the colors used for the search list above. The search list is provided to the left of the diagram with their associated colors as a legend. The labels for each color include the search ID number and the number of crosslink or looplink UDRs found in each of the merged searches. The total number of crosslink or looplink UDRs resulting from the merge is presented in the header above the legend next to "Merged Crosslinks" or "Merged Looplinks".

10.5 View Looplinks

By default, the table shows crosslinks. To switch to looplinks, click the [View Looplinks] link at the top of the table. To view crosslinks again, click the [View Crosslinks] link that appears at the top of the table.

10.6 Download Data

All crosslinks and looplinks that meet the current filtering criteria may be downloaded as tab-delimited text by cliking the [Download Data (#)] link above the table. # indicates the number of rows in the table.

10.7 Download UDRs

UDR stands for "unique distance restraint", which takes its name from 3D modelling terminology. A UDR, in proxl, is any specific position in a protein linked to a specific position in another protein, whether it is a crosslink or a looplink. The [Download UDRs (#)] link downloads a non-redundant tab-delimited text table of these UDRs consolidated from the crosslinks and looplinks. The # is the number of UDRs.

10.8 Table Description

The table presents columns describing the proteins/UDRs and indicates in which of the merged searches they were found. There is one row per UDR. Each row in the table may be clicked on to expand and view the protein-level data by search. Each of these searches may then be clicked on to view peptides, PSMs and spectra from those searches.

10.8.1 Columns

The columns are described below. Note that all column headers may be clicked to toggle between ascending and descending sorting of that column. Holding the shift key while clicking column headers allow sorting on multiple columns.

Search Columns

The first 1-3 columns will be labeled with search ID numbers as headers, and provide an indication for whether or not the UDR in that row was found in that search. If found in that search, the cell for this search in this row will be shaded the same color associated with that search in the Euler diagram and search list at the top of the page. The column will also contain an asterisk. If not found, this cell is empty.

Searches

The number of the merged searches that contain this UDR. The [+] icon indicates that the row may be clicked on to be expanded to show underlying searches in which this UDR was found, the peptides and their statistics, and PSMs and associated spectra.

Protein 1 and 2 (Crosslink-only)

In the case of crosslinks, these are the crosslinked proteins

Position (Crosslink-only)

This is the crosslinked position in the respective proteins, where the first residue is counted as position 1.

Protein (Looplink-only)

In the case of looplinks, this is the looplinked protein

Position 1 and 2 (Looplink-only)

In looplinked proteins, these are the positions in the protein that are linked.

PSMs

The total number of PSMs (peptide spectrum matches) meeting the cutoff that identified either crosslinked (crosslink view) or looplinked (looplink view) peptides that mapped to the reported proteins and positions.

Peptides

The total number of identified crosslinked (crosslink view) or looplinked (looplink view) peptides meeting the filtering critiera that mapped to the reported proteins and positions.

Note: The individual peptides may be viewed by clicking a row in the table to view a table of peptides. Rows in that peptide table may also be viewed to view the underlying PSMs and view spectra.

Unique Peptides

Of the # of peptides, the total number that uniquely mapped to this protein pair (crosslink view) or protein (looplink view).

Best PSM- and Peptide-level Scores

Separate columns, color-coded for each search, display the best PSM- and peptide-level scores currently being used to filter the data from each search.

10.9 View Search-level summary

Clicking on a row for a UDR will expand that row and present search-level data for that UDR–such as in which search(es) it was found, how many peptides were found for it, how many PSMs, and PSM- and peptide-level scores. Clicking on the search rows will expand to reveal underlying peptides.

10.10 View Peptides

All peptides that meet the filtering critiera that were mapped to a protein-level crosslink or looplink may be seen by clicking on the respective row in the search-level summary. Additionally, all rows of this peptide table may clicked to view all PSMs associated with that peptide identification.

10.10.1 Columns

The peptides appear in a table with the following columns:

Reported peptide

The peptide identificaton as it was reported by the respective search program.

Peptide 1 and 2 (Crosslink-only)

The sequences of the two crosslinked peptides.

Pos (Crosslink-only)

The positions in the respective peptides that were crosslinked (starting at 1).

Peptide (Looplink-only)

The sequence of the looplinked peptide.

Pos 1 and 2 (Looplink-only)

The positions in the peptide that were looplinked.

Peptide Scores

The peptide-level scores for this peptide from this search.

PSMs

The number of PSMs that meet the cutoff criteria that identified this peptide.

Best PSM-level Scores

The best PSM-level scores for this search for this peptide for the PSM-level scores currently being used as filtering criteria.

10.11 View PSMs

All PSMs meeting the current filtering criteria may be viewed for a peptide by clicking on a peptide's row.

10.11.1 Columns

The PSMs appear in a table with the following columns:

Scan Num.

The scan number from the spectral file (e.g., mzML file)

Charge

The predicted charge state of the precursor ion.

Obs. m/z

The observed m/z of the precursor ion.

RT (min)

The retention time in minutes.

Scan Filename

The filename of the scan file.

Scores

Each PSM-level score for this PSM from this search are displayed in separate columns.

10.11.2 View Spectra

The annotated mass spectrum may be viewed for any PSM by clicking the "View Spectrum" link. For help on our spectrum viewer, see the *Spectrum Viewer* page.

10.12 Sort Data

All column headers may be clicked to toggle between ascending and descending sorting of that column. Holding the shift key while clicking column headers allow sorting on multiple columns.

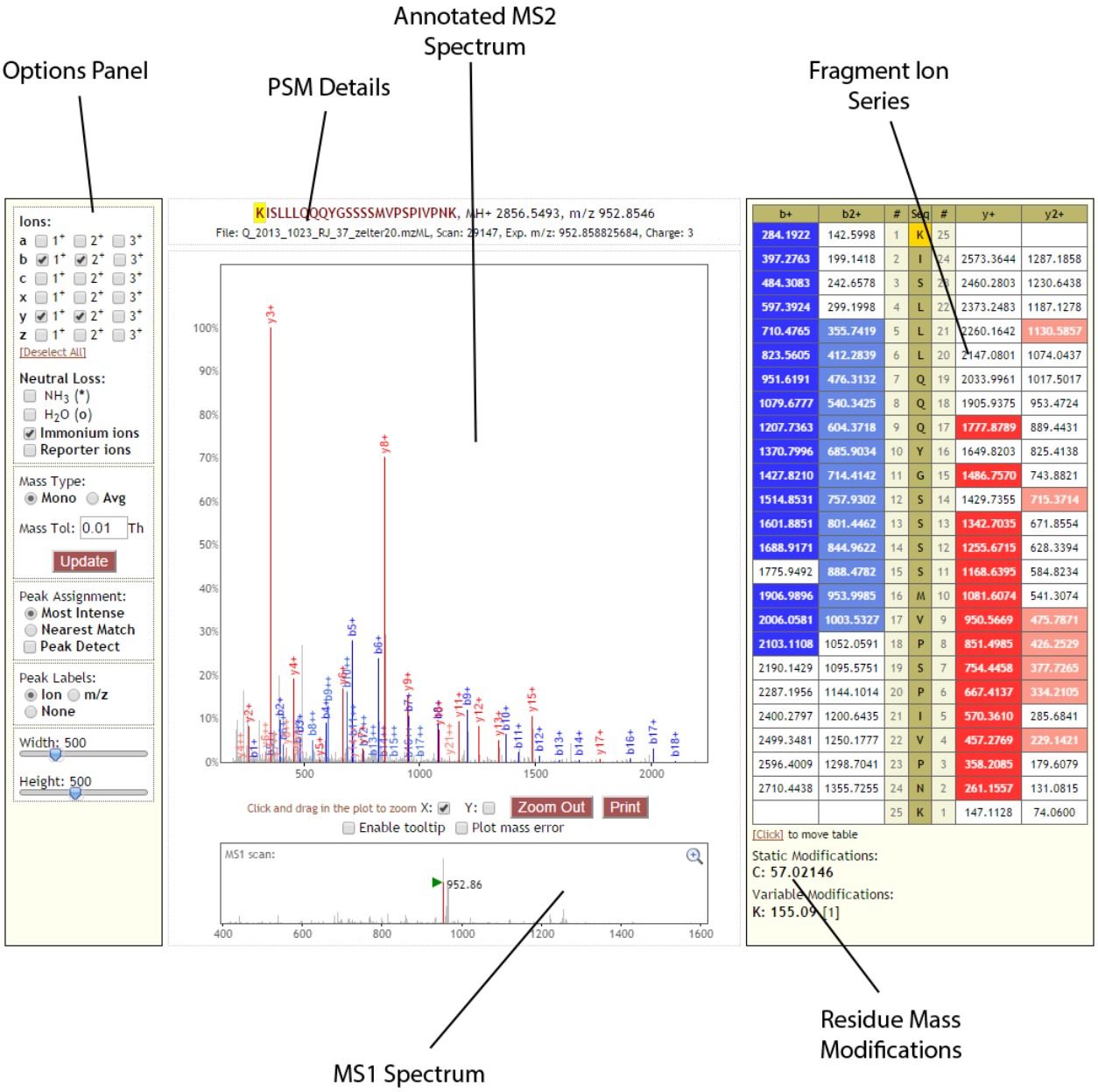
CHAPTER 11

Spectrum Viewer

Proxl uses a version of the Lorikeet spectrum viewer that has been modified to support crosslinked and looplinked spectra. Lorikeet is a pure-HTML and Javascript viewer that requires no 3rd party plugins to use. (Source code of the customized version of Lorikeet may be found here.) Some knowledge of interpreting tandem mass spectra is assumed in this document.

