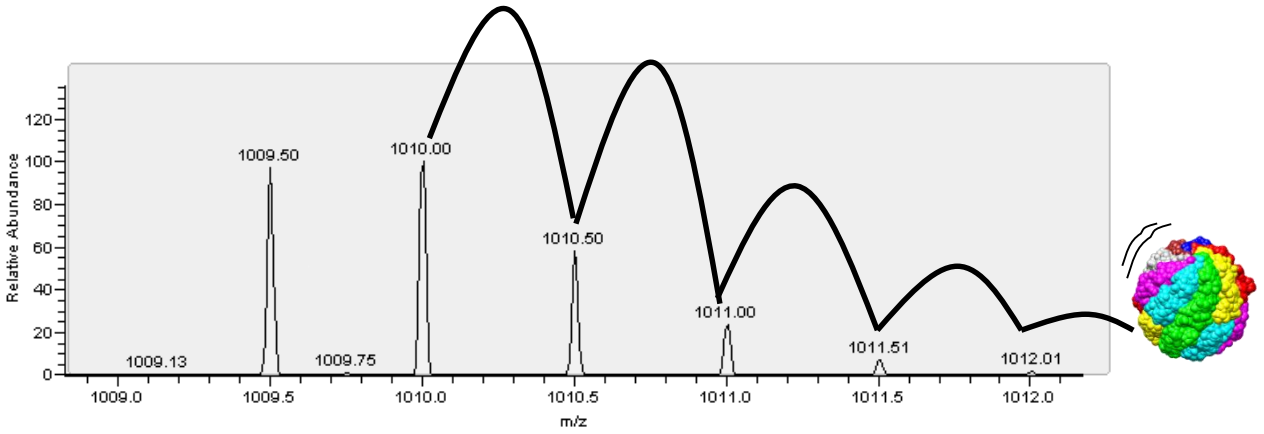
Basics of Protein Mass Spectrometry



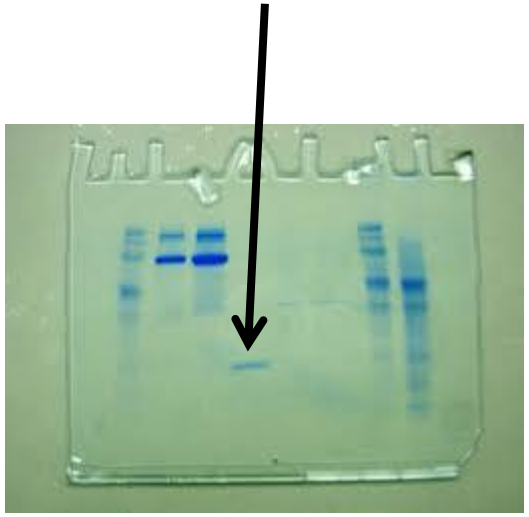
Nir Kalisman

Dept. of Biological Chemistry

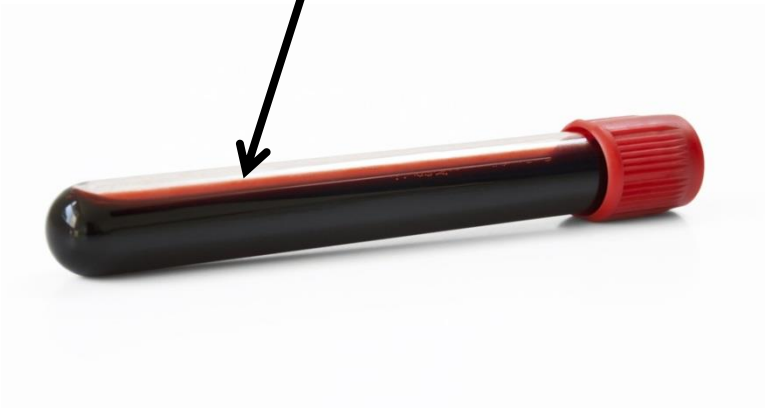
Hebrew University of Jerusalem

Proteomics – Mass-Spectrometry

What is this protein?

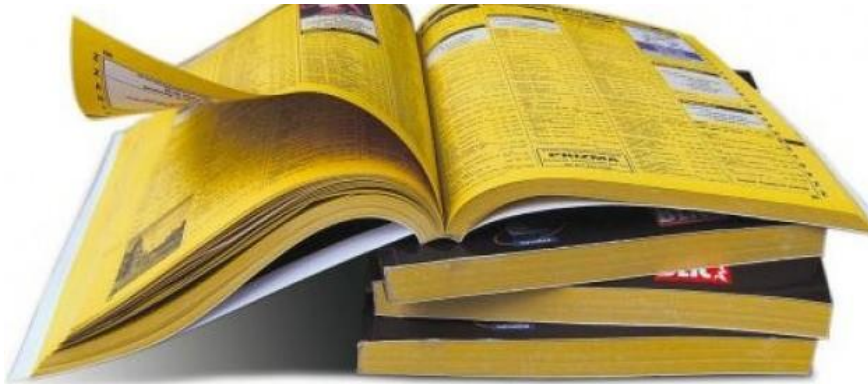


What are ALL the proteins in this blood sample?



