User manual

CLMSVault

Protein cross-linking bioinformatics analysis platform

Written by Mathieu Courcelles

Institute for research in immunology and cancer

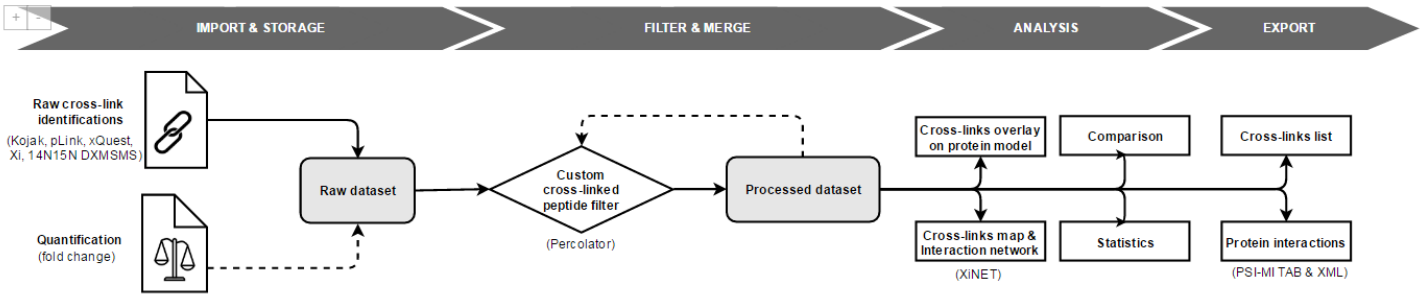
Université de Montréal

Last update: 2017/04/10

Table of Contents

1. Introduction

CLMSVault is a bioinformatics platform for protein cross-linking mass spectrometry analyses. An overview of its functionalities is presented in this manual. The diagram below present the data processing workflow provided by CLMSVault. CLMSVault takes over the analysis next to the cross-link identification step.



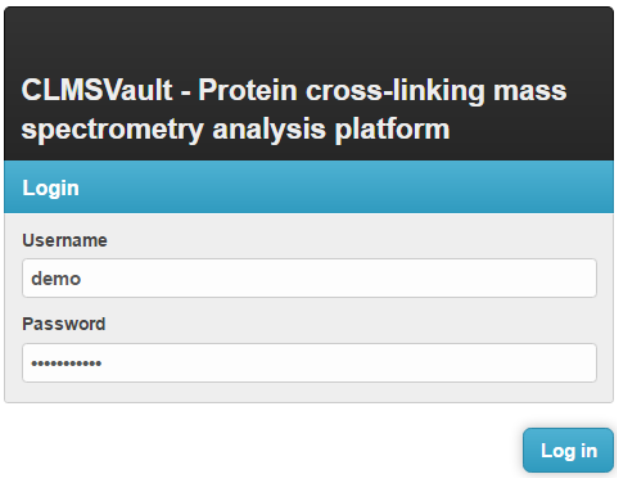
2. Installation

CLMSVault is an open source project available here: https://gitlab.com/courcelm/clmsvault.

Please refer to installation note file to deploy it under Linux or Windows.

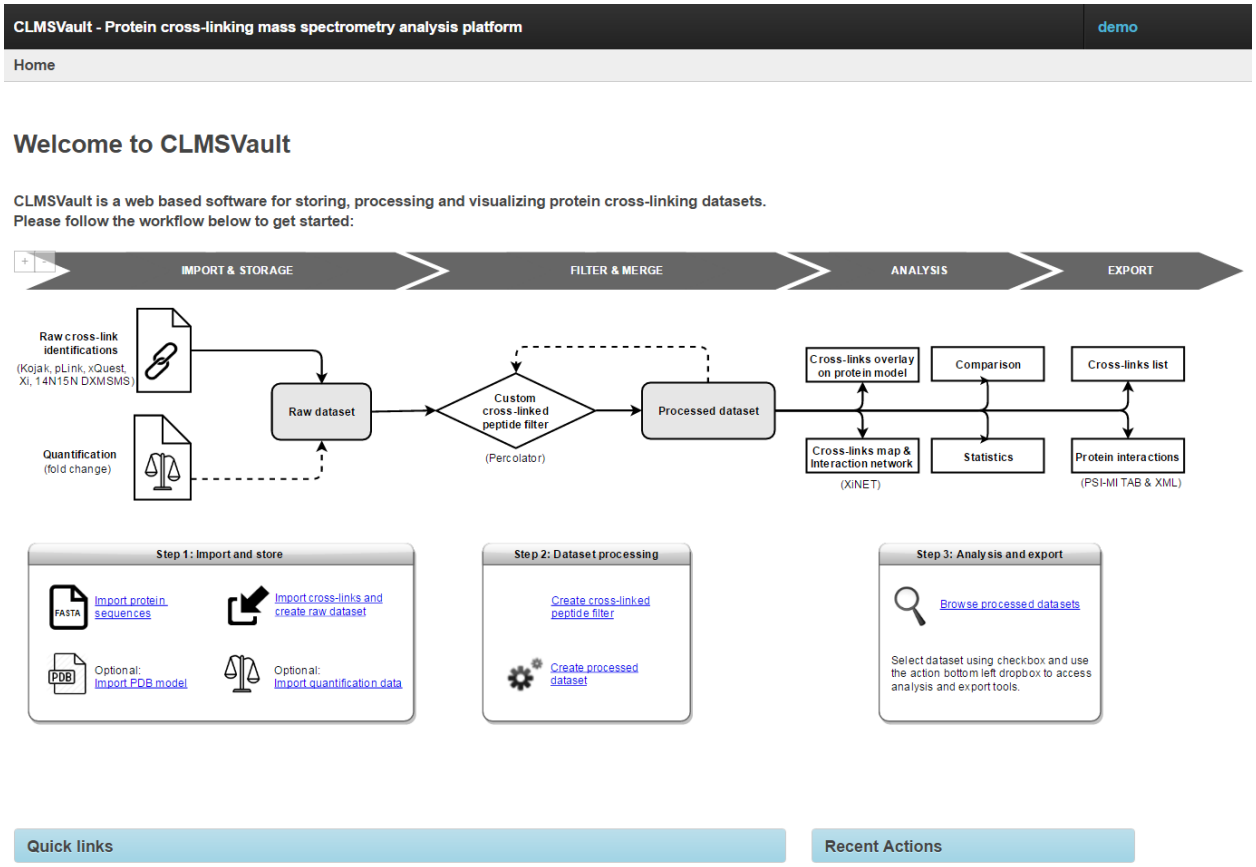
3. First login

To access CLMSVault, open your web browser of choice and enter the URL to where it was deployed (default: http://localhost:8000). The login view should be displayed if the installation worked successfully.



Log in using this user/password combination: user name: clmsvault, password: clms123

Refer to section 7.1 to add more users. After login, you will be presented CLMSVault workflow view from which you can access its functionalities.



CLMSVault processing workflow is divided into three main steps:

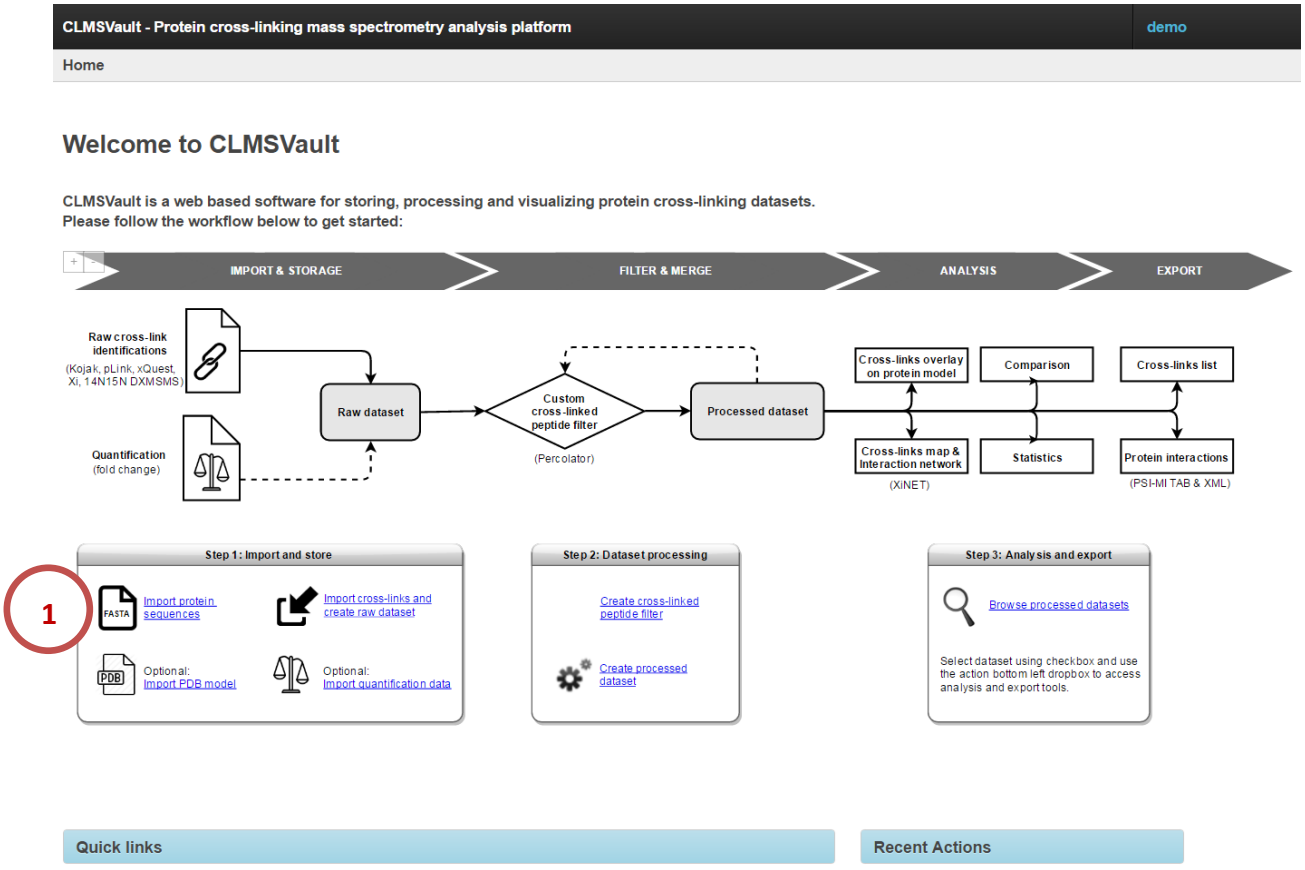
- Data import and storage (section 4)
- Cross-link datasets filtering and merging (section 5)
- Analysis, visualization, and export (section 6)

4. Data import and storage

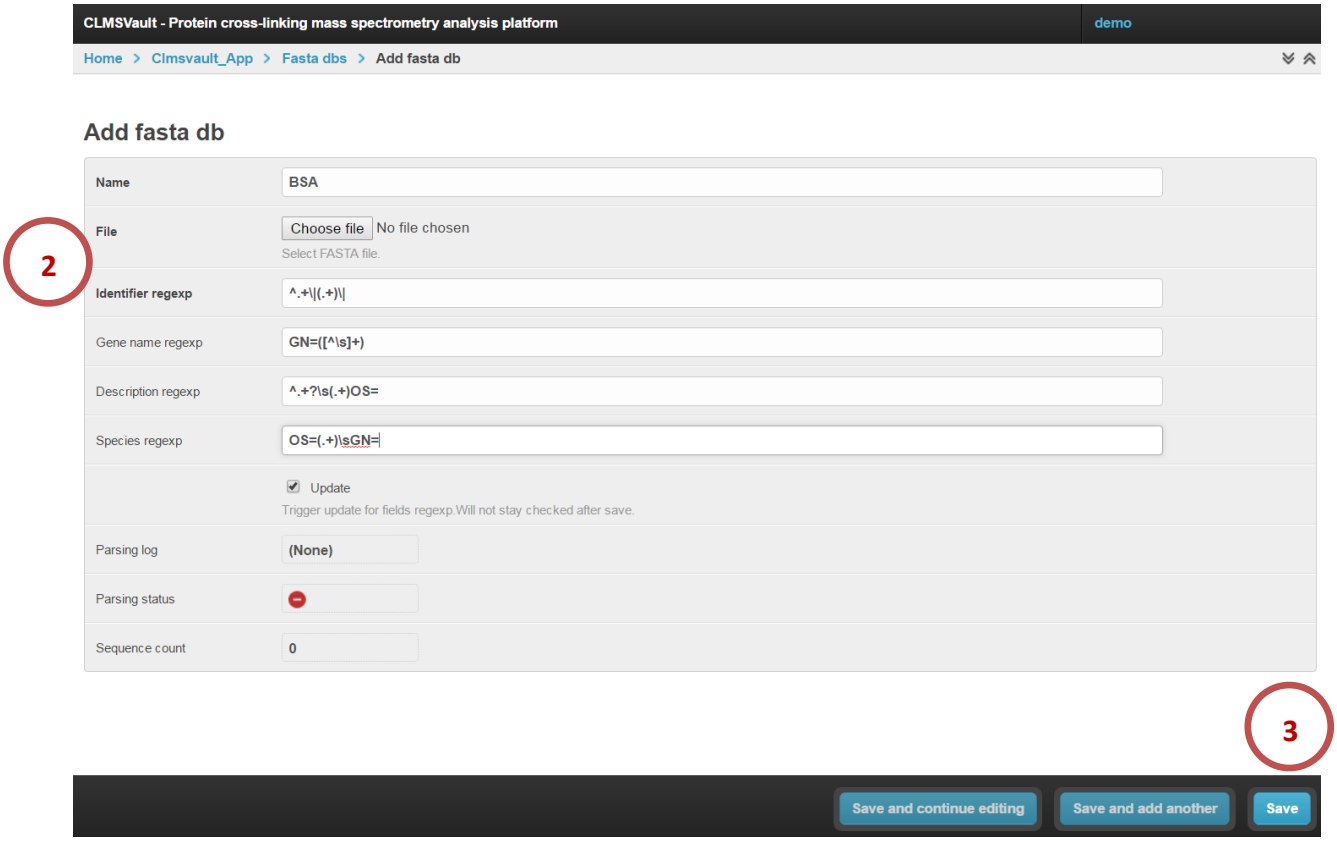
4.1. Protein sequence databases

CLMSVault requires protein sequences for its different visualizations. The user must first imports protein sequences in FASTA format before loading any cross-link dataset.

1. Click on “Import protein sequences” from the main screen.

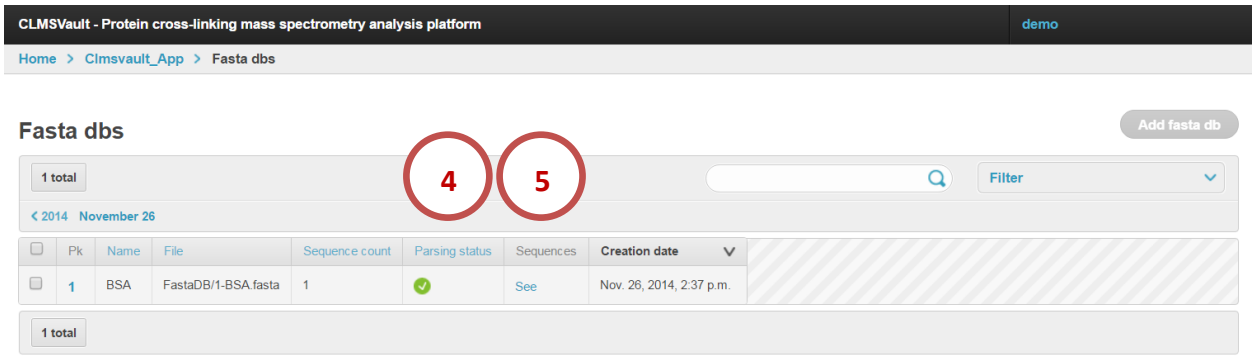


2. Fill out the form. Regular expressions are used to extract identifier (mandatory), gene name, description and species from the FASTA file. If you are unfamiliar with regular expression or you would like to test them before, we recommend using the following tool: https://regex101.com/#python

