LUMSProT

MATLAB Toolbox

For

Top-down Proteomics

Version 1.0.0.0

User Manual

Abdul Rehman Basharat[2,†], Zohra Bibi[1,†], Rashad Hussain[1], Bibi Amina[1], Humnah Gohar Kabir[1], Anam Shahid[1], Maheen Humayun[1], Amna Riaz[1], Hira Azmat Hayat[3], Muhammad Mustafa[1], Muhammad Ali Shoaib[1], Zakir Ullah[5], Shamshad Zarina[6], Sameer Ahmed[1], Emad Uddin[7], Sadia Hamera[1], Fayyaz Ahmad[4], Safee Ullah Chaudhary[1,*]

BIOMEDICAL INFORMATICS RESEARCH LABORATORY, DEPARTMENT OF BIOLOGY

http://biolabs.lums.edu.pk/BIRL

Table of Contents

Table of figures

Introduction to LUMSProT

This chapter introduces the user to the LUMSProT application and describes its basic features.

About LUMSProT

LUMSProT, a top-down proteomics toolbox based on the popular mathematical computing platform MATLAB, is presented in this work. LUMSProT provides a richly featured environment for searching and identifying proteins from top down proteomics data obtained from high resolution mass spectrometers. The toolbox stands to fill the critical gaps in availability of open source platforms for TDP data analysis as well as the MATLAB® Bioinformatics Toolbox which currently only provides limited spectral data analysis features. The proposed toolbox can be employed by computational proteomics instructors in their educational and training endeavors. LUMSProT comprises of a front-end GUI and the back-end search "engine'.

Features

The salient features of the toolbox are summarized below:

- **Graphical User Interface** - A set of rich and intuitive graphical user interfaces has been developed for setting up the search parameters as well as for integrating the main components of the engine.
- **Whole Protein Molecular Weight Estimation** - The protein identification begins with the tuning of precursor protein's monoisotopic MW (MS1) as guided by its fragmentation spectra (MS2). Relative abundances and mass/charge (m/z) ratios are used to calculate the consensus MW which is then employed in the search and scoring process.
- **Peptide Sequence Tag Extractor** - Peptide sequence tag ladders (PST) are extracted from the spectra by enumerating successive peaks having MW differences equal to an amino acid and within the user specified mass tolerance. Protein database is then filtered for proteins reporting these PSTs. The length of PST ladders, cumulative mass off-sets and relative abundances are used in calculating the PST scores.

- ***In silico* fragmentation** – *In silico* fragments of candidate proteins are generated by the user selected fragmentation techniques. *In vitro* and *in silico* spectral comparisons are performed and scored.
- **Post-translational Modification (PTM) Search** - Support for predicting typical PTMs has been provided in the toolbox. Users can select and search variable and fixed PTMs of their choice by simply selecting them from the GUI.
- **Multifactorial Composite Scoring System** - A multifactorial candidate protein scoring scheme incorporating the aforementioned algorithms has been developed. User customization of the parameters and weights in the scoring function is admitted via a GUI.
- **Single and Batch Processing Mode Search** - Towards an automated batch processing of multiple spectral data files (e.g. peaklists, MGF and mzXML), a batch processing mode has also been implemented. The experimental spectra, search parameters and results are automatically stored in the project directory for further processing and visualization.

Getting Started with LUMSProT Toolbox

LUMSProT source code, manual, samples and issues database is freely available (under the MIT open license) at (https://github.com/BIRL/TDProteomicsToolbox)

To download, click on 'Download ZIP'.

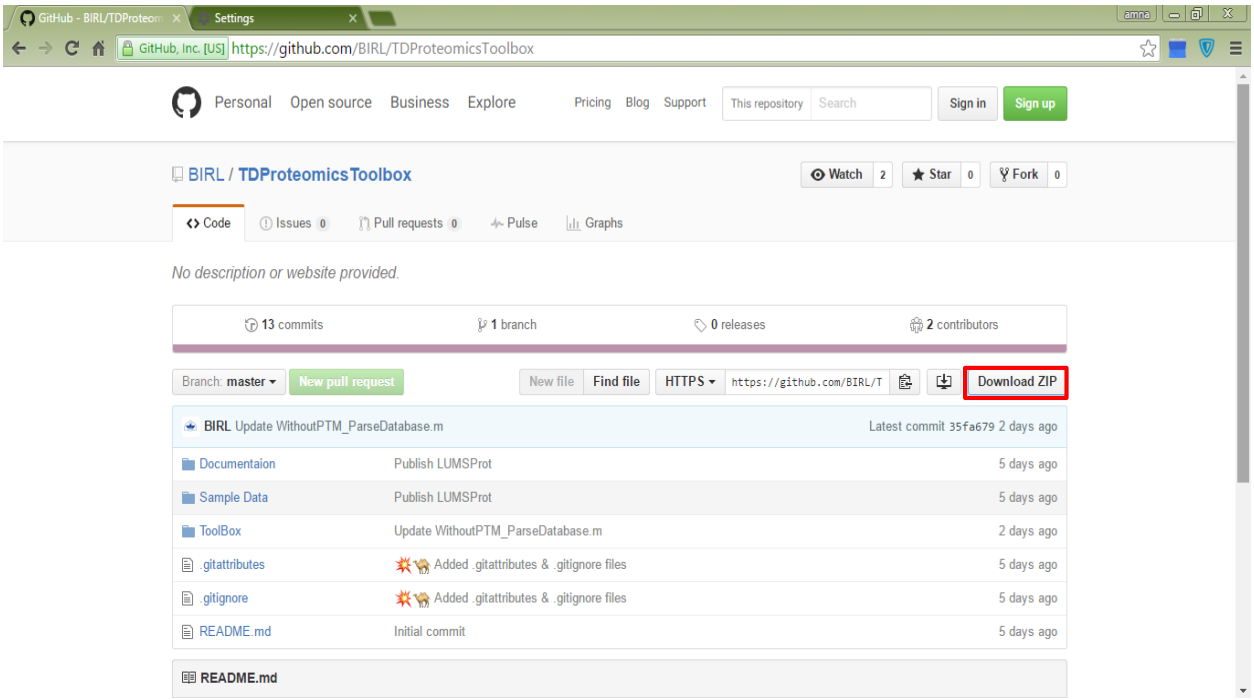


Figure 1. LUMSProT homepage

Once downloaded, unzip/extract the files from the folder.

Figure 2. Downloading LUMSProT toolbox.

GUI description:

This chapter presents the interface overview for user facilitation.

Customizing parameters:

WINDOW 1: LUMSProT MATLAB Toolbox for Top-down Proteomics

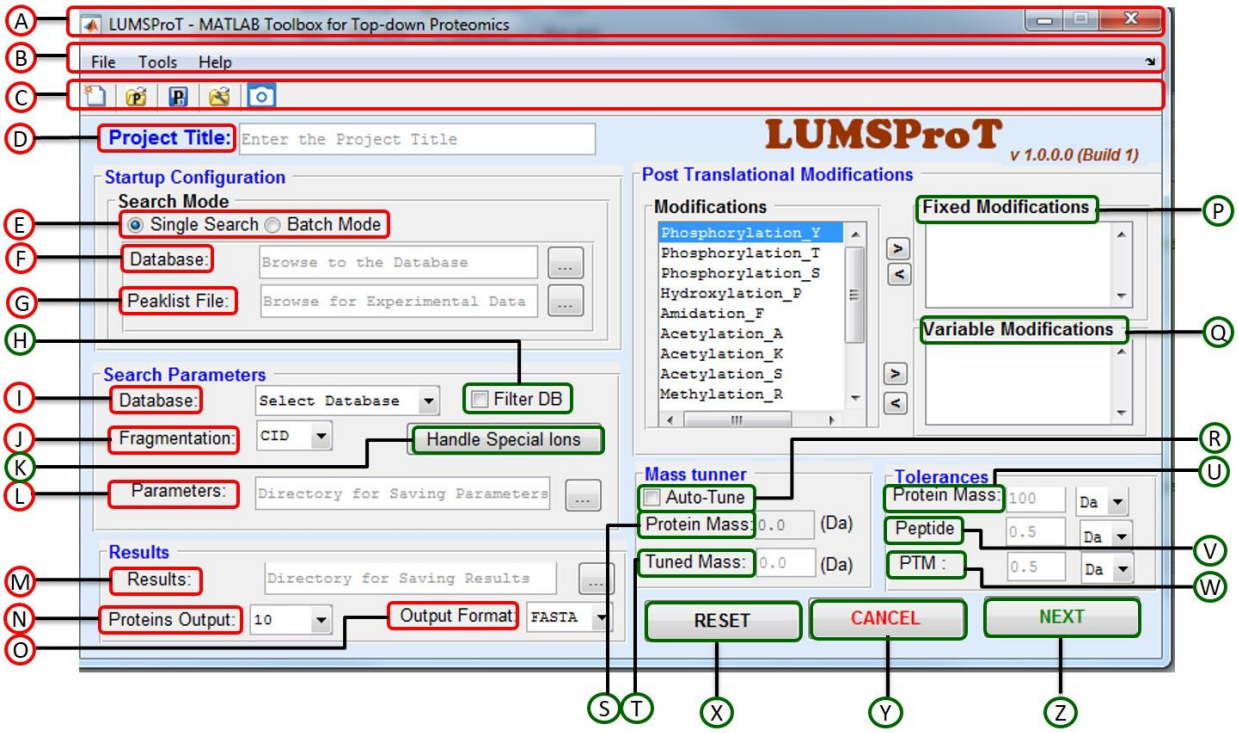


Figure 3. LUMSProT window 1- Overview of User Interface

A. At the top of the LUMSProT interface is the 'Title bar', which displays the name of the program and the title of the current document.

B. Just below the title bar, there is a 'Menu bar' which lists the heading for each drop-down menu. It allows the user to perform all basic functions (reset, load, save or browse the default/saved parameters). Commands are grouped under each of these menu headings according to function.

C. Under the Menu bar, 'Toolbar' contains shortcuts to some of the most frequently used commands from the menu bar.

D. In order to start a project, user has to enter the 'Project title' first (for My project)

E. Select the search mode according to your requirement (i.e. single search or batch mode). Batch mode option will lead to drop down menu of 'file type'.

F. Browse and select the protein database
G. Browse and upload experimental data (Peaklist file for Single mode; .mzxml/ .mgf/.txt file for Batch mode)
H. Select Database from the drop down menu
I. User can filter database by checking the option -'Filter DB'
J. Select the 'Fragmentation technique' from drop down menu.
K. 'Handle Special Ions' option will lead to a new window (Fig. 6) where user can select the fragmentation ions option (i.e. a', b', y', z", a*, b*,y*, z' ions)
L. User can select a file to store the selected parameters
M. User can select a file to store the results
N. Set the protein output (for example 10, 20, 30 etc.)
O. Select the 'Output Format' from the drop down menu
P. User can opt for required fixed 'Post translation Modifications' from the list of modifications
Q. Similarly, various 'Variable Modifications' are also selected from the list
R. Check the option 'Auto-tune' for tuning protein mass and proceed to the next step
S. After uploading the MS input file, user can see the 'Protein mass' in the box
T. Tuned mass can be seen when user proceed to the next step
U. Set the tolerance value for protein mass (Unit can be changed by clicking on the drop down arrow)
V. Set the tolerance value for Peptide
W. Set the tolerance value for Post Translational Modification (PTM)
X. User can clear the already uploaded data by clicking on 'Reset' option
Y. Click on ' Cancel' option to close the window
Z. After filing all the requirements, user can proceed further by clicking on 'Next' option below.

In silico fragmentation

Window 2: Select Fragmentation Ions

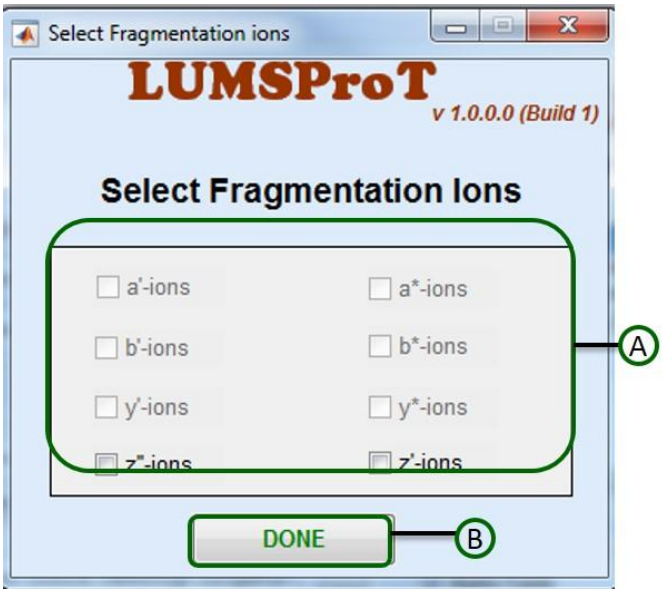


Figure 4. Selecting fragmentation ions

A. 'Handle Special Ions' in the previous window, option leads to a new window where user can select the fragmentation ions option (i.e. a', b', y', z", a*, b*,y*, z' ions)

B. Click on 'Done' to record the selection

Mass tuning Protein

Window 3: Intact Protein Mass Tuner

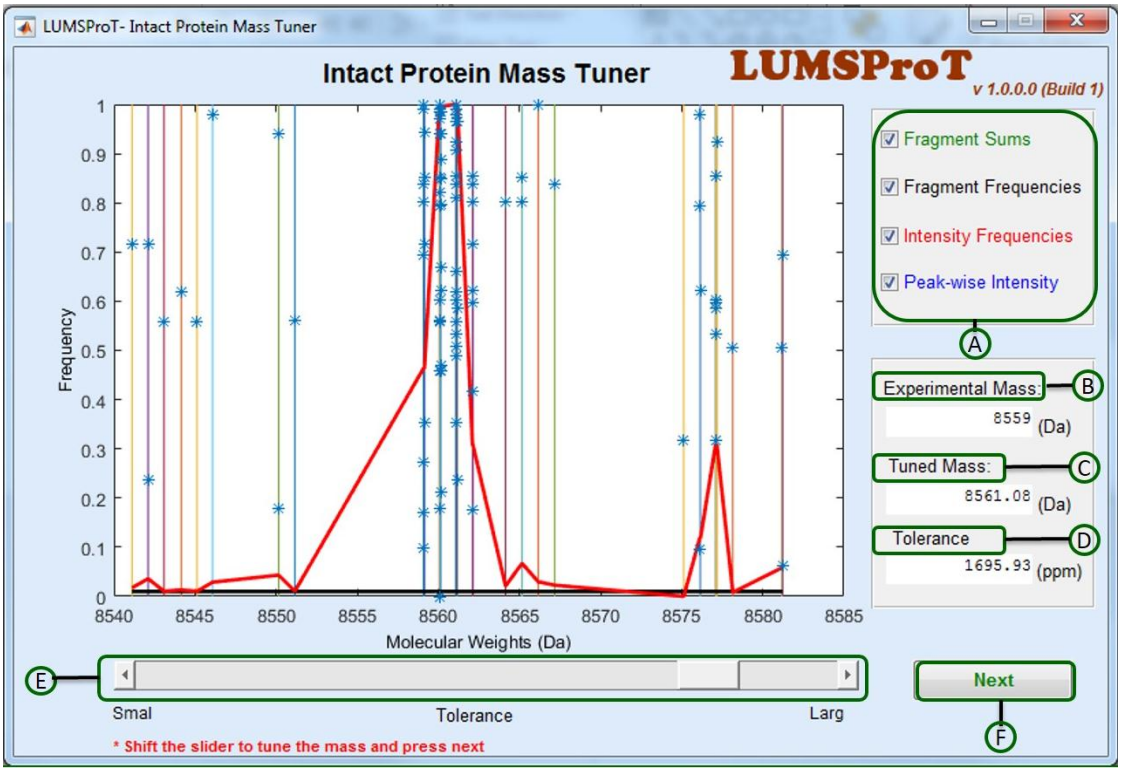


Figure 5. Mass tuning a protein

A. Graph between 'Frequency' on Y-axis and 'Molecular weight' on X-axis is shown when a used selected 'Auto-tune' option from the previous window. Graph represents Fragment Sums, Fragment Frequencies, Intensity Frequencies and Peak-wise Intensity displayed by green, black, red and blue colors respectively. User can select from the list of Attributes to be represented in the graph.

B. 'Experimental mass' shows the mass of respective protein

C. 'Tuned mass' shows represents more accurate and precise protein mass

D. 'Tolerance' shows the value you set by shifting the slider

E. User can tune the mass by shifting the slider below, to left or right and press next

F. To proceed further, press 'Next'

Extracting Peptide Sequence Tag (PSTs)

Window 4: Peptide Sequence Tags (PSTs)-LUMSProT

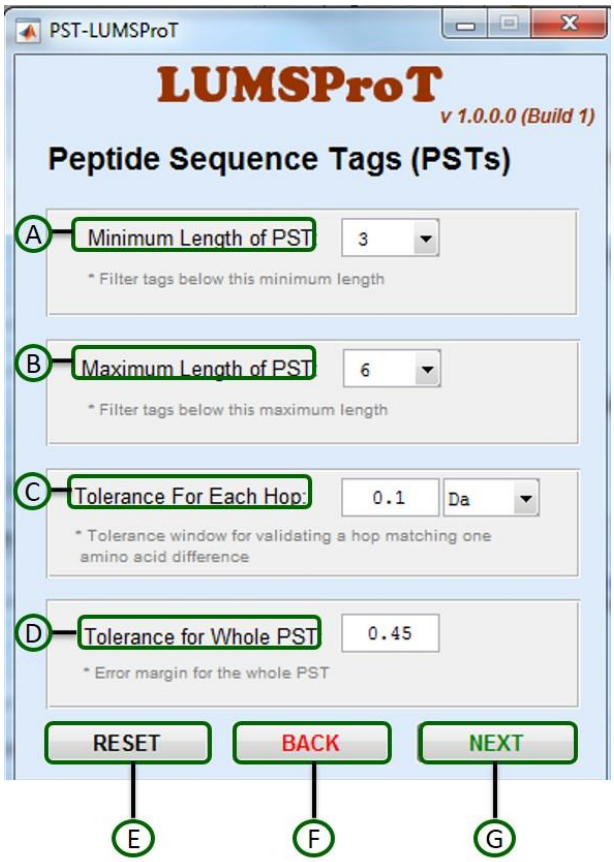


Figure 6. Filtering results reporting selected Protein Sequence Tag (PST)

A. Tags will be filtered below the minimum length of PST selected from the drop down menu by the user

B. Tags will be filtered below the maximum length of PST selected from the drop down menu by the user

C. Set the 'Tolerance for each Hop'

D. 'Tolerance for Whole PST' shows error margin for the whole PST

E. Click 'Reset' to clear the already selected data

F. Click on 'Cancel' option to close the window

G. Click 'Next' to proceed to the next window

Window 5: LUMSProT-Components Score

Customization of Parameters and weights

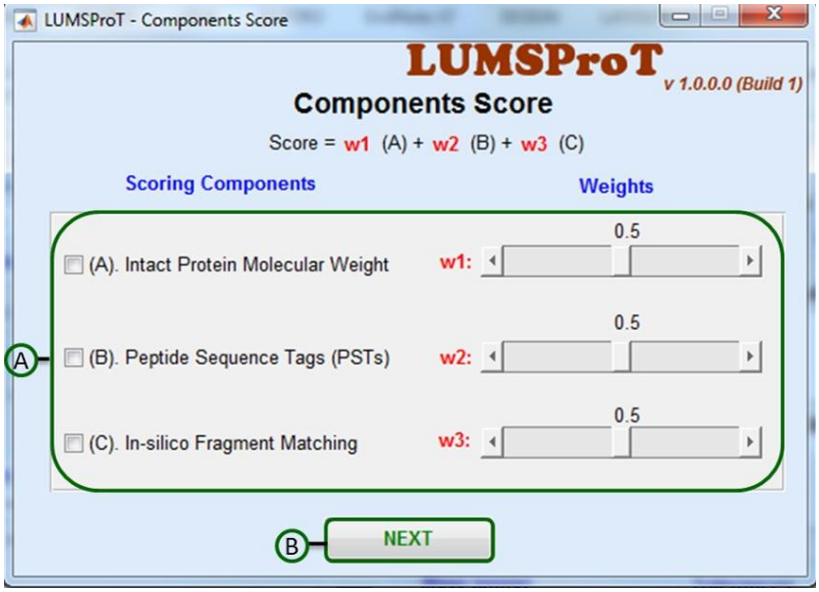


Figure 7. Customizing Scoring Components and balancing their respective weights

A. Check the Scoring Components from the list and set their respective weights by shifting the slider left or right accordingly

B. Click 'Next' to proceed

Window 6: LUMSProT- Candidate proteins

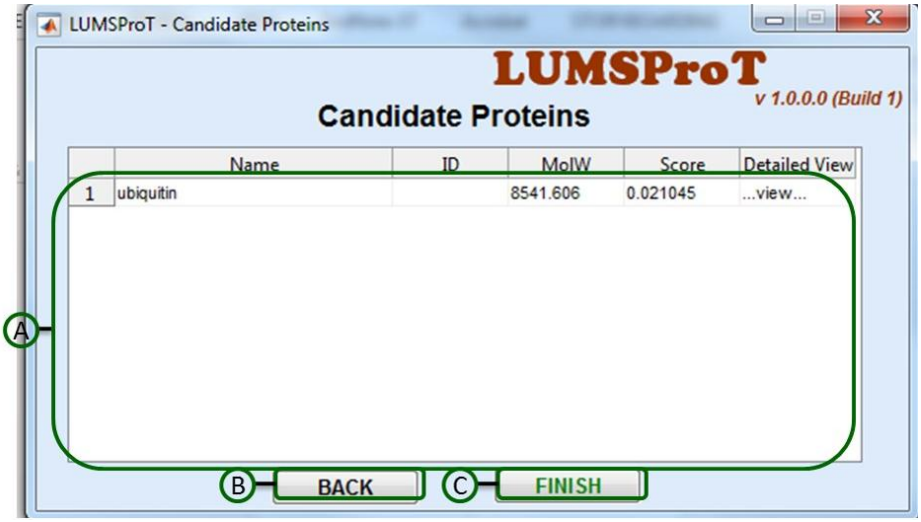


Figure 8. Result window showing candidate proteins

A. Name of the resultant protein along with molecular weight, score and detailed view according to the uploaded and selected data is represented under the list of 'Candidate Proteins'.

B. Click 'Back' to revert to the previous window

C. Click 'Finish' to end the process

Details of Candidate proteins:

Window 7: LUMSProT-Detailed Protein View

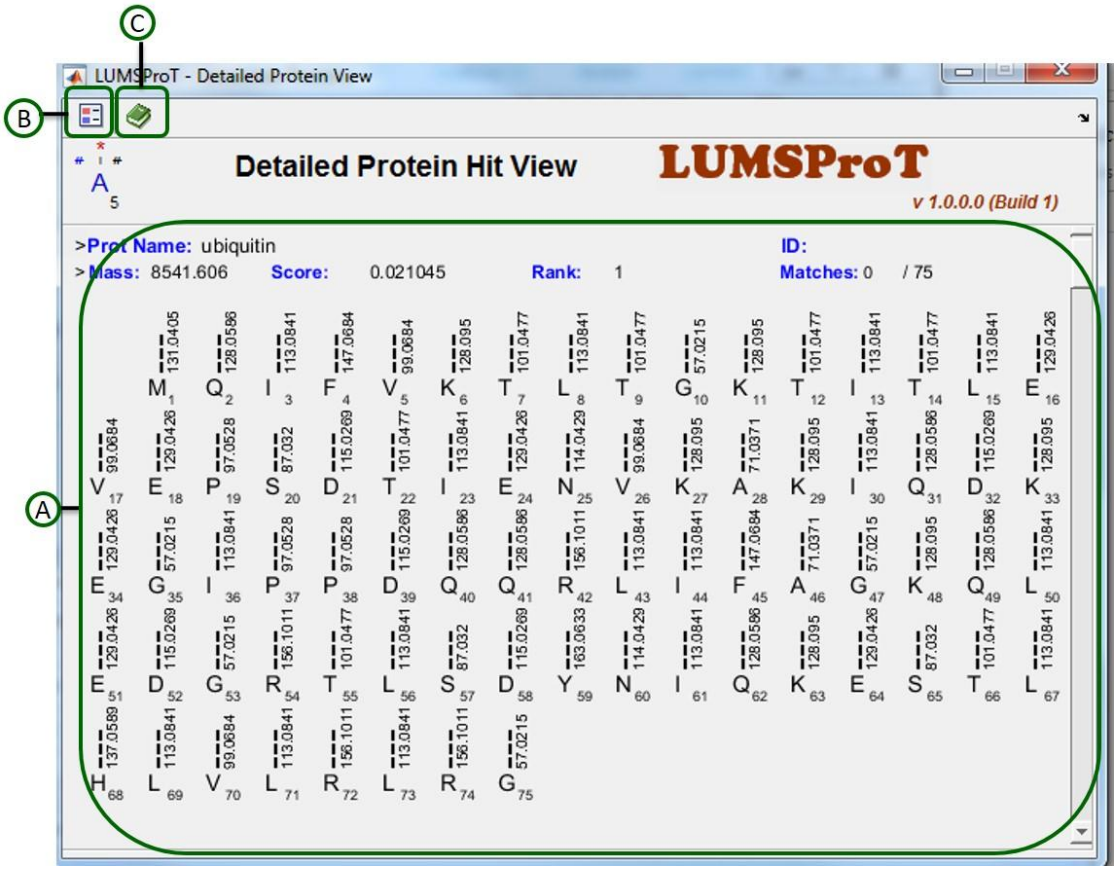
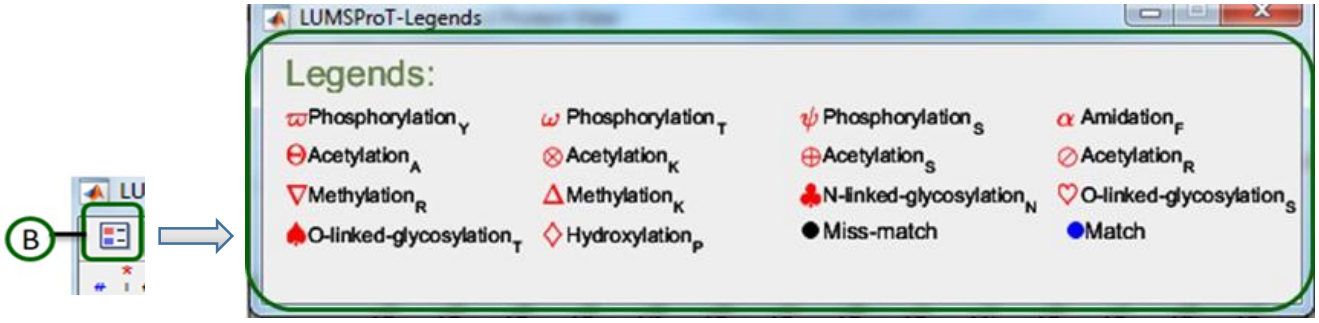


Figure 9. Detailed view of resultant protein

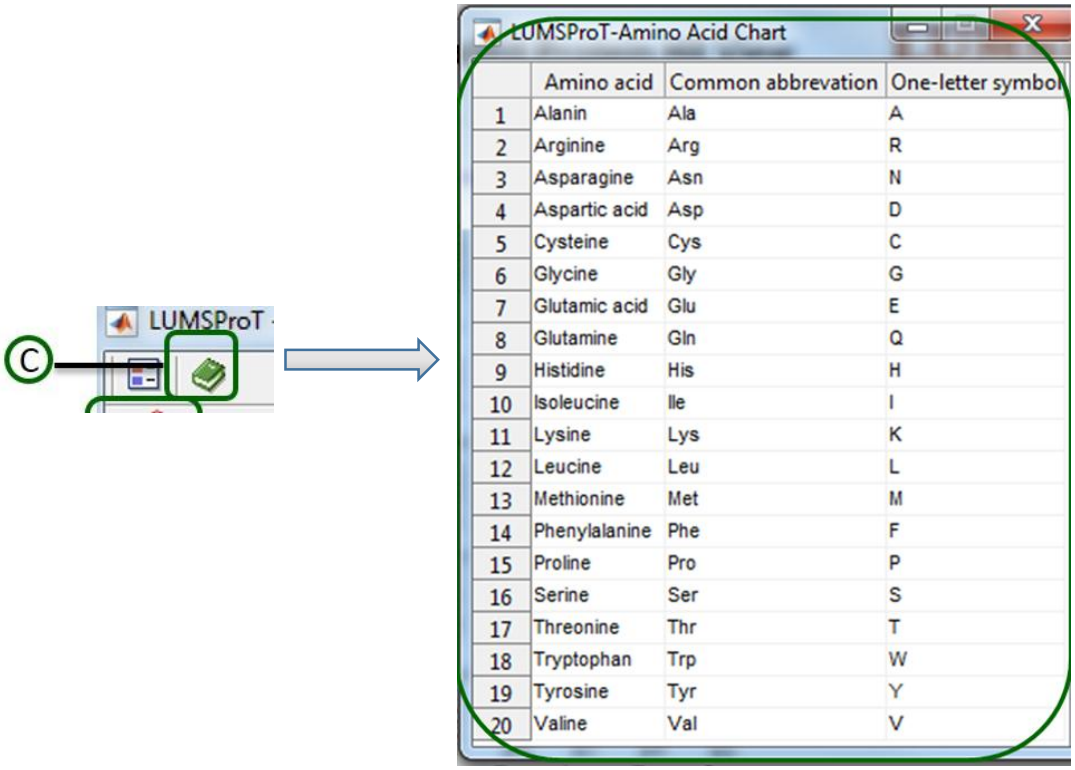


A. 'Detailed Protein View' window shows proteins name, mass, score, rank, matches and amino acids (along the mass) present in the protein.

Figure 10. Legends list (might be represented within peptide sequence)

B. Open the window of legends (i.e. LUMSProT-Legends), which might be present in peptide sequence of candidate proteins.

C. Leads to a window (i.e. LUMSProT-Amino acid Chart) containing list of full form of Amino acids along with abbreviations and one-letter symbol of protein to facilitate the user.



	Amino acid	Common abbrevation	One-letter symbol
1	Alanin	Ala	A
2	Arginine	Arg	R
3	Asparagine	Asn	N
4	Aspartic acid	Asp	D
5	Cysteine	Cys	C
6	Glycine	Gly	G
7	Glutamic acid	Glu	E
8	Glutamine	Gln	Q
9	Histidine	His	H
10	Isoleucine	Ile	I
11	Lysine	Lys	K
12	Leucine	Leu	L
13	Methionine	Met	M
14	Phenylalanine	Phe	F
15	Proline	Pro	P
16	Serine	Ser	S
17	Threonine	Thr	T
18	Tryptophan	Trp	W
19	Tyrosine	Tyr	Y
20	Valine	Val	V

Figure 11. Amino acid chart- Full form and abbreviation of amino acids

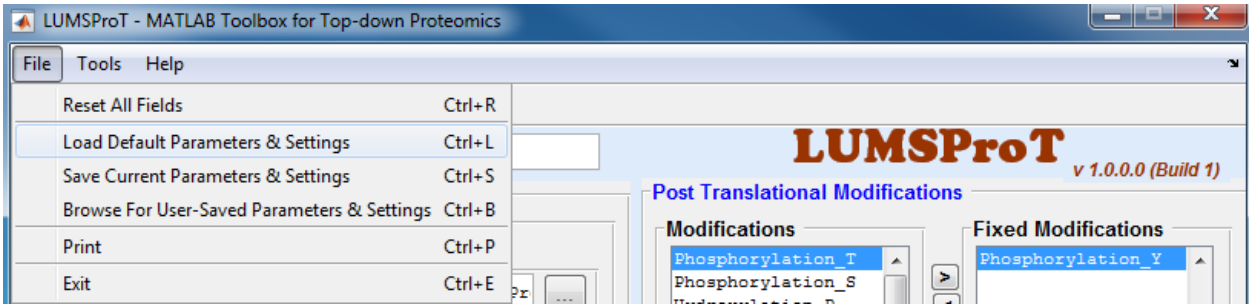
Search

Parameters:

LUMSProT works when all parameters are set. Two kind of parameters can be used by the user which includes (i) Default Parameters and (ii) Passed/selected parameters.

How to load default Parameters?

To submit the job using default parameters, go to 'File' tab on the menu bar. Once done, user will be left with different commands. Select 'Load Default Parameters & Settings'.



One shortcut is to simply press the second icon on the toolbar.

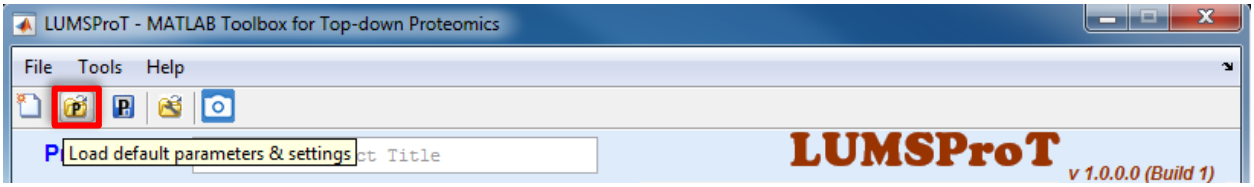


Figure 12. Loading default parameters through (a) menu bar and (b) shortcut toolbar

How to save passed/selected Parameters?

Parameters can be passed according to user requirement and can be reused later. For this purpose, user has to save the selected parameters, click on the 'File' menu and select 'Save Current Parameters & Settings' or simply press the highlighted icon below.