Proteomics - Mass-Spectrometry (2)

A complex sample



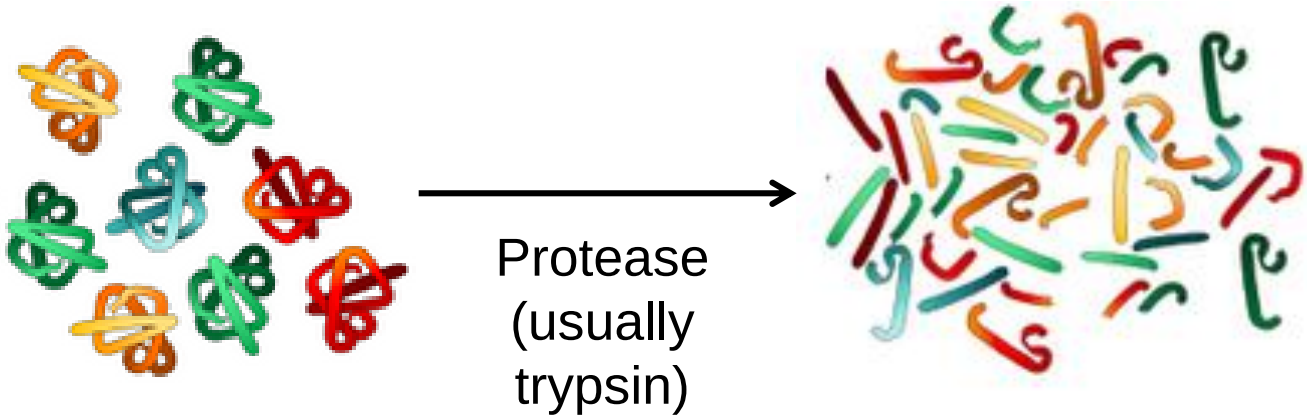
The Mass-of-Individual directory

An accurate scale

Mass

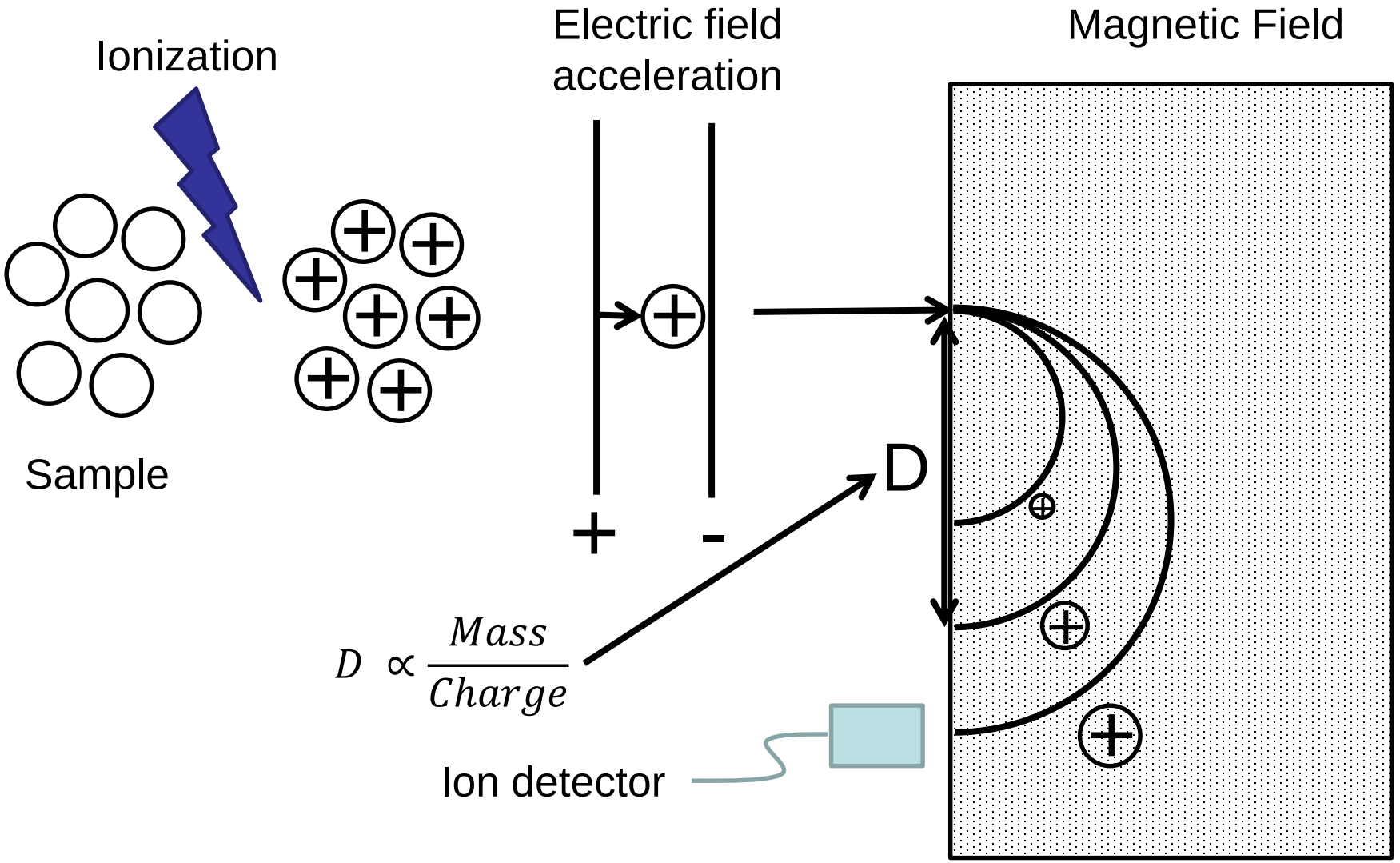
