ProXL (Protein Cross-linking Database): a platform for analysis, visualization, and sharing of protein cross-linking mass spectrometry data.

Michael Riffle[1,2,*], Daniel Jaschob[1], Alex Zelter[1], and Trisha N. Davis[1]

[1] Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

[2] Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA

[*]Corresponding author:

UW Box 357350

1705 NE Pacific St.

Seattle WA 98195-7350

Phone: 206-685-3740

Email: mriffle@uw.edu

Table of Contents:

Page S-2: Figure S-1. Entity relationship diagram (ERD) of ProXL database schema.

Page S-3: Figure S-2. Diagram of the ProXL XML schema.

Page S-5: Figure S-3. Screenshots of the quality control plots from ProXL

Page S-7: Figure S-4. Screenshot from the ProXL "Merged Proteins" page

Page S-9: Figure S-5. Screenshot from the ProXL project page

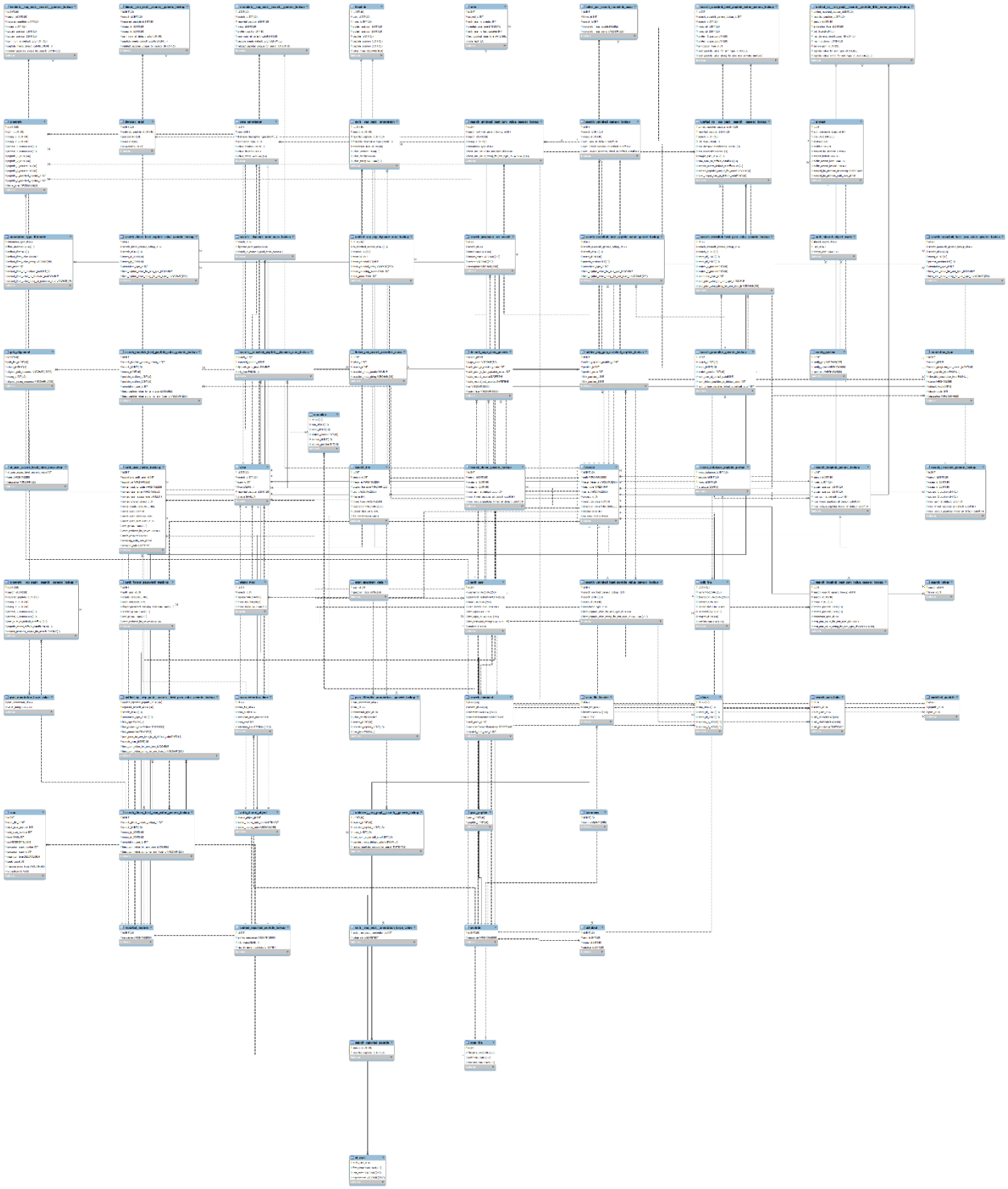


Figure S-1. An entity relationship diagram showing the MySQL database schema used by ProXL.

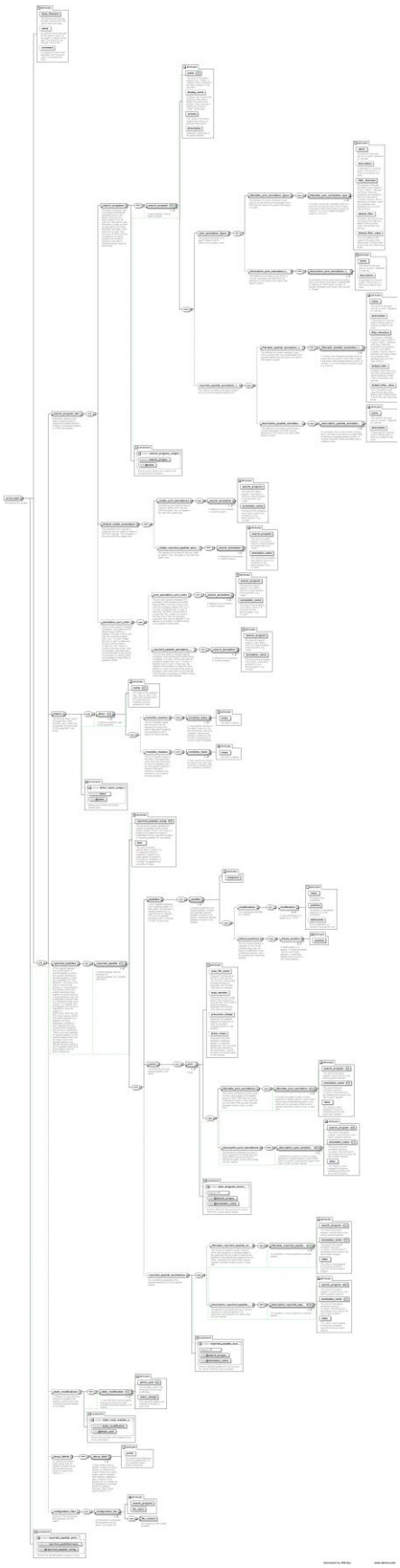
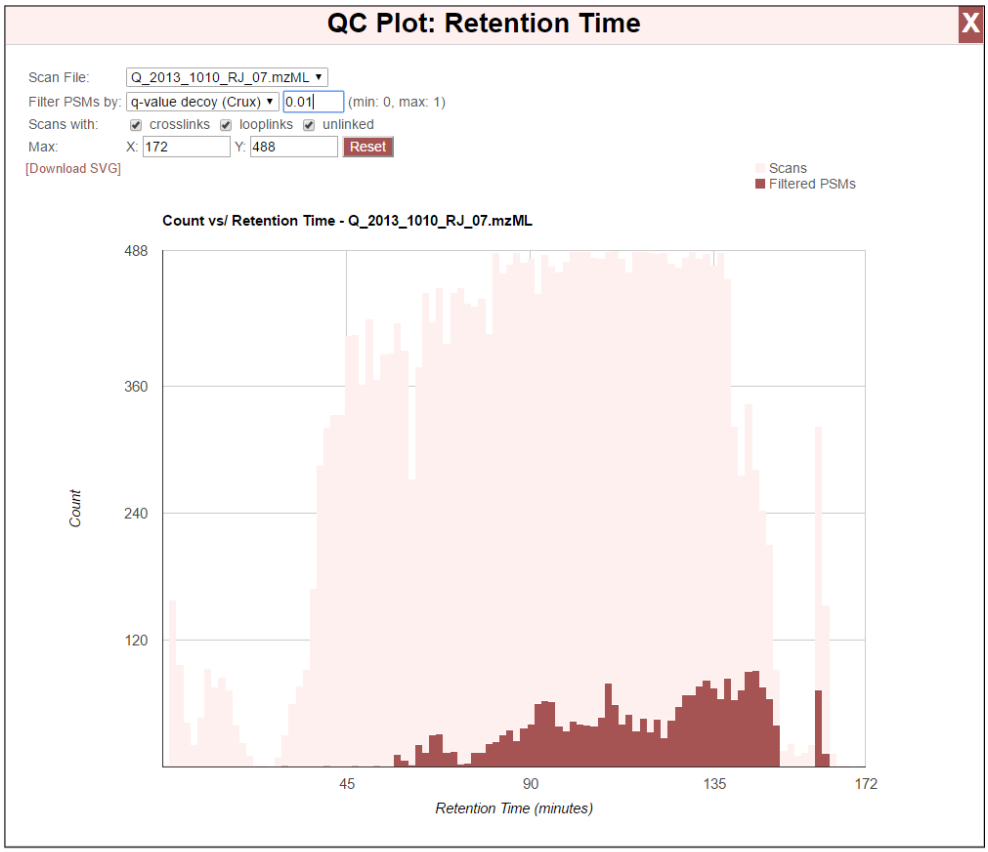


Figure S-2. An image displaying the entire ProXL XML v1.0 schema. The root element is "proxl_input" and connecting lines show the hierarchical layout of a ProXL XML document. Items with dashed borders are optional, items with solid borders are required. Documentation for the element types and attributes are included within the image. More information, including a higher resolution image, may be found at http://proxl-web-app.readthedocs.io/en/latest/install/converter_guide.html.

(A)


QC Plot: Retention Time
Scan File: Q_2013_1010_RJ_07.mzML
Filter PSMs by: q-value decoy (Crux) 0.01 (min: 0, max: 1)
Scans with: crosslinks looplinks unlinked
Max: X: 172 Y: 488 Reset
[Download SVG]
Scans
Filtered PSMs
Count vs/ Retention Time - Q_2013_1010_RJ_07.mzML
488
360
240
120
Count
45
90
135
172
Retention Time (minutes)

(B)

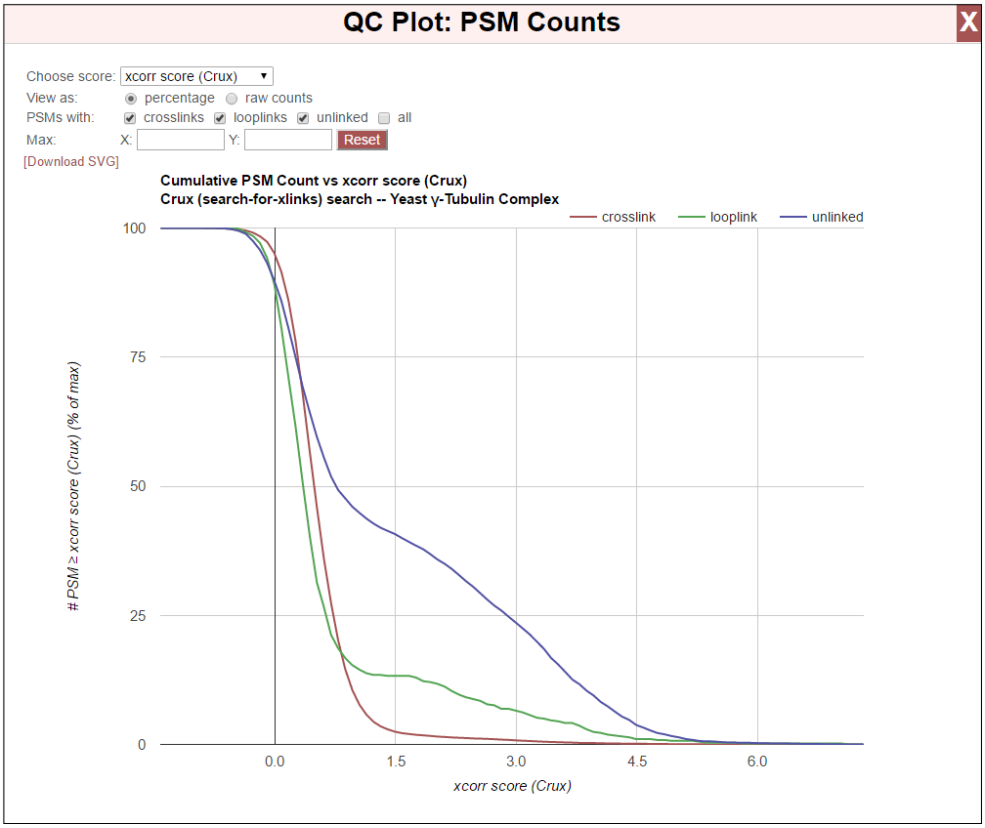
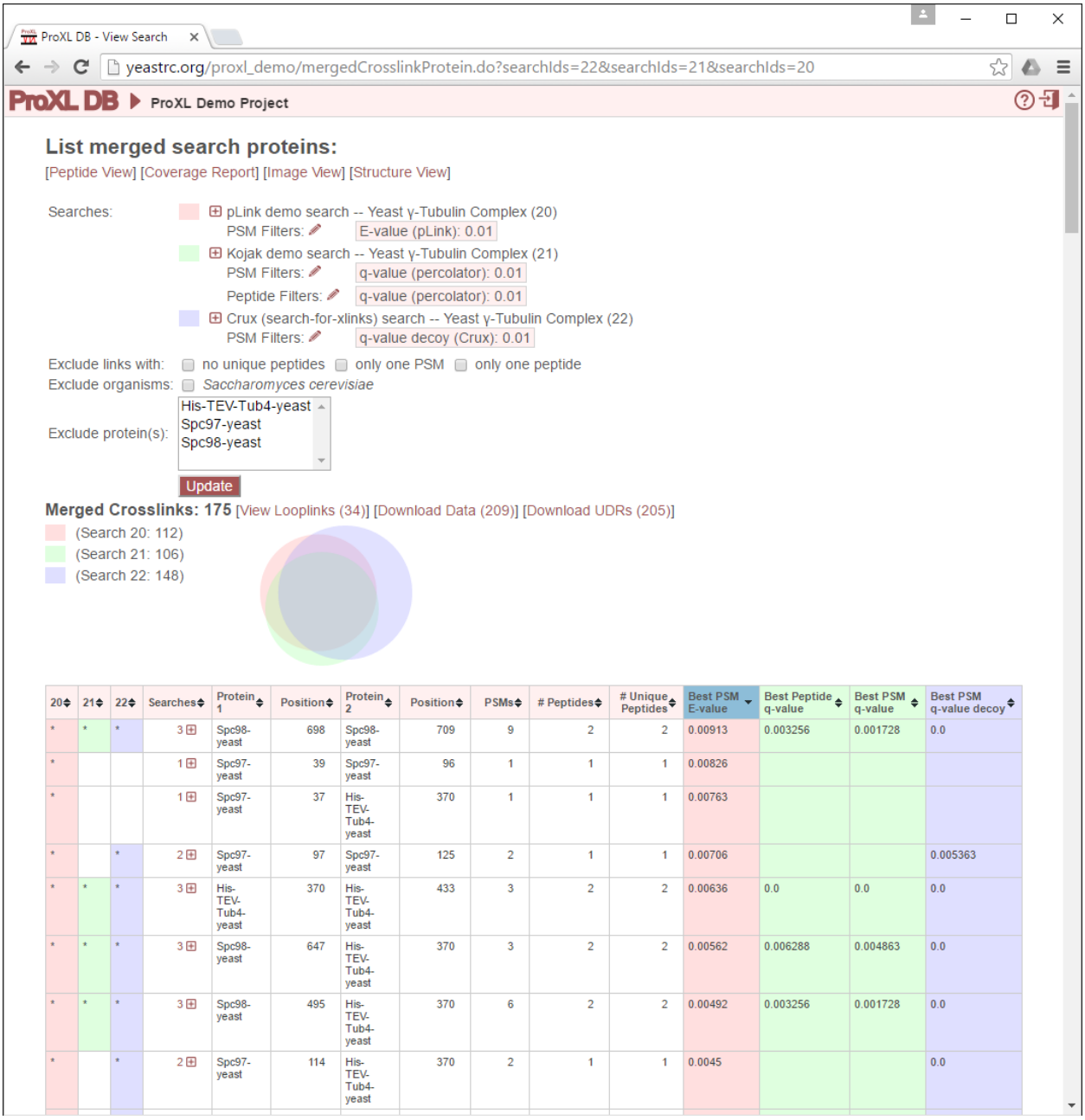