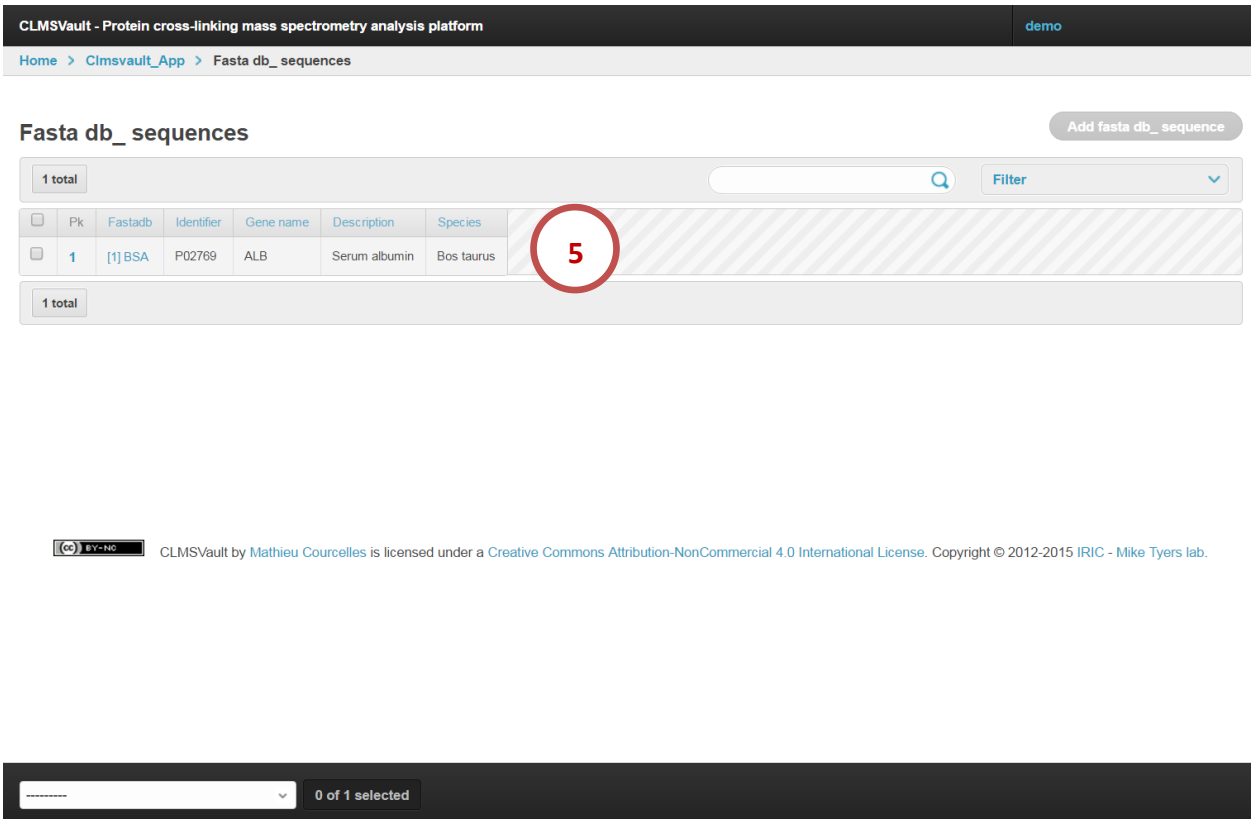


3. Click “Save”.

4. Check parsing status to verify that protein sequences were imported correctly.



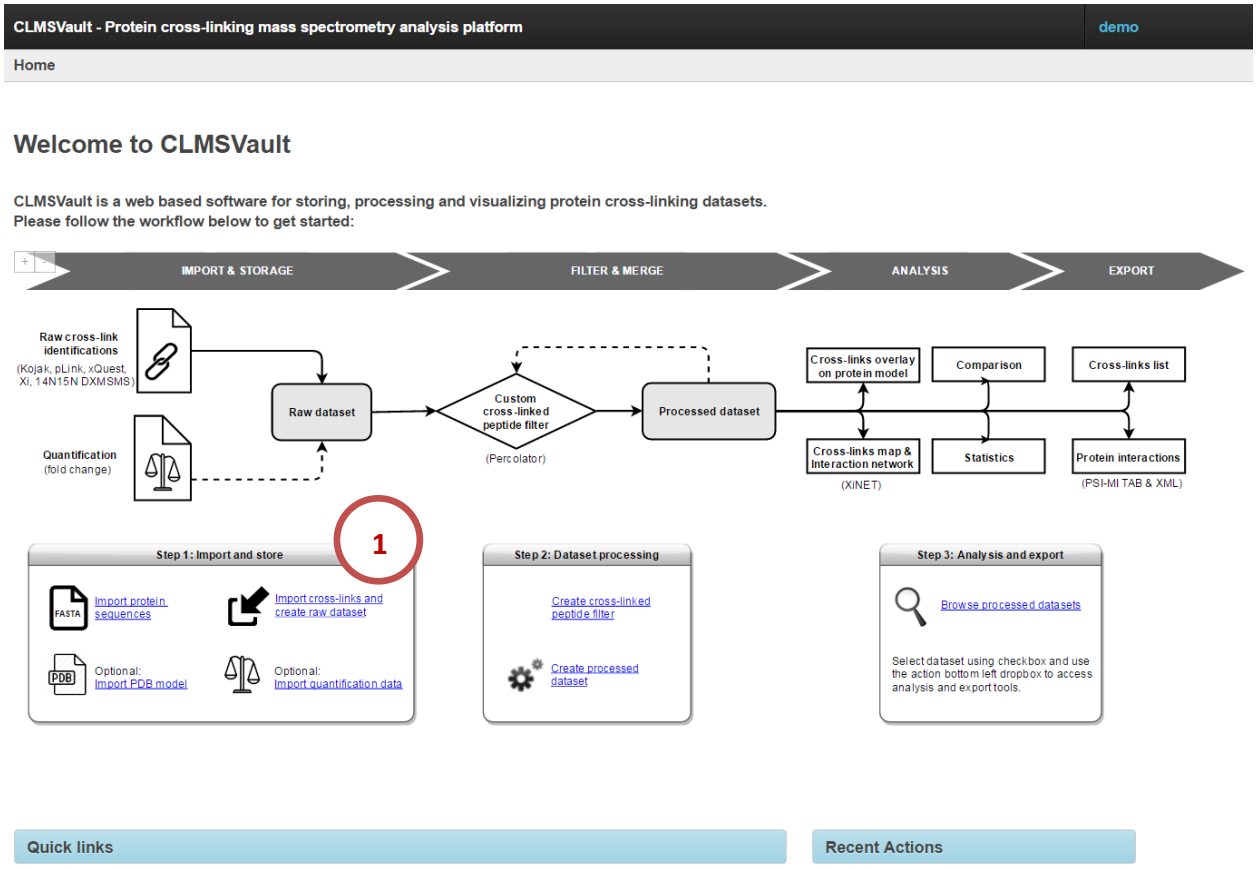
1. Click “See” to double check protein import by listing imported proteins.



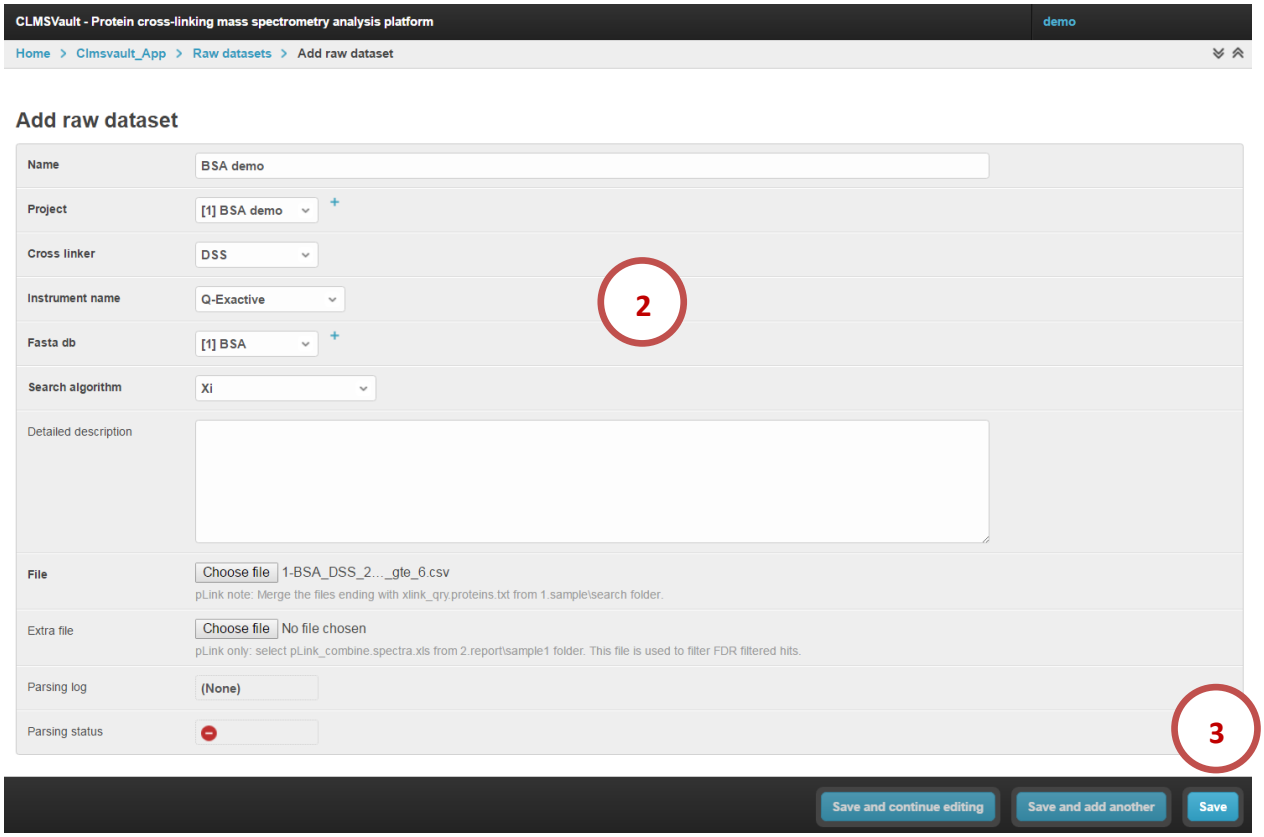
4.2. Cross-link datasets

CLMSVault imports cross-link results from the following search engine: Kojak, pLink, xQuest, Xi, 14N 15N DXMSMS. Note: some older or recent version of these softwares might not be yet supported. Cross-link datasets are organized into project to facilitate data access.

1. Click on “Import cross-links and create raw dataset” from the main screen.



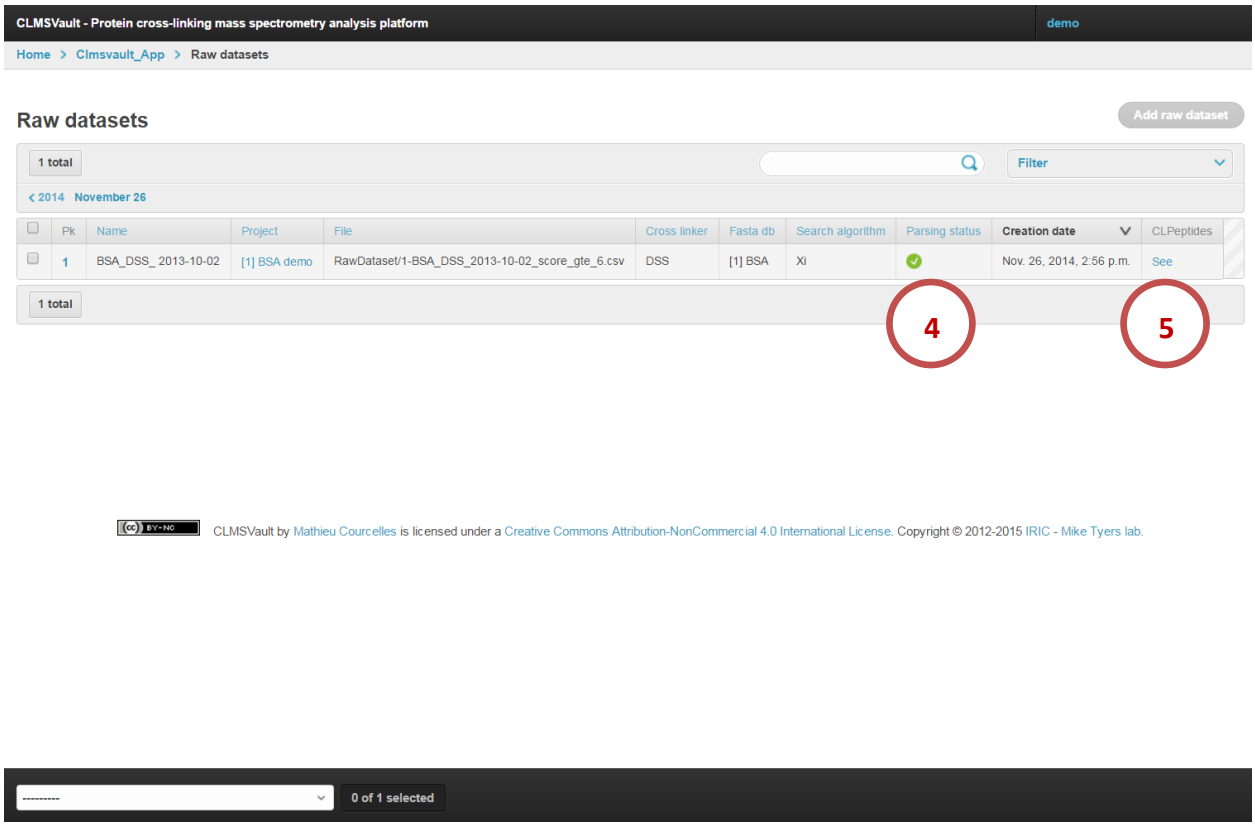
2. Fill out the form.



3. Click “Save”.

4. Verify that dataset import was successful.

5. Click “See” to view imported cross-linked peptides.



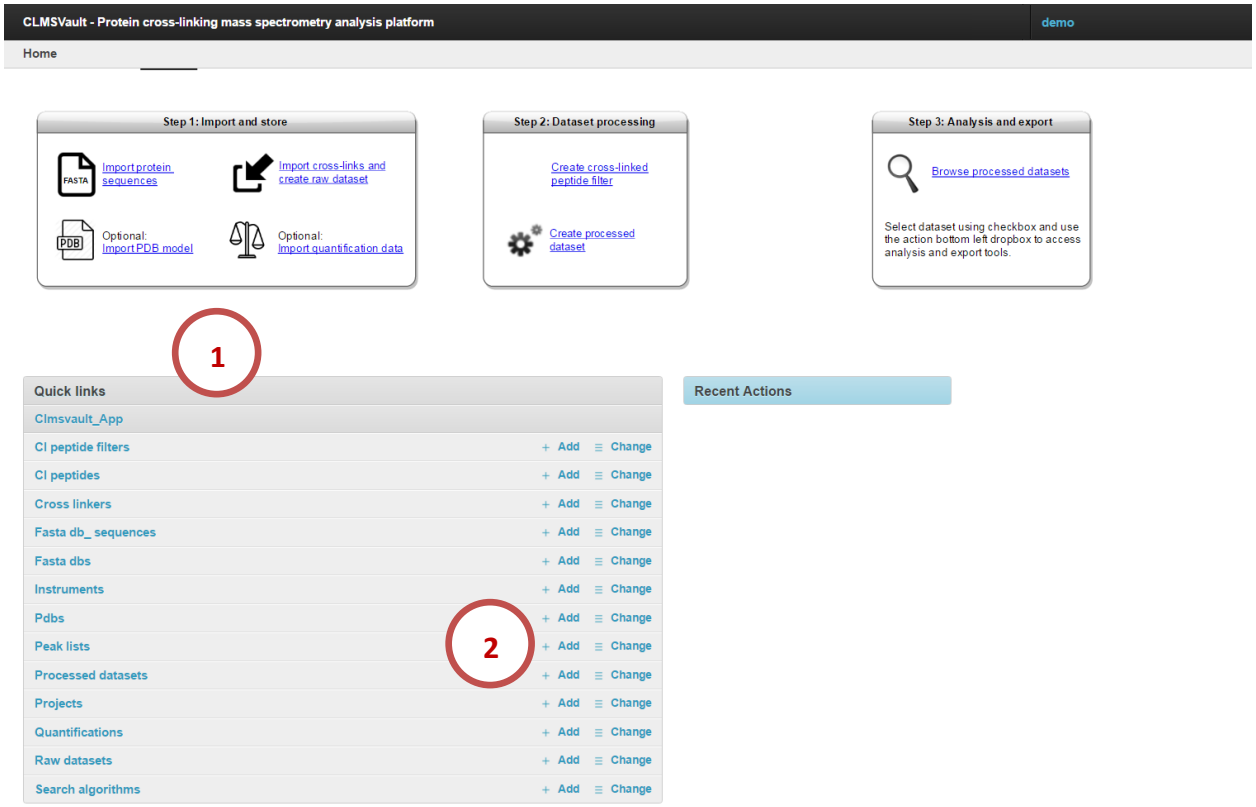
Once the cross-links dataset is successfully imported, it is possible to either create a processed dataset (refer to section 5) or apply different actions to analyze it directly (refer to section 6).

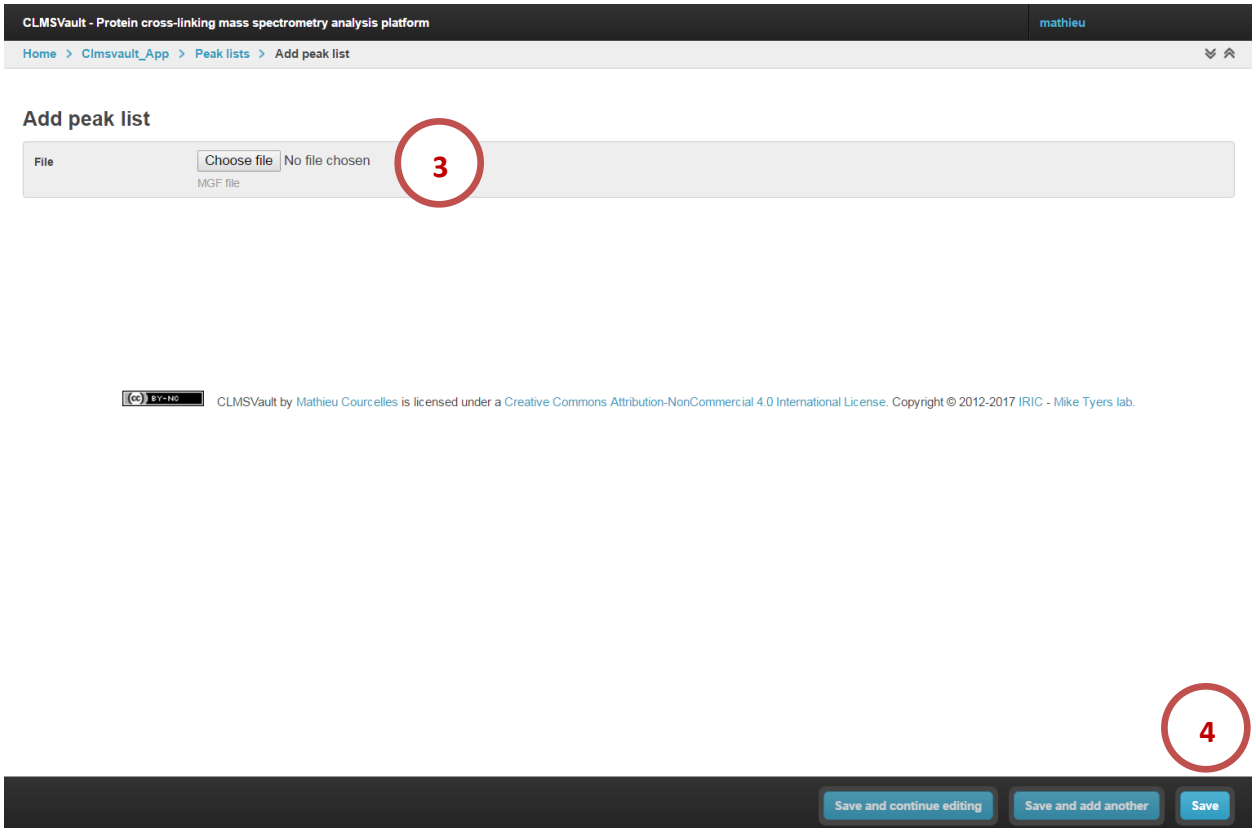
4.3. MS/MS spectra

MS/MS spectrum of cross-linked peptides can be viewed with the embeded Xi spectrum viewer (refer to section X). MS/MS spectra file are imported into CLMSVault via MGF file which has been generated by msconvert tool from the Proteowizard (http://proteowizard.sourceforge.net/tools.shtml). Note that MS/MS spectra must be imported after cross-link identifications.