Name

Proteomics is based on peptide identification



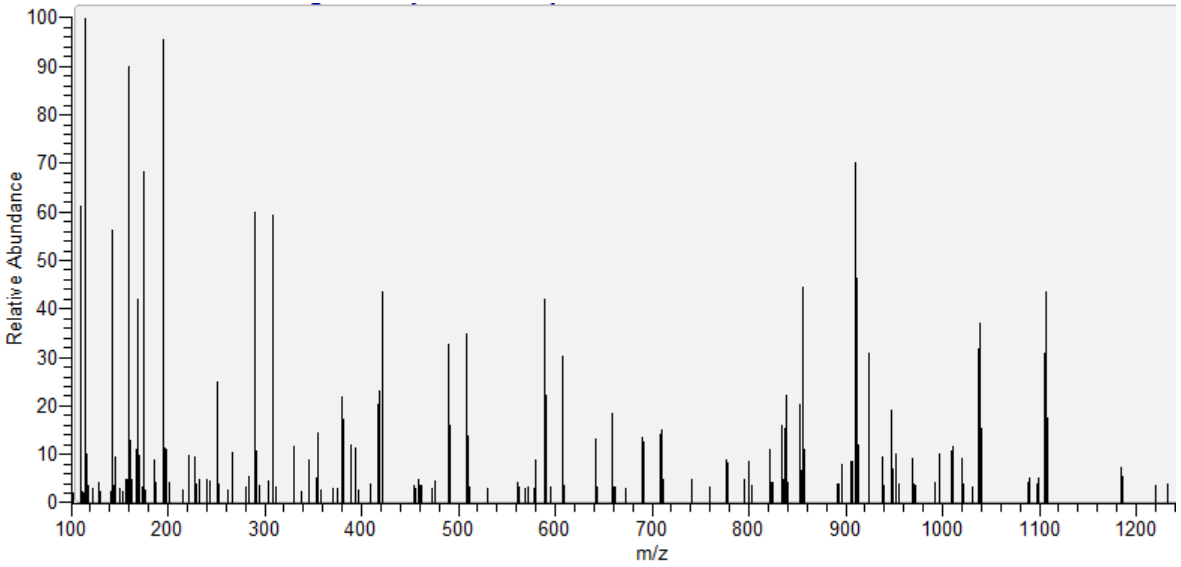
QLEYSRDCFCEDHCDKCSVVLTLQAFGESESTTNVYSKDLVIVSNLMG

How to measure the mass of an ion?