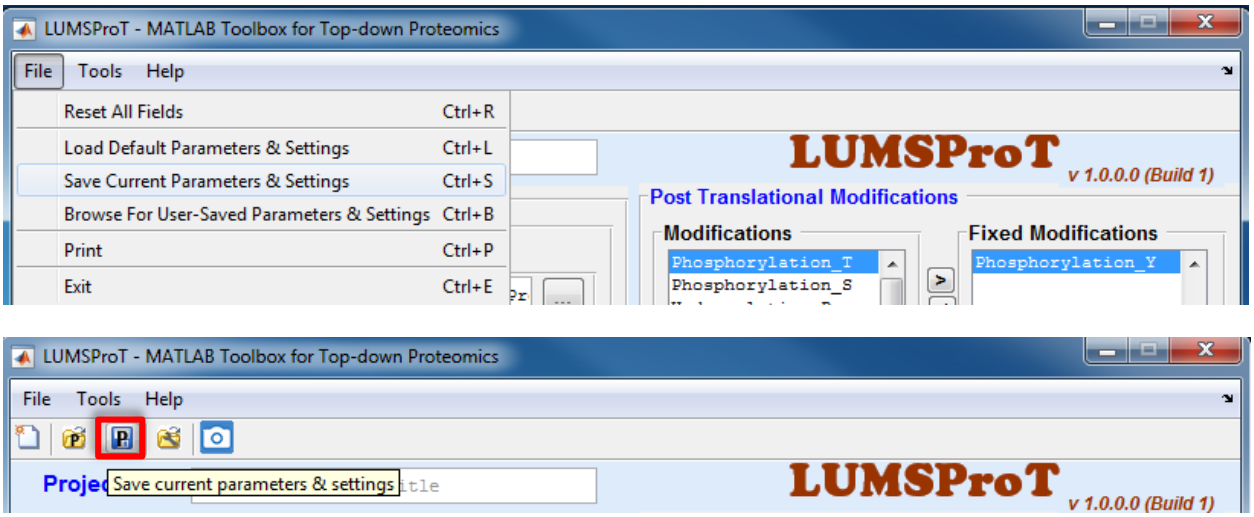


Figure 13. Saving user selected parameters through (a) menu bar and (b) shortcut toolbar

How to load saved parameters?

To browse for the above saved parameters, select the option 'Browse for User-Saved Parameters & Settings' or click on forth icon highlighted below.

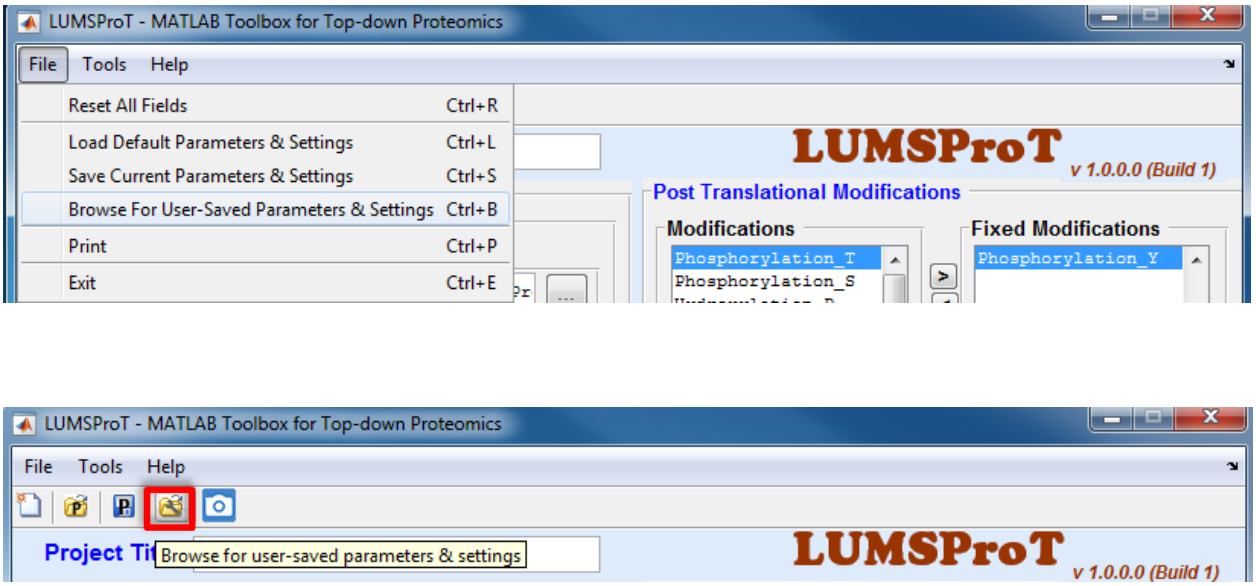


Figure 14. Loading user-saved parameters through (a) menu bar and (b) shortcut toolbar

Databases

SwissProt database is included in LUMSProT download package by default. User can take any protein sequence from other databases (such as Uniprot) in fasta format.

Modes

User can select 'Search mode' that best fits the need. LUMSProT works with two search modes.

(i) Single Search Mode

(ii) Batch Mode

Databases and Peaklist files are uploaded according to the selected mode. Single search mode demands experimental data in .txt file whereas Batch mode supports .plk, .mzxml, .mgf files.

Batch mode takes more processing time as it deals with larger data.

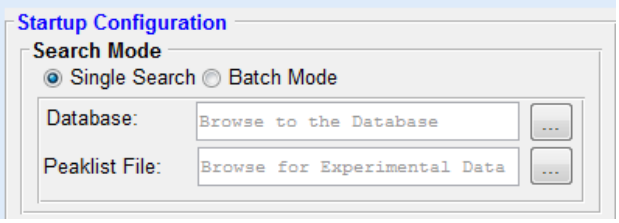


Figure 15. Selecting Search mode

The experimental spectra, search parameters and results are automatically stored in the project directory for further processing and visualization.