1. Click on “Expand” from the main screen.
2. Click on “Add” in the “Peak lists” section.

3. Select your MGF file.
4. Click “Save”.

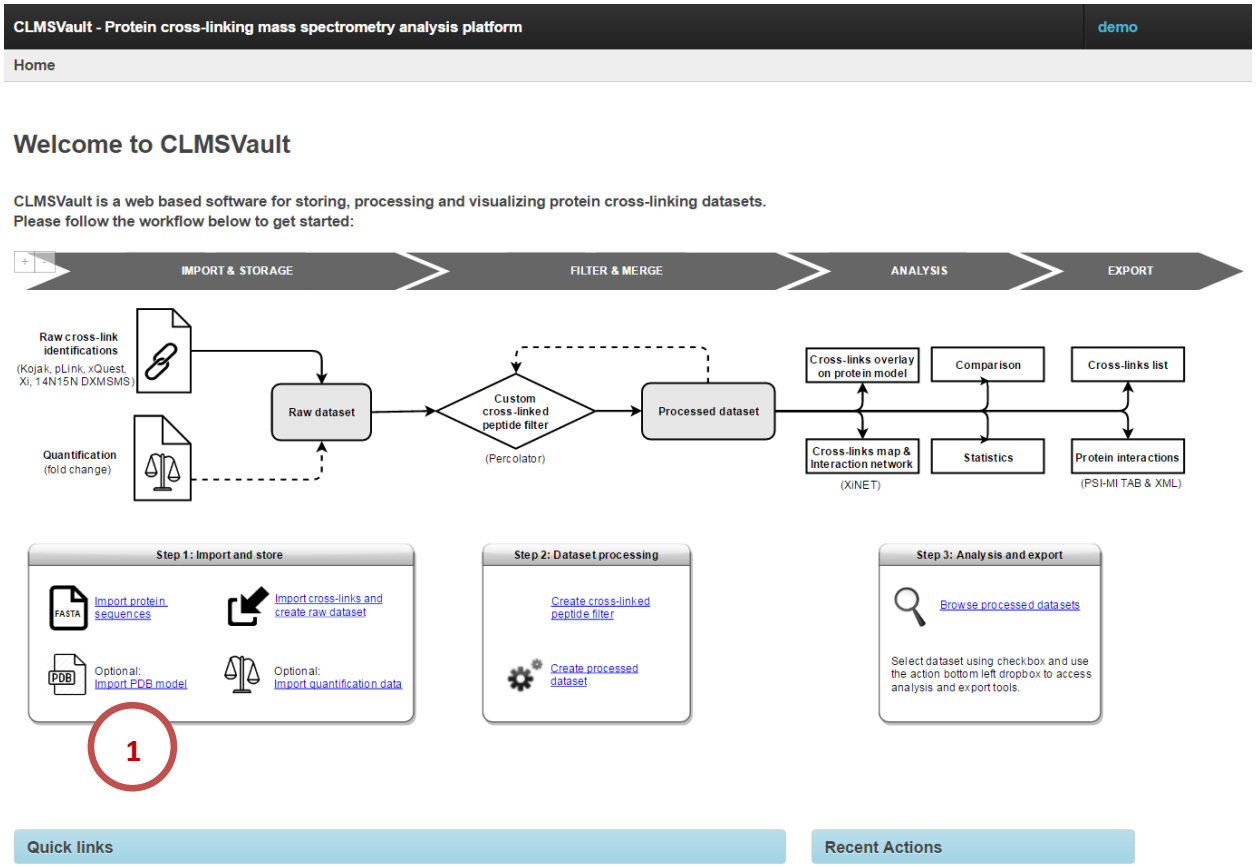




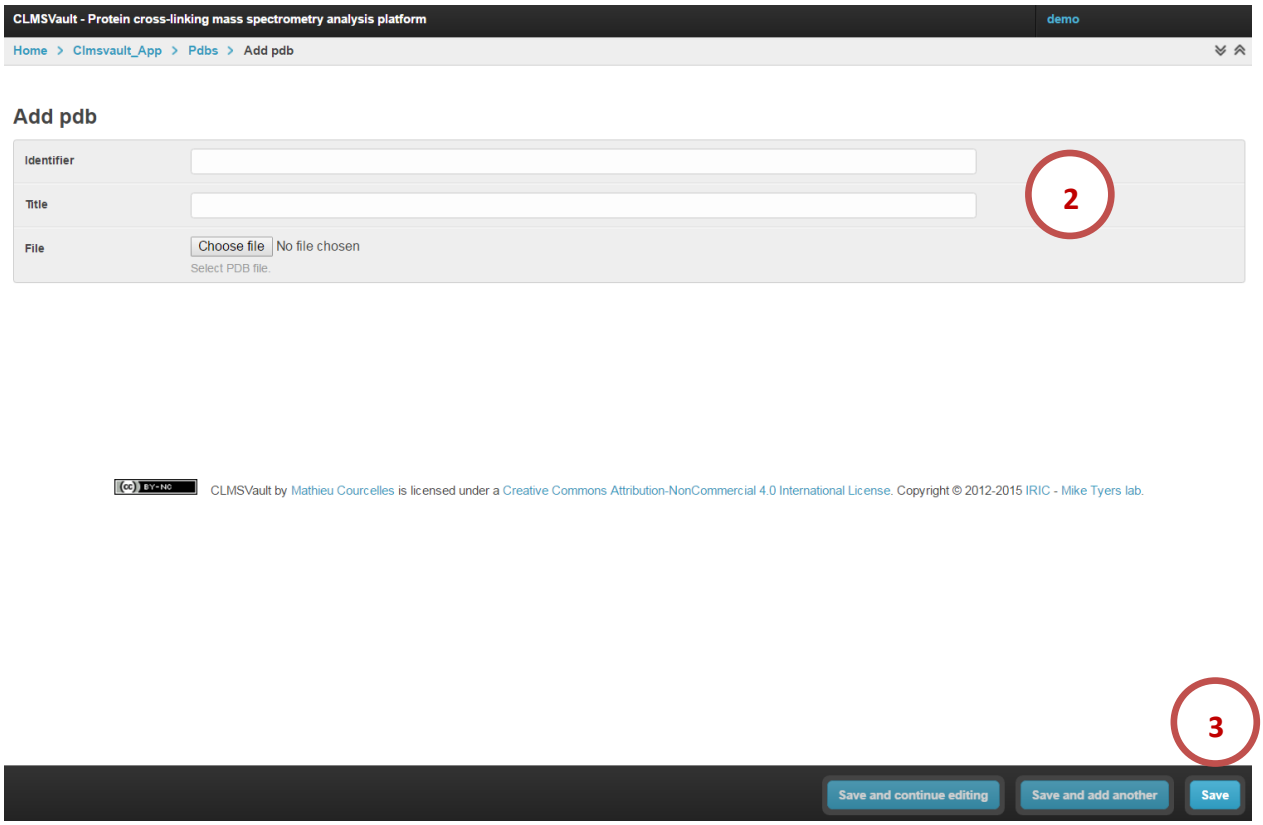
4.4. PDB models

PDB models are imported into CLMSVault for 3D visualization of cross-links. The model can be imported automatically from RCSB Protein data bank with an identifier (refer to section 6.5). Custom in-house model can be imported with this procedure:

1. Click on “Import PDB model” from the main screen.



2. Fill the form and select your PDB file.



3. Click “Save”.

4.4. Quantification datasets

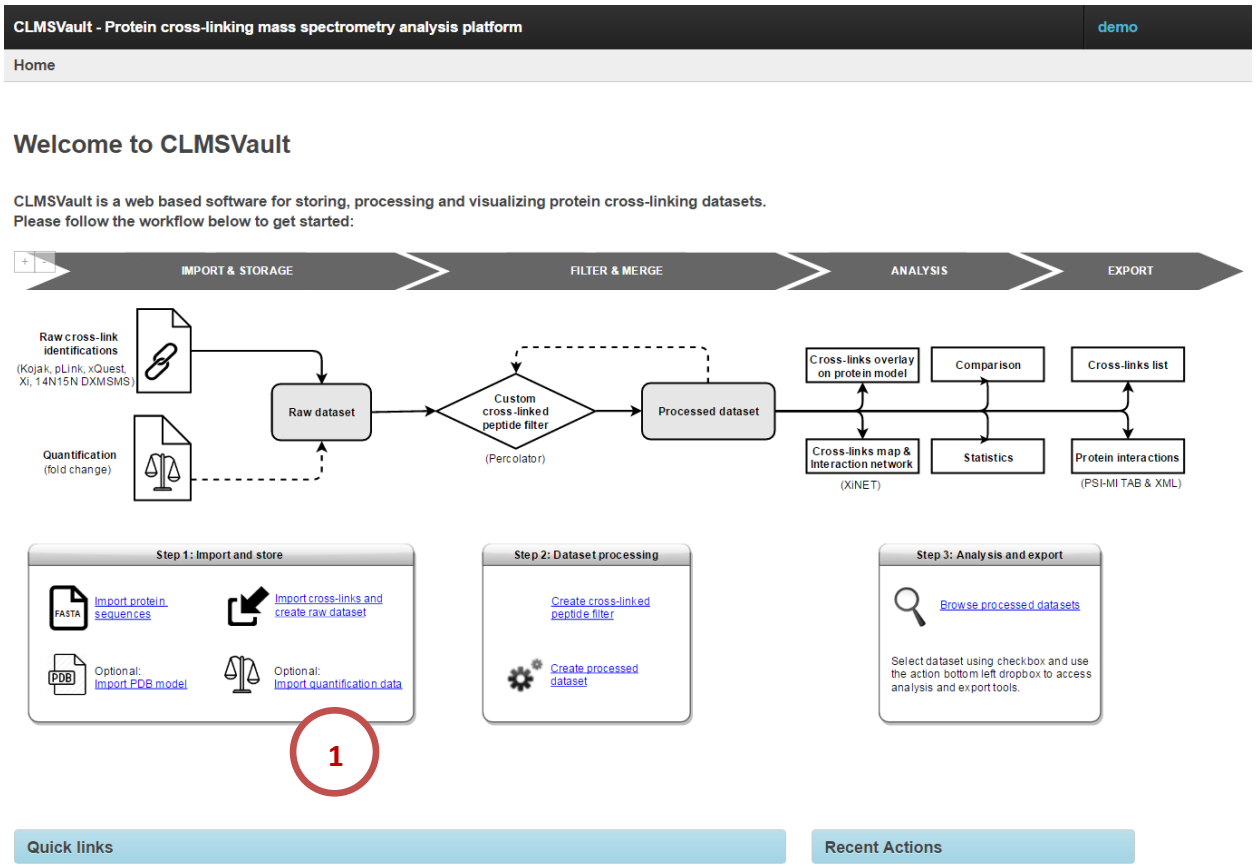
Quantification data can be linked to cross-linked peptide identification within CLMSVault for further filtering or visualization. Quantification data can be imported from a CSV file. There are two CSV file formats accepted. Those file formats must have one of the following headers and the respective data associated with it in each column:

1. `"File","ScanNumber","FoldChange"`
2. `"CLPeptideId","FoldChange"`

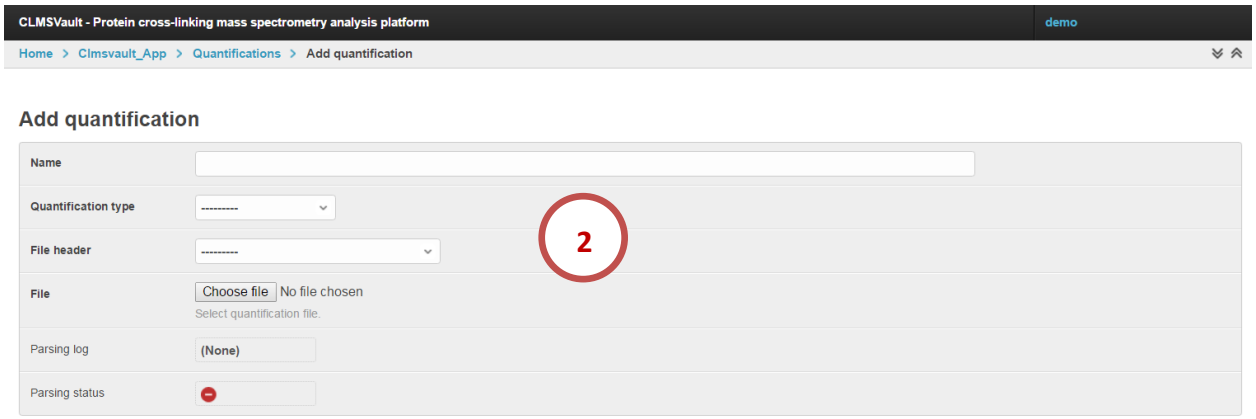
The first format should allow import from most quantification software after reformatting the results. The second format allows direct linking to CLMSVault internal cross-link identifier (CLPeptideId). Those are obtained by exporting cross-links from CLMSVault before quantification. Two files named “CF_quantfile_test.csv” and “FSF_quantfile_test.csv” are available in the CLMSVault archive to demonstrate the format.

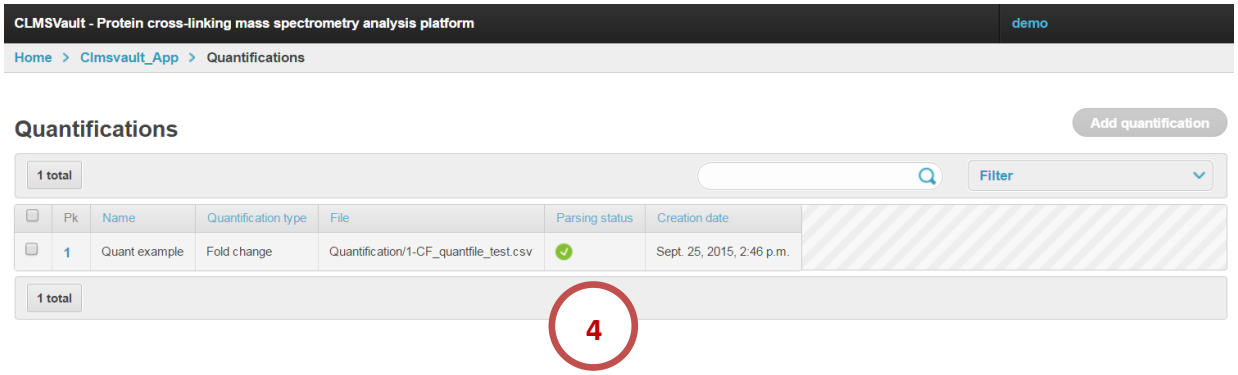
Import of quantification dataset is done with the following steps:

1. Click on “Import quantification data” from the main screen.



2. Fill the form.
3. Click “Save”.
4. Check parsing status.





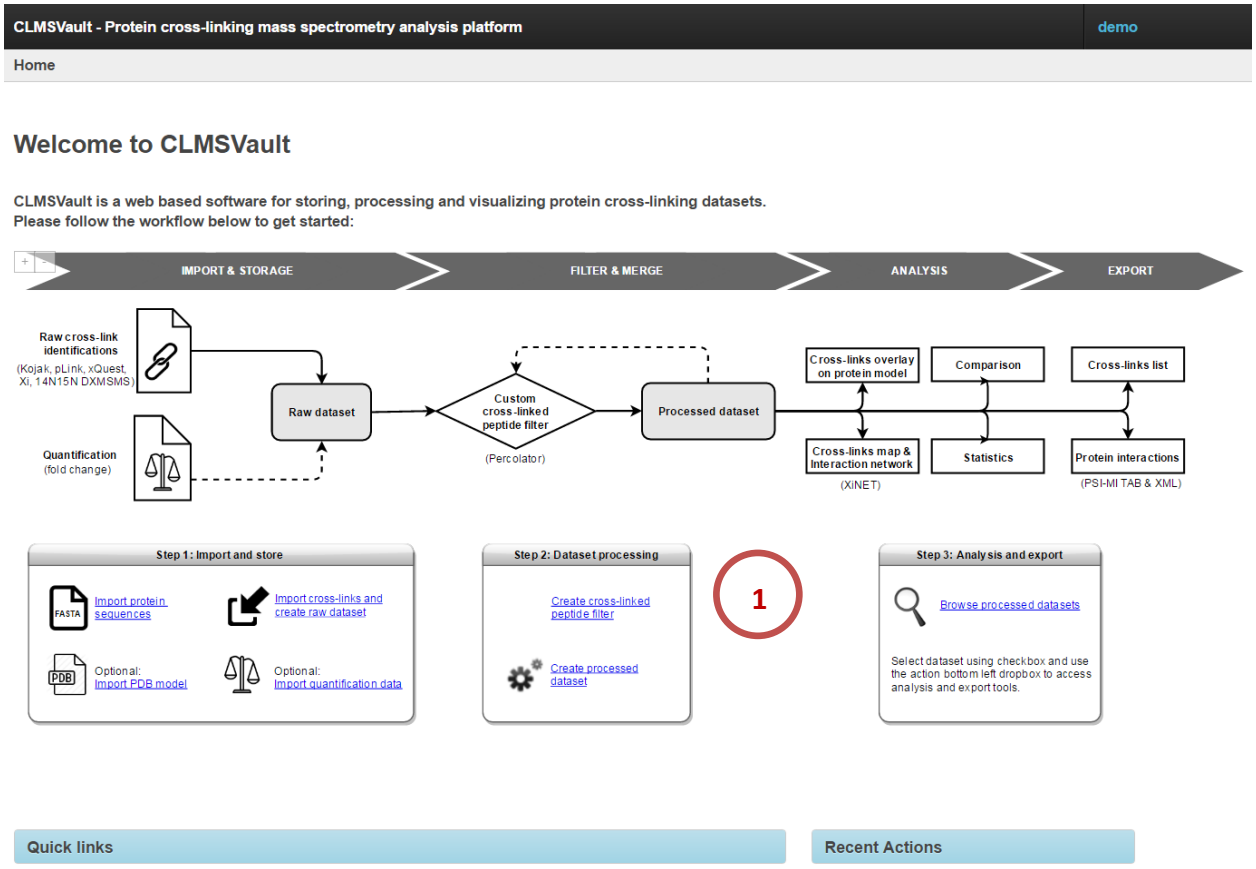
5. Cross-link datasets filtering and merging

Cross-link datasets are presented in two forms in CLMSVault: raw and processed. A raw dataset is the list of all cross-linked peptides imported from the search engine into CLMSVault database (refer to section 4.2). A processed dataset is a subset of cross-linked peptides. It is created by applying various filters to a raw or processed dataset. It can also be created by merging multiple dataset from different experiments. Processed datasets are used for data analysis and visualization.

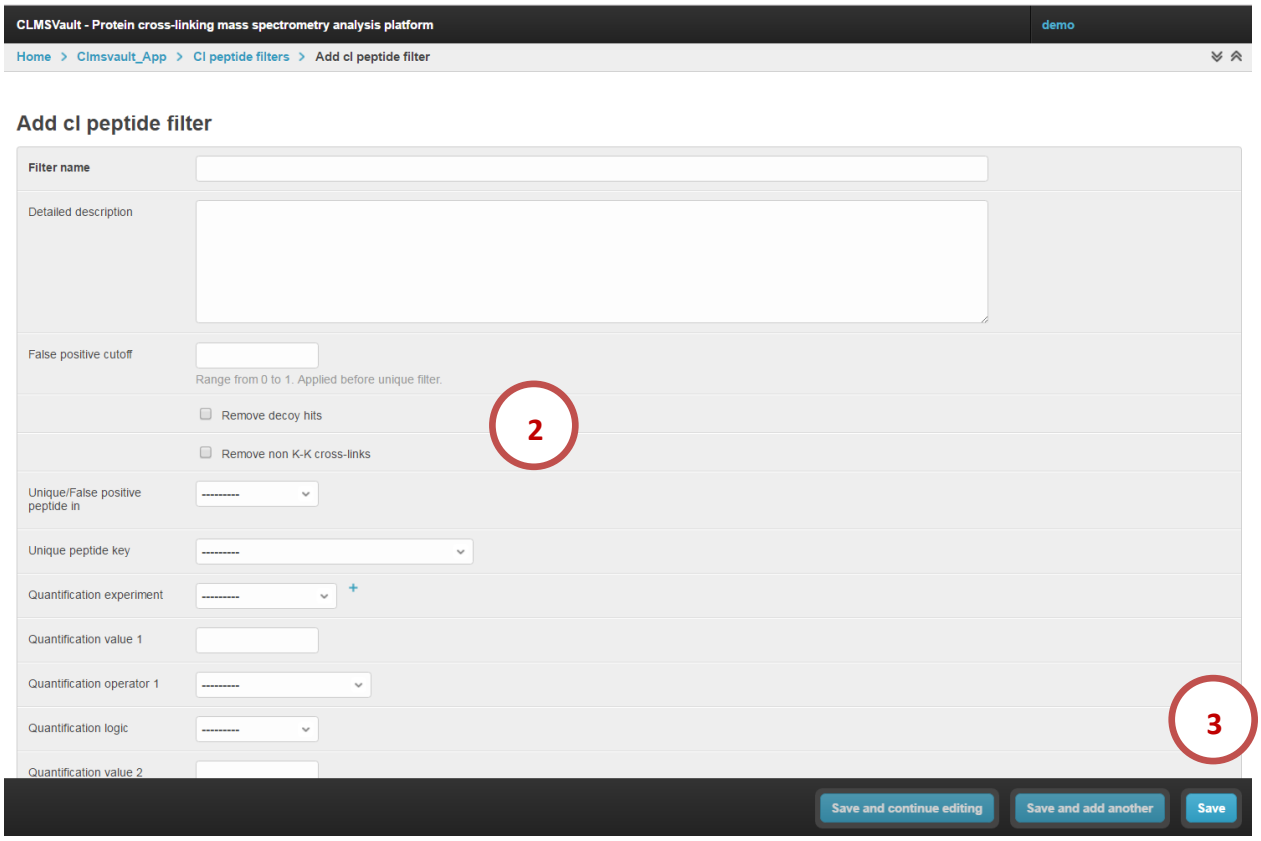
5.1. Filters

A filter in CLMSVault is a re-usable entity that is used to filter or exclude peptides from a dataset to create a processed dataset.