11.1 Overview

Below is a labeled screen shot of the Lorikeet showing the various panels that make up the basic spectrum viewer. In this example, Lorikeet is presenting an unlinked spectrum (not a crosslink or looplink) that has a monolink on the first residue.



b+	b2+	#	Seq	#	y+	y2+
284.1922	142.5998	1	K	25		
397.2763	199.1418	2	I	24	2573.3644	1287.1858
484.3083	242.6578	3	S	23	2460.2803	1230.6438
597.3924	299.1998	4	L	22	2373.2483	1187.1278
710.4765	355.7419	5	L	21	2260.1642	1130.5857
823.5605	412.2839	6	L	20	2147.0801	1074.0437
951.6191	476.3132	7	Q	19	2033.9961	1017.5017
1079.6777	540.3425	8	Q	18	1905.9375	953.4724
1207.7363	604.3718	9	Q	17	1777.8789	889.4431
1370.7996	685.9034	10	Y	16	1649.8203	825.4138
1427.8210	714.4142	11	G	15	1486.7570	743.8821
1514.8531	757.9302	12	S	14	1429.7355	715.3714
1601.8851	801.4462	13	S	13	1342.7035	671.8554
1688.9171	844.9622	14	S	12	1255.6715	628.3394
1775.9492	888.4782	15	S	11	1168.6395	584.8234
1906.9896	953.9985	16	M	10	1081.6074	541.3074
2006.0581	1003.5327	17	V	9	950.5669	475.7871
2103.1108	1052.0591	18	P	8	851.4985	426.2529
2190.1429	1095.5751	19	S	7	754.4458	377.7265
2287.1956	1144.1014	20	P	6	667.4137	334.2105
2400.2797	1200.6435	21	I	5	570.3610	285.6841
2499.3481	1250.1777	22	V	4	457.2769	229.1421
2596.4009	1298.7041	23	P	3	358.2085	179.6079
2710.4438	1355.7255	24	N	2	261.1557	131.0815
		25	K	1	147.1128	74.0600

11.1.1 Options Panel

The options panel includes options for deciding which ions should be drawn, how peaks should be matched, dimensions of the viewer.

Ions

The Ions Panel allows the user to decide which types of ions should be matched and annotated on the spectrum. For example, checking the 1+ and 2+ checkboxes next to "b" be will annotate single and doubly charged b-ions. By default, all b- and y-ions up to the precursor charge minus one (up to +3) are checked. [Deselect All] will un-check all checkboxes.

Neutral Loss

The Neutral Loss panel allows for the optional annotation of ions resulting from the loss of neutrally charged molecules (water or ammonia) from fragment ions. Reporter ions and Immonium ions may also be labeled.

Mass Type

This panel determines how masses are calculated for the theoretical ions to match to the spectrum. "Mono" uses the monoisotopic mass of the fragment ion, "Avg" uses the average mass given natural isotopic abundances. (More Information.)

"Mass Tol" is the tolerance used to find peaks in the spectrum that correspond to the calculated masses of the theoretical fragment ions. A value of 0.01 will search the spectrum within plus or minus 0.01 Th of the calculated mass for matches to the theoretical fragment ion.

Peak Assignment

If more than one peak is found in the spectrum within the specified tolerance, this panel determines which of those peaks will be annotated as the fragment ion. "Most Intense" will label the most intense peak in the window and "Nearest Match" will label the peak with a m/z closest to the calculated m/z for the fragment ion.

Enabling "Peak Detect" will apply a filtering algorithm to the spectrum prior to attempting to match peaks. The displayed spectrum remains unchanged, this algorithm is applied for peak-matching purposes only. This algorithm is:

1. Keep 50 most intense peaks
2. If a peak is the most intense peak +/- 50m/z of itself, and there are fewer than 11 peaks in that window, keep it.
3. If this peak's intensity is >= 2 standard deviations above the mean in this window, keep it.

Peak Labels

These options determine how annotated peaks in the spectrum are labeled. Choosing "Ion" labels peaks according to fragment ion type, number, and charge (e.g., y8++ would be y-ion 8 with a +2 charge). Choosing "m/z" labels peaks according to the m/z of that peak in the spectrum. Choosing "None" removes the labels.

Width and Height

These sliders adjust the dimensions of the viewer.

11.1.2 PSM Details

The top row of this panel first gives the sequences of the matched peptide. Residues containing modifications are highlighted. Then the calculated MH+ (+1 charge) mass of the peptide is given. Then, the calculated m/z of the peptide is given, given the computed charge.

The bottom row gives the name of the spectral file in which this spectrum was found, the scan number in that file corresponding to this spectrum, the observed m/z of the parent ion in the MS1 scan, and the calculated charge.

11.1.3 Annotated MS2 Spectrum

This panel contains a rendering of the tandem mass spectrum, with m/z on the X axis and the intensity (relative to the most intense peak) on the y axis. Peaks that were matched to predicted ions are labeled and colored according to their type and charge, which matches the coloring in the Fragment Ion Series Panel.

Zooming

It is possible to zoom on the X and Y axis. By default, users may click and drag on the X axis to define a new range for the X axis. To define a new Y axis, check the "Y:" checkbox and click and drag to define a new Y axis maximum. Note: if both X and Y are checked, clicking and dragging will simultaneously define a new X range and a new Y maximum.

Print

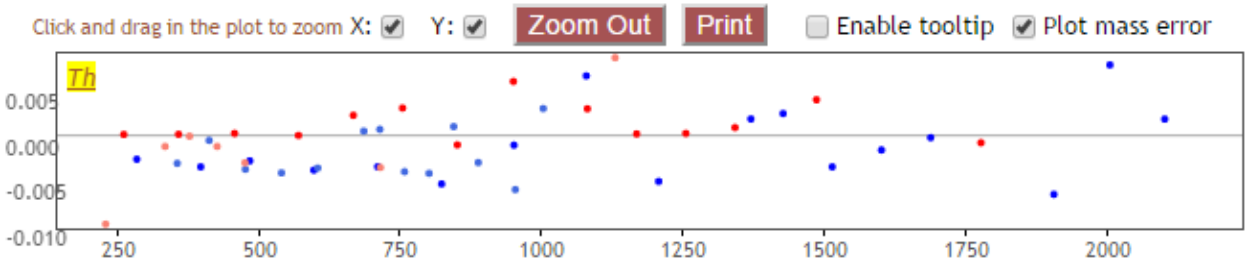
Clicking the "Print" button opens the system print dialog containing a rendering of the PSM Details, Annotated MS2 Spectrum, and MS1 Spectrum panels.

Enable tooltip

If this option is selected, a tooltip will appear when mousing over peaks in the spectrum that shows m/z and intensity of that peak.

Plot mass error

If this option is selected, a plot will be presented below the checkbox that shows the difference of actual and theoretical mass of the matched ions:



11.1.4 MS1 Spectrum

This panel shows which peak from the MS1 scan was chosen for fragmentation resulting in the shown MS2 spectrum.

Zooming

It is possible to zoom on the X and Y axis. Click and drag to define a new range for the X-axis and a new maximum for the Y-axis.

