

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

Utilities for Mass Spectrometry Analysis of Proteins
Copyright © 2017 Kenny Bravo Rodriguez.
All Rights Reserved.

Contents

List of Figures	III
List of Tables	IV
1 Introduction	1
1.1 Citing Utilities for Mass Spectrometry Analysis of Proteins	1
1.2 Acknowledgments	2
2 Obtaining and Installing Utilities for Mass Spectrometry Analysis of Proteins	3
2.1 Obtaining Utilities for Mass Spectrometry Analysis of Proteins	3
2.2 Installing Utilities for Mass Spectrometry Analysis of Proteins	3
2.3 Uninstalling Utilities for Mass Spectrometry Analysis of Proteins	4
3 Workflow in Utilities for Mass Spectrometry Analysis of Proteins	5
3.1 The input files	6
3.2 Using Utilities for Mass Spectrometry Analysis of Proteins	6
3.3 Navigating through Utilities for Mass Spectrometry Analysis of Proteins	7
3.3.1 Keyboard shortcuts	8
3.4 The output files	8
3.5 Backward compatibility	8

List of Figures

3.1	The main window of UMSAP	5
-----	------------------------------------	---

List of Tables

3.1	List of built-in keyboard shortcuts	8
-----	---	---

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 in the right side list. A complete description of each module and utility is given in the following chapters.

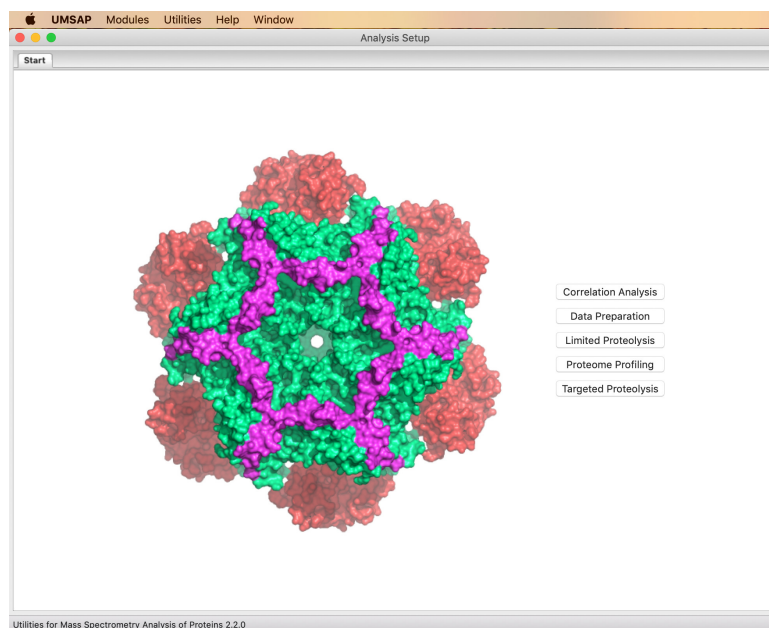


Figure 3.1: The main window of UMSAP. From this window users can access all the available Modules and Utilities.

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 limitation 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 may use fasta files containing the sequences of the recombinant and native proteins used in the experiments. The fasta files must contain only one protein sequence. Alternatively, instead of a fasta file, a plain text file can be used. In this case, the plain text file must contain only the sequence of a single protein using the one letter code for amino acids. The Targeted Proteolysis module may also use a pdb file. The sequence and pdb files can be directly downloaded from UNIPROT or PDB, respectively, by supplying the correct codes.

Certain output files generated by UMSAP can be used as input files as well. Files that can be used this way will contain the results from previous data post-processing. It is important that these files remain unchanged to ensure the correct display of the results contained in the files by UMSAP.

3.2 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 window, fill in the needed information and hit the Start analysis button in the bottom right corner. 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. In each window performing an analysis, there is a progress bar near the Start analysis button that gives a rough guess of the remaining time needed to complete the current analysis.

Currently, the text boxes showing the path to a file do not have auto-complete capabilities. In addition, relative paths are not supported. Therefore, the full paths to files are needed. In the case of output files and folders if the text boxes are left empty, appropriate default values will be provided, based on the main input file path.

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 in the bottom left corner of the module/utility window

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.

Special care have been taken to handle errors that may appear during the processing of the input files. Errors will be reported to users using dialog boxes with plain English messages so users may correct the errors right away and continue with the analysis. In the case of an unforeseen error occurring, the message error will be more code oriented. It will be helpful if users send a crash report to umsap@umsap.nl so we can correct the error and/or create an appropriate error message.

In general, windows showing a graphical representation of results will be allowed to resize while module/utilities windows will not. The Tools menu will have options specific for each window. In most of the windows the functionalities in the Tools menu can be also accessed with the right mouse button.

For modules, UMSAP will create a .user file after finishing the analysis. This is an input file that can be used to run an analysis one more time without users having to type in again all the needed information. Once the .user file is selected, using the Run Input File entry in the Script menu, the appropriate module will be launched and the information found in the .user file will be used to fill the fields in the interface of the module. At this point, users may decide to update the analysis by changing some of the parameters in the interface of the module. Clicking the Start analysis button will start the analysis of the input data files. More details are given in ??.

The helper window designed to facilitate the definition of the control and various experiments in each module will try to read the values from a .user, .txt file or the given text in the corresponding text entry of a module interface. If something goes wrong while reading the given values, then the helper window will default to show the corresponding empty fields so users can define the experiments and control for the module.

3.3 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 *UMSAP - Utilities or Module Name - Name of the window*. For example, the window with name *UMSAP - Util - Correlation coefficients (t1.corr)* will be displaying the correlation coefficient matrix saved in the file *t1.corr*.

3.3.1 Keyboard shortcuts

The following table contains the built-in keyboard shortcuts.

Shortcut	Action	Module
Ctrl/Cmd+D	Close current window	All
Ctrl/Cmd+I	Read UMSAP .uscr file	All
Ctrl/Cmd+Q	Quit UMSAP	All
Ctrl/Cmd+R	Read UMSAP output file	All
Alt+Cmd+L	Launch the Limited Proteolysis module	All
Alt+Cmd+P	Launch the Proteome Profiling module	All
Alt+Cmd+T	Launch the Targeted Proteolysis module	All
Alt+Cmd+U	Launch the Utilities window	All
Ctrl/Cmd+L	Change selection mode	Limited Proteolysis
Ctrl/Cmd+Z	Remove last applied filter	Proteome Profiling

Table 3.1: List of built-in keyboard shortcuts.

3.4 The output files

In order to minimize user input, the names for output files and output folders can be left unspecified. In this case, UMSAP will use default names for files and folders. When names for files and folders are given, it is recommended to use names with no spaces, e.g. my-output-file or myoutputfile. The software will avoid overwriting previous output files or folders. If a selected folder already exists and it is not empty, UMSAP will create a new folder to write in the results. The location of the new folder will be the same as the selected folder. The name of the new folder will be the same as the selected folder plus the current date and time to the second, for example, selectedfolder-20190425092904. The same is true for files unless the user forces UMSAP to overwrite a file.

In general, the output file will be saved using a json format. If needed the data can be exported to a csv format with tab delimited columns, see ??.

3.5 Backward compatibility

UMSAP is capable to read the .tarprot file from previous versions. Other files generated by UMSAP are version dependent and cannot be read in with different versions of UMSAP. Nevertheless, two utilities allows to quickly generate any UMSAP file based on a .tarprot file from a previous version of UMSAP. These utilities are described in ?? and ?. How to generate a .tarprot file is explained in ??.