1. Click on “Create cross-linked peptide filter” from the main screen to expand the menu.

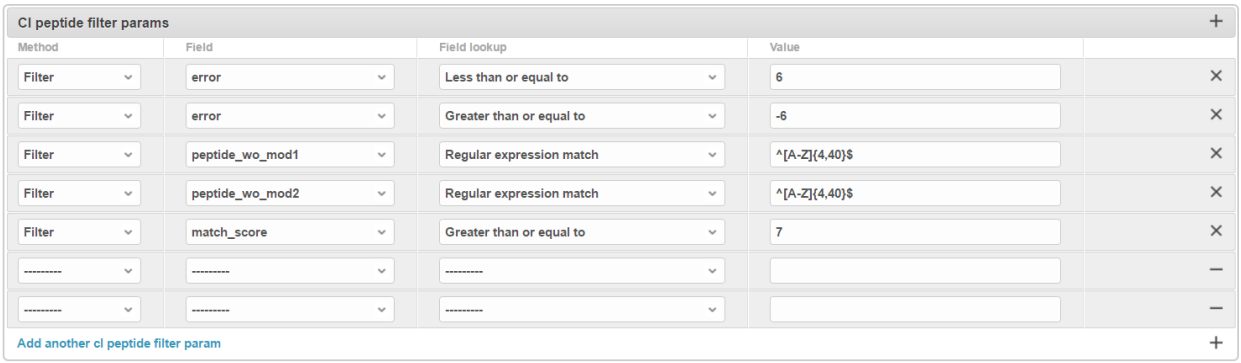


2. Fill the form. The form is divided into two parts. The first part has a global filter such false positive cutoff, removal of decoys and unique cross-link. For unique cross-link, a key must be selected to specify unique definition. Unique can be applied per MS run file or dataset. This section also contains a filter for quantification. The bottom part specifies a filter for individual cross-link peptide.



3. Click “Save”.

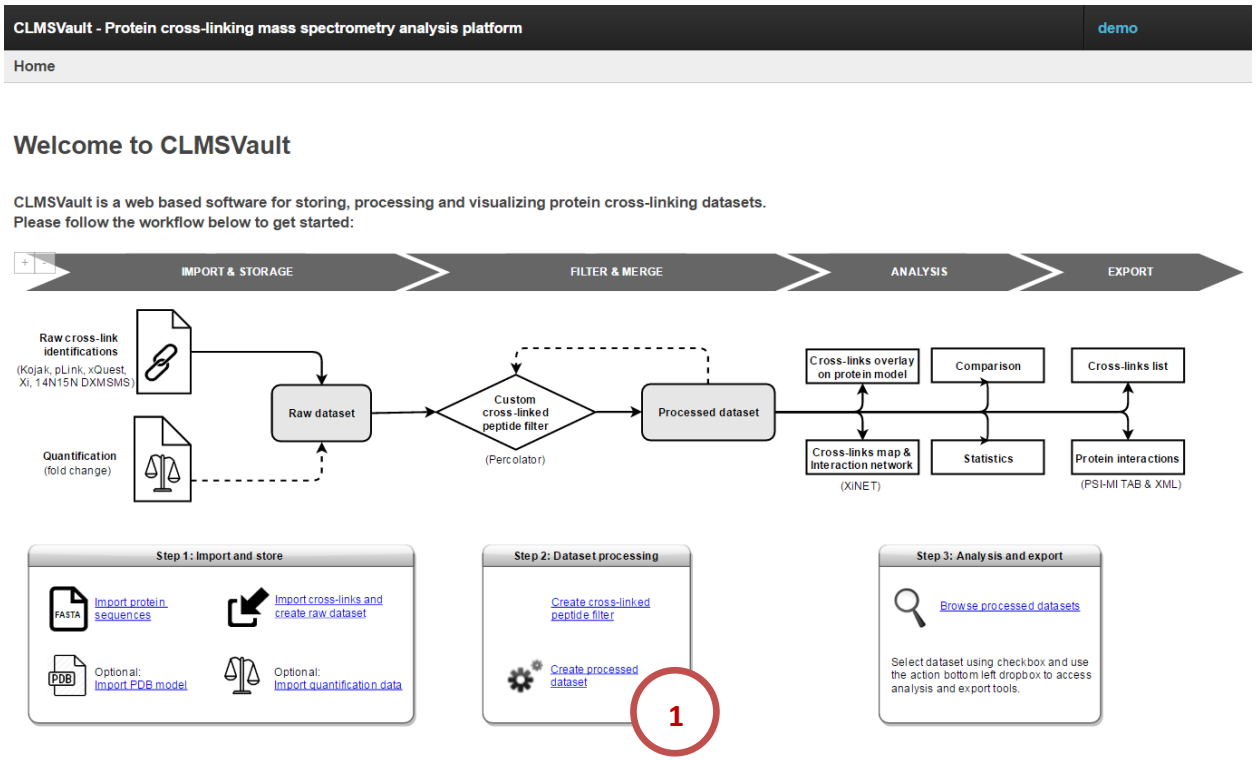
For example, the following filter will keep peptides with a mass error within 6 ppm, the length between 4-40 residues and with a score above 7.



5.2. Creation of processed dataset

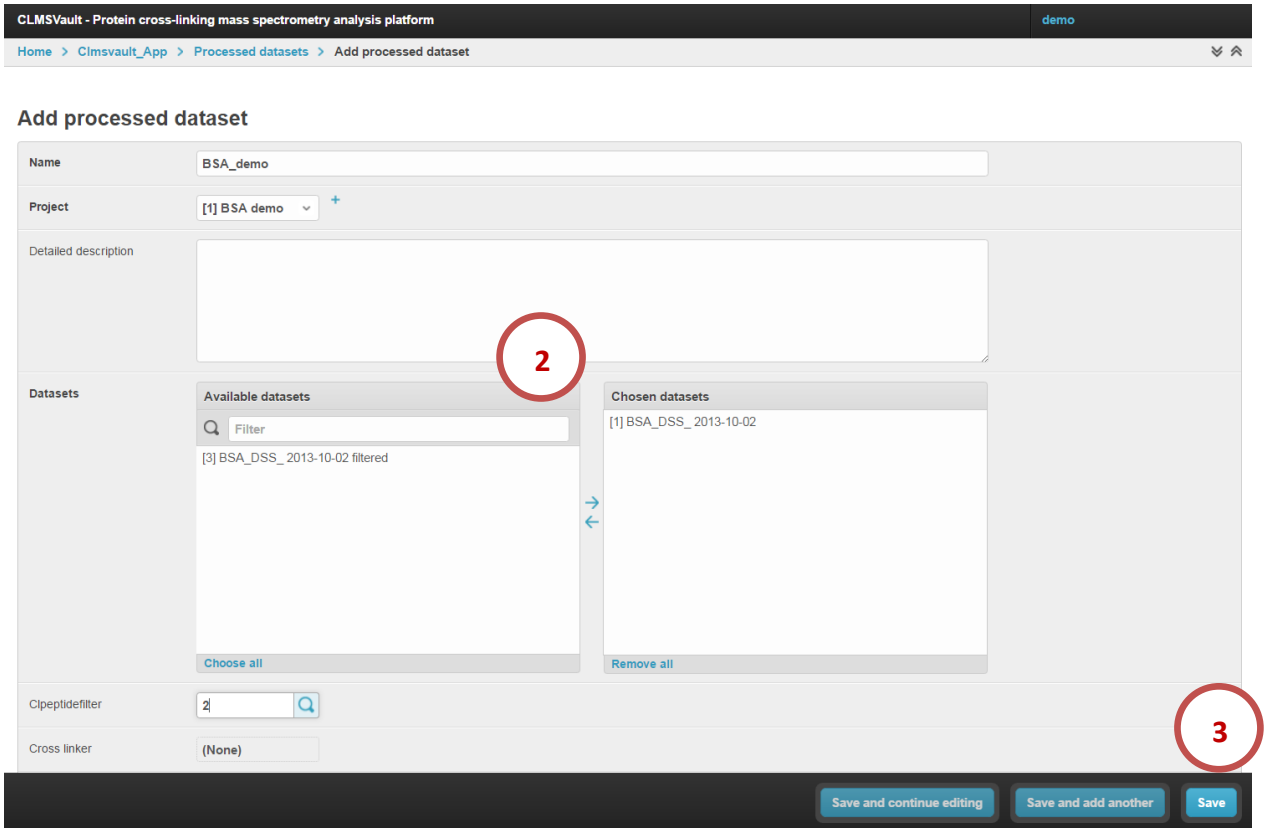
Once filters have been created, they can be applied to filter raw dataset to create a processed dataset. The filter can be applied iteratively to dataset until the desired list of cross-linked peptides is obtained.

1. Click on “Create processed dataset” from the main screen to expand the menu.



2. Select one or multiple datasets and the desired filter to apply.

3. Click “Save”.

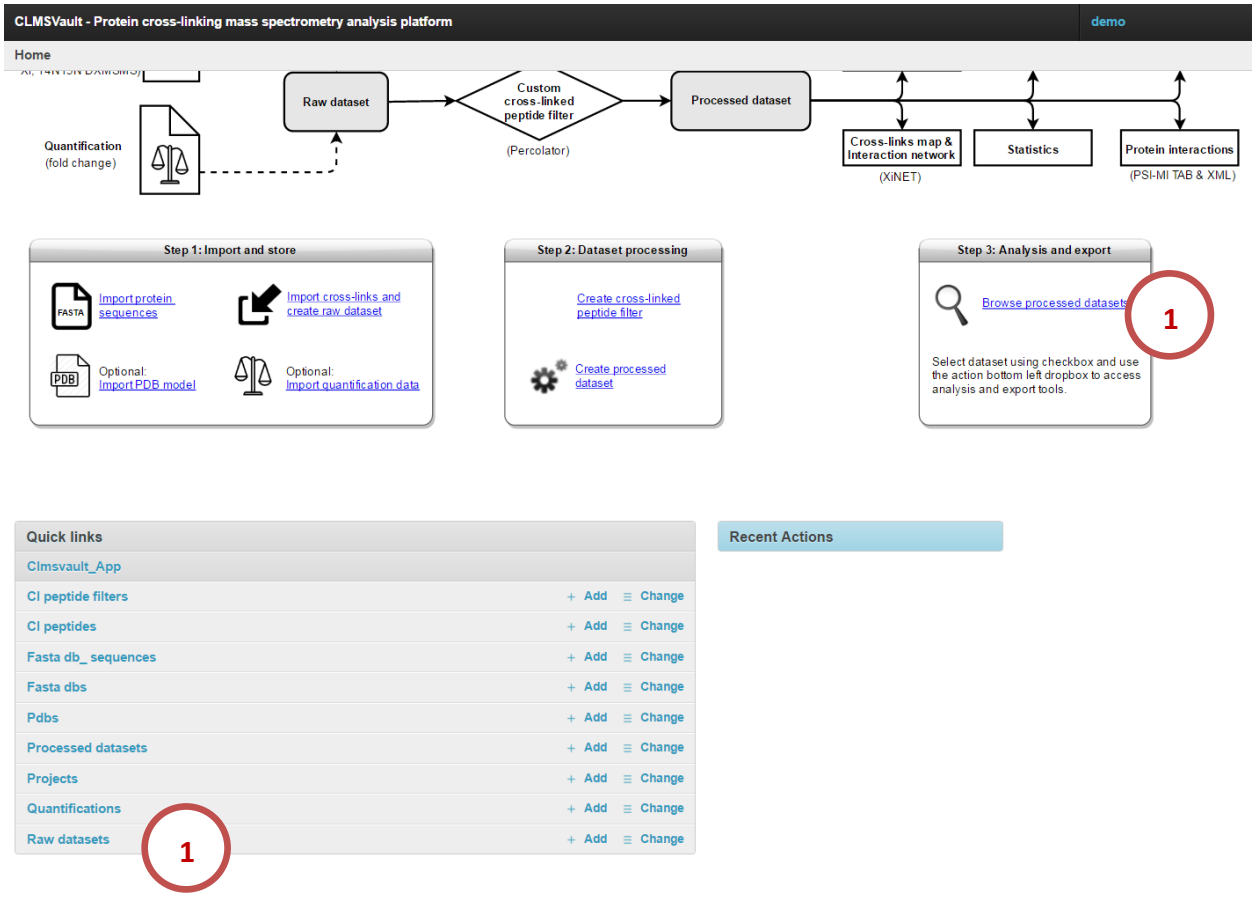


Refer to section 6 to analyze your processed dataset.

5.3. Creation of processed dataset with Percolator

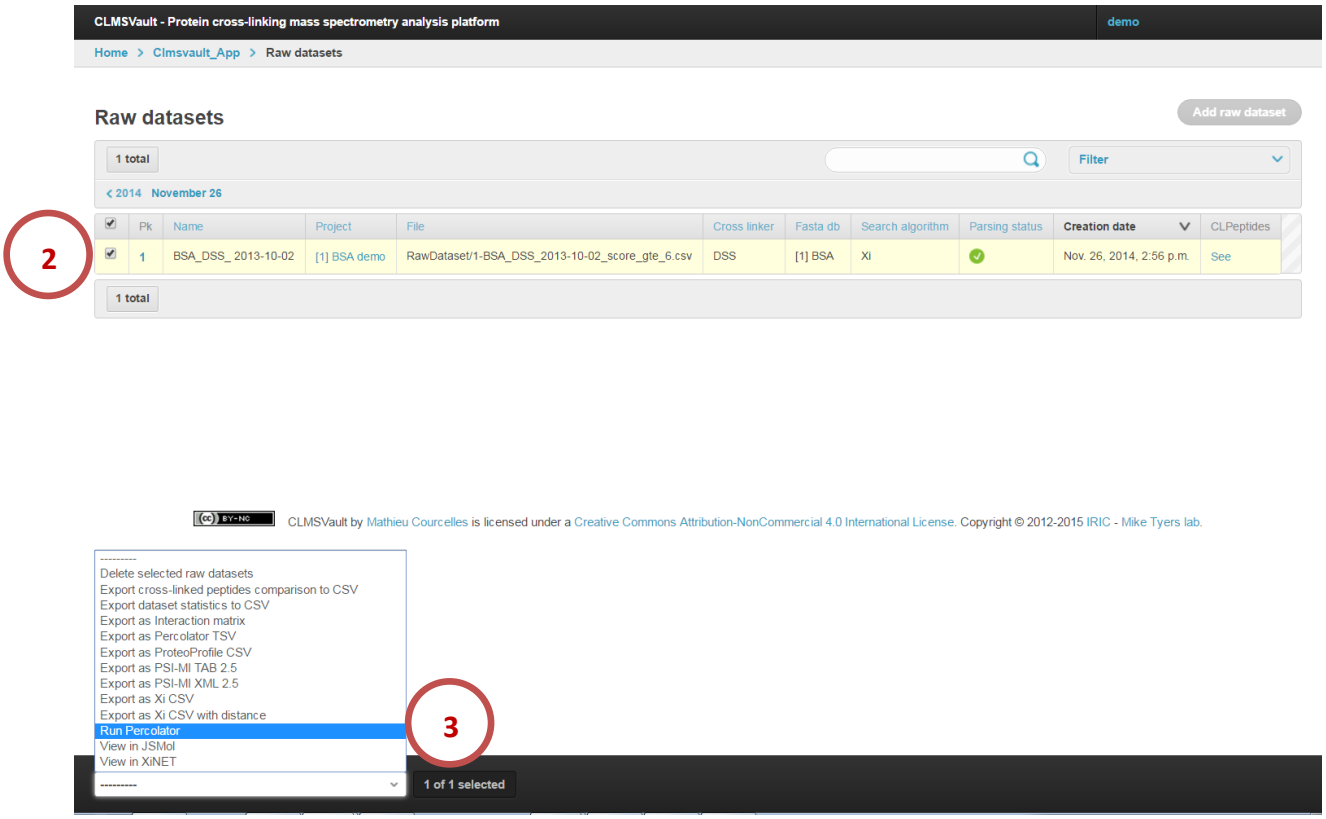
Percolator can be used to create a processed dataset with a defined false discovery rate. Be sure that the dataset was searched with a target/decoy database.

1. Go to either “Raw datasets” or “Processed datasets” view from the main screen.



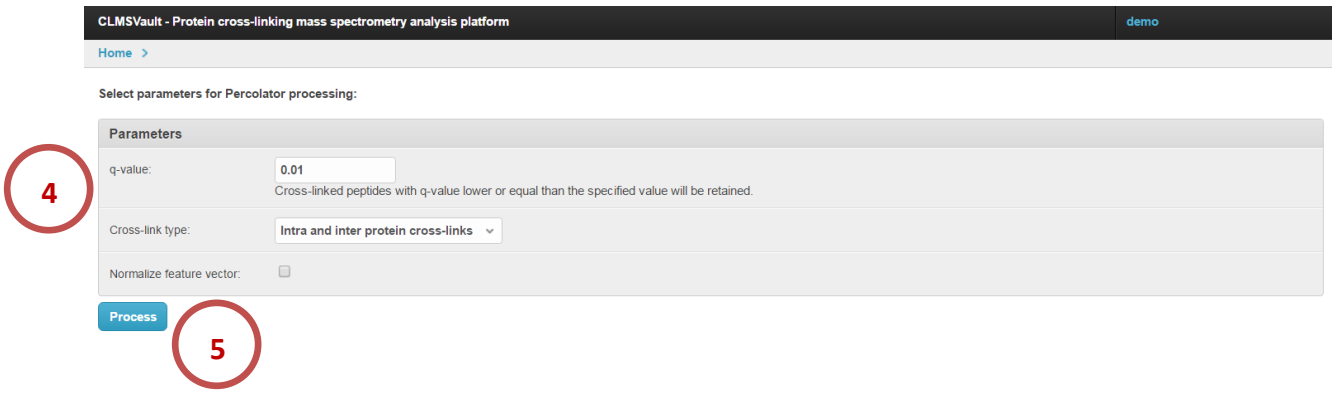
2. Click on the checkbox to select the dataset that you want to process with Percolator.

