Protein Ligand Analyzer Guide

Basic Information

- Errors will show up if the input has the wrong type. (rational instead of whole numbers, etc.) The error messages can be found in the text field at the end of the window in red.
- Black messages in this text field explain the steps of the analysis as they happen.
- The analysis is going to take some time, scaling with number of scans and different mass-over-charge values.
- It is normal that it needs a second to show the first graph

Standard Input

Most of the input fields have already been filled with some example values. They can easily be changed by just entering another value.

Mode	
Untargeted:	Analyze all ligands (and protein) in the mass spectrometry data without prior selection. The mode specific inputs are explained later.
Targeted:	Focus analysis on specific ligands selected by the user. The mode specific inputs are explained later.
Upload:	
Upload Mass Spectrometry Data:	Select the .ms1 file that should be analyzed. See layout restrictions for more information.
Input	
Protein m/z	Specify the mass-to-charge ratio of the protein to focus on.
Protein charge state	Define the charge state of the protein for accurate filtering and analysis.
Energy function protein	Choose between the energy modes for the protein. Mostly “Low” = 1 and “High” = 2
Energy function ligands	Choose between the energy modes for the ligands. Mostly “Low” = 1 and “High” = 2
Different range Input	
Protein sampling range	Specify the m/z range around the protein's actual m/z value to average them.
Ligand sampling range	Specify the m/z range around the (possible) ligands' actual m/z values to average them.
Similarity settings	
DTW threshold	Set the similarity threshold for curve comparison using Dynamic Time Warping (DTW). The threshold is between 0 and positive infinity, the DTW value having to be below the threshhold to count as similar.
Pearson threshold	Determine the minimum Pearson correlation coefficient for similarity checks. The threshhold is between -1.0 and 1.0 (0 to 1 can be interpreted as a percentage correlation), the value having to be above the threshhold to count as similar.
Analyzing range settings	
Analysis x-start	Set the starting scan number for analysis and visualization

Analysis x-end	Set the final scan number for analysis and visualization. . This value is automatic set after the .ms1 file is selected. It can be changed after this selection.
Output settings	
Result Folder	Select the folder where the output PDF and the .csv files will be saved.
Protein & Ligand graphs in:	One plot: both graphs are in a single plot with different colors. Two plots: each graph has one plot. Every page first has the protein graph and then a ligand graph underneath.
Creation of .csv files?	NO: the data of the matching ligands is not saved as .csv files. YES: the data of each matching ligand is saved in one .csv file.

Mode: Targeted

In this mode only the selected ligands will be considered.

Select ligands	Choose ligands for targeted analysis by uploading a list. See input information for more details.

Mode: Untargeted

In this mode all ligands will be considered.

Charge exclusion range	Exclude specific charge states of the protein to refine the analysis. Enter 0 if no other charge state of the protein should be considered. If the range is odd: there is one step more above Example: range =5 and protein charge state 5 than 3,4,6,7, 8are excluded.
Start m/z	Starting m/z value for the analysis. This value is automatic set after the .ms1 file is selected. It can be changed after this selection.
End m/z	Final m/z value for the analysis. This value is automatic set after the .ms1 file is selected. It can be changed after this selection.
Ligand m/z grouping range	Define the grouping range for ligands to group them by their m/z values. Out of each group the curve with the highest similarity is considered for further analysis. Example: range=2 and m/z-values= 499, 501, 502 Group1= [499, 501], Group2=[502]
Protein exclusion window	Specify a window to exclude m/z values near the protein to reduce noise.

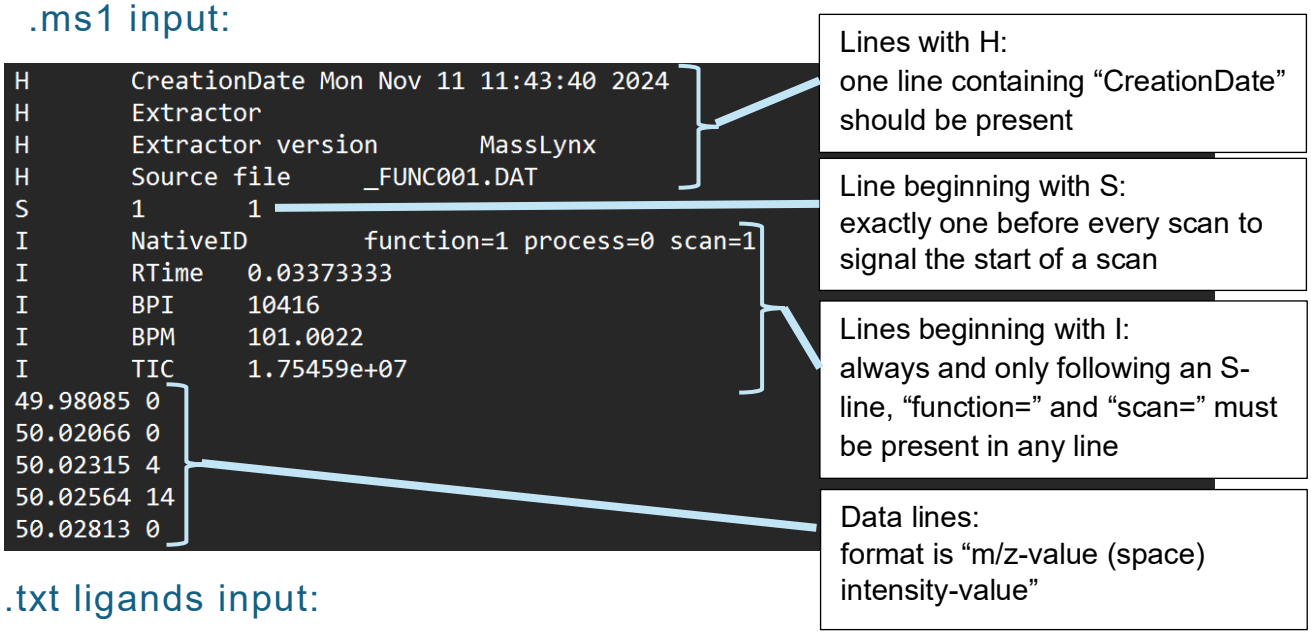
Advanced settings:

These settings should be changed carefully.

Savitzky-Golay filter window length	Specifies the number of data points used in smoothing to reduce noise while preserving signal shape.
Savitzky-Golay filter polyorder	Determines the degree of the polynomial used for fitting data within the specified window to smooth the signal.
Num of parse processes	Set the number of processes for parallel file parsing to improve performance. If >1: very RAM intensive.
Num of analysis processes	Specify the number of processes for parallel data analysis tasks.
Max Cache Size	Define the maximum cache size for storing intermediate data.
Cache Use	Enable (YES) or disable (NO) the use of cache to optimize runtime efficiency on multiple analyses of the same scan set.
Protein Charge state sum range	Through this mechanism the intensity values of the different charge states of the protein are summed. This new curve is used for analysis. Example: charge state 6 range 3 Sum of charge state= 5, 6, 7, 8

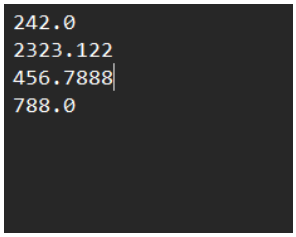
Input information:

.ms1 input:



.txt ligands input:

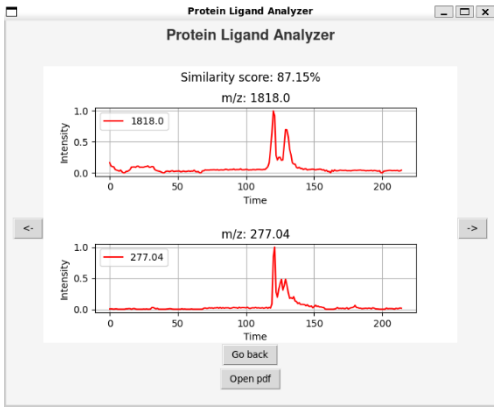
- Just enter every ligand that should be considered in a new line.
- They need to be floats



Output information:

Output in the program:

- The first plot is always the protein and the plot below one of the matching ligands. Both show their m/z value as identification.
- The similarity score shows how similar this protein curve is compared with the protein curve.
- The buttons “->” and “<-“ iterate through the matching ligands.
- The button “Go back” closes the result area and returns the app to the main interface.
- “Open pdf” opens the PDF result file.

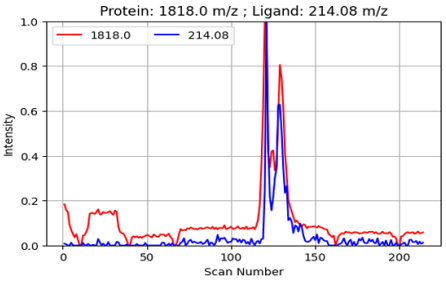


Output:

- This folder is created in the selected folder (see above). If there is no folder selected, it is saved on the desktop.
- The name contains: Analysis date and time, but also used scan date and time
- There are two output types saved in the folder

PDF:

- First page contains metadata about analysis and scan time
- Contains plots like the GUI output with two modes (protein and ligand in a single plot or in two separate ones)



Intensity graph comparing protein and ligand 1

.csv (optional):

- The name of each csv file corresponds to the pdf ligand numbering and contains the m/z value
- The files contain two columns consisting of scan numbers and intensities, the values corresponding to the plots in the graphical output

(Note: Certain machines use different separators in Excel for csv files, resulting in both columns being merged into one. Should this occur, research changing the “csv delimiter/separator” for your application)

	A	B
1	Scan Num	Intensity
2	1	0.007814
3	2	0.004883
4	3	0
5	4	0
6	5	0.01186
7	6	0
8	7	0
9	8	0
10	9	0
11	10	0
12	11	0
13	12	0.010988
14	13	0
15	14	0