11.1.5 Fragment Ion Series

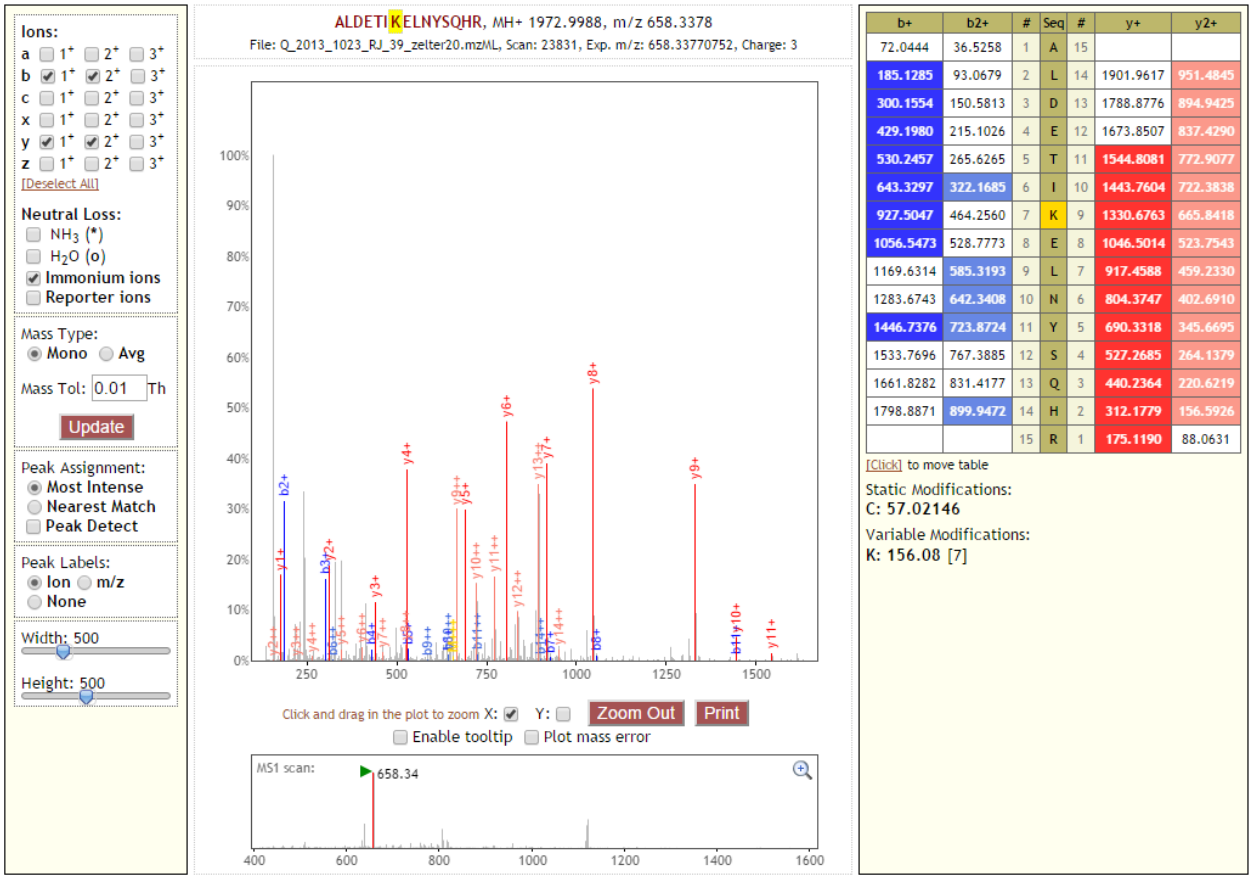
This panel displays the calculated theoretical masses for the currently-selected ion types for the current peptide. The peptide sequence is displayed top-to-bottom for N-to-C terminus. Modified residues are highlighted. Ion types corresponding to the N-terminal side of fragmentation are displayed on the left side of the sequence, and the C-terminal side on the right-side of the sequence. Cells corresponding to matched peaks are colored, with those colors based on the type and charge of the ion, and match the colors in the annotated MS2 spectrum.

11.1.6 Residue Mass Modifications

Modifications present in the PSM are shown in this panel. "Static Modifications" are mass modifications applied to all instances of the given residue in the PSM search. "Variable Modifications" mass modifications that may or may not be present on the given residue during the search. The position of the variable modification in the peptide sequence is given in brackets.

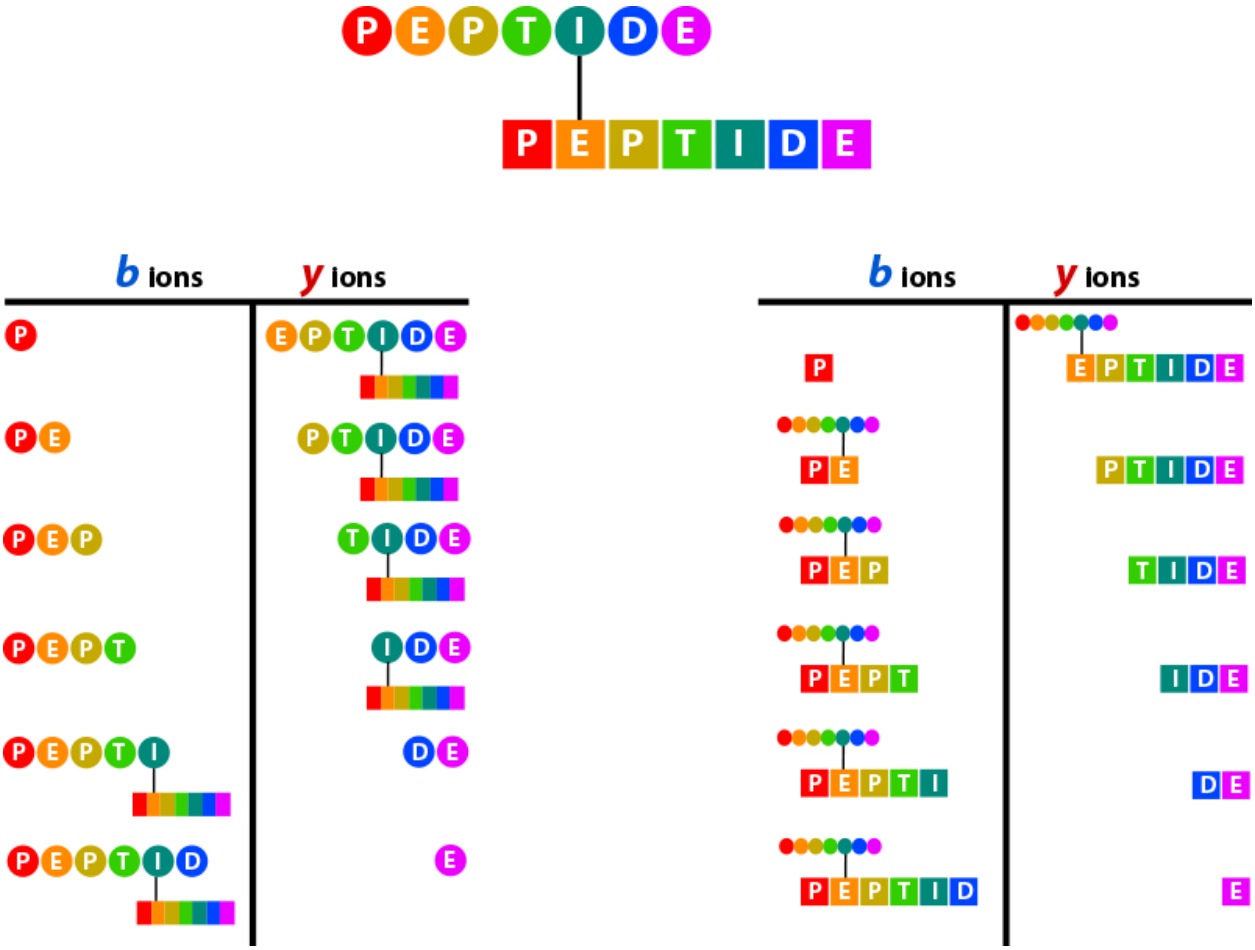
11.2 Monolink Spectrum

Monolinks in proxl are treated as modifications on residues in the same way as other post-translational modifications. There may be multiple monolinks present in the sequence, and monolinks may appear in unlinked, crosslinked or looplinked peptides. In the example below, the modification on the lysine at position 7 has a mass of 156.08, the mass of the crosslinker used in this experiment when it has bound to an amino acid on one end (but not the other). This residue is highlighted in the "PSM Details" area and in the sequence presented in the "Fragment Ion Series" panel. The modification is also listed in the "Residue Mass Modifications" below the "Fragment Ion Series" panel.