3. From the action menu, select “Run Percolator”.



4. Choose q-value threshold and cross-link type.

5. Click “Process”.



6. Check the following report to see if any error occurred in Percolator. Your new processed dataset is listed at the bottom.

Percolator version 2.09, Build Date Apr 3 2015 09:42:26
Copyright (c) 2006-9 University of Washington. All rights reserved.
Written by Lukas Käll (lukall@u.washington.edu) in the
Department of Genome Sciences at the University of Washington.
Issued command:
/usr/bin/percolator -t 0.01 -F 0.01 -U /tmp/tmp0R98zY
Started Wed Jan 25 09:09:12 2017
Hyperparameters selectionFdr=0.01, Cpos=0, Cneg=0, maxNiter=10
Reading Tab delimited input from datafile /tmp/tmp0R98zY
Features:
Score dScore Charge Mass PPM LenShort LenLong LenSum
Train/test set contains 66 positives and 16 negatives, size ratio=4.125 and pi0=1
Warning : the number of negative samples read is too small to perform a correct classification.

selecting cpos by cross validation
selecting cneg by cross validation
Selected feature number 1 as initial search direction, could separate 7 positives in that direction
Selected feature number 7 as initial search direction, could separate 29 positives in that direction
Selected feature number 7 as initial search direction, could separate 10 positives in that direction
Estimating 24 over q=0.01 in initial direction
Reading in data and feature calculation took 0 cpu seconds or 0 seconds wall time
---Training with Cpos selected by cross validation, Cneg selected by cross validation, fdr=0.01
Iteration 1 : After the iteration step, 39 target PSMs with q<0.01 were estimated by cross validation
Iteration 2 : After the iteration step, 44 target PSMs with q<0.01 were estimated by cross validation
Iteration 3 : After the iteration step, 45 target PSMs with q<0.01 were estimated by cross validation
Iteration 4 : After the iteration step, 45 target PSMs with q<0.01 were estimated by cross validation
Iteration 5 : After the iteration step, 45 target PSMs with q<0.01 were estimated by cross validation
Iteration 6 : After the iteration step, 45 target PSMs with q<0.01 were estimated by cross validation
Iteration 7 : After the iteration step, 45 target PSMs with q<0.01 were estimated by cross validation
Iteration 8 : After the iteration step, 45 target PSMs with q<0.01 were estimated by cross validation
Iteration 9 : After the iteration step, 45 target PSMs with q<0.01 were estimated by cross validation
Iteration 10 : After the iteration step, 45 target PSMs with q<0.01 were estimated by cross validation
Obtained weights (only showing weights of first cross validation set)
first line contains normalized weights, second line the raw weights
Score dScore Charge Mass PPM LenShort LenLong LenSum m0
1.11 -0.5211 0.2377 -1.0643 -0.0420 1.4890 -0.9398 -0.3025 -0.6322
0.7519 -0.0000 0.2312 -0.0011 -0.0198 0.4555 -0.1174 -0.0344 -4.5409
After all training done, 29 target PSMs with q<0.0100 were found when measuring on the test set
Found 29 target PSMs scoring over 1.0000% FDR level on testset
Merging results from 3 datasets
Selecting pi_0=0.4204
Calibrating statistics - calculating q values
New pi_0 estimate on merged list gives 29 PSMs over q=0.0100
Calibrating statistics - calculating Posterior error probabilities (PEPs)
Processing took 0.22 cpu seconds or 1 seconds wall time

A new Processed dataset was created with Percolator results: [4] BSA_DSS_ 2013-10-02 filtered (percolated: q-value <=0.01, cl-type=IntraInter, nv=False)

Processed datasets

Add processed dataset

2 total

Filter

2014 2017

	Pk	Name	Project	Datasets	Filter	Cross linker	Fasta db	Search algorithm	Creation date	CLPeptides
	4	BSA_DSS_ 2013-10-02 filtered (percolated: q-value <=0.01, cl-type=IntraInter, nv=False)	[1] BSA demo	[3] BSA_DSS_ 2013-10-02 filtered		DSS	[1] BSA	Xi	Jan. 25, 2017, 9:09 a.m.	See
	3	BSA_DSS_ 2013-10-02 filtered	[1] BSA demo	[1] BSA_DSS_ 2013-10-02	[2] Match score >= 7, Error: +-6 ppm, Pep. Length: 4-40 a.a. Unique CL positions	DSS	[1] BSA	Xi	Nov. 26, 2014, 3:32 p.m.	See

2 total

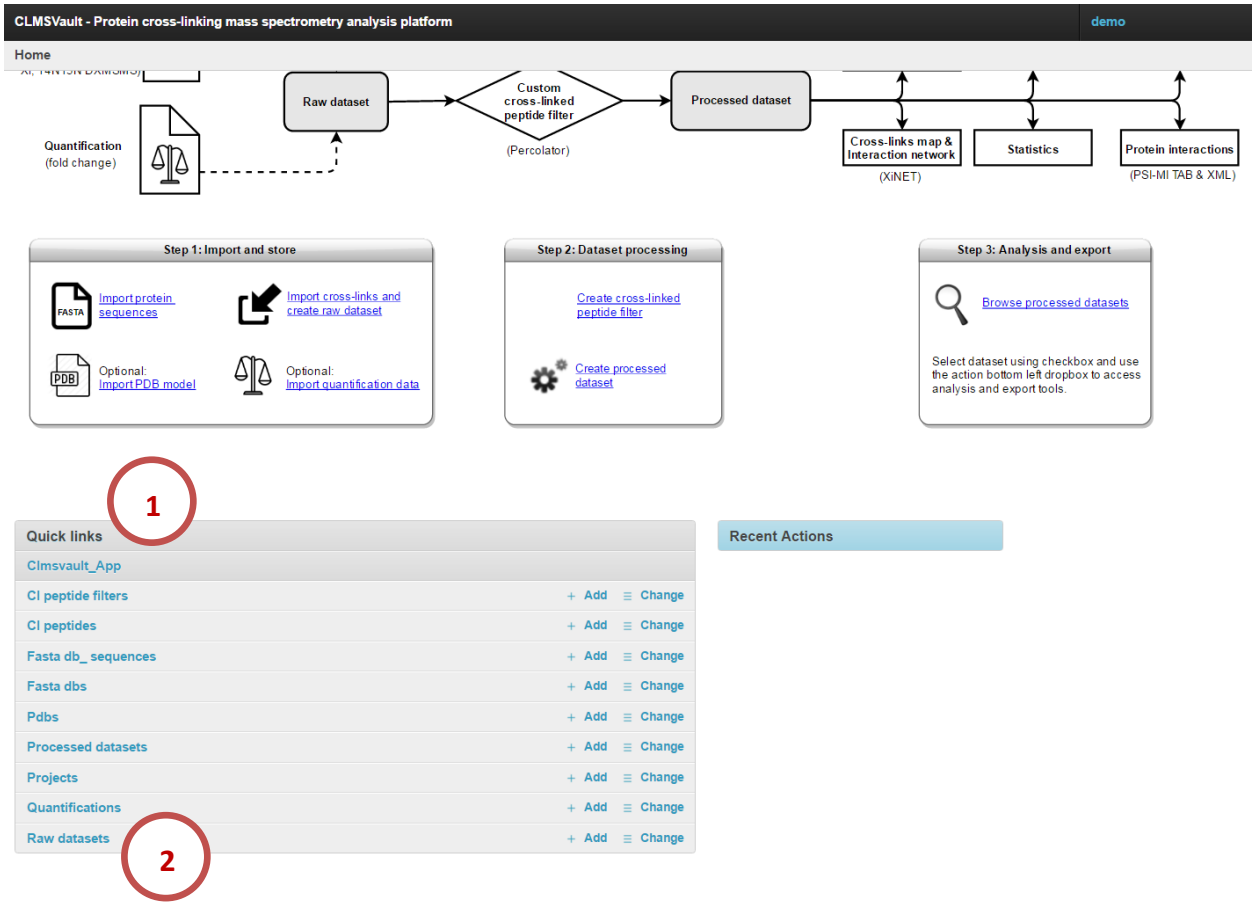
6. Analysis, visualization, and export

6.1. List, view, and delete

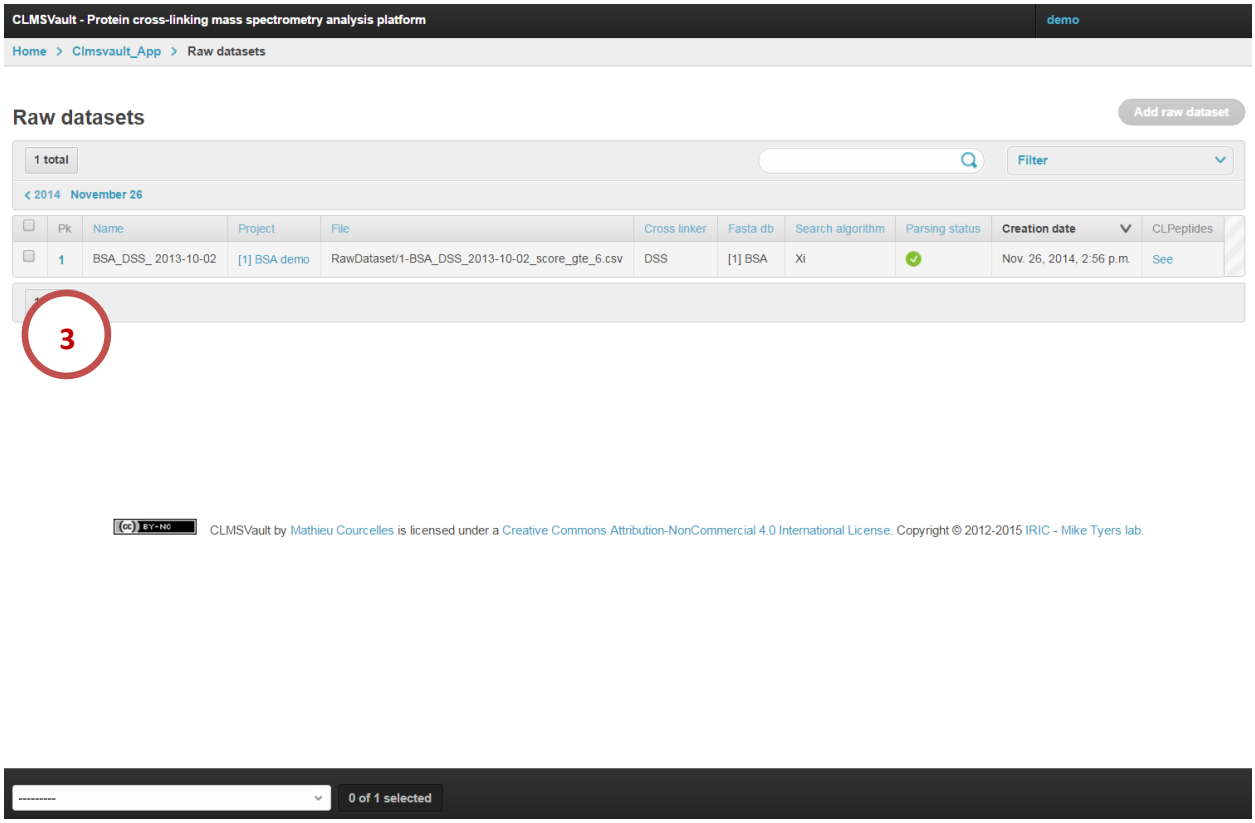
To list, view or delete any object stored in CLMSVault, use the following instructions. Here is an example for raw dataset object:

1. Click on “Quick links” from the main screen to expand the menu.

2. Click on “Raw datasets” to list of raw datasets stored in CLMSVault.

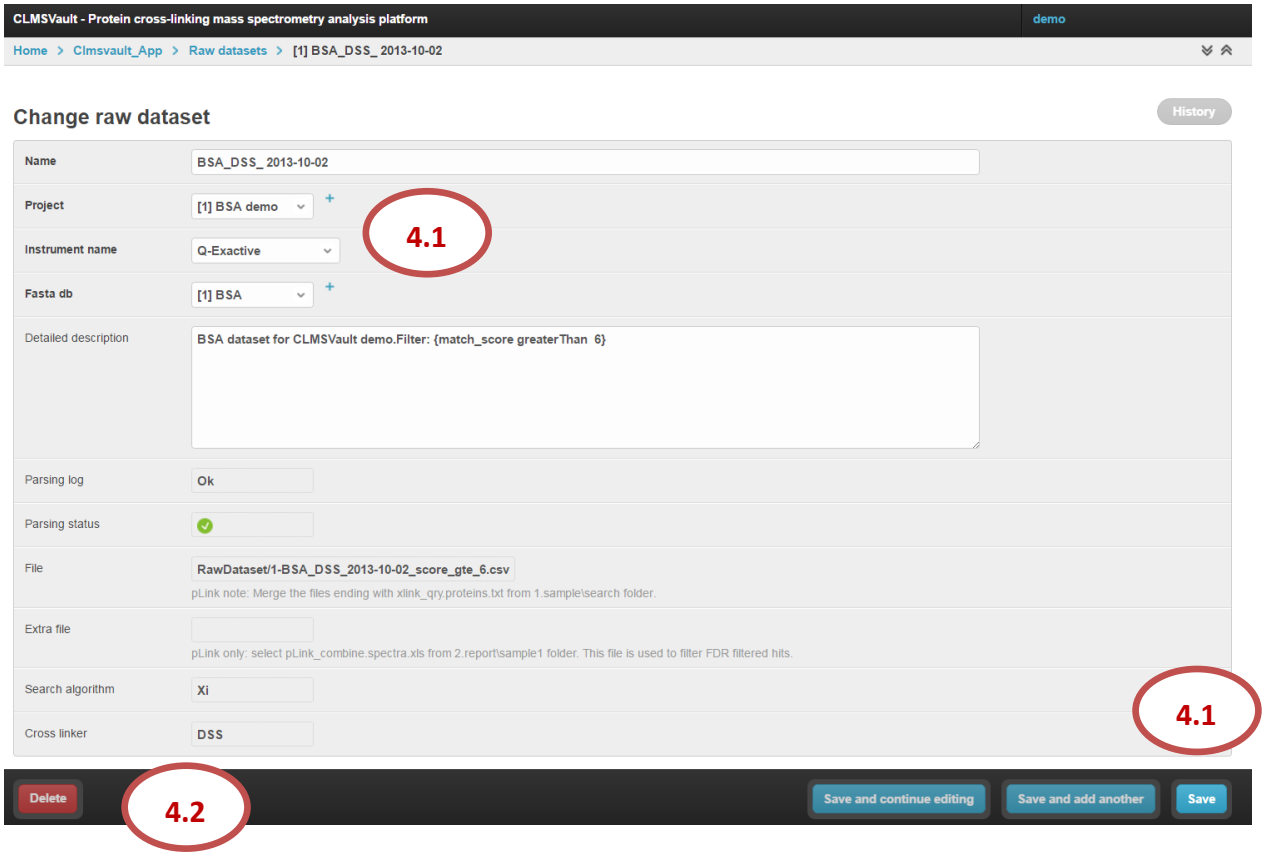


3. To view and edit raw dataset details, click on the dataset numeric identifier in the “Pk” column.



4. From the “Change” view, you can either:

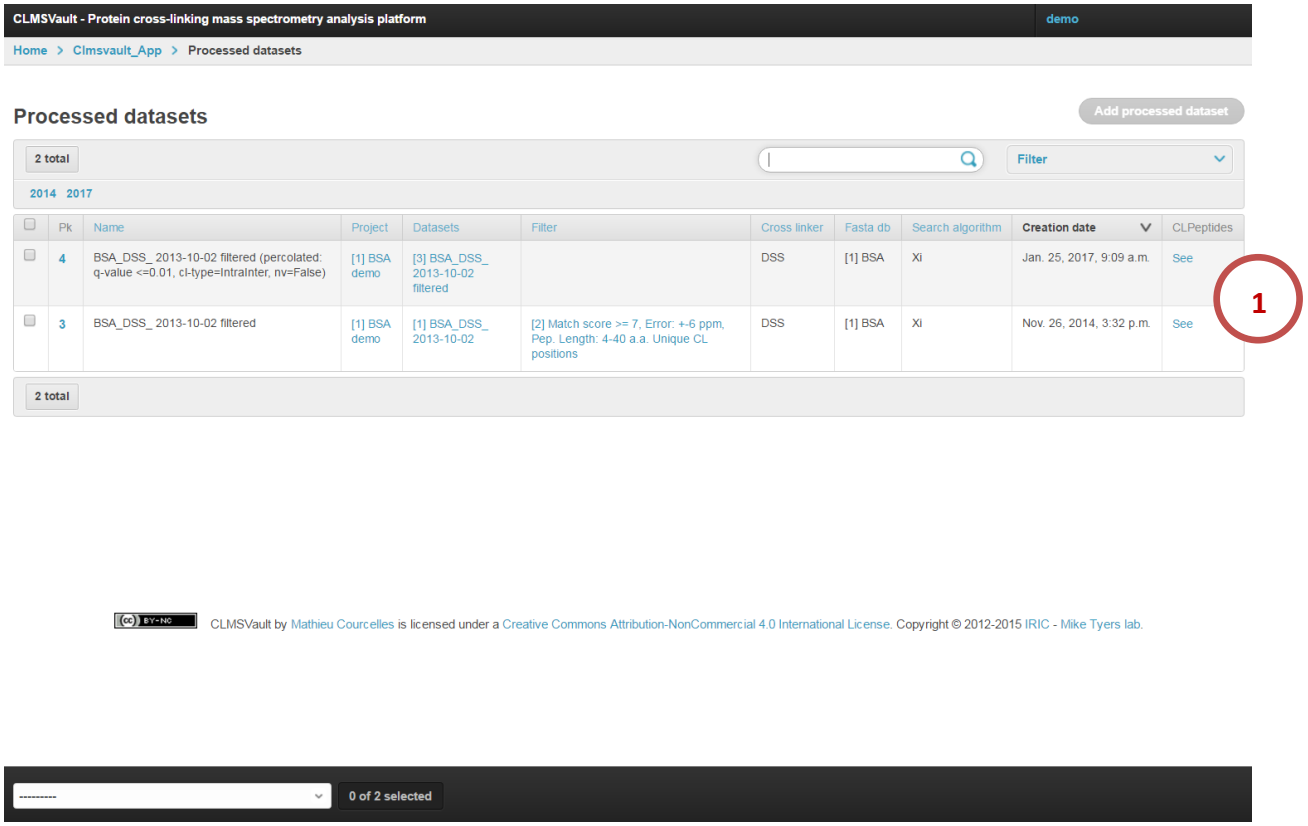
   4.1. Modify the raw dataset (modification of certain fields are restricted) and click “Save”.
   4.2. Delete the raw dataset (deletion can be restricted for some object).



