

For this demo, we will analyze explore proteomics data and run a database search.

Important: Proteomics tools may not run on Mac OS

Getting the data

1. **MaxQuant** can be downloaded from <https://www.maxquant.org/>.
2. Mass spectrometry data (.raw files) and protein database (.fasta) can be downloaded from <https://figshare.com/account/projects/179178/articles/24288913>
 - a. The data originally came from <https://zenodo.org/record/4274987>
 - b. They are proteomics data of healthy human serum samples that were digested with Trypsin and analyzed on a Q-Exactive Plus mass spectrometer (Orbitrap mass analyzer).
3. **MZmine** tool for visualizing mass spectra can be downloaded from <https://github.com/mzmine/mzmine3/releases/tag/v3.2.8> (select the _portable file if you don't want to install it)
4. A proteomics paper where we can learn to spot key protocol details: <https://www.nature.com/articles/s41586-020-2402-x>.

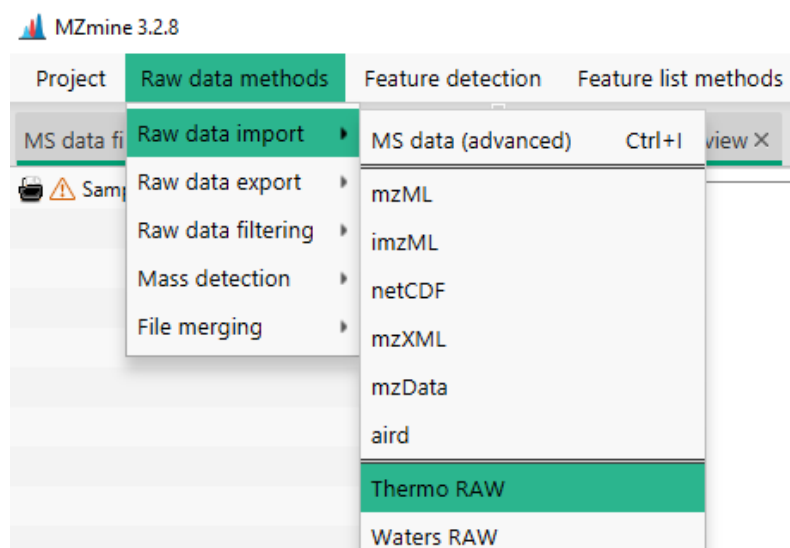
Setting up software

Just unzip **MaxQuant** and **MZmine** (if you selected the _portable version). When launching **MaxQuant**, if there is an error about **Microsoft .NET framework**, please follow the link to download and install .NET on your PC.