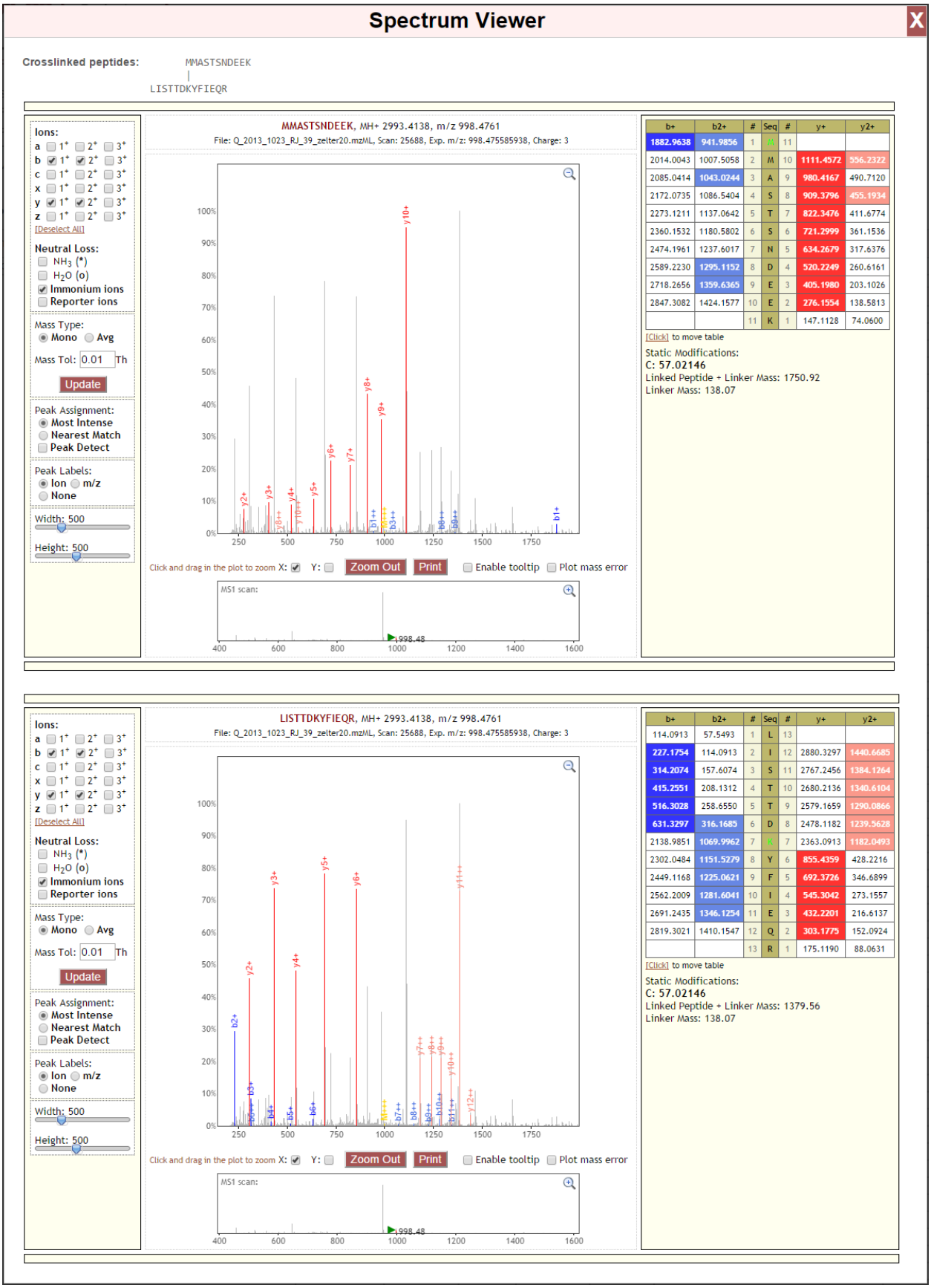


11.3 Crosslink Spectrum

Crosslinks between peptides may, from the point of view of each peptide, be thought of a large mass modification on the linked residue equal to the mass of the reacted crosslinker plus the mass of the other peptide. This is illustrated in the figure below. The hypothetical b- and y-ion series for the "circle" peptide and "square" peptide are given. Note that the opposite peptide is present as a modification on the linked residue in each peptide.



Proxl displays the ion series for each of the linked peptides separately. The sequences and positions of the crosslink are presented graphically at the top of the window. In the sequence presented in the "Fragment Ion Series" panel for each peptide, the linked residue is highlighted green.

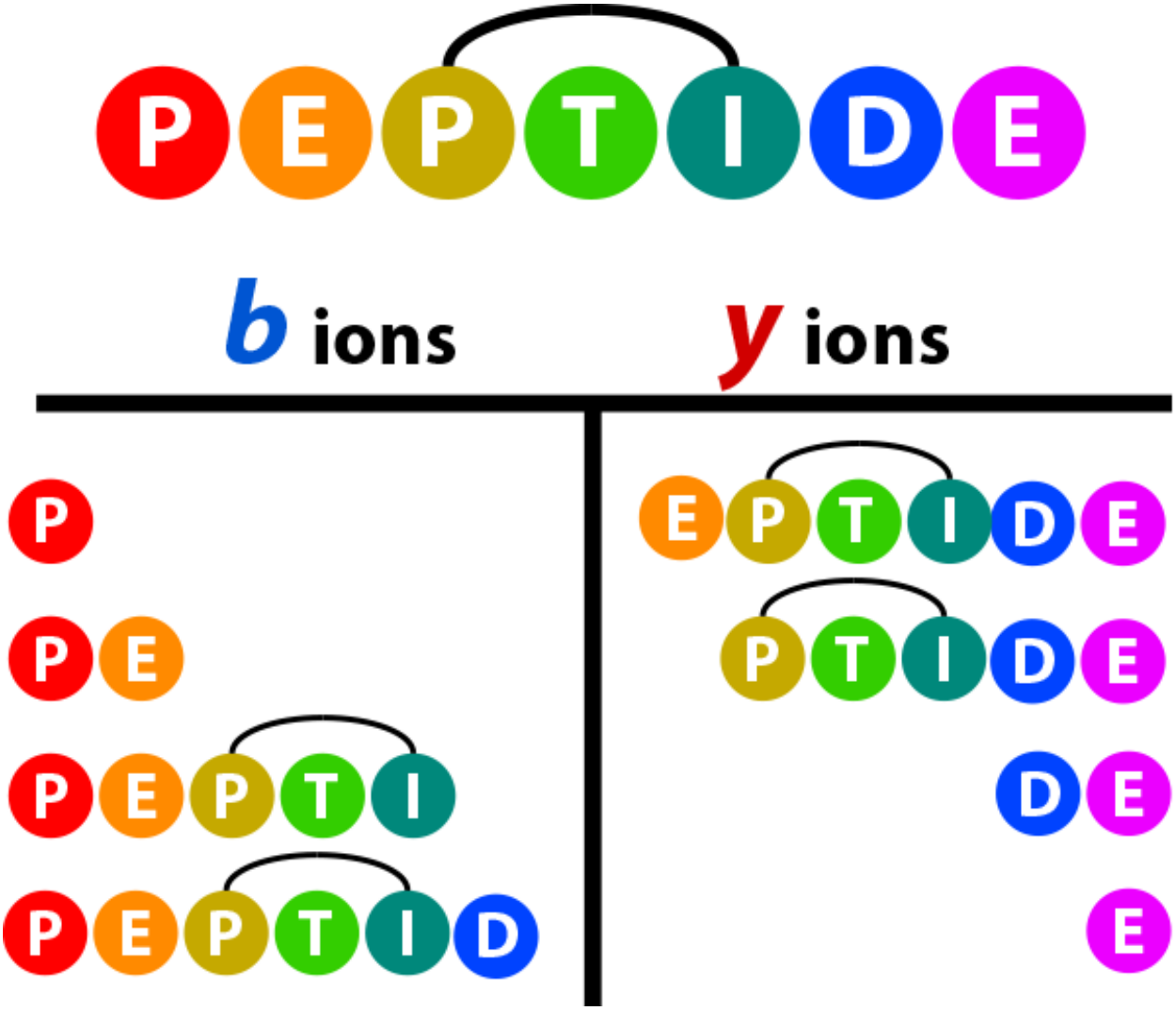


b+	b2+	#	Seq	#	y+	y2+
1882.9638	**941.9856**	1	M	11		
2014.0043	1007.5058	2	M	10	**1111.4572**	**556.2322**
2085.0414	**1043.0244**	3	A	9	**980.4167**	490.7120
2172.0735	1086.5404	4	S	8	**909.3796**	**455.1934**
2273.1211	1137.0642	5	T	7	**822.3476**	411.6774
2360.1532	1180.5802	6	S	6	**721.2999**	361.1536
2474.1961	1237.6017	7	N	5	**634.2679**	317.6376
2589.2230	**1295.1152**	8	D	4	**520.2249**	260.6161
2718.2656	**1359.6365**	9	E	3	**405.1980**	203.1026
2847.3082	1424.1577	10	E	2	**276.1554**	138.5813
		11	K	1	147.1128	74.0600

