Utilities for Mass Spectrometry Analysis of Proteins

User's Manual

Version 2.2.0

May 2022

To download Utilities for Mass Spectrometry Analysis of Proteins visit:
www.umsap.nl

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Utilities for Mass Spectrometry Analysis of Proteins (UMSAP) is a graphical user interface (GUI) designed to speed up the post-processing of data obtained during mass spectrometry studies involving proteins. The program is not intended to analyze a mass spectrum or a mass chromatogram, neither to identify the peaks in a mass spectrum. The main objective is the fast post-processing of the vast amount of data generated in mass spectrometry experiments involving proteins after peak identification have been performed.

The program is organized in modules with each module performing a single type of data post-processing. The reason for this clear separation is the high dependency between the type of mass spectrometry experiment performed and the way in which the resulting data must be post-processed. The modules are designed in such a way that the required user input is minimized but still users can control every aspect of the analysis. Currently, the software contains three modules, but several others are already planned.

1.1 Citing Utilities for Mass Spectrometry Analysis of Proteins

If results obtained with UMSAP are published in any way, please acknowledge the use of UMSAP by including the following sentence:

"Utilities for Mass Spectrometry Analysis of Proteins was created by Kenny Bravo Rodriguez at the University of Duisburg-Essen and is currently developed at the Max Planck Institute of Molecular Physiology."

Any published work, which uses UMSAP, should include the following reference:

Kenny Bravo-Rodriguez, Birte Hagemeier, Lea Drescher, Marian Lorenz, Michael Meltzer, Farnusch Kaschani, Markus Kaiser and Michael Ehrmann. (2018). Utilities for Mass Spectrometry Analysis of Proteins (UMSAP): Fast post-processing of mass spectrometry data. Rapid Communications in Mass Spectrometry, 32(19), 1659–1667.

Electronic documents should include a direct link to the official web page of UMSAP at: www.umsap.nl

1.2 Acknowledgments

I would like to thank all the persons that have contributed to the development of UMSAP, either by contributing ideas and suggestions or by testing the code. Special thanks go to: Dr. Farnusch Kaschani, Dr. Juliana Rey, Dr. Petra Janning and Prof. Dr. Daniel Hoffmann.

In particular, I would like to thank Prof. Dr. Michael Ehrmann.

Chapter 2

Obtaining and Installing Utilities for Mass Spectrometry Analysis of Proteins

2.1 Obtaining Utilities for Mass Spectrometry Analysis of Proteins

UMSAP is distributed free of charge for anyone interested in using it. To obtain a copy of the software just register at www.umsap.nl and go to the Download page.

No extra software or packages are needed for UMSAP to properly work. So far, UMSAP have been tested in macOS 10.14.6 and 12.3 and Windows 10. Support for some Linux distributions will be available in the future.

2.2 Installing Utilities for Mass Spectrometry Analysis of Proteins

Windows

Unzip the file you just downloaded from www.umsap.nl. Then, copy the folder UMSAP to the location in your file system where you want to keep it. Finally, create a shortcut to the executable file UMSAP.exe found inside the main folder UMSAP. That is all. You are now ready to use UMSAP.

macOS

Unzip the file you just downloaded from www.umsap.nl. Then, just move the UMSAP.app folder to /Applications/. That is all. You are now ready to use UMSAP.

Depending on the security settings in macOS, it may be needed to explicitly allow UMSAP to be opened the first time the app is used.

2.3 Uninstalling Utilities for Mass Spectrometry Analysis of Proteins

UMSAP will not create any installation file in your computer. Therefore, the only thing you need to do, to completely uninstall UMSAP, is to delete the folder UMSAP.app in macOS or UMSAP in Windows. You should also delete any shortcut pointing to the executable file of UMSAP and the configuration file .umsap_config.json in your home folder. That is all.

Chapter 3

Workflow in Utilities for Mass Spectrometry Analysis of Proteins

When you start UMSAP, the program will display the main window (Figure 3.1). From this window you can access all the modules and utilities either by the menu entries: Modules and Utilities or by the corresponding buttons on the right side list. A complete description of each module and utility is given in the following chapters.

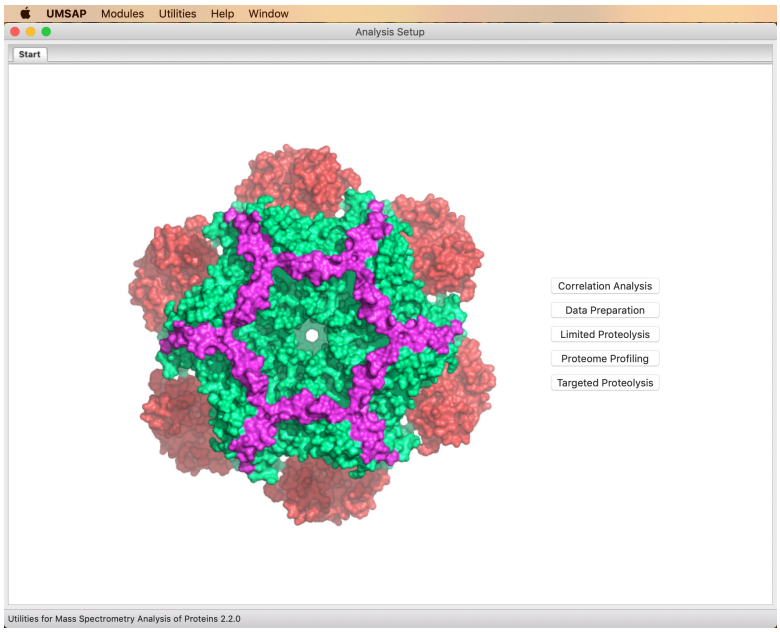

Figure 3.1: The main window of UMSAP. From this window users can access all the available Modules and Utilities.

3.1 The input files

UMSAP has two main input files. One file contains the detected peptide sequences after all peak assignments have been completed, and the other file contains the detected proteins. The program expects these files to be plain text files containing a table with the data. Columns in the files are expected to be tab separated. The first row in the files is expected to contain only the names of the columns. There is no limit in the amount and type of data present in the Data files. However, each module will expect certain columns to be present. Columns not needed by the modules will simply be ignored.

In addition, certain modules use other input files as well. The modules Targeted Proteolysis and Limited Proteolysis use fasta files containing the sequences of the recombinant and native proteins used in the experiments. The first sequence found in the fasta file is assumed to be the sequence of the recombinant protein. The second sequence found in the fasta file is assumed to be the sequence of the native protein. All other sequences found in the fasta file are discarded.

The Targeted Proteolysis module may also use a local PDB file.

3.2 The output files

Results generated by UMSAP will be saved in two folders and a file with extension .umsap (Figure 3.2). Direct manipulation of the umsap file and files within these folders should be avoided. UMSAP provides a way to manage them through the UMSAP Ctrl window (**??**). Nevertheless, all the files created by UMSAP are plain text files with json or cvs (tab separated) format, in order for users to be able to read their content. Changing the content of the files is highly discouraged as this will lead to errors in the reliability and visualization of the results with UMSAP.

The folder Input_Data_Files contains a copy of the input files used for the analysis in the project. When adding a new analysis to the project, the new input files used will be copied to the Input_Data_Files folder. The date and time of the analysis will be added to the name of the file to avoid overwriting existing files inside the folder.

The folder Steps_Data_Files contains a folder for each analysis in the project. These folders contain the main results for the analysis as well as a step by step account of the calculations and any further analysis performed after the main results were created.

The .umsap file contains information about all the analysis in the project and allows managing the project and the visualization of the results.

3.3 Using Utilities for Mass Spectrometry Analysis of Proteins

Once the input files are ready to be analyzed, using UMSAP is straightforward. Just open the program and select a module or utility. In the new tab, fill in the needed

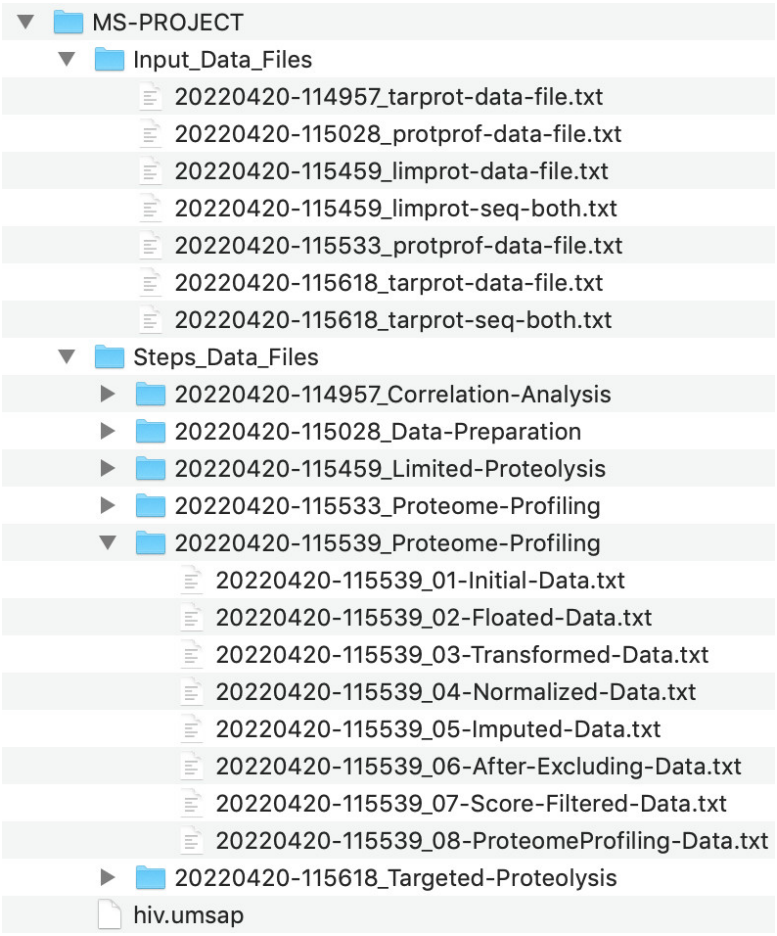

Figure 3.2: Structure of the output generated by UMSAP. Results are saved in the Steps_Data_Files folder. The .umsap file allows managing and visualizing the results.

information and hit the Start Analysis button at the bottom of the tab. Depending on the amount of data and the complexity of the analysis to perform it may take a few minutes for the program to complete the task at hand. While the analysis is running, a window, containing a progress bar, will appear. This window will give a rough guess of the remaining time needed to complete the current analysis and will report any error encountered. It will be helpful if users send a crash report to umsap@umsap.nl, so we can correct them.

In order to make the program as user-friendly as possible help messages will pop up from buttons and labels. The help messages will contain a brief description of what is the button or label for and what input is expected from the user. In this way, users can find basic information about a particular element of the interface without needing to go to the manual or online tutorials. If more information is needed, users may consult the manual or click the Help button at the bottom of the module/utility tab to read an online tutorial.

Depending on the module or utility just run, new windows will be created to show a graphical representation of the results.

3.4 Navigating through Utilities for Mass Spectrometry Analysis of Proteins

The entries Modules and Utilities will be available in the menu of every window. The Modules entry in the menu gives direct access to all modules. The same is true for the Utilities entry. These menu entries are the fastest way to access all the functions in UMSAP. In a typical UMSAP session, users will work with different independent windows simultaneously. The windows have descriptive names, so users can quickly guess the content of any window. The scheme of the windows name is *File Name - Utilities or Module Name - ID of the Analysis*. For example, the window with name *hiv.umpap - Target Proteolysis - 20220420-115618 - Cleavage Sites* will be displaying the Targeted Proteolysis analysis with ID 20220420-115618 - Cleavage Sites from file hiv.umsap.

A list of current shortcuts is given in Table 3.1.

3.5 Backward compatibility

Unfortunately, UMSAP 2.2.0 is not capable to read any file generated with previous versions of UMSAP.

Shortcut	Action	Window
Alt+Cmd+L	Create the Limited Proteolysis tab	All
Alt+Cmd+P	Create the Proteome Profiling tab	All
Alt+Cmd+T	Create the Targeted Proteolysis tab	All
Cmd+R	Read umsap file	All
Cmd+C	Copy	Text and List boxes
Cmd+X	Cut	Text and List boxes
Cmd+P	Paste	Text and List boxes
Cmd+A	Select all	Text and List boxes
Cmd+P	Show Data Preparation results	Results plot
Cmd+D	Duplicate result window	Results plot
Cmd+E	Export data	Results plot
Cmd+I	Export image	Results plot
Cmd+K	Clear all selections	Results plot
Cmd+A	Add analysis	UMSAP Ctrl
Cmd+X	Delete analysis	UMSAP Ctrl
Cmd+E	Export analysis	UMSAP Ctrl
Cmd+U	Reload file	UMSAP Ctrl
Cmd+Z	Reset the zoom on a plot	Selected plot
Alt+Shift+I	Export all images	Multiple plots
Alt+Shift+Z	Reset all zooms	Multiple plots
Shift+I	Export main plot image	Multiple plots
Shift+Z	Reset main plot zoom	Multiple plots
Alt+I	Export secondary plot image	Multiple plots
Alt+Z	Reset secondary plot zoom	Multiple plots
Cmd+A	Show all peptides	Limited Proteolysis
Cmd+L	Toggle Band/Lane selection mode	Limited Proteolysis
Cmd+S	Export sequence alignments	Limited Proteolysis
Shift+A	Add label to Volcano plot	Proteome Profiling
Shift+P	Toggle Pick label / Select protein	Proteome Profiling
Shift+Cmd+A	Apply all Filters	Proteome Profiling
Shift+Cmd+F	Auto apply all Filters	Proteome Profiling
Shift+Cmd+R	Remove selected Filters	Proteome Profiling
Shift+Cmd+Z	Remove last applied Filter	Proteome Profiling
Shift+Cmd+X	Remove all Filters	Proteome Profiling
Shift+Cmd+C	Copy Filters	Proteome Profiling
Shift+Cmd+P	Paste Filters	Proteome Profiling
Shift+Cmd+S	Save Filters	Proteome Profiling
Shift+Cmd+L	Load Filters	Proteome Profiling
Shift+Cmd+E	Export filtered data	Proteome Profiling
Cmd+S	Export sequence alignments	Targeted Proteolysis

Table 3.1: List of built-in keyboard shortcuts.