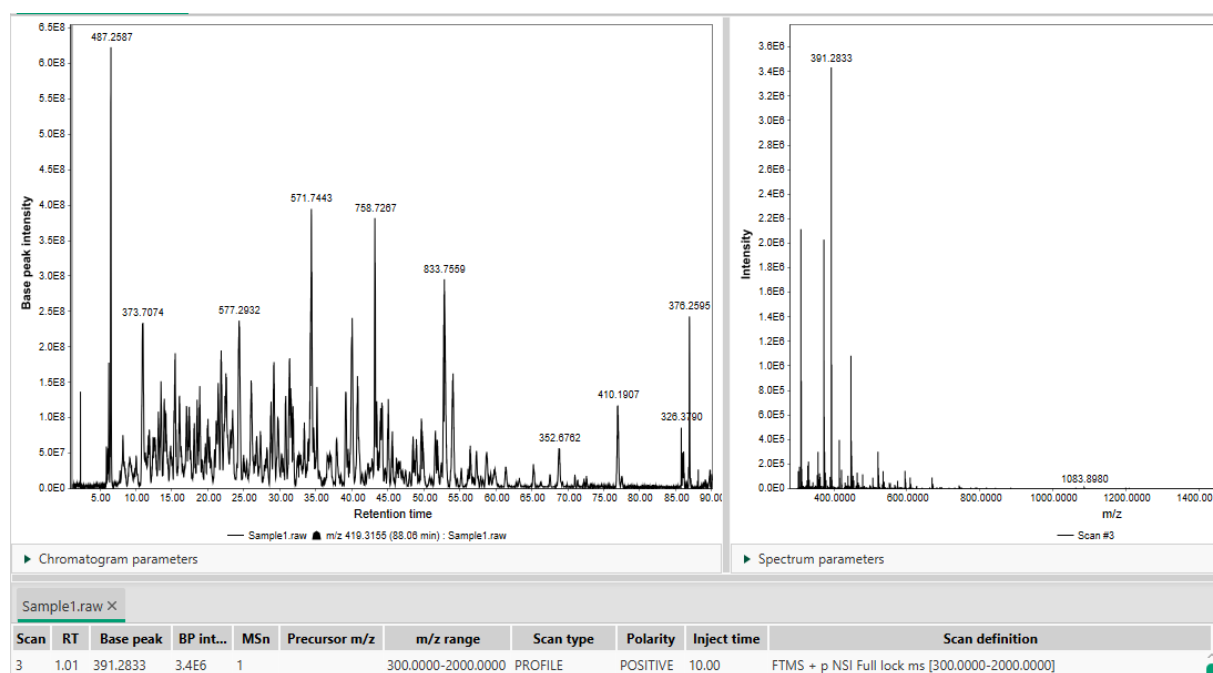
Running the demo

Section 1: MS data exploration

1. The .raw files came from Thermo Scientific's mass spectrometer machine. We can view the spectra in **MZmine** by specifying the import format as **Thermo RAW** as shown below.



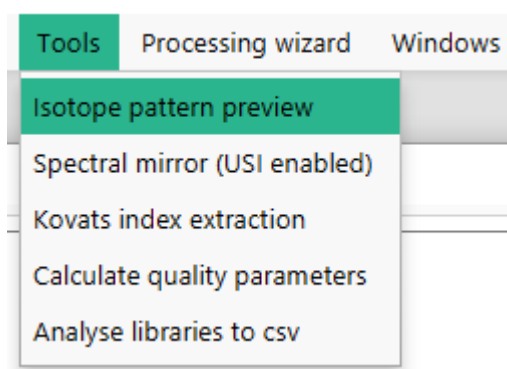
- Once imported, you will be shown the **Chromatogram** (left panel). This shows the most intense ions that eluted out from liquid chromatography into the mass spectrometer.