QC Plot: PSM Counts
Choose score: xcorr score (Crux)
View as: percentage raw counts
PSMs with: crosslinks looplinks unlinked all
Max: X: Y: Reset
[Download SVG]
Cumulative PSM Count vs xcorr score (Crux)
Crux (search-for-xlinks) search -- Yeast γ-Tubulin Complex
crosslink
looplink
unlinked
100
75
50
25
0
PSM ≥ xcorr score (Crux) (% of max)
0.0
1.5
3.0
4.5
6.0
xcorr score (Crux)

Figure S-3. Screenshots of the quality control plots from ProXL. (A) A plot of the number of PSMs (shown in pink) and the number of "quality" PSMs (shown as red) as a function of retention time. The filter for "quality" PSMs may be changed by the user to any value for any filterable annotation type specified in the ProXL XML document used to import the data. (B) The cumulative PSM count as a function of a score (XCorr in this case). Any filterable annotation type from the ProXL XML file may be selected. The separate curves represent different classes of PSM—blue is unlinked, green is loop-links, and red is cross-links. Scores for which larger values are more significant, the y-value for a given score will be the number of PSMs with a score greater than or equal to that score. Scores for which lower values are more significant, the y-value will be the number of PSMs less than or equal to that score.



20	21	22	Searches	Protein 1	Position	Protein 2	Position	PSMs	# Peptides	# Unique Peptides	Best PSM E-value	Best Peptide q-value	Best PSM q-value	Best PSM q-value decoy
*	*	*	3	Spc98-yeast	698	Spc98-yeast	709	9	2	2	0.00913	0.003256	0.001728	0.0
*			1	Spc97-yeast	39	Spc97-yeast	96	1	1	1	0.00826			
*			1	Spc97-yeast	37	His-TEV-Tub4-yeast	370	1	1	1	0.00763			
*		*	2	Spc97-yeast	97	Spc97-yeast	125	2	1	1	0.00706			0.005363
*	*	*	3	His-TEV-Tub4-yeast	370	His-TEV-Tub4-yeast	433	3	2	2	0.00636	0.0	0.0	0.0
*	*	*	3	Spc98-yeast	647	His-TEV-Tub4-yeast	370	3	2	2	0.00562	0.006288	0.004863	0.0
*	*	*	3	Spc98-yeast	495	His-TEV-Tub4-yeast	370	6	2	2	0.00492	0.003256	0.001728	0.0
*		*	2	Spc97-yeast	114	His-TEV-Tub4-yeast	370	2	1	1	0.0045			0.0

Figure S-4. A screenshot from the ProXL "Merged Proteins" page showing the comparison of three searches of the same spectral data with the same FASTA file, but with three distinct software pipelines. Each search is color coded, and in this example pink represents a pLink search, green a Kojak search, and blue a Crux search. Each search is automatically filtered according to its native scoring mechanism using default cutoffs specified in the ProXL XML

document. The filtering criteria may be changed independently for each search. The Euler diagram indicates the overlap between the three searches. The table lists each distinct pair of linked proteins and positions and indicates in which of the three searches that cross-link was found at the given cutoffs. Each row may be expanded to view underlying peptides, PSMs, and spectra.

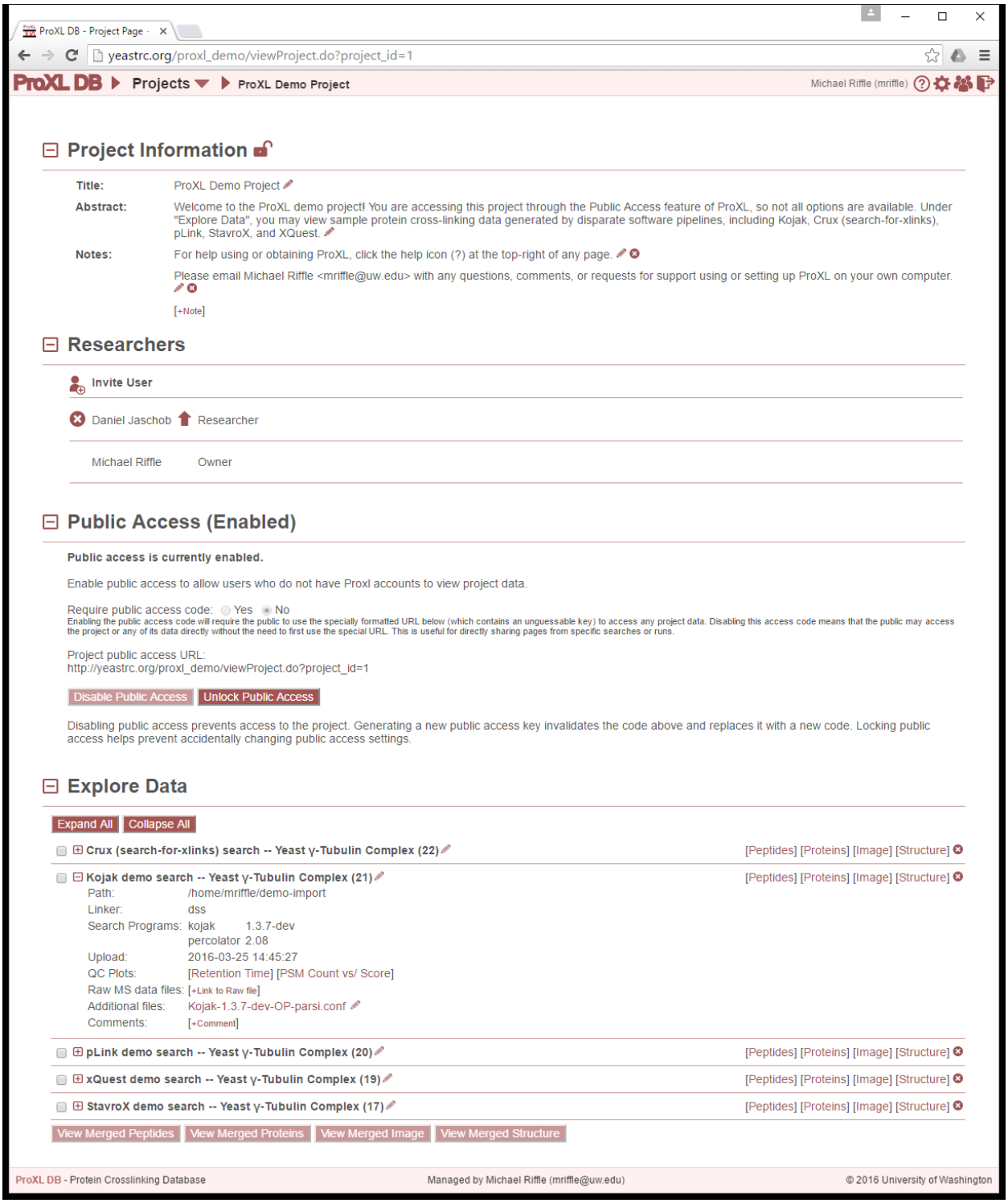


Figure S-5. A screenshot from the ProXL project page. From this page, users can manage all aspects of the project and click links to view data uploaded to the project. In the “Project

Information" section, users can change the title or abstract of the project, and add notes visible to other members of the project. In the "Researchers" section, users can review who has access to the project and project owners can add or remove users from the project. In the "Public Access" section, project owners can enable or disable public access and alter public access settings. In the "Explore Data" section, users can review data associated with individual searches (such as the linker or search program used), view QC plots and add comments. Users can view data from individual searches by clicking the "[Peptides]", "[Proteins]", "[Image]", or "[Structure]" links next to each search. Users may also combine and compare multiple searches by clicking the check boxes next to searches and clicking the "View Merged" buttons at the bottom of the list.