The result is a **SPECTRUM**

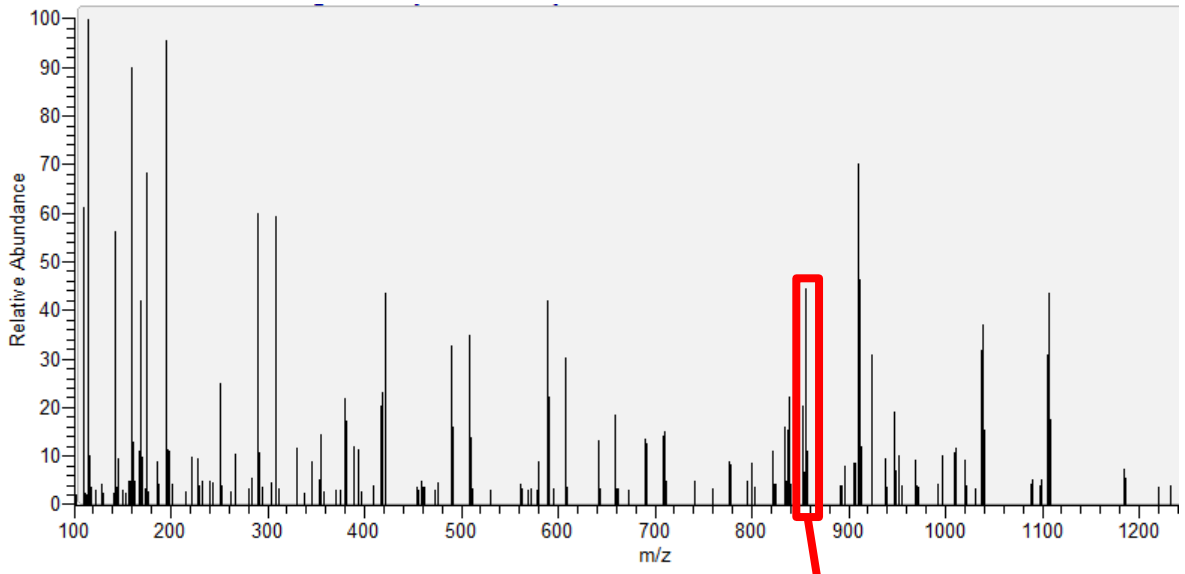
Intensity



$$\frac{Mass}{Charge}$$

Mass Spectrum

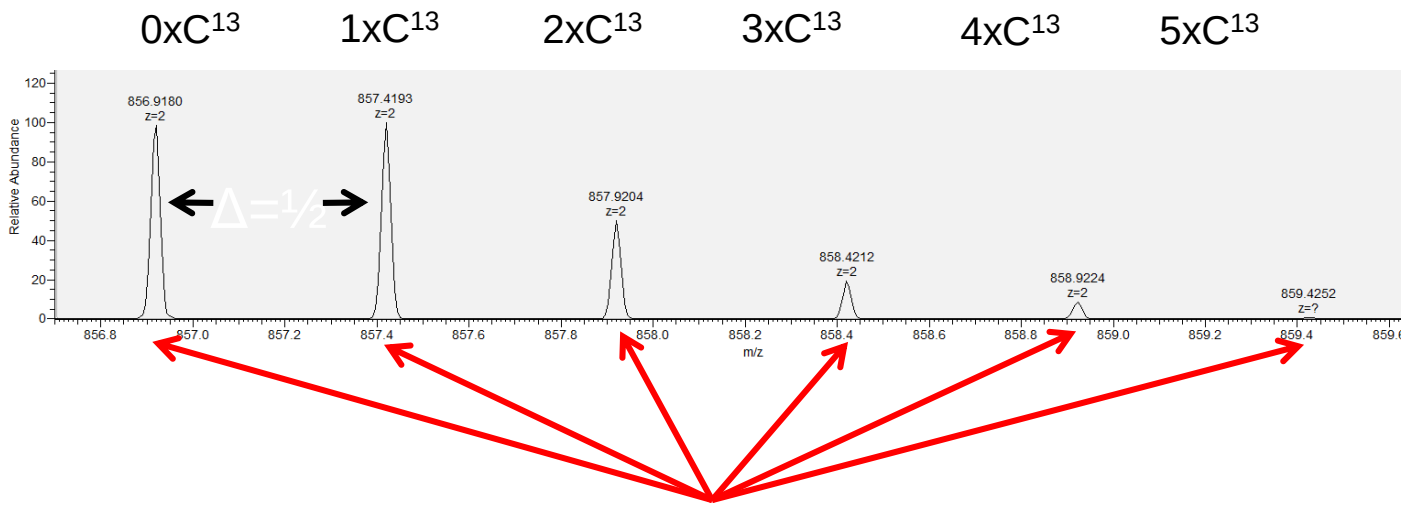
Intensity



$$\frac{Mass}{Charge}$$

Let's zoom here…

Mass Spectrum - Zoom



- This peptide has 75 carbon atoms
- About 1 carbon in 100 is C^{13}

$(M+1)/Z - M/Z = 1/Z = 1/2$

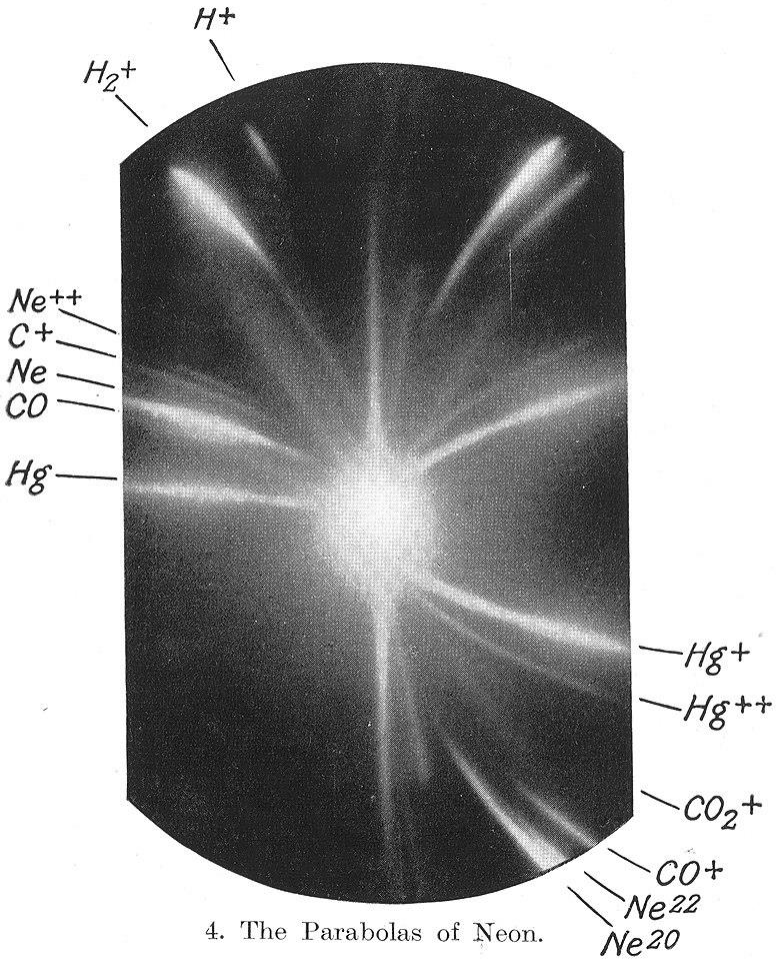
$Z = 2$

Isotopes of elements in organic compounds

Element	Common isotope	Next common isotope
Hydrogen	^{1}H (99.98%)	^{2}H (0.02%)
Carbon	^{12}C (98.1%)	^{13}C (1.1%)
Nitrogen	^{14}N (99.6%)	^{15}N (0.4%)
Oxygen	^{16}O (99.8%)	^{18}O (0.2%)
Sulfur	^{32}S (95.0%)	^{34}S (4.25%)

- Carbon is both occurring in high numbers and has a second isotope with relatively high prevelance.