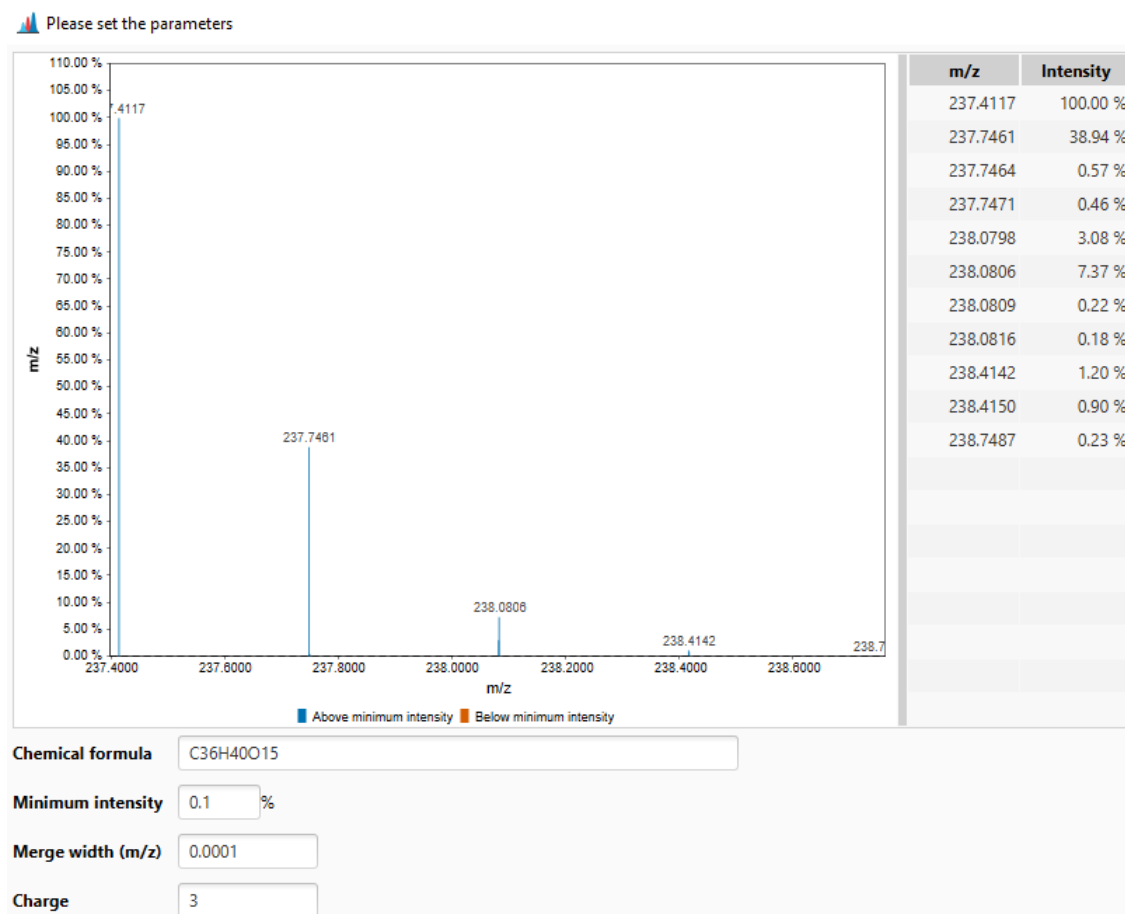
- At the bottom, you will see the list of all spectra (sorted by scan number, or time). The **MSn** column tells us whether the spectrum is **MS1** (of intact peptides) or **MS2** (of fragmented peptides). **MS1** is like a survey of existing peptides while **MS2** provides signature ions that can be matched to protein/peptide databases.

Sample1.raw X										
Scan	RT	Base peak	BP int...	MSn	Precursor m/z	m/z range	Scan type	Polarity	Inject time	Scan definition
1	1.00	391.2833	3.6E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
2	1.01	391.2833	4.0E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
3	1.01	391.2833	3.4E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
4	1.02	391.2832	3.8E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
5	1.02	310.0276	6.7E4	2	388.0055	80.6667-1210.0000	CENTROIDED	POSITIVE	80.00	FTMS + c NSI d Full ms2 388.0055@hcd25.00 [80.6667-1210.0000]
6	1.02	391.2834	3.9E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
7	1.03	391.2850	3.4E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
8	1.03	391.2833	3.7E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
9	1.04	391.2834	3.7E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
10	1.04	391.2834	3.9E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
11	1.05	391.2833	3.9E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
12	1.05	391.2834	3.6E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
13	1.06	391.2834	3.5E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
14	1.06	391.2833	3.9E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
15	1.07	391.2833	3.7E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
16	1.07	391.2835	5.2E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
17	1.08	391.2834	3.8E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]
18	1.08	97.0412	6.5E3	2	381.8745	53.3333-800.0000	CENTROIDED	POSITIVE	80.00	FTMS + c NSI d Full ms2 381.8745@hcd25.00 [53.3333-800.0000]
19	1.09	298.9213	2.4E4	2	376.8994	78.3333-1175.0000	CENTROIDED	POSITIVE	80.00	FTMS + c NSI d Full ms2 376.8994@hcd25.00 [78.3333-1175.0000]
20	1.09	391.2833	3.4E6	1		300.0000-2000.0000	PROFILE	POSITIVE	10.00	FTMS + p NSI Full lock ms [300.0000-2000.0000]

4. We can visualize isotopic envelope pattern using the tools shown below.



5. Given a chemical formula, $C_xH_yN_zO_w$, we can compute the expected abundance of all isotopes of this molecule based on the frequencies of naturally occurring C^{13} , C^{14} , N^{15} , and O^{18} .