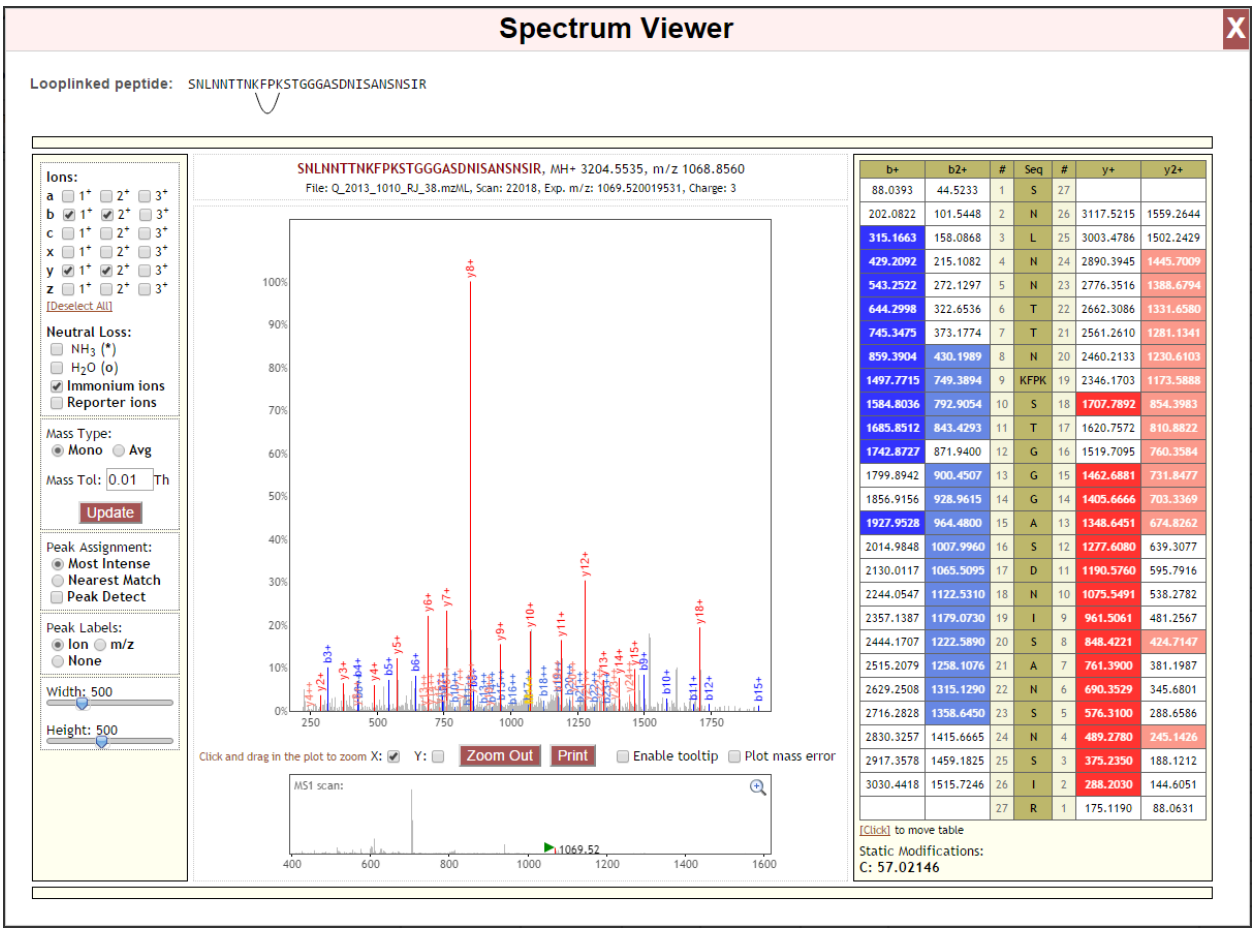
b+	b2+	#	Seq	#	y+	y2+
114.0913	57.5493	1	L	13		
227.1754	114.0913	2	I	12	2880.3297	**1440.6685**
314.2074	157.6074	3	S	11	2767.2456	**1384.1264**
415.2551	208.1312	4	T	10	2680.2136	**1340.6104**
516.3028	258.6550	5	T	9	2579.1659	**1290.0866**
631.3297	**316.1685**	6	D	8	2478.1182	**1239.5628**
2138.9851	**1069.9962**	7	R	7	2363.0913	**1182.0493**
2302.0484	**1151.5279**	8	Y	6	**855.4359**	428.2216
2449.1168	**1225.0621**	9	F	5	**692.3726**	346.6899
2562.2009	**1281.6041**	10	I	4	**545.3042**	273.1557
2691.2435	**1346.1254**	11	E	3	**432.2201**	216.6137
2819.3021	1410.1547	12	Q	2	**303.1775**	152.0924
		13	R	1	175.1190	88.0631

11.4 Looplink Spectrum

Looplink peptides contain a crosslinker that has linked two residues in the single peptide. When calculating ion series, proxl treats the sub-sequence between the linked residues (inclusive) as a single unit, as cleavages between the linked residues would result in crosslinked peptides–not a looplinked peptide. As a result, a hypothetical b- and y-ion series for a looplinked peptide would look be as follows:



The subunit "PTI" is treated as a single residue normally would be when calculating the theoretical ion series. Below is an example spectrum displayed in Lorikeet that treats the looplinked subunit as a single entity. Note when moving to b9/y19, the mass is increased by the sum of KFPK plus the crosslinker.


Spectrum Viewer
Looplinked peptide: SNLNNTTNKFPKSTGGGASDNISANSNSIR
Ions:
a 1+ 2+ 3+
b 1+ 2+ 3+
c 1+ 2+ 3+
x 1+ 2+ 3+
y 1+ 2+ 3+
z 1+ 2+ 3+
[Deselect All]
Neutral Loss:
NH3 (*)
H2O (o)
Immonium ions
Reporter ions
Mass Type:
Mono Avg
Mass Tol: 0.01 Th
Update
Peak Assignment:
Most Intense
Nearest Match
Peak Detect
Peak Labels:
Ion m/z
None
Width: 500
Height: 500
SNLNNTTNKFPKSTGGGASDNISANSNSIR, MH+ 3204.5535, m/z 1068.8560
File: Q_2013_1010_RJ_38.mzML, Scan: 22018, Exp. m/z: 1069.520019531, Charge: 3
Click and drag in the plot to zoom X: Y: Zoom Out Print Enable tooltip Plot mass error
MS1 scan:
1069.52
b+ b2+ # Seq # y+ y2+
88.0393 44.5233 1 S 27
202.0822 101.5448 2 N 26 3117.5215 1559.2644
315.1663 158.0868 3 L 25 3003.4786 1502.2429
429.2092 215.1082 4 N 24 2890.3945 1445.7009
543.2522 272.1297 5 N 23 2776.3516 1388.6794
644.2998 322.6536 6 T 22 2662.3086 1331.6580
745.3475 373.1774 7 T 21 2561.2610 1281.1341
859.3904 430.1989 8 N 20 2460.2133 1230.6103
1497.7715 749.3894 9 KFPK 19 2346.1703 1173.5888
1584.8036 792.9054 10 S 18 1707.7892 854.3983
1685.8512 843.4293 11 T 17 1620.7572 810.8822
1742.8727 871.9400 12 G 16 1519.7095 760.3584
1799.8942 900.4507 13 G 15 1462.6881 731.8477
1856.9156 928.9615 14 G 14 1405.6666 703.3369
1927.9528 964.4800 15 A 13 1348.6451 674.8262
2014.9848 1007.9960 16 S 12 1277.6080 639.3077
2130.0117 1065.5095 17 D 11 1190.5760 595.7916
2244.0547 1122.5310 18 N 10 1075.5491 538.2782
2357.1387 1179.0730 19 I 9 961.5061 481.2567
2444.1707 1222.5890 20 S 8 848.4221 424.7147
2515.2079 1258.1076 21 A 7 761.3900 381.1987
2629.2508 1315.1290 22 N 6 690.3529 345.6801
2716.2828 1358.6450 23 S 5 576.3100 288.6586
2830.3257 1415.6665 24 N 4 489.2780 245.1426
2917.3578 1459.1825 25 S 3 375.2350 188.1212
3030.4418 1515.7246 26 I 2 288.2030 144.6051
27 R 1 175.1190 88.0631
[Click] to move table
Static Modifications:
C: 57.02146

CHAPTER 12

Installing Proxl

Follow these steps to set up your own installation of Proxl on your own computer. These instructions include running Proxl with default settings and require only minimal configuration by the user. If you already have access to Proxl (e.g. at https://www.yeastrc.org/proxl_public/), you do not need to do this to use Proxl.

This tutorial assumes you have Docker installed on your system. Please see our Docker Installation Tutorial to get Docker installed.

Important: System Requirements: Proxl will consume a large amount of RAM, particularly when uploading data. You should have at least 6 gigabytes of RAM available on your system.

12.1 1. Open a Terminal

On Linux and MacOS, open a normal terminal. On Windows, if you followed our instructions for installing Docker, follow the directions on our Docker Installation Tutorial to open a Linux terminal.