Historical side note: MS and the discovery of isotopes



4. The Parabolas of Neon.

Mass-Spectrometry - Analysis

Measurements

Ideally

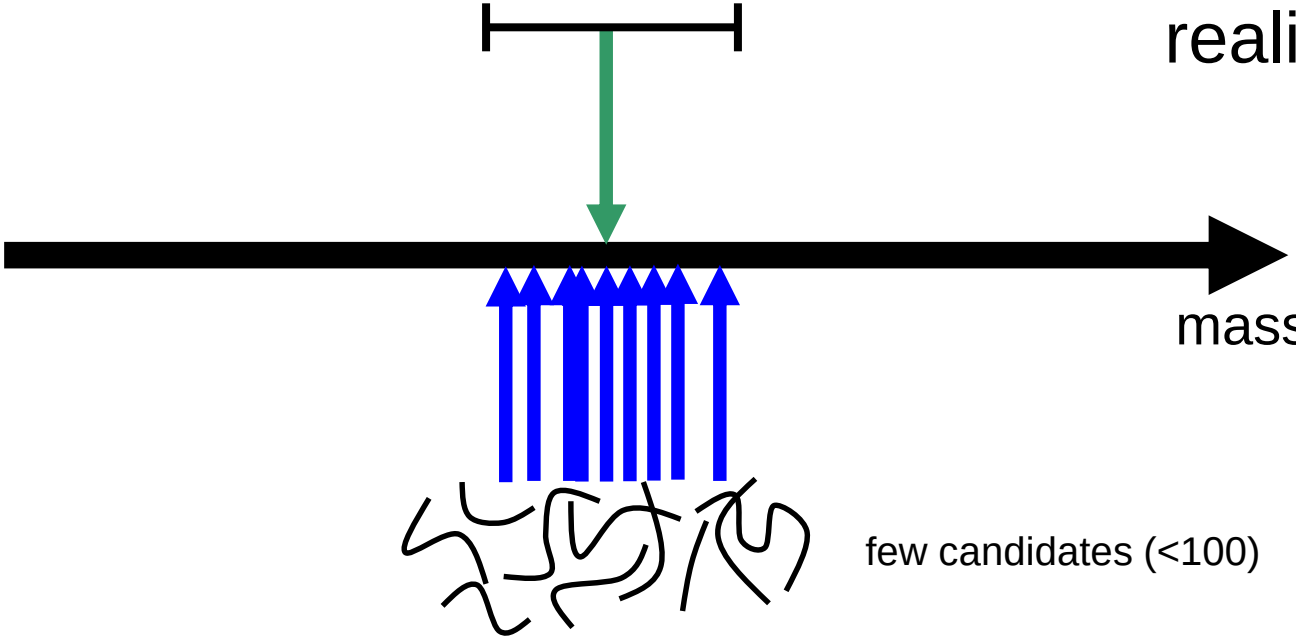
mass

In-silico
digestion

Mass-Spectrometry – Analysis (2)

Measurements

1000 ± 0.001 Da

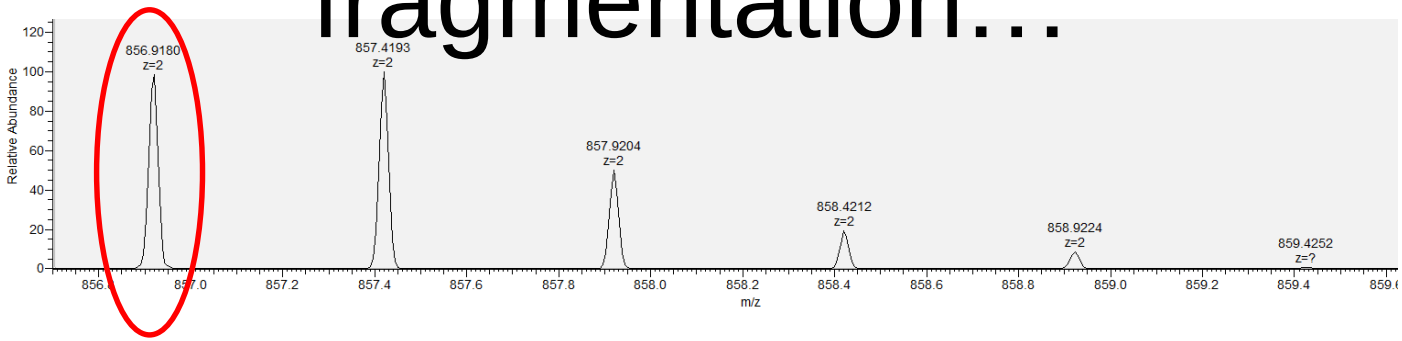


In-silico digestion

few candidates (<100)