6. To visualize the process of peptide fragmentation, we use an online tool <https://proteomicsresource.washington.edu/cgi-bin/fragment.cgi>. Let's input a peptide HGGLASKPW.

PEPTIDE FRAGMENTATION

Enter sequence here:

[Clear sequence](#) [Click here to paste in a sample peptide](#)

HGGLASKPW

FRAGMENT!

Set fragmentation parameters:

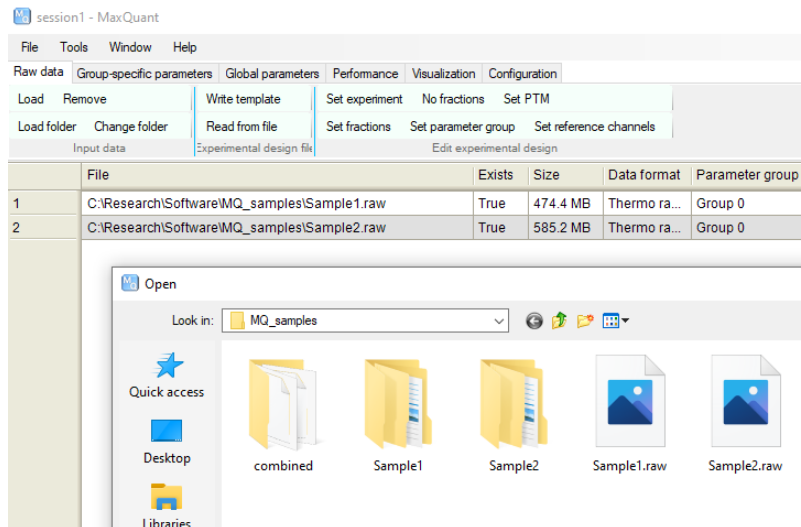
- mass type: ☒ mono ☐ avg
- product charge: ☒ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10
- fragment ion type: ☐ a ☒ b ☐ c ☐ x ☒ y ☐ z ☐ z⁺

7. The tool will show all b- and y-ions that could be produced from this peptide. The default setting is to show the monoisotopic masses (without extra neutron) and to show the masses at charge state of +1.
- b-1 ion (from the N-terminus) consists of only a Histidine. Histidine's molecular weight is 137.06. The weight shown is Histidine with an addition of a proton = 138.066.
 - b-ions at +1 charge = 1 + total amino acid molecular weight
 - y-1 ion (from the C-terminus) consists of only a Tryptophan. Tryptophan's molecular weight is 186.08. The weight shown is Tryptophan with an addition of H₂O + proton = 205.097.
 - y-ions at +1 charge = 19 + total amino acid molecular weight
 - The mass difference between adjacent ions is equal to amino acid mass
 - Glycine's weight is 57.02 = 195.088 – 138.066 = 252.109 – 195.088

