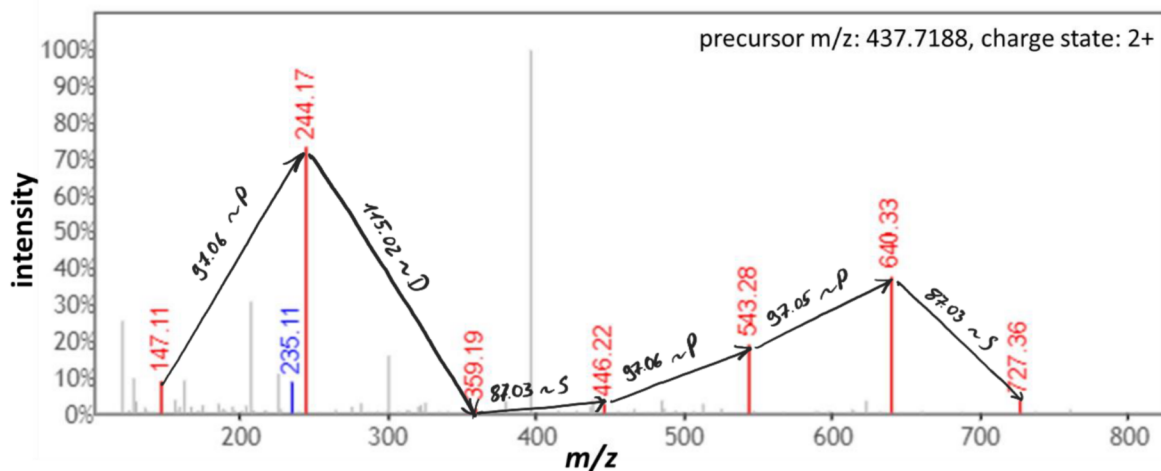


Exercise Sheet 8 – Computational mass spectrometry Computational Proteomics

1. Peptide Identification (11.0 points)

a) Consider the following tandem mass spectrum of a doubly charged precursor at an m/z of 437.7188. Which of the following peptides is likely the correct identification: FSPSPDPK, AEGILTDR, VLEGETAR? Explain your choice based on the precursor mass _and_ fragments. (3 point)



As the sequence given these m/z ratios is SPPSDP (K)
the most likely peptide FSPSPDPK
The precursor m/z of 437.7188 also fits

b) Write a small program/script/function that merges the experimental spectrum (csv file) provided on Moodle with the output of your program/script/function from Exercise 7 (2c). How many of the calculated monoisotopic b- and y-fragment peaks of the peptide "LHVPLEAGVLLFK" have a corresponding peak in the experimental spectrum when using a tolerance of 20 ppm on fragment level? Provide your source code and its output (as PDF at the end) and the output here (fill in blank line). (4.0 points)

Number of matching fragment peaks: _____

c) Use the combination of programs/scripts/functions written in Exercise 7 (2a and b) and this Exercise 8 (1b) to identify which tryptic peptide sequence from the protein "Q13740-1" matches best to the experimental spectrum (csv file from 2b). For this purpose, you have to 1) generate all possible tryptic peptides for the given protein, 2) generate the corresponding fragment mass lists for all peptides, 3) merge the mass list of each peptide with the experimental spectrum (ignoring precursor mass), 4) calculate the number of matching fragments and 5) pick the most likely correct peptide sequence matching the experimental spectrum. Provide your source code and its output (as PDF at the end) and the output here (fill in blank line) (4.0 points)

Peptide sequence matching the experimental spectrum best is: _____

2. False discovery rate estimation (4 points)

a) Calculate the estimated FDR (q-value) for each spectrum of the following list of peptide spectrum matches. How many different target peptide sequences do we identify at an FDR < 25%. (4 points).

Spectrum	Peptide sequence	Score	Target/Decoy	#Targets	#Decoys	FDR
1	ECCEKPLLEK	100	T	1	0	0
2	LVNEVTEFAK	90	T	2	0	0
3	CCTESLVNR	68	T	3	0	0
4	NECFLQHK	67	D	3	1	0.33
5	SLHTLFGDK	50	T	4	1	0.25
6	LVAASQAALGL	49	T	5	1	0.2
7	ETCFAEEGK	30	D	5	2	0.4
8	QTALVELVK	29	D	5	3	0.6
9	TYETTLEK	28	T	6	3	0.5
10	FQNALLVR	27	D	6	4	0.66

We can find up to 5 targets at an FDR < 0.25