In reality

We solve this ambiguity by fragmentation…

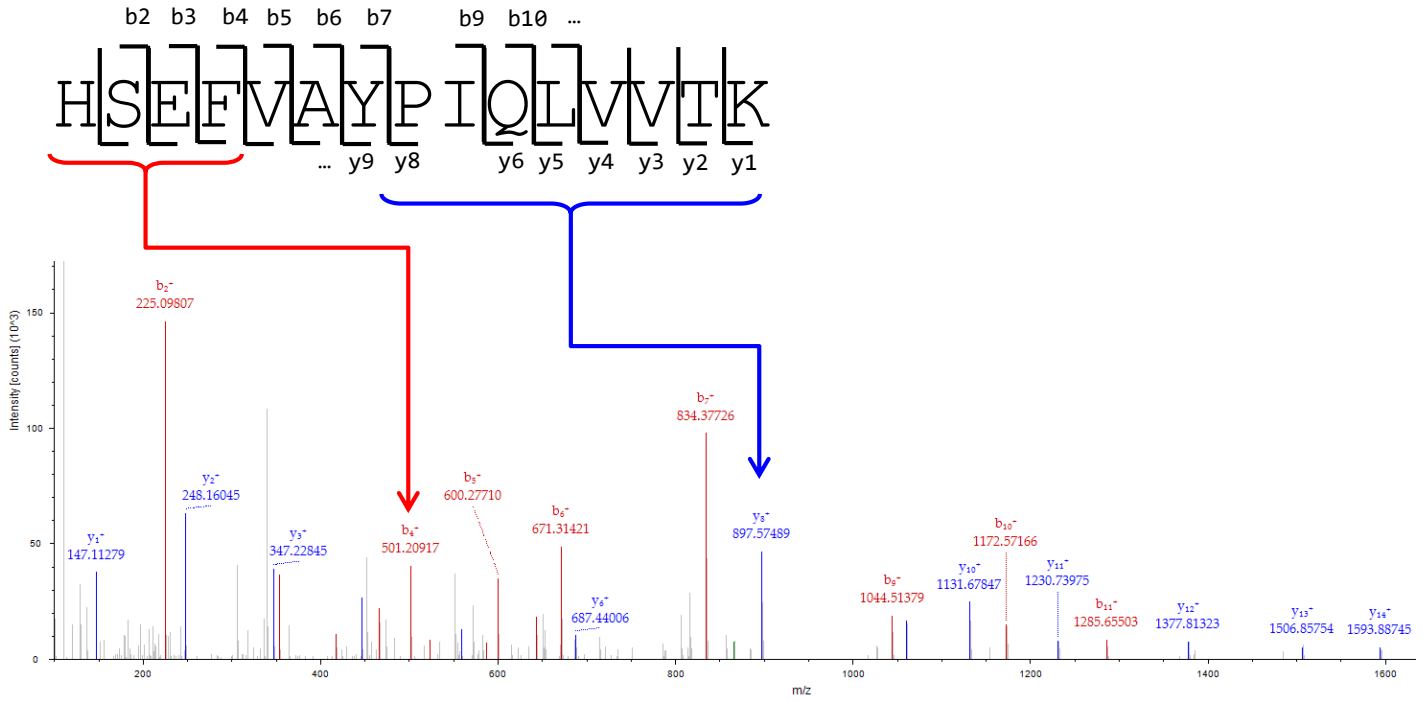


Only the ions with this m/z are selected

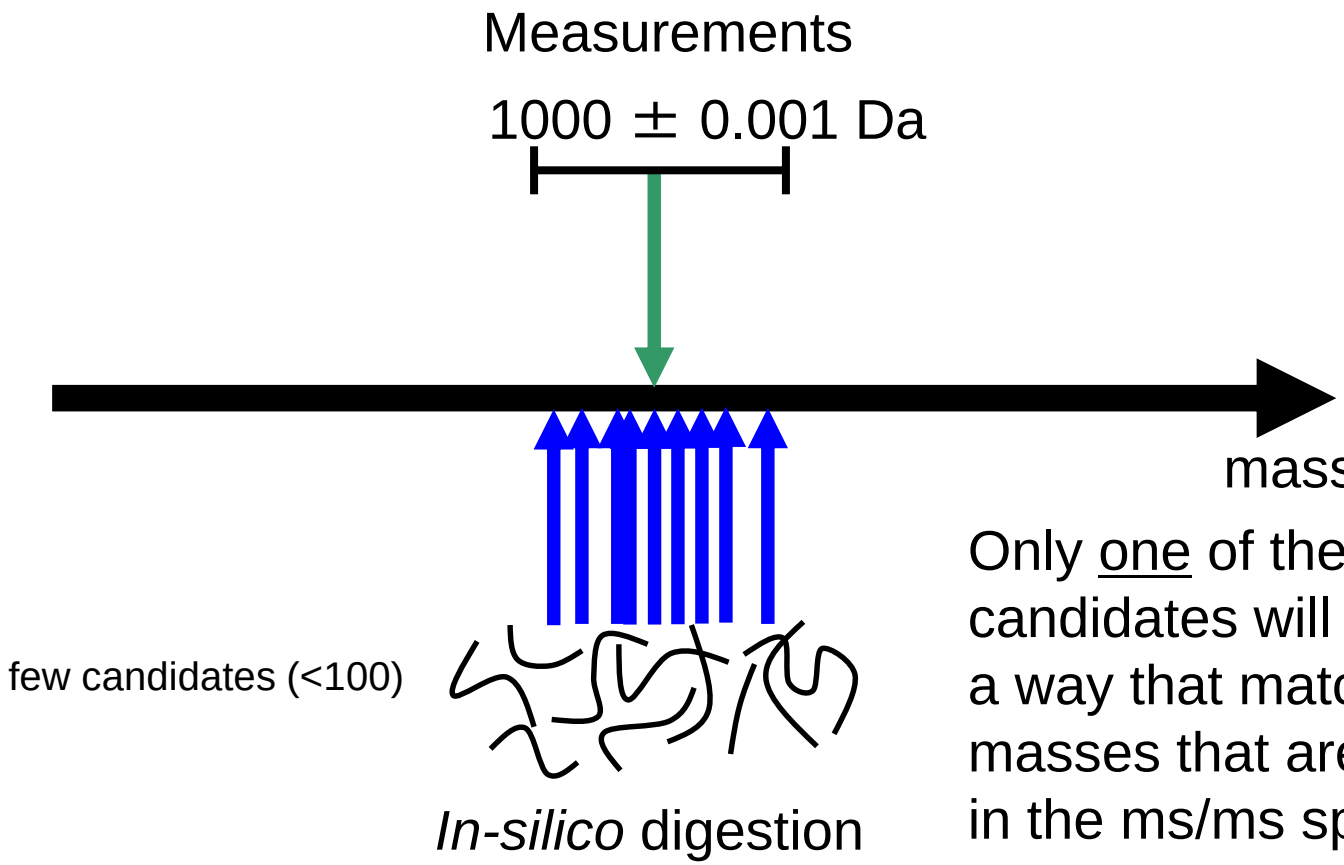
Made to collide with dilute N_2 gas → The fragments are subjected to a second mass spectrometry analysis

This method of fragmentation is called HCD (High-energy Collisional Dissociation).