12.2 2. Install Docker Compose

Docker Compose is an official add-on to Docker that greatly simplifies running applications that have multiple parts. Proxl has several parts, including a database, multiple web applications, and running programs. Docker Compose allows you to run a single command to launch and correctly stitch all of those components together into a working system. This all happens inside of Docker and does not install the software elsewhere on your computer.

If you are on MacOS, you will likely already have Docker Compose installed. If you are on Linux (including Windows users who installed Docker according to our instructions), test if Docker Compose is installed by typing `docker-compose`. If the command is not found, please install Docker Compose by typing the following:

```
sudo curl -L "https://github.com/docker/compose/releases/download/1.29.1/
↪docker-compose-$(uname -s)-$(uname -m)" -o /usr/local/bin/docker-compose
sudo chmod +x /usr/local/bin/docker-compose
```

If you run into any trouble installing Docker Compose, please see Docker's official Docker Compose installation instructions.

12.3 3. Download Proxl Install Files

First set up a Proxl project directory:

```
mkdir ~/proxl
cd ~/proxl
```

Now, download the Proxl Docker Compose files:

```
# Download Proxl Docker Compose files
curl -L "https://github.com/yeastrc/proxl-web-app/releases/latest/download/
↪docker-compose-files.tgz" -o docker-compose-files.tgz

# Expand the archive
tar -xvzf docker-compose-files.tgz
```

Note: If you prefer to download a ZIP file or if you prefer to download the file another way, the latest release can be found on GitHub at https://github.com/yeastrc/proxl-web-app/releases/latest

12.4 4. Configure Proxl

Copy the sample configuration file into place:

```
cp docker/env-sample ./.env
```

What follows below in this step is optional. Your installation of proxl will be more secure if you change the default passwords. The instructions below describe how to do this.

The `.env` file holds all of the necessary configuration for Proxl. It is recommended (but not required) that you change the first two lines of the file, which contain passwords to be used for the MySQL database.

The `.env` file should look something like this:

```
# .env file for supplying settings to initializing proxl using docker-compose

# Change these passwords.
MYSQL_ROOT_PASSWORD=change_this_password
MYSQL_PASSWORD=change_this_password

# Can change the mysql user proxl uses, but not necessary
MYSQL_USER=proxl_db_user

# name of the proxl database
PROXL_DATABASE_NAME=proxl
```

(continues on next page)

(continued from previous page)

```
# Used by importer
PROXL_WEB_APP_BASE_URL=http://proxl:8080/proxl/

# This manages the memory usage of components of proxl
IMPORTER_JAVA_OPTIONS=-Xmx3g -Xms500m
WEBAPP_JAVA_OPTIONS=-Xms2024m -Xmx2024m

# This manages optimization settings for MySQL
MYSQL_OPTIONS=--max-connections=500 --skip-ssl

# Settings for setting up sending of emails by proxl
SMTP_HOST=smtp.example.com
SMTP_PORT=587
SMTP_USERNAME=smtp_username
SMTP_PASSWORD=smtp_password
```

These can be changed using your favorite text editor. On Linux (including Docker on Windows), we'll assume that is `nano`. To edit the file, type:

```
nano .env
```

Change the passwords and type `Control-o`, `<ENTER>`, and `Control-x` to save and exit.

Important: By default, Docker manages where data are stored on your disk. If you would like to customize where proxl stores data, please follow our *Customize Data Locations* tutorial. This should be done before continuing on to Step 6 below. Once that is complete, proceed to Step 6.

12.5 5. Starting and Stopping Proxl

At this point, starting and stopping Proxl should be straight forward.

To start Proxl:

```
sudo docker-compose up --detach
```

To stop Proxl:

```
sudo docker-compose down
```

Note: If you are using **Windows**, ensure Docker is running by typing:

```
sudo service docker start
```

You should now be able to start Proxl.

Note: The first time you start Proxl, all of the components will download and the database will initialize. This may take a few minutes, depending on your download speed. Subsequent startups of Proxl will not require these steps and will be faster.

Note: These commands must be typed while you are in the project code directory. If you followed these instructions, you can ensure you are in this directory by typing:

```
cd ~/proxl
```

12.6 6. Connect to Your Proxl Installation

Point your web browser to to access Proxl running on your own computer!

Note: If this is the first time bringing up Proxl, it may take a minute for the database to initialize. If you see message saying there was a problem with your request, try again in about a minute.

12.6.1 Login with Default User

By default, you can log in using `admin` as the username and `changeme` as the password.

12.6.2 (Optional) Change Default User Information

For security reasons, you should change the default log in information of the admin user. To change the default information click on the Account Settings icon in the top right of the page:

Change the name, username, and password in the form to your liking.

12.6.3 Start Using Proxl

That's it, you are ready to use Proxl!

12.7 7. (Optional) Set up SMTP For Emails

Some functions of Proxl require sending email to users. Examples of this include inviting new users to projects, resetting forgotten passwords, and notifications that data uploads have been completed. Although it's not required that you set up SMTP, the above features will not be enabled unless you do. If you would like to enable these features, please see our *Proxl SMTP Setup Guide*.

If you do not set up SMTP, you must use the administrative interface to add new users to proxl. The manage users icon can be found at the top-right after logging in as an administrative user.

CHAPTER 13

Customize Data Locations

Note: This tutorial assumes you have completed our *Installing Proxl* tutorial through step 4.

By default, our installation tutorial will allow Docker to manage where proxl stores its data. This includes things like where MySQL stores its data files, where uploaded scans are stored, and working directories for processing uploaded data. On Linux (including Windows running Ubuntu), these data will mostly likely be kept under `/var/lib/docker/`.

It is recommended that you let Docker manage the data directories if you can. However, if you would like to customize where the data are stored for proxl, follow the steps below.

13.1 1. Create data directories

You will need to create five directories for proxl to store its data.

1. MySQL data directory. This is the directory used to store the database.
2. Spectr upload directory. This the directory used for spectra processing.
3. Spectr storage directory. This the directory used to store spectra.
4. Proxl upload directory. This is the directory where uploads are temporarily stored.
5. Proxl cache directory. Cache some results on disk to speed up the web site.

Important: If you are using WSL2 on Windows, specifying a Windows filesystem drive (e.g., `/mnt/d/`) for your data directories is not supported.

For example, if you would like store store all data in the `/data/proxl-data` directory, you would type the following:

```
# make a parent directory for proxl data
sudo mkdir -p /data/proxl-data

# make the five directories for storing data
sudo mkdir /data/proxl-data/mysql
sudo mkdir /data/proxl-data/spectr-upload
sudo mkdir /data/proxl-data/spectr-storage
sudo mkdir /data/proxl-data/proxl-upload
sudo mkdir /data/proxl-data/proxl-cache
```

13.2 2. Update `.env` with data storage locations

The `.env` configuration file will need to be updated to include the locations of the data directories. Open this file using your favorite text editor. On Linux (including Docker on Windows), we'll assume that is `nano`. To edit the file, type:

```
# ensure you are in correct directory. if you followed tutorial type:
cd ~/proxl

# edit the file
nano .env
```

Add the following lines to the end of the file. Substitute the actual directories with directories you chose above. This example uses the example directory names:

```
MYSQL_DATA_DIRECTORY=/data/proxl-data/mysql
SPECTR_UPLOAD_DIRECTORY=/data/proxl-data/spectr-upload
SPECTR_STORAGE_DIRECTORY=/data/proxl-data/spectr-storage
PROXL_UPLOAD_DIRECTORY=/data/proxl-data/proxl-upload
PROXL_CACHE_DIRECTORY=/data/proxl-data/proxl-cache
```

Type `Control-o`, `<ENTER>`, and `Control-x` to save and exit `nano`.

13.3 3. Starting and Stopping Proxl

Important: The commands below are different than the commands for starting and stopping Proxl on our *Installing Proxl* tutorial! You must always use these commands if you have customized the data locations.

At this point, starting and stopping proxl should be straight forward.

To start proxl:

```
sudo docker-compose -f docker-compose-custom-data.yml up --detach
```

To stop proxl:

```
sudo docker-compose -f docker-compose-custom-data.yml down
```

Note: If you are using **Windows**, ensure Docker is running by typing:

```
sudo service docker start
```

You should now be able to start Proxl.

Note: The first time you start proxl, all of the components will download and the database will initialize. This may take a few minutes, depending on your download speed. Subsequent startups of proxl will not require these steps and will be faster.

Note: These commands must be typed while you are in the project code directory. If you followed these instructions, you can ensure you are in this directory by typing:

```
cd ~/proxl
```

13.4 4. Proceed with installation

You should now proceed to step 6 in our *Installing Proxl* tutorial. However, recall that your command for stopping and starting is different than that listed in the tutorial. (See above.)

CHAPTER 14

Proxl SMTP Setup Guide

This guide assumes you have completed all the steps in the *Installing Proxl* tutorial.

Setting up SMTP allows proxl to send emails to users. This enables you to invite researchers to projects using their email, users to reset forgotten passwords and to receive notifications when their file uploads are complete. Setting up SMTP is not required to run proxl, but these features will not be available.

14.1 1. Acquire SMTP Relay Server Information

This is, potentially the most complex part of enabling SMTP in proxl. Acquiring SMTP relay server information means finding the **host name**, **port**, **username**, and **password** to use to send email through an email server. Below we present some options on how to locate or set up SMTP relay server information.

14.1.1 Check With Your Organization

The best place to begin this process is checking with your organization's IT department or your internet service provider to see if this service is available to you. If it is, it would potentially look something like:

Table 1: Hypothetical SMTP Relay Server Information

Host Name	smtp.organization.com
Port	587
Username	my_username
Password	my_password

14.1.2 SendGrid

SendGrid is a free service that allows applications like proxl to send emails through their servers. The free tier allows you to send up to 100 emails per day. Here are the rough steps for setting up email sending capabilities:

1. Go to https://sendgrid.com/.

2. Create a free account
3. Establish a single sender, which is an email address from which you will be allowed to send emails through SendGrid. SendGrid will require that you verify you can receive emails at this address. **This does not give SendGrid access to your email.** It only allows you to send email **from this address** using SendGrid's servers. See SendGrid's Guide for Single Sender Verification.
4. Click on the `Email API` navigation option on the left and choose `Integration Guide`. Click on the `SMTP Relay` option that appears on the page.
5. Follow the directions to acquire your SMTP relay server information.

Your SendGrid SMTP relay server information will look something like:

Table 2: SendGrid SMTP Relay Server Information

Host Name	smtp.sendgrid.net
Port	587
Username	apikey
Password	YOUR_API_KEY

14.1.3 Google SMTP Relay

Google allows your Google Workspace account to use their servers as a SMTP relay. To enable this, follow Google's Guide for setting up SMTP relay. In the `Authentication` section, you want to enable `Require SMTP Authentication`.

Your Google SMTP relay server information will look something like:

Table 3: Google SMTP Relay Server Information

Host Name	smtp-relay.gmail.com
Port	587
Username	Google username
Password	Google password

14.1.4 Other Options

There are other service on the internet that provide SMTP relay server information. Any of them should work, so long as you have a **host name**, **port**, **username**, and **password**.

14.2 2. Update Your `.env` File

The `.env` file you set up during the *Installing Proxl* tutorial should contain the following lines (among others):

```
# Settings for setting up sending of emails by proxl
SMTP_HOST=smtp.example.com
SMTP_PORT=587
SMTP_USERNAME=smtp_username
SMTP_PASSWORD=smtp_password
```

Open this file using your favorite text editor. On Linux (including Docker on Windows), we'll assume that is `nano`. To edit the file, type:

```
# ensure you are in correct directory. if you followed tutorial type:
cd ~/proxl

# edit the file
nano .env
```

Update these lines to reflect the SMTP relay server information from part 1. If you used SendGrid for your SMTP relay server, your information would be something close to:

```
SMTP_HOST=smtp.sendgrid.net
SMTP_PORT=587
SMTP_USERNAME=apikey
SMTP_PASSWORD=your API KEY goes here
```

Type `Control-o`, `<ENTER>`, and `Control-x` to save and exit `nano`.

14.3 3. Update Email Address for Sender in proxl

1. Log into proxl and click the `Manage Proxl Configuration` icon in the top right (shaped like a gear). You must be logged in as an administrator user, such as the initial user created when you followed the *Installing Proxl* tutorial.
2. Edit the field for `From Address for emails sent`. This is the email address from which emails sent by proxl will appear to come. You may be restricted by what email address you can use here by the SMTP server you are using. For example, if you set up SMTP relay service with SendGrid, this email must match the verified sender you set up.
4. Click the `Save` button to save the changes.

14.4 4. Restart proxl

Proxl must be restarted to use the new configuration settings in the `.env` file. Type the following into your terminal to restart proxl:

```
# ensure you are in correct directory. if you followed tutorial type:
cd ~/proxl

# shutdown proxl
sudo docker-compose down

# startup proxl
sudo docker-compose up --detach
```

14.5 5. Investigating Problems

If after following this guide, emails are not being sent, you can view the logs of the SMTP server by typing the following into a terminal:

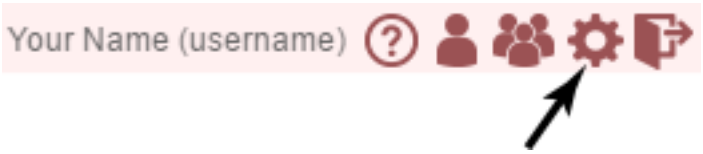
```
sudo docker logs proxl-smtp
```

Carefully read this log and look for error messages, such as an authentication failure or other reasons the message may have been rejected.

CHAPTER 15

Proxl Configuration

If you are an administrator running your own instance of proxl, you can manage the proxl configuration by clicking the gear-shaped icon in the top-right of the site:



15.1 Configuration options:

15.1.1 Allow Account Registration WITHOUT Invite

By default, only users invited to projects by existing users of proxl may create accounts. If this check box is checked, any user visiting your proxl server will be able to register and create accounts.

15.1.2 Google Recaptcha (Not used if either not configured):

Google reCAPTCHA is a free service that verifies that a human is interacting with your site instead of a robot. More information here: https://www.google.com/recaptcha/ If you allow account registration without an invite, and if both values below are set, a reCAPTCHA block will be presented to users who register.

Site key

Upon registering your site with reCAPTCHA, enter your site key here.

Secret key

Upon registering your site with reCAPTCHA, enter your secret key here. Do not share this.

15.1.3 HTML to put at center of bottom of web page

This line of HTML will appear in the bottom center of the footer throughout proxl. For example: `Managed by Joe McDowadle (<a href="mailto:jmcdow@university.edu" target="_top">jmcdow@university.edu</a>)`

15.1.4 From Address for emails sent

All emails sent from proxl will be sent from this email address. For example: `do_not_reply@university.edu`.

15.1.5 SMTP Server URL for emails sent

The SMTP server to use for sending emails. This is typically `localhost` on a Linux server. A SMTP server is also typically provided by your institution, possibly as `smtp.university.edu`–please consult your local IT support resources if you have questions. **Note:** Functions in proxl, such as inviting users to projects or resetting forgotten passwords, will not work unless this configuration option is set correctly.

15.1.6 Google Analytics Tracking Code

Google Analytics provides statistics about visitors to your web site. For more information visit: http://analytics.google.com/. Once you have signed up and registered your website with Google Analytics, enter your tracking code here to track visits to your installation of proxl.

15.1.7 Protein Annotation Service URL

This is the URL to the PAWS service that provides protein sequence annotations (such as disordered region predictions or secondary structure predictions) for supplied protein sequences. This service is used on the protein image view to retrieve the displayed annotation information. Currently, the YRC provides this service and it is not recommended that you change this setting.

15.1.8 Protein Listing Service URL

This is the URL used to try to find more widely-recognized names for proteins found in experiments, regardless of what naming database was used to generate the respective FASTA file used in a search. This name appears when mousing over protein names in proxl. By default, this is set to the YRC PDR's web service designed for this purpose. It is not recommended that you change this setting. Making this field blank will remove the tooltip from appearing when mousing over protein names in proxl.

15.1.9 Submit Search Upload on Website

If configured, users may upload their data using proxl's web interface. Please see our *Installing Proxl* for more information on how to set up the helper application to handle user uploads.

Run Importer Workspace

This is the directory on the local machine (same machine running the proxl web application) where the processing of user uploads will take place. The user running Tomcat must have read/write permission to this directory.

Allow Scan file Upload

If checked, users will be able to upload scan data with their data. This may greatly increase storage requirements.

Delete uploaded files after Successful Import

If checked, data on disk will be deleted after being successfully imported into proxl.

15.1.10 Terms of Service Management

If enabled, Terms of Service (ToS) is a text document that users must view and accept before they may use an installation of proxl. To enable the ToS requirement, click "Change Terms of Service" to edit the document, then "Enable Terms of Service". To disable the ToS requirement, click "Disable Terms of Service".

Note that if ToS is enabled, and if the text document is changed in any way, all users will have to view and agree to the new ToS before continuing to use this installation of proxl.

CHAPTER 16

Proxl XML Converter Development Guide

This guide provides information for how to develop software to convert the output of a cross-linking proteomics pipeline to proxl XML suitable for import into proxl.