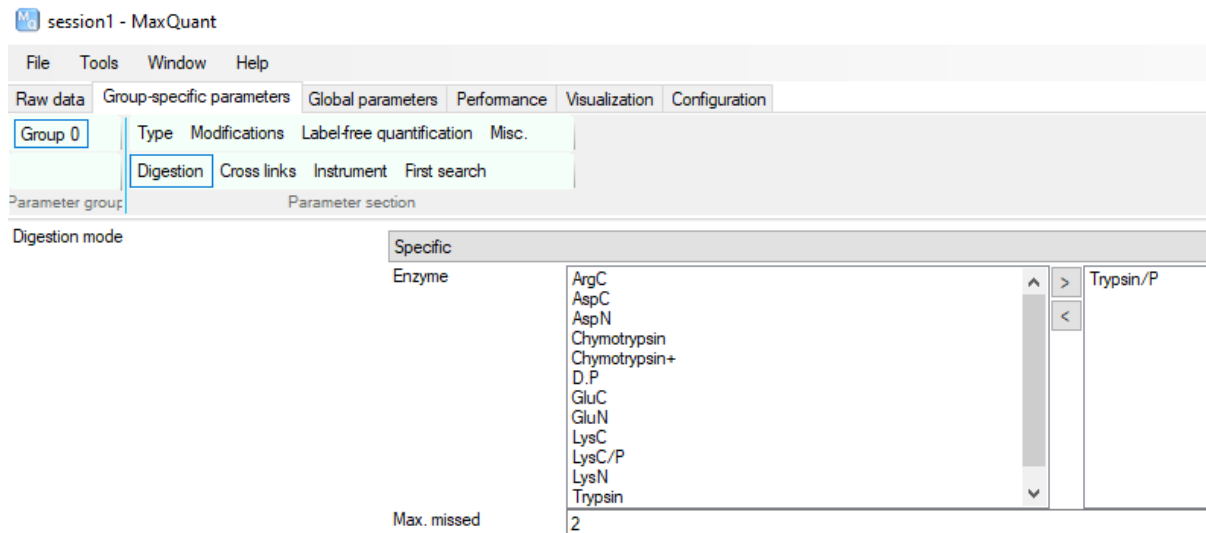
b ⁺		AA		y ⁺
138.066188	→ 1	H	9	
195.087652	2	G	8	815.441015
252.109116	3	G	7	758.419551
365.193180	4	L	6	701.398087
436.230294	5	A	5	588.314023
523.262322	6	S	4	517.276910
651.357285	7	K	3	430.244881
748.410049	8	P	2	302.149918
	9	W	1	← 205.097154

Section 2: Protein/peptide search with MaxQuant

1. Load individual .raw file or a folder containing multiple .raw files using the **Load** and **Load folder** buttons.



2. Next, let us explore the vast settings in **MaxQuant** (we will do this in class). Proteomics analysis is highly specific to the experimental conditions and mass spectrometer setups. So, if you want to re-analyze public data, you need to understand the entire protocol.



- We can also add new post translational modification or chemical modification to be included in the protein/peptide search. This functionality is in the **Configuration** tab.

session1 - MaxQuant

File	Tools	Window	Help		
Raw data	Group-specific parameters	Global parameters	Performance	Visualization	Configuration
Modifications	Add	Duplicate	Move down	Discard changes	Modify table
Proteases	Crosslinks	Remove	Move up	Save changes	
Data type	Table actions				Actions
	Name	Description	Composition	Position	
1	Acetyl (K)	Acetylation	C(2) H(2) O	Not C-term	
2	Acetyl (Protein N-term)	Acetylation of the protein N-terminus	C(2) H(2) O	Protein N-term	
3	Carbamidomethyl (C)	Iodoacetamide derivative	C(2) H(3) N O	Anywhere	
4	Oxidation (M)	Oxidation	O	Anywhere	
5	Phospho (STY)	Phosphorylation	H O(3) P	Anywhere	
6	GlyGly (K)	Ubiquitination residue	H(6) C(4) N(2) O(2)	Anywhere	
7	RGG (K)	Ubiquitination residue (chymotrypsin)	H(18) C(10) N(6) O(3)	Anywhere	
8	Methyl (KR)	Methylation	C H(2)	Anywhere	
9	Dimethyl (KR)	di-Methylation	H(4) C(2)	Anywhere	
10	Trimethyl (K)	tri-Methylation	C(3) H(6)	Anywhere	
11	Pro5	Proline-13C5	Cx(5) C(-5)	Anywhere	
12	Pro6	Proline-13C515N1	Cx(5) Nx C(-5) N(-1)	Anywhere	