The result is the fragment spectrum – MS/MS – MS^2

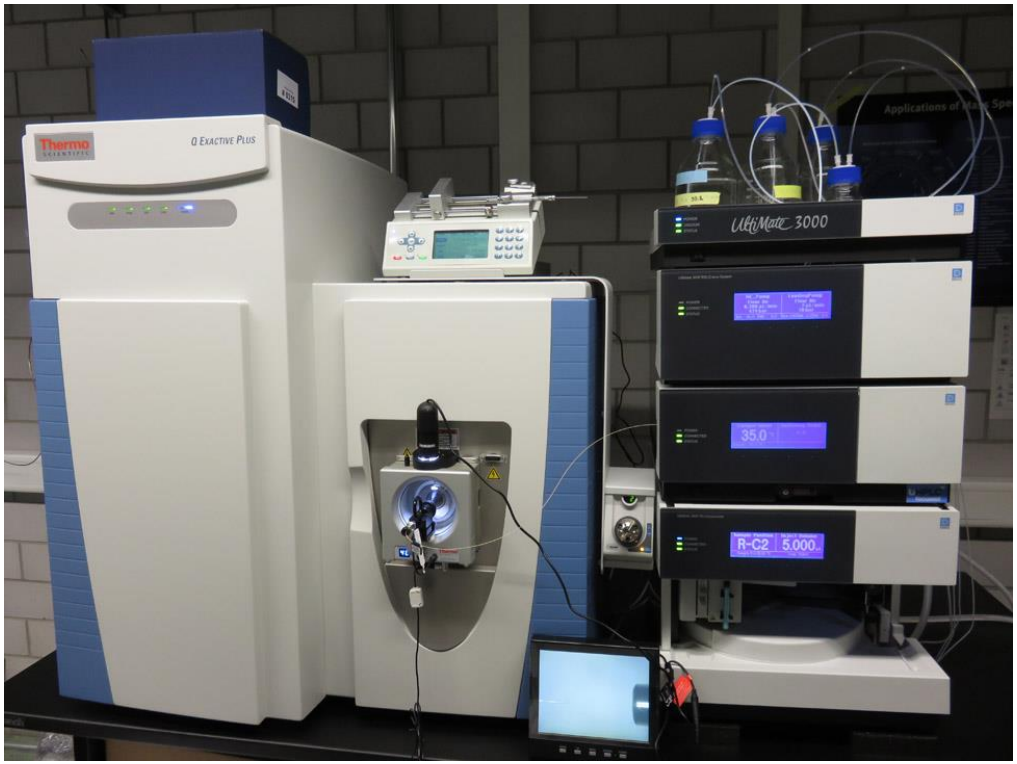


The fragment spectrum resolves the ambiguity



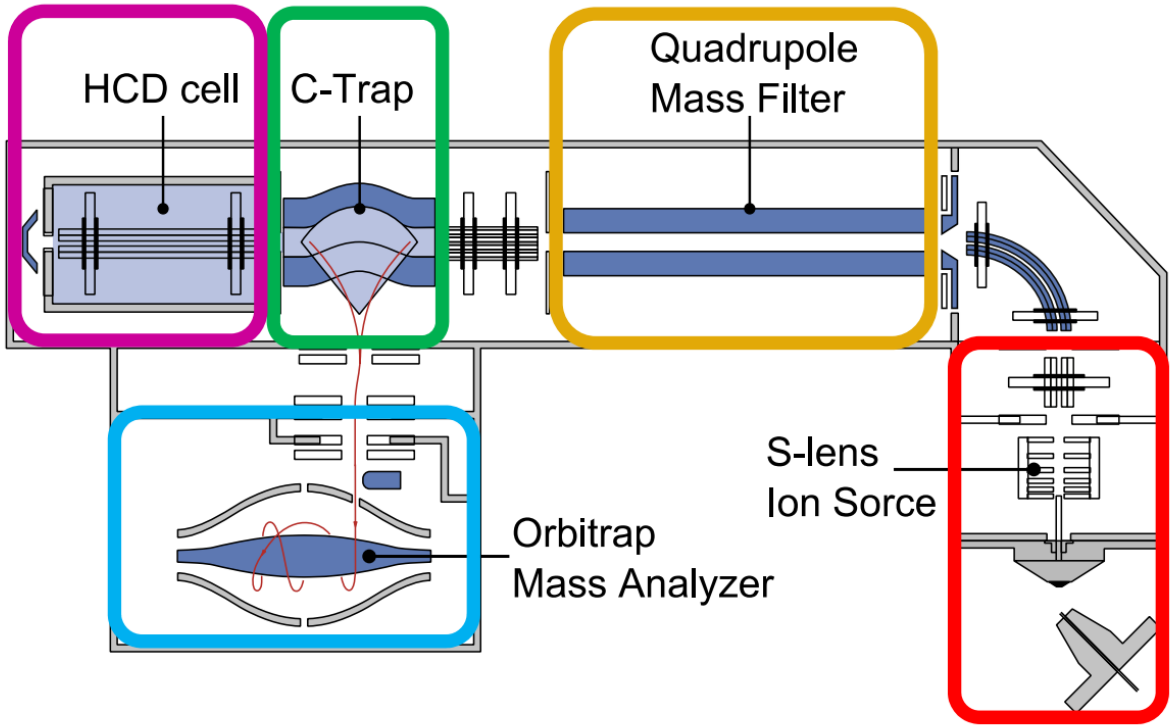
Only <u>one</u> of these candidates will fragment in a way that matches to new masses that are observed in the ms/ms spectrum.