16.1 Proxl XML schema

Proxl XML files must adhere to the proxl XML schema, which may be found at https://github.com/yeastrc/proxl-import-api/tree/master/xsd. In addition to the assigned scores, descriptions of those scores (names, descriptions, and how to treat them) are incorporated into the schema, which allows proxl XML to describe data generated by nearly any pipeline.

HTML documentation for the schema can be found at http://yeastrc.org/proxl-xml-documentation/1.4/proxl-xml-v1.4.html. The root element is proxl_input. You may click on any attribute or element in the diagrams to navigate the schema.

An image containing the entire schema structure and documentation can be found at https://raw.githubusercontent.com/yeastrc/proxl-import-api/master/xsd/docs/proxl-input.png.

16.2 Example proxl XML file

Below is a very shortened, but valid, proxl XML file to illustrate the structure of the proxl XML schema. In this example StavroX was the analysis program and two peptides were identified, one of of which includes a cross-link.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<proxl_input fasta_filename="gTuSC-parsimonious-plusRev.fasta">
    <search_program_info>
        <search_programs>
            <search_program name="StavroX" display_name="StavroX" version="3">
                <psm_annotation_types>
                    <filterable_psm_annotation_types>
                        <filterable_psm_annotation_type name="score" description=
↪"StavroX Score" filter_direction="above" default_filter="false" default_filter_
↪value="50"/>
```

(continues on next page)

(continued from previous page)

```
                        <filterable_psm_annotation_type name="FDR" description="False␣
↪discovery rate" filter_direction="below" default_filter="true" default_filter_value=
↪"0.01"/>
                        <filterable_psm_annotation_type name="rank" description="Rank␣
↪of PSM for scan" filter_direction="below" default_filter="true" default_filter_
↪value="1"/>
                    </filterable_psm_annotation_types>
                    <descriptive_psm_annotation_types>
                        <descriptive_psm_annotation_type name="m/z" description="m/z"/
↪>
                        <descriptive_psm_annotation_type name="obs. mass" description=
↪"obs. mass"/>
                        <descriptive_psm_annotation_type name="cand. mass"␣
↪description="cand. mass"/>
                        <descriptive_psm_annotation_type name="deviation" description=
↪"deviation"/>
                        <descriptive_psm_annotation_type name="scan num." description=
↪"scan num."/>
                    </descriptive_psm_annotation_types>
                </psm_annotation_types>
            </search_program>
        </search_programs>
        <default_visible_annotations>
            <visible_psm_annotations>
                <search_annotation search_program="StavroX" annotation_name="scan num.
↪"/>
                <search_annotation search_program="StavroX" annotation_name="rank"/>
                <search_annotation search_program="StavroX" annotation_name="score"/>
                <search_annotation search_program="StavroX" annotation_name="FDR"/>
                <search_annotation search_program="StavroX" annotation_name="m/z"/>
                <search_annotation search_program="StavroX" annotation_name="obs. mass
↪"/>
                <search_annotation search_program="StavroX" annotation_name="cand.␣
↪mass"/>
                <search_annotation search_program="StavroX" annotation_name="deviation
↪"/>
            </visible_psm_annotations>
        </default_visible_annotations>
    </search_program_info>
    <linkers>
        <linker name="dss">
            <crosslink_masses>
                <crosslink_mass mass="138.0681"/>
            </crosslink_masses>
        </linker>
    </linkers>
    <reported_peptides>
        <reported_peptide reported_peptide_string="[SFSLEKISAPDQR](K6)" type="unlinked
↪">
            <peptides>
                <peptide sequence="SFSLEKISAPDQR">
                    <modifications>
                        <modification mass="138.0681" position="6" isMonolink="true"/>
                    </modifications>
                </peptide>
            </peptides>
            <psms>
```

(continues on next page)

(continued from previous page)

```
                <psm scan_file_name="Q_2013_1010_RJ_07.mzML" scan_number="25982"␣
↪precursor_charge="3">
                    <filterable_psm_annotations>
                        <filterable_psm_annotation search_program="StavroX"␣
↪annotation_name="score" value="23"/>
                        <filterable_psm_annotation search_program="StavroX"␣
↪annotation_name="rank" value="1"/>
                        <filterable_psm_annotation search_program="StavroX"␣
↪annotation_name="FDR" value="0.0200"/>
                    </filterable_psm_annotations>
                    <descriptive_psm_annotations>
                        <descriptive_psm_annotation search_program="StavroX"␣
↪annotation_name="m/z" value="545.2862"/>
                        <descriptive_psm_annotation search_program="StavroX"␣
↪annotation_name="obs. mass" value="1633.8440"/>
                        <descriptive_psm_annotation search_program="StavroX"␣
↪annotation_name="cand. mass" value="1633.8432"/>
                        <descriptive_psm_annotation search_program="StavroX"␣
↪annotation_name="deviation" value="0.4885"/>
                        <descriptive_psm_annotation search_program="StavroX"␣
↪annotation_name="scan num." value="25982"/>
                    </descriptive_psm_annotations>
                </psm>
            </psms>
        </reported_peptide>
        <reported_peptide reported_peptide_string="[KDVLLR](K1)--
↪[TNQSSQEDFNNFMDSMKNESSLHLR](K17)" type="crosslink">
            <peptides>
                <peptide sequence="KDVLLR">
                    <linked_positions>
                        <linked_position position="1"/>
                    </linked_positions>
                </peptide>
                <peptide sequence="TNQSSQEDFNNFMDSMKNESSLHLR">
                    <linked_positions>
                        <linked_position position="17"/>
                    </linked_positions>
                </peptide>
            </peptides>
            <psms>
                <psm scan_file_name="Q_2013_1010_RJ_07.mzML" scan_number="29178"␣
↪precursor_charge="5" linker_mass="138.0681">
                    <filterable_psm_annotations>
                        <filterable_psm_annotation search_program="StavroX"␣
↪annotation_name="score" value="28"/>
                        <filterable_psm_annotation search_program="StavroX"␣
↪annotation_name="rank" value="1"/>
                        <filterable_psm_annotation search_program="StavroX"␣
↪annotation_name="FDR" value="0.0000"/>
                    </filterable_psm_annotations>
                    <descriptive_psm_annotations>
                        <descriptive_psm_annotation search_program="StavroX"␣
↪annotation_name="m/z" value="768.7731"/>
                        <descriptive_psm_annotation search_program="StavroX"␣
↪annotation_name="obs. mass" value="3839.8366"/>
                        <descriptive_psm_annotation search_program="StavroX"␣
↪annotation_name="cand. mass" value="3839.8327"/>
```

(continues on next page)

(continued from previous page)

```
                        <descriptive_psm_annotation search_program="StavroX"␣
↪annotation_name="deviation" value="1.0017"/>
                        <descriptive_psm_annotation search_program="StavroX"␣
↪annotation_name="scan num." value="29178"/>
                    </descriptive_psm_annotations>
                </psm>
            </psms>
        </reported_peptide>
    </reported_peptides>
    <matched_proteins>
        <protein sequence=
↪"MEIKEVDDRAELLRYTNNIPLLGKLVNHQPLWSTNPKLKSFSLEKISAPDQRRVQEALVVKDLLNVLIGLEGTYIRYFNDYEPSDPETPIEFKIAKKN
↪">
            <protein_annotation name="Spc97-yeast"/>
        </protein>
        <protein sequence=
↪"MELEPTLFGIIEALAPQLLSQSHLQTFVSDVVNLLRSSTKSATQLGPLIDFYKLQSLDSPETTIMWHKIEKFLDALFGIQNTDDMVKYLSVFQSLLPS
↪">
            <protein_annotation name="Spc98-yeast"/>
        </protein>
        <protein sequence=
↪"MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFTMGGEIITLQAGQCGNHVGKFLWSQLAKEHAIGTDGLSQLPDSSTERDDDTKPFFRENSRNKFTPE
↪">
            <protein_annotation name="His-TEV-Tub4-yeast"/>
        </protein>
    </matched_proteins>
    <static_modifications>
        <static_modification amino_acid="C" mass_change="57.0215"/>
    </static_modifications>
    <decoy_labels>
        <decoy_label prefix="random_seq"/>
    </decoy_labels>
</proxl_input>
```

16.3 Writing the conversion software

Your software will read in the native output of your respective cross-linking search software and re-represent it as proxl XML. The converter software may be written in any programming or scripting language, so long as valid proxl XML is produced.

We have developed several open-source converters in Java, which may be used as examples. Follow the links below to access the source code associated with each converter:

- iProphet (TPP)
- Kojak
- Crux
- pLink 1.x pLink 2.x
- StavroX
- MetaMorpheus
- xQuest

For any assistance developing converters, please email us at proxl-help@yeastrc.org.

CHAPTER 17

PAWS Guide

Coming soon.