6.2. CL Peptides view

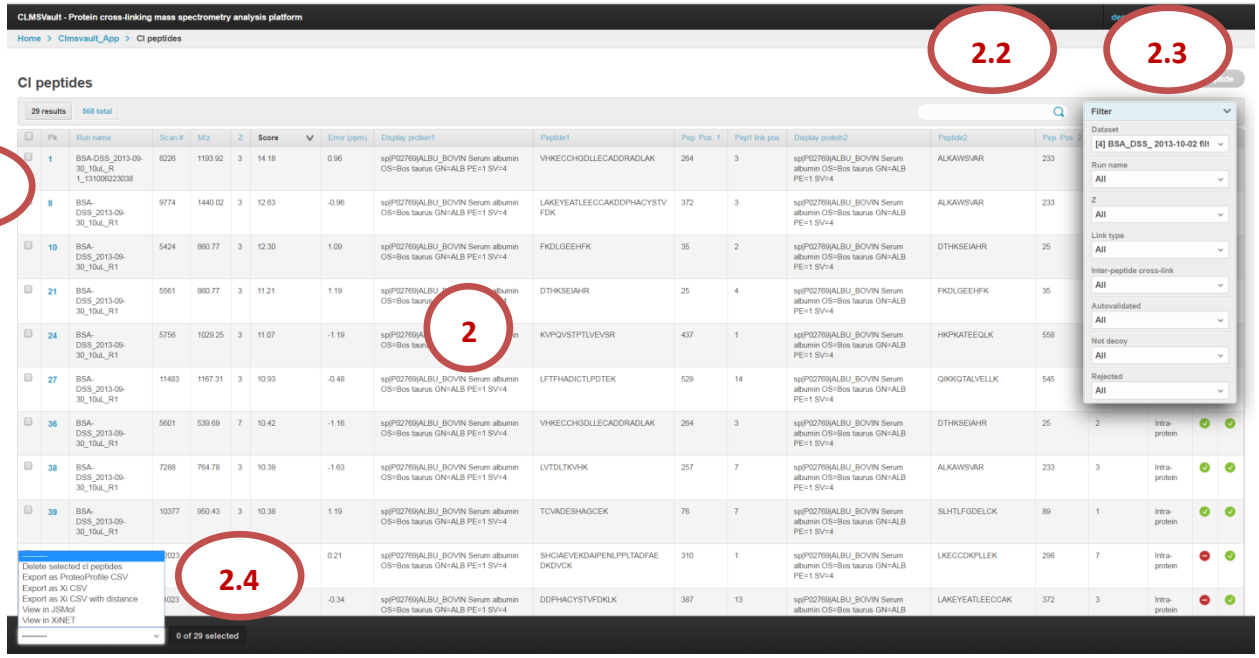
The “CL Peptides view” provides the list of all cross-linked peptides associated to a dataset.

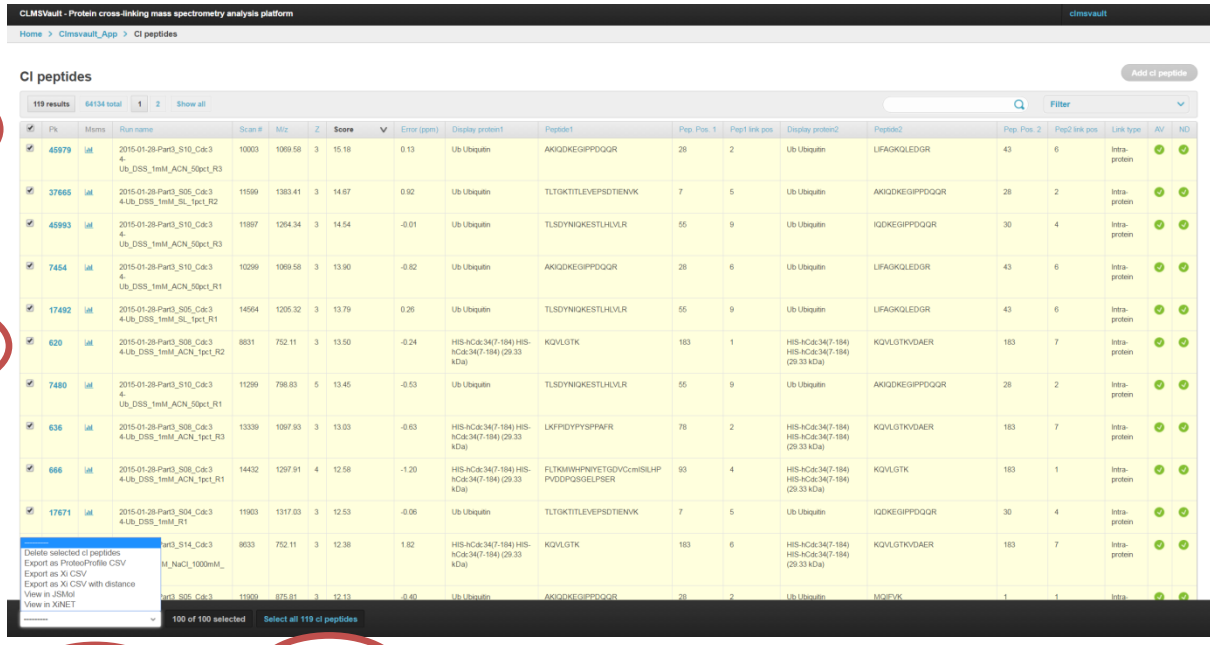
1. To access it, click on “See” from the CL Peptides column in the raw or processed dataset view.



2. The view provides details for the individual cross-linked peptide.

   2.1. Each individual peptide can be viewed, edited or delete by clicking on their identifier in the “pk” column.
   2.2. Keyword search is available to restrict the list of displayed peptides.
   2.3. The filter drop-down menu allows further filtering with various categories.
   2.4. The action drop-down menu allows to delete peptides or to run different analysis on a subset of peptides. Actions are applied on selected items.
      2.4.1. Check the boxes to select the subset of peptides
      2.4.2. If you want all peptides, you must click the top box
      2.4.3. Click “Select all” button at the bottom.
      2.4.4. Select your action

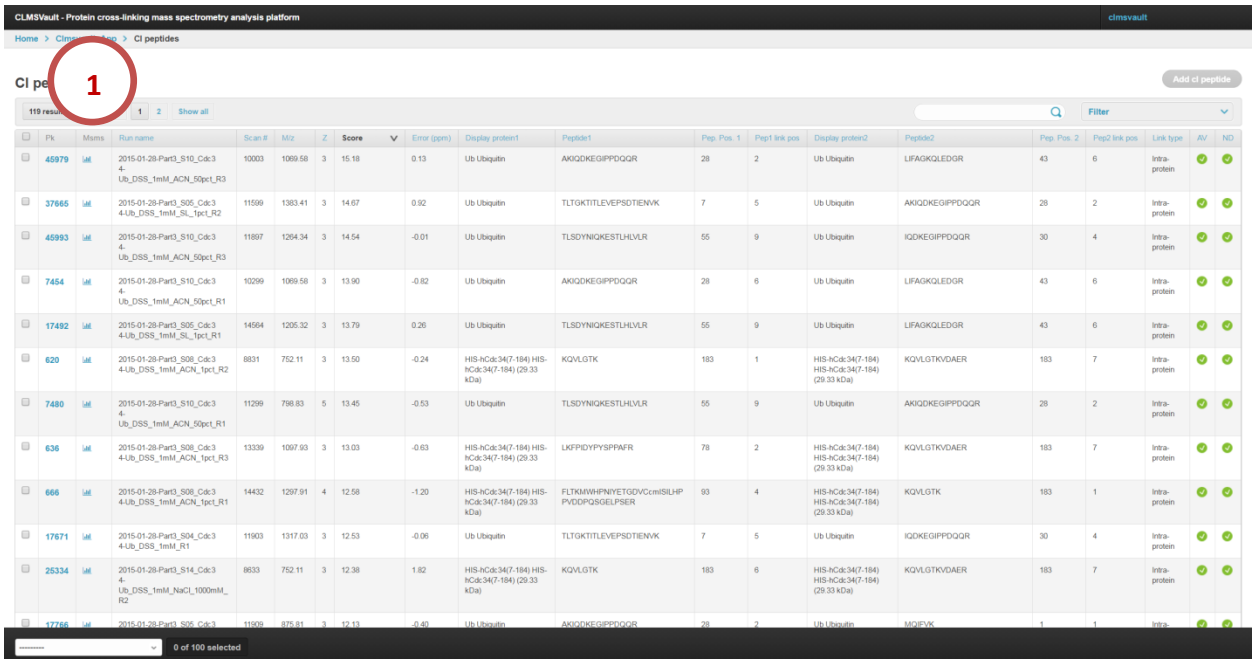

CLMSVault - Protein cross-linking mass spectrometry analysis platform
Home > Clmsvault_App > CI peptides
CI peptides
29 results
568 total
Filter
Dataset
[4] BSA_DSS_ 2013-10-02 fil
Run name
Z
Link type
Inter-peptide cross-link
Autovalidated
Not decoy
Rejected
Delete selected cl peptides
Export as ProteoProfile CSV
Export as Xi CSV
Export as Xi CSV with distance
View in JSMol
View in XiNET
0 of 29 selected
2.1
2
2.2
2.3
2.4


CLMSVault - Protein cross-linking mass spectrometry analysis platform
clmsvault
Home > Clmsvault_App > CI peptides
CI peptides
Add cl peptide
119 results
64134 total
Show all
Filter
Delete selected cl peptides
Export as ProteoProfile CSV
Export as Xi CSV
Export as Xi CSV with distance
View in JSMol
View in XiNET
100 of 100 selected
Select all 119 cl peptides
2.4.
2.4.
2.4.
2.4.

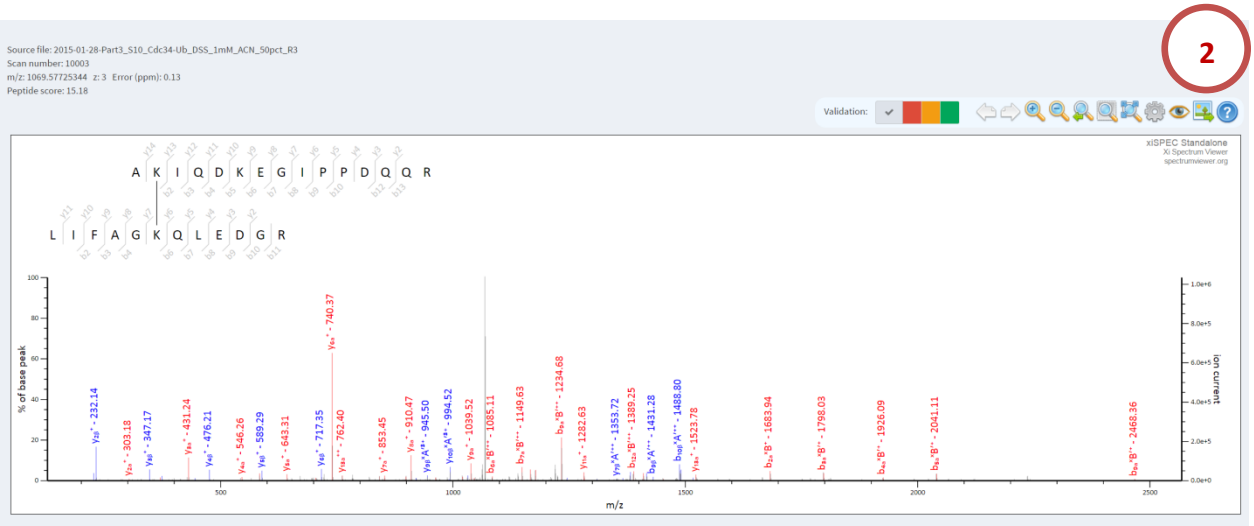
6.3. Spectrum viewer

MS/MS spectrum of cross-linked peptides can be viewed with the embedded Xi spectrum viewer MS/MS spectra file are imported into CLMSVault via MGF file which has been generated by msconvert tool from the Proteowizard. Refer to section 4.3 for spectra import.

1. From the “CL Peptides” screen, click on the icon in the “MSMS” column. This will show the spectrum in the interactive Xi spectrum viewer.



2. Spectrum can be exported to SVG file.



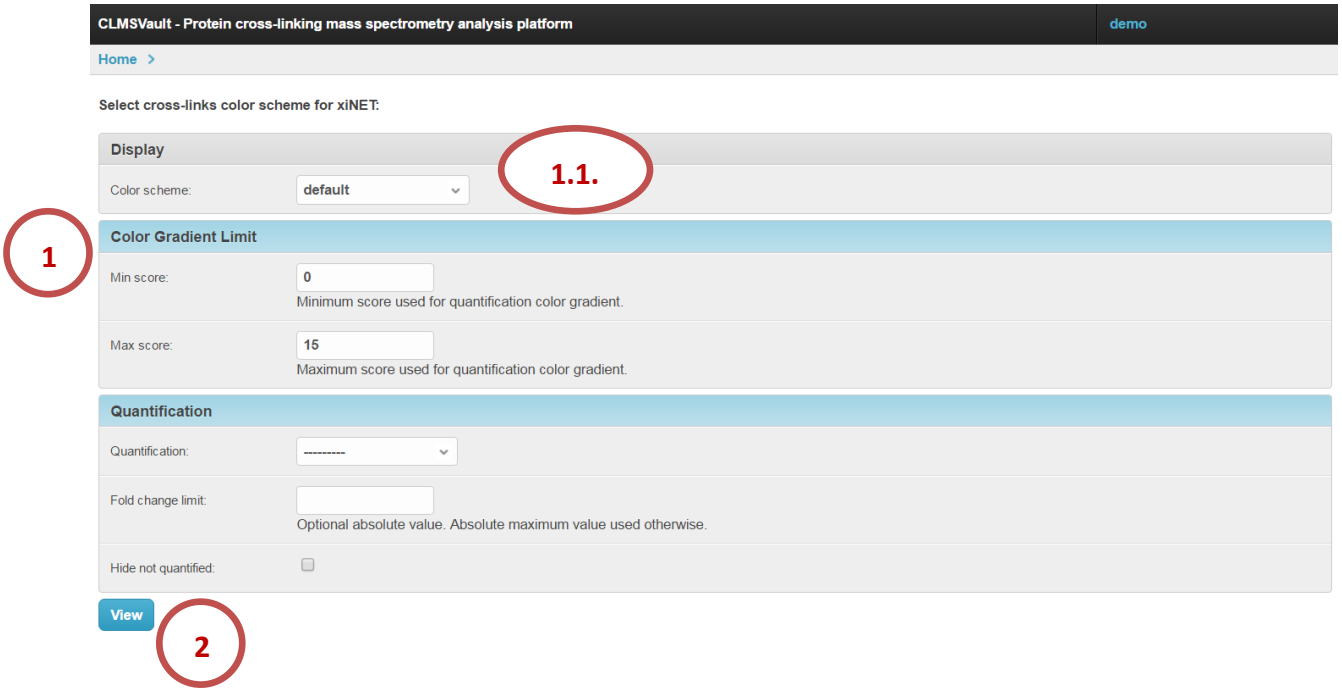
6.4. View in XiNET

XiNET displays cross-link map of intra or inter-protein cross-links. This action is accessed from either the “CL Peptides” or the “Dataset” views after selecting the desired objects.

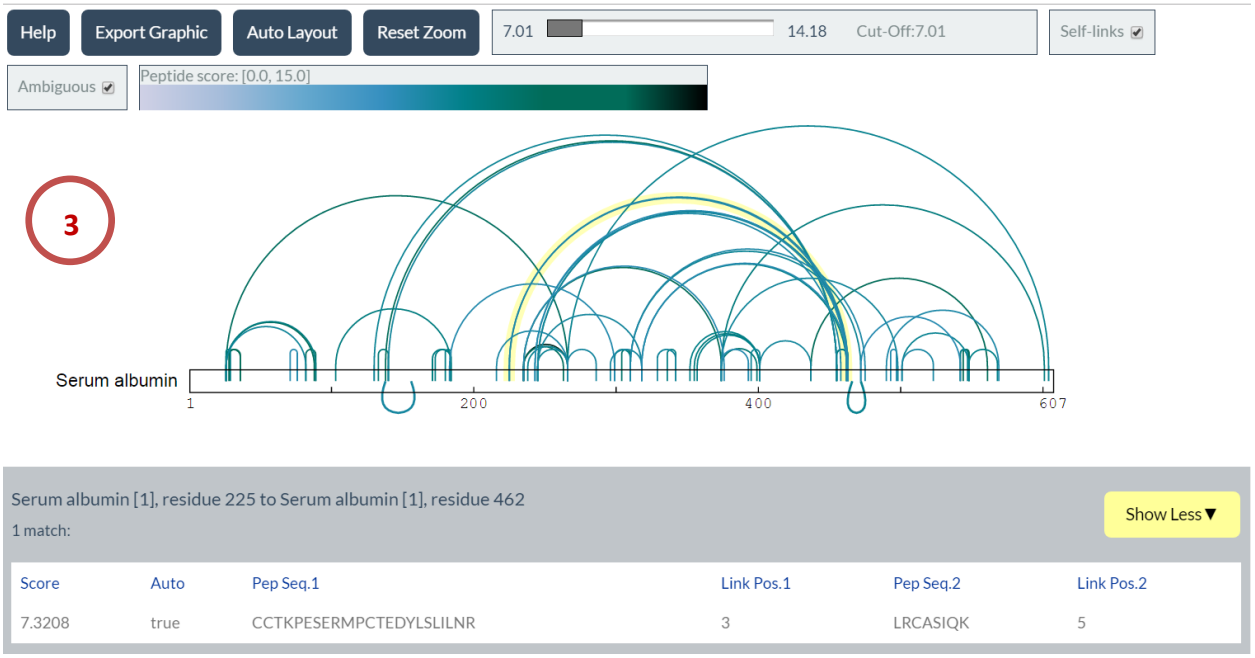
3. Select the desired parameters.

1.1. By default, XiNET colors cross-link by their cross-link type (inter, intra, dead-end). In CLMSVault, you can also color cross-links by their peptide score or by fold-change.

4. Click “View”.



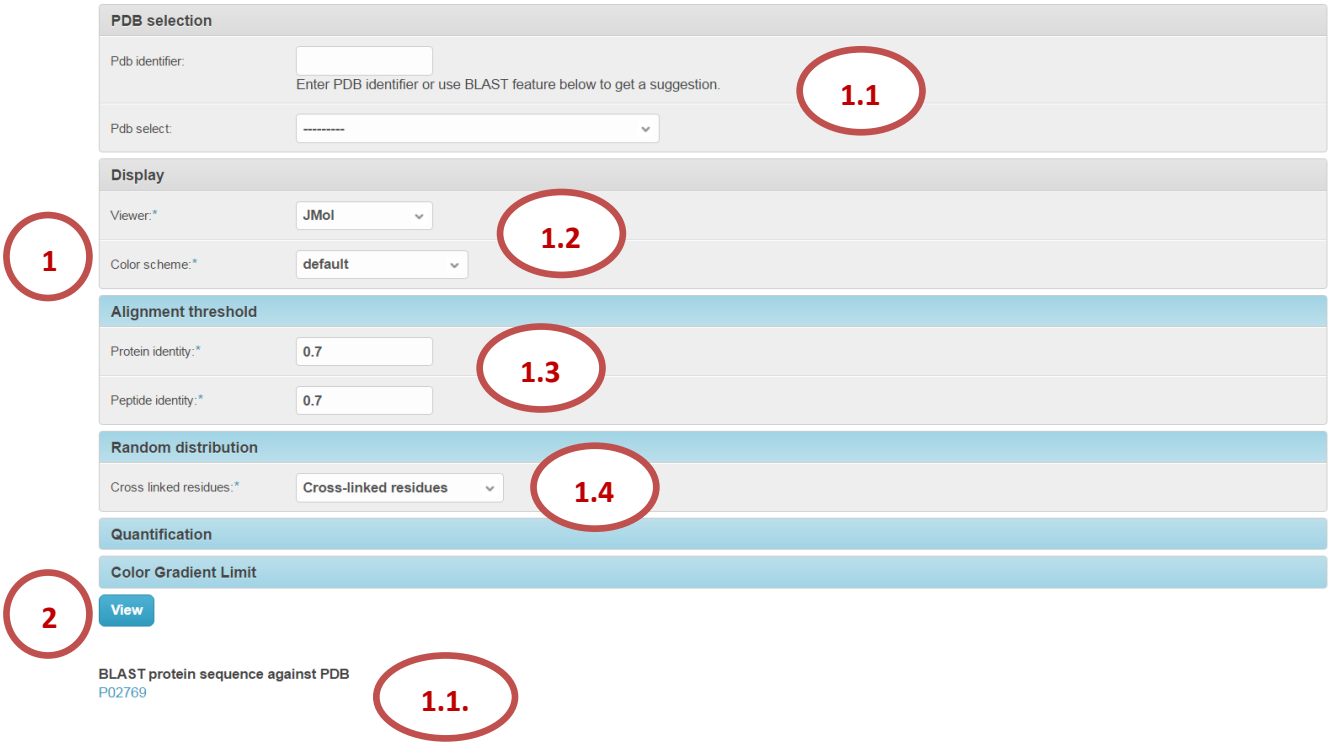
5. Interact with the XiNET to visualize cross-linked regions.



6.5. View in JSMol

Identified cross-links can be displayed on PDB model to visualize their spatial distribution. This action is accessed from either the “CL Peptides” or the “Dataset” views after selecting the desired objects.

1. Select the desired parameters.
    1.1. PDB selection: Select PDB file stored in CLMSVault or enter PDB identifier to automatically retrieve it from RCSB Protein Data bank.
        1.1.1. For each protein, a link is provided to run a protein Blast to query identifiers of available PDB models.
    1.2. Display: JMol and JSMol viewers are available to display the model. JMol requires Java installed. Rendering speed can differ between the two viewers.
    1.3. Alignment threshold: CLMSVault uses sequence alignments to match and align protein sequences used for cross-links identification and protein sequences in the PDB model. The alignment is also used to position the cross-linked peptides in the model. Protein and peptide thresholds are thus used to control this. These parameters allow matching protein and peptides in a model with residues substitution and sequence gap. It should also allow mapping to homologous protein model.
    1.4. Random distribution: CLMSVault plots distribution of inter-residues cross-link distances. It also computes random distance distributions to evaluate if cross-links are random hits for this model. To compute this distribution, the user must specify which residues can be cross-linked.
2. Click “View”.



3. The next view presents the 3D viewer. It is separated into four parts:

3.1. JSMol/JMol viewer:
Left-click allows rotating the molecule and the right-click displays the menu to access all functionalities provided by JSMol/JMol viewer. Check boxes are provided to toggle the display of chain.

3.2. Cross-linked peptides list:

This part list for each cross-link, the mapping of cross-linked residues between your identifications and the PDB model. Cross-linked residues can be labeled and highlighted in red by clicking the label in the PDB residue columns. Cross-link inter-residue distance is reported in the “Distance” column. PI columns report the peptide identity value between the sequence from the MS/MS search and the PDB model. “S/H” checkbox toggles the display of the cross-links on the viewer. At the bottom of this section, there is also a color gradient for displayed cross-links for either peptide score or quantification values. There are also some controls to automatically display cross-links with a length lower than a specified threshold. Additional controls are provided to control JSMol/JMol image quality and rendering speed.

3.3. Cross-link inter-residues distance distributions:

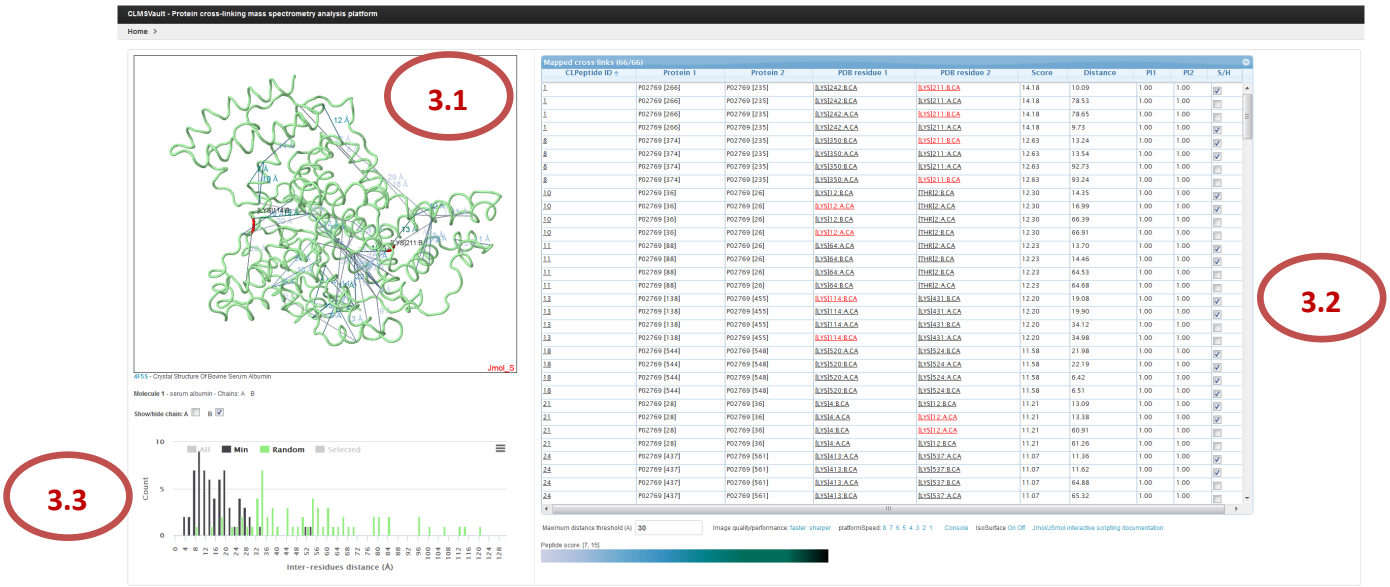
Four cross-link inter-residue distance distributions are plotted:

- **All:** all possible cross-links
- **Minimum:** In the case where cross-links are mapped multiple in the model (e.g. dimer), it will keep only the distance with the minimum value.
- **Random**: Random distribution based on user-selected definition (see 6.4-1.4) and the size of cross-link list.
- **Selected**: Distribution of user selected cross-links.

A Mann-Whitney rank test is done to indicate if the identified cross-link distance distribution differs from thousand random distributions.

3.4. Sequence alignments:

This section displays protein sequences alignments and indicates which protein sequences have been paired between the MS/MS sequences and the PDB model.



Mann-Whitney rank test

Two-sided p-value for 1000 draw of sample with size of 66.

p-value <= 6.8e-11

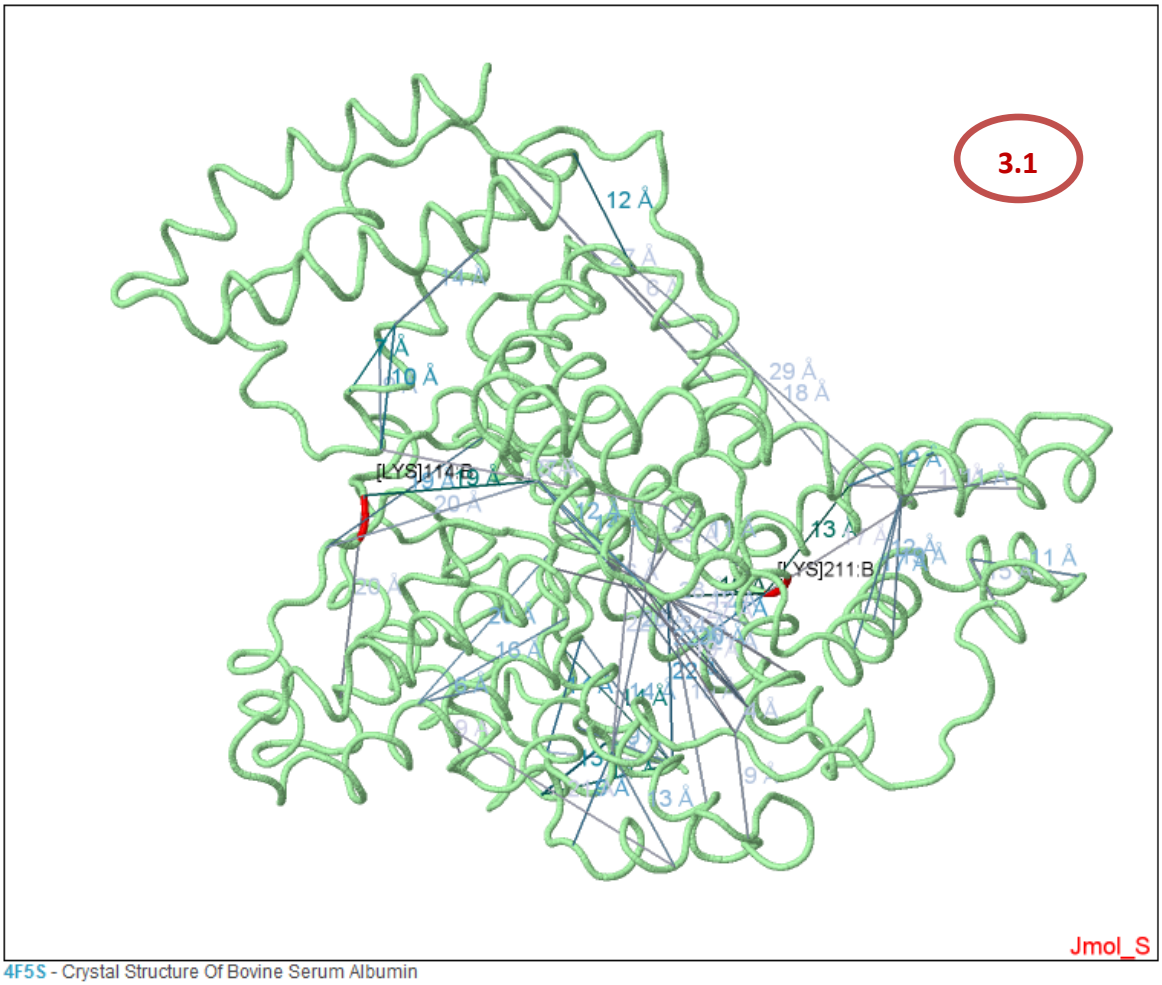
A small p-value indicates that the observed cross-links distance distribution differs from a random selection distribution. Please select the appropriate random distribution for your cross-linker in the previous page.

Alignments

```
P02769 - Chain A
MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEPERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKIETMREKVLASSARQRLRCASIQKFGERALKAWSVARLSQKFPKAEFV
------------------------DTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEPERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKIETMREKVLTSSARQRLRCASIQKFGERALKAWSVARLSQKFPKAEFV
Protein identity: 0.998

P02769 - Chain B
MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEPERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKIETMREKVLTSSARQRLRCASIQKFGERALKAWSVARLSQKFPKAEFV
------------------------DTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPFDEHVKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEPERNECFLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYYANKYNGVFQECCQAEDKGACLLPKIETMREKVLTSSARQRLRCASIQKFGERALKAWSVARLSQKFPKAEFV
Protein identity: 0.998
```

3.4



4F5S - Crystal Structure Of Bovine Serum Albumin

Molecule 1 - serum albumin - Chains: A B

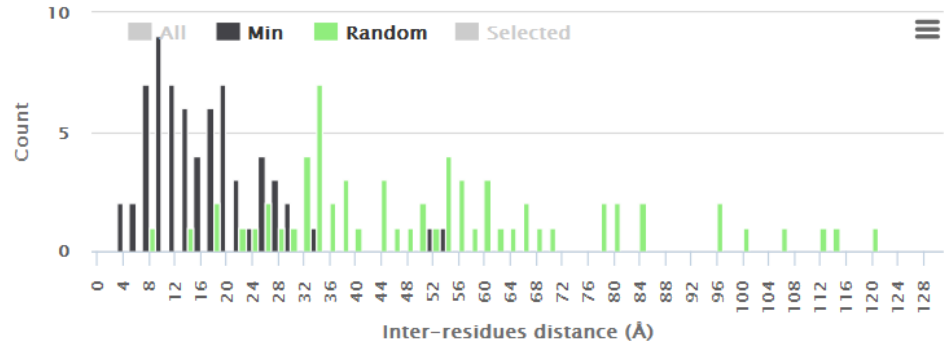
Show/hide chain: A ☐ B ☑

3.2

Mapped cross-links (66/66)

CLPeptide ID	Protein 1	Protein 2	PDB residue 1	PDB residue 2	Score	Distance	PI1	PI2	S/H
1	P02769 [266]	P02769 [235]	[LYS]242:B.CA	[LYS]211:B.CA	14.18	10.09	1.00	1.00	☑
1	P02769 [266]	P02769 [235]	[LYS]242:B.CA	[LYS]211:A.CA	14.18	78.53	1.00	1.00	☐
1	P02769 [266]	P02769 [235]	[LYS]242:A.CA	[LYS]211:B.CA	14.18	78.65	1.00	1.00	☐
1	P02769 [266]	P02769 [235]	[LYS]242:A.CA	[LYS]211:A.CA	14.18	9.73	1.00	1.00	☑
8	P02769 [374]	P02769 [235]	[LYS]350:B.CA	[LYS]211:B.CA	12.63	13.24	1.00	1.00	☑
8	P02769 [374]	P02769 [235]	[LYS]350:A.CA	[LYS]211:A.CA	12.63	13.54	1.00	1.00	☑
8	P02769 [374]	P02769 [235]	[LYS]350:B.CA	[LYS]211:A.CA	12.63	92.73	1.00	1.00	☐
8	P02769 [374]	P02769 [235]	[LYS]350:A.CA	[LYS]211:B.CA	12.63	93.24	1.00	1.00	☐
10	P02769 [36]	P02769 [26]	[LYS]12:B.CA	[THR]2:B.CA	12.30	14.35	1.00	1.00	☑
10	P02769 [36]	P02769 [26]	[LYS]12:A.CA	[THR]2:A.CA	12.30	16.99	1.00	1.00	☑
10	P02769 [36]	P02769 [26]	[LYS]12:B.CA	[THR]2:A.CA	12.30	66.39	1.00	1.00	☐
10	P02769 [36]	P02769 [26]	[LYS]12:A.CA	[THR]2:B.CA	12.30	66.91	1.00	1.00	☐
11	P02769 [88]	P02769 [26]	[LYS]64:A.CA	[THR]2:A.CA	12.23	13.70	1.00	1.00	☑
11	P02769 [88]	P02769 [26]	[LYS]64:B.CA	[THR]2:B.CA	12.23	14.46	1.00	1.00	☑
11	P02769 [88]	P02769 [26]	[LYS]64:A.CA	[THR]2:B.CA	12.23	64.53	1.00	1.00	☐
11	P02769 [88]	P02769 [26]	[LYS]64:B.CA	[THR]2:A.CA	12.23	64.68	1.00	1.00	☐
13	P02769 [138]	P02769 [455]	[LYS]114:B.CA	[LYS]431:B.CA	12.20	19.08	1.00	1.00	☑
13	P02769 [138]	P02769 [455]	[LYS]114:A.CA	[LYS]431:A.CA	12.20	19.90	1.00	1.00	☑
13	P02769 [138]	P02769 [455]	[LYS]114:A.CA	[LYS]431:B.CA	12.20	34.12	1.00	1.00	☐
13	P02769 [138]	P02769 [455]	[LYS]114:B.CA	[LYS]431:A.CA	12.20	34.98	1.00	1.00	☐
18	P02769 [544]	P02769 [548]	[LYS]520:A.CA	[LYS]524:B.CA	11.58	21.98	1.00	1.00	☑
18	P02769 [544]	P02769 [548]	[LYS]520:B.CA	[LYS]524:A.CA	11.58	22.19	1.00	1.00	☑
18	P02769 [544]	P02769 [548]	[LYS]520:A.CA	[LYS]524:A.CA	11.58	6.42	1.00	1.00	☑
18	P02769 [544]	P02769 [548]	[LYS]520:B.CA	[LYS]524:B.CA	11.58	6.51	1.00	1.00	☑
21	P02769 [28]	P02769 [36]	[LYS]4:B.CA	[LYS]12:B.CA	11.21	13.09	1.00	1.00	☑
21	P02769 [28]	P02769 [36]	[LYS]4:A.CA	[LYS]12:A.CA	11.21	13.38	1.00	1.00	☑
21	P02769 [28]	P02769 [36]	[LYS]4:B.CA	[LYS]12:A.CA	11.21	60.91	1.00	1.00	☐
21	P02769 [28]	P02769 [36]	[LYS]4:A.CA	[LYS]12:B.CA	11.21	61.26	1.00	1.00	☐
24	P02769 [437]	P02769 [561]	[LYS]413:A.CA	[LYS]537:A.CA	11.07	11.36	1.00	1.00	☑
24	P02769 [437]	P02769 [561]	[LYS]413:B.CA	[LYS]537:B.CA	11.07	11.62	1.00	1.00	☑
24	P02769 [437]	P02769 [561]	[LYS]413:A.CA	[LYS]537:B.CA	11.07	64.88	1.00	1.00	☐
24	P02769 [437]	P02769 [561]	[LYS]413:B.CA	[LYS]537:A.CA	11.07	65.32	1.00	1.00	☐

Maximum distance threshold (A) 30 Image quality/performance: faster sharper platformSpeed: 8 7 6 5 4 3 2 1 Console IsoSurface On Off Jmol/JSmol interactive scripting documentation



3.3

Mann-Whitney rank test

Two-sided p-value for 1000 draw of sample with size of 66.

p-value <= 6.8e-11

A small p-value indicates that the observed cross-links distance distribution differs from a random selection distribution. Please select the appropriate random distribution for your cross-linker in the previous page.

6.6. Export dataset statistics to CSV

This action is accessed from the "Dataset" views after selecting the datasets. It generates a CSV file with statistics of cross-link datasets. Note that redundancy must be filtered prior using statistics if required.

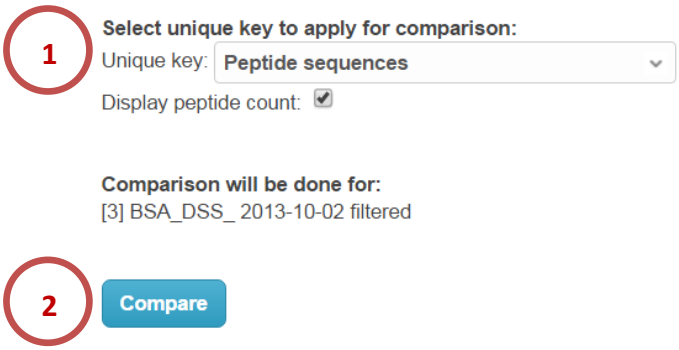
The CSV file contains the following columns:

Column name	Description
Dataset	Name of the dataset
Run	Name of MS file
PSM	Number of peptide-spectrum matches
Inter-protein	Number of inter-protein cross-links
Inter-protein(NoDimerNoDecoy)	Number of inter-protein cross-links without cross-link from the same protein
Intra-protein	Number of intra-protein cross-links
Intra-peptide	Number of intra-peptide cross-links
Dead-end	Number of dead-end peptides
Linear peptide	Number of linear peptides
Min. match score	Minimum peptide score
Average match score	Average peptide score
Max. match score	Maximum peptide score
Autovalidated	Number of auto validated cross-links (Xi only)
Decoy	Number of decoy matches
False positive ratio	Decoy matches divided by total matches
z1-9	Number of peptides per charge state

6.7. Export cross-linked peptides comparison to CSV

This action is accessed from the "Dataset" views after selecting the datasets.

1. The user is first asked to select how the selected datasets are compared. A unique key must be selected for this purpose (peptide sequence, protein cross-linked positions, etc.).
2. Click "Compare" to generate the CSV file.



The CSV file contains the following columns:

Column name	**Description**
Unique Key	Key used for comparison
Dataset or run name	Contains peptide score and peptide counts
run_name	MS run name
scan_number	Scan number
precursor_mz	Precursor m/z
precursor_charge	Precursor charge
Precursor_intensity	Not available
rank	Not available
match_score	Peptide score
spectrum_intensity_coverage	Ratio of MS/MS matched peaks intensity by total intensity
total_fragment_matches	Number of matched fragments
delta	Delta score between the first and second matches
error	Precursor mass error in ppm
peptide1	Peptide sequence
display_protein1	Protein identifier
peptide_position1	Peptide position in the protein
pep1_link_pos	Cross-linked position in the peptide
peptide2	Peptide sequence
display_protein2	Protein identifier
peptide_position2	Peptide position in the protein
pep2_link_pos	Cross-linked position in the peptide
autovalidated	True or false value (Xi only)
validated	Manual validation (Xi only)
rejected	Manual validation (Xi only)
notes	Notes

Note that only a single peptide-spectrum match is displayed per cross-link. A summary of peptide count and peptide score sum is provided at the bottom for each dataset.

6.8. Export as Interaction matrix

This action is accessed from the "Dataset" views after selecting the desired datasets. It generates a CSV file that contains a matrix of cross-links count and the sum of cross-linked peptide scores between proteins.

6.9. Export as Percolator TSV

This action is accessed from the "Dataset" views after selecting the desired datasets. It generates a TSV file to be used with Percolator (see section 5.3 for automated Percolator filtering).

The TSV file contains the following columns:

Column name	**Description**
ClPeptideId	CLMSVault peptide identifier
Label	1 (target) or -1 (decoy)
ScanNr	Scan number
Score	Peptide score
dScore	Delta score between the first and second matches
Charge	Peptide charge
Mass	Peptide mass
PPM	Mass error in ppm
LenShort	Length of the shortest peptide
LenLong	Length of the longest peptide
LenSum	Sum of peptide lengths
Peptide	Cross-linked peptide sequences and positions
ProteinId1	Protein identifier
ProteinId2	Protein identifier

6.10. Export as Xi CSV

This action is accessed from either the “CL Peptides” or the “Dataset” views after selecting the desired objects. It generates a CSV file of cross-linked peptides in the Xi format.

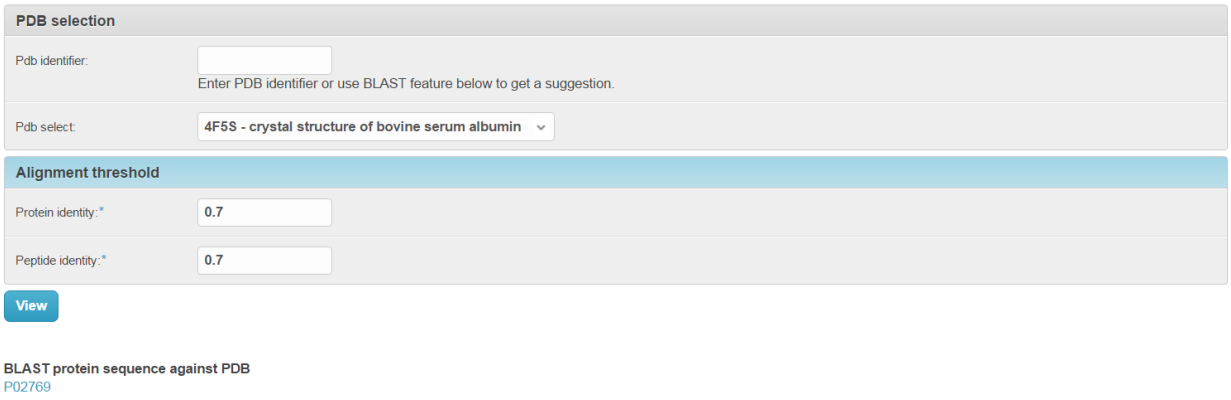
The CSV file contains the following columns:

Column name	Description
run_name	MS run name
scan_number	Scan number
precursor_mz	Precursor m/z
precursor_charge	Precursor charge
Precursor_intensity	Not available
rank	Not available
match_score	Peptide score
spectrum_intensity_coverage	Ratio of MS/MS matched peaks intensity by total intensity
total_fragment_matches	Number of matched fragments
delta	Delta score between the first and second matches
error	Mass error in ppm
peptide1	Peptide sequence
display_protein1	Protein identifier
peptide_position1	Peptide position in the protein
pep1_link_pos	Cross-link position in the peptide
peptide2	Peptide sequence
display_protein2	Protein identifier
peptide_position2	Peptide position in the protein
pep2_link_pos	Cross-link position in the peptide
autovalidated	True or false value (Xi only)
validated	Manual validation (Xi only)
rejected	Manual validation (Xi only)
notes	Not available

6.11. Export as Xi CSV with distance

This action is accessed from either the “CL Peptides” or the “Dataset” views after selecting the desired objects. It generates a CSV file of cross-linked peptides in the Xi format with cross-link inter-residue distance.

1. Select your desired PDB models and set your alignment threshold.
2. Click “View”.



The CSV file contains these additional columns:

Column name	Description
min_distance	Minimum inter-residues distance
distances	All inter-residues distances

6.12. Export as PSI-MI TAB 2.5

This action is accessed from the “Dataset” views after selecting the desired objects. It exports the list of protein interactions in PSI-MI TAB 2.5 format.

6.13. Export as PSI-MI XML 2.5

This action is accessed from the “Dataset” views after selecting the desired objects. It exports the list of protein interactions in PSI-MI XML 2.5 format.

6.14. Export as xTract CSV

This action is accessed from either the “CL Peptides” or the “Dataset” views after selecting the desired objects. It generates a CSV file to be used for label-free quantification with xTract (http://proteomics.ethz.ch/cgi-bin/xtract_cgi/index.cgi). Note that retention times are extracted from peak list for some search algorithms (refer to section 4.3).

6.15. Export as ProteoProfile CSV

This action is accessed from either the “CL Peptides” or the “Dataset” views after selecting the desired objects. It generates a CSV file to be used for label-free quantification with ProteoProfile (https://proteomics.iric.ca/tools/ProteoProfile/).

The CSV file contains the following columns:

Column name	Description
Search Log Num	Not available
FileName	MS run file name
idIpeptide	CLMSVault peptide identifier
UniProt ID	Protein identifiers
UniProt URL	Not available
PIR URL	Not available
EntrezID	Not available
Entrez URL	Not available
Protein Description	Protein description
Species	Species name
Mass	Not available
Num of Peptides	Not available
Peptide QueryNum	Not available
Peptide Sequence	Peptide sequences

Pep Modification	Cross-link type and position
Protein Assignments	Not available
Peptide Start	Peptide start position in protein 1 dot-peptide start position in protein 2
Peptide End	Peptide end position in protein 1 dot-peptide end position in protein 2
Pep Score	Peptide score
Pep Rank	Not available
Pep Observed Mz	Peptide observed m/z
Pep Calc Mass	Not available
Peptide Observed Mass	Peptide Observed Mass
Peptide Charge	Peptide charge
Pep Elution Time	Not available
Pep Sample File	Not available
Peptide URL	Not available
Protein Score	Not available
Prot. PI	Not available
Prot. Seq Length	Not available
Seq Coverage	Not available
PubMedID	Not available
MedLineID	Not available
Scan Number	Scan Number

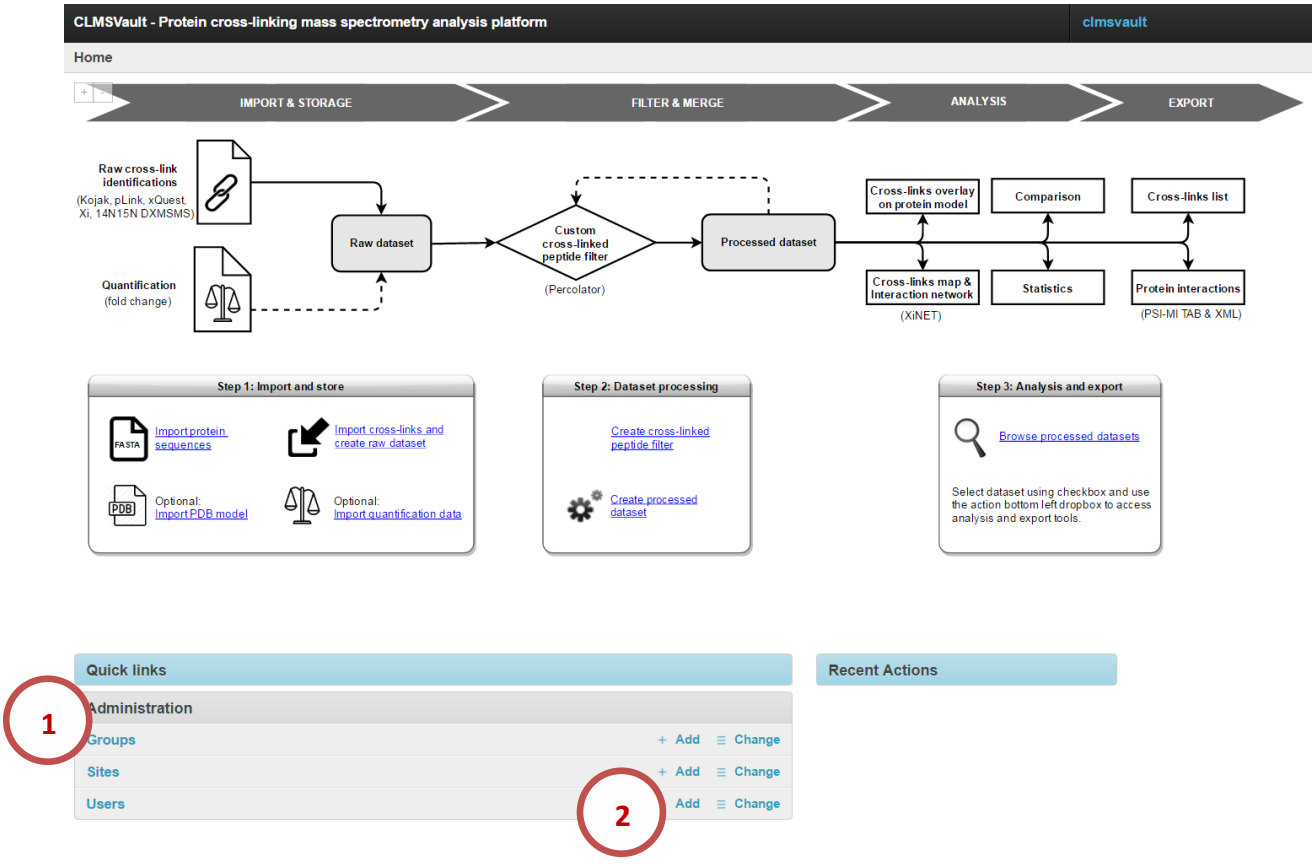
7. Administrative task

7.1. Add new user

To add a new user to CLMSVault, use the following steps as an administrator user:

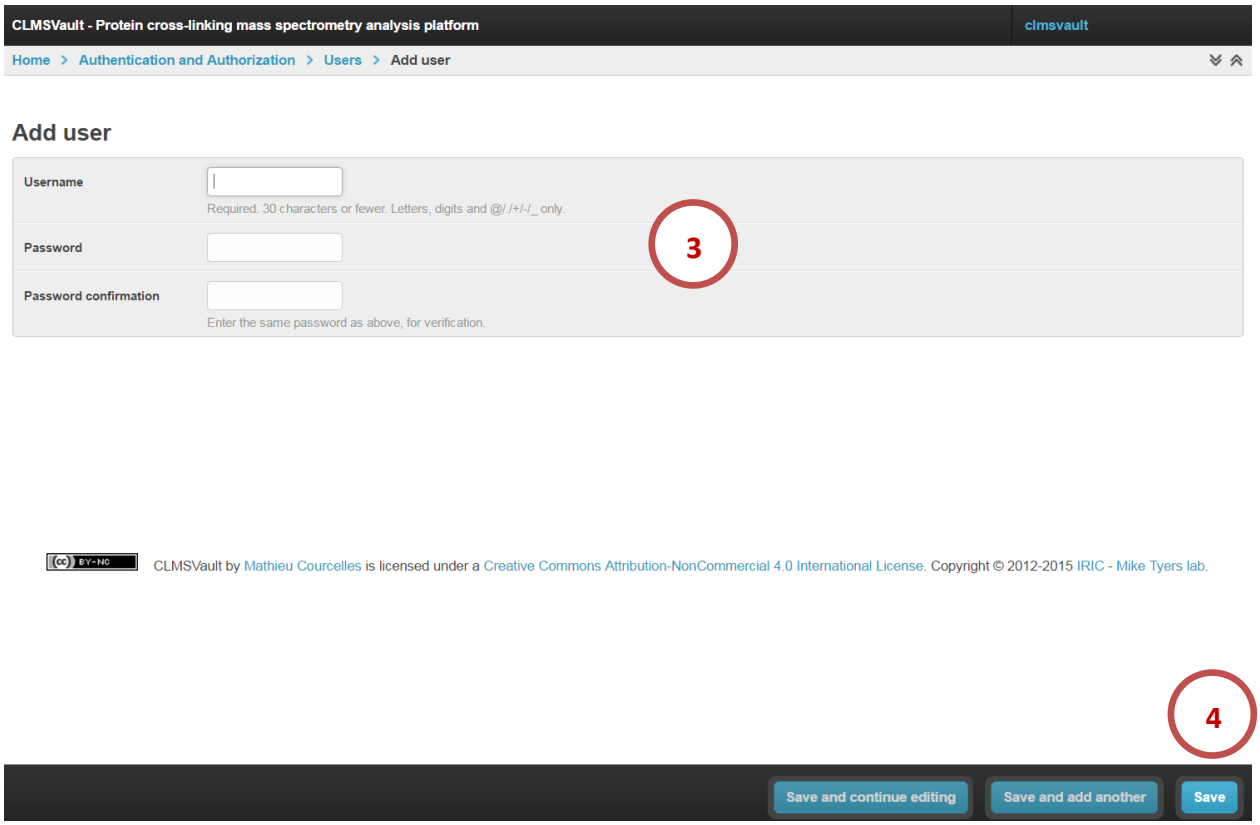
1. From the main screen, click on “Administration” to expand the menu.

2. Click on “Add” in the Users row.



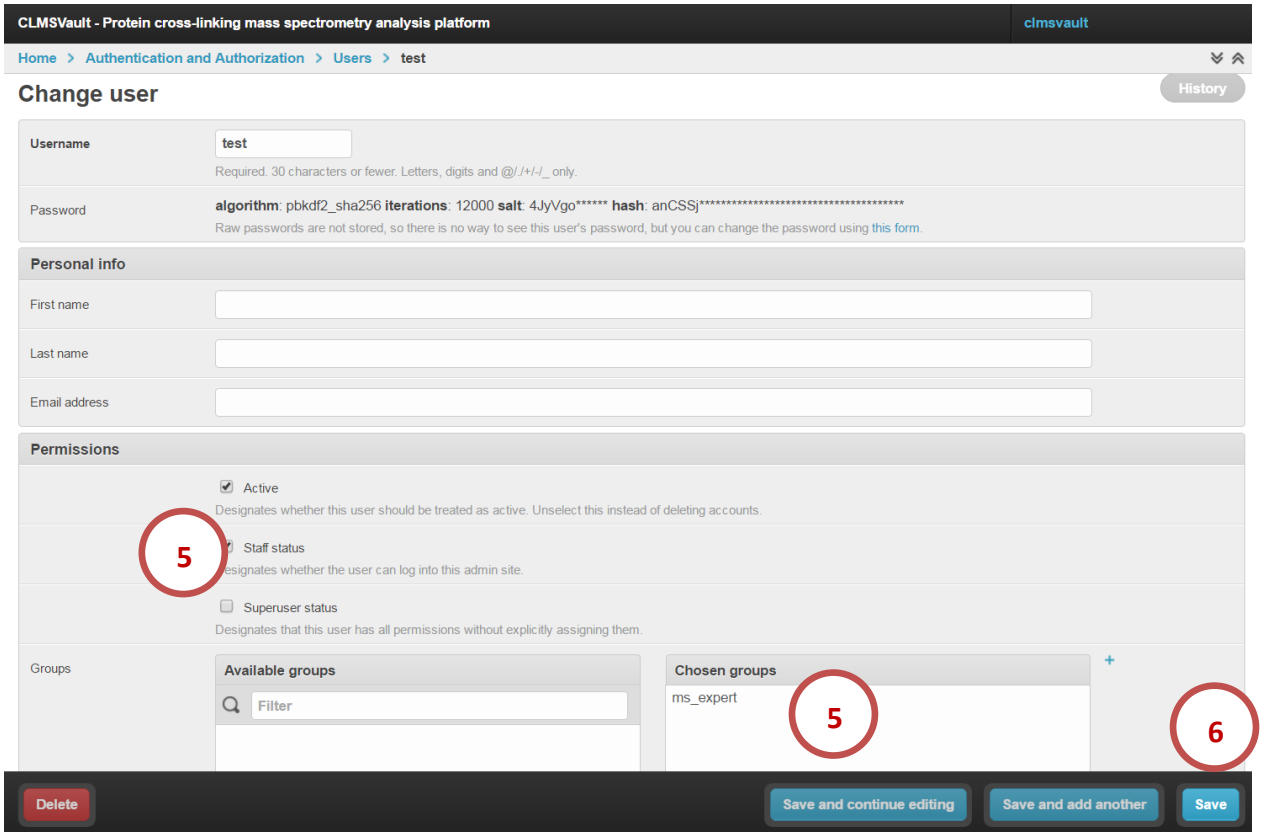
3. Fill the form.

4. Click “Save”.



5. Fill the next form to add further details or permissions to the user. Every user should have "Staff" status activated and be in the "ms_expert" group.

6. Click "Save" to finish.


CLMSVault - Protein cross-linking mass spectrometry analysis platform
clmsvault
Home > Authentication and Authorization > Users > test
Change user
History
Username
test
Required. 30 characters or fewer. Letters, digits and @/./+/-/_ only.
Password
algorithm: pbkdf2_sha256 iterations: 12000 salt: 4JyVgo****** hash: anCSSj**********
Raw passwords are not stored, so there is no way to see this user's password, but you can change the password using this form.
Personal info
First name
Last name
Email address
Permissions
Active
Designates whether this user should be treated as active. Unselect this instead of deleting accounts.
Staff status
Designates whether the user can log into this admin site.
Superuser status
Designates that this user has all permissions without explicitly assigning them.
Groups
Available groups
Filter
Chosen groups
ms_expert
5
5
6
Delete
Save and continue editing
Save and add another
Save