The actual instrument…

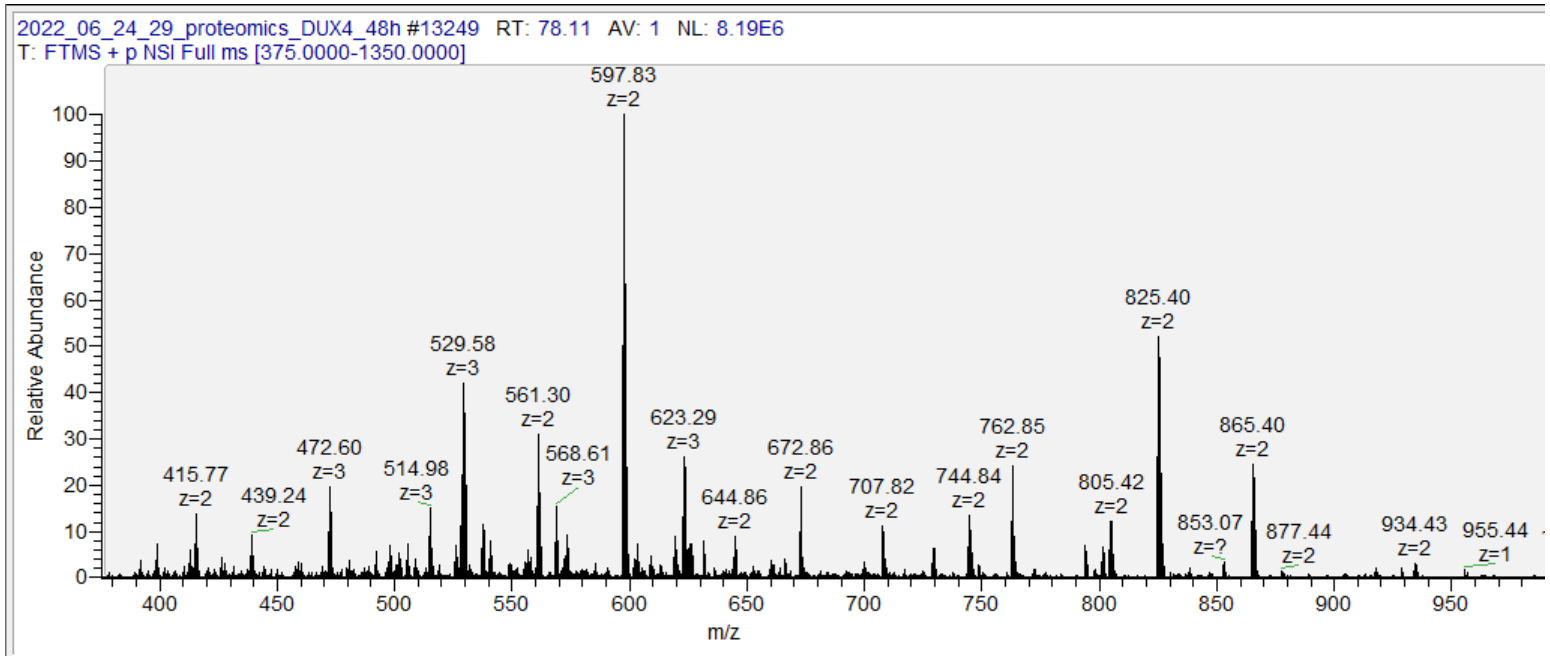


Q-Exactive mass spectrometer + Liquid Chromatography system

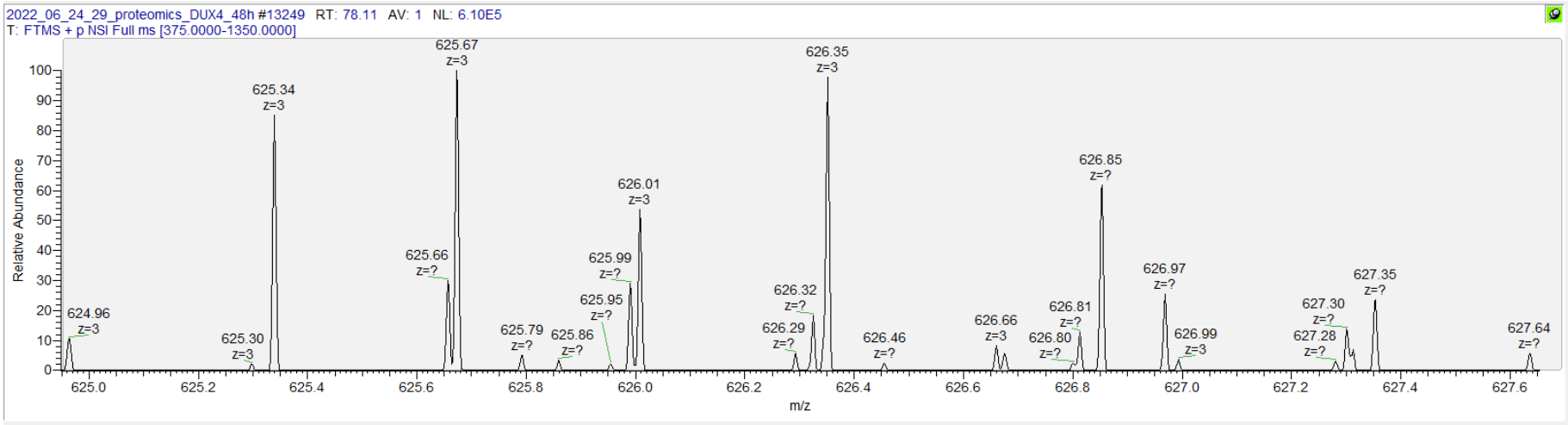
The Q-Exactive Plus



Additional problems: (1) Dynamic range

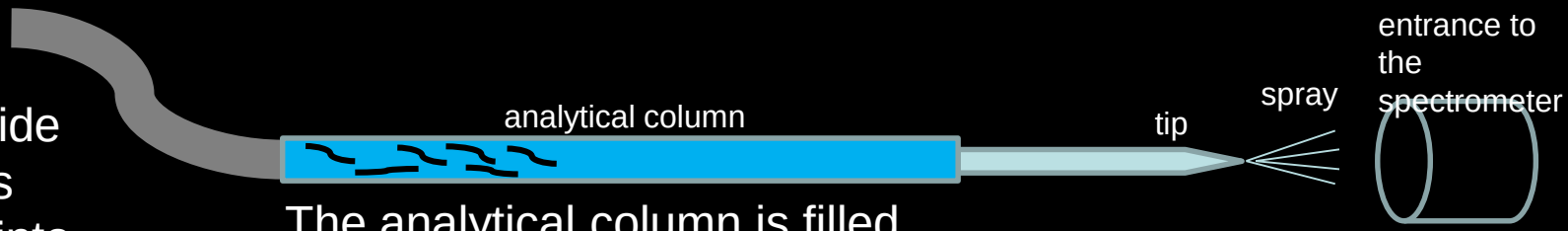


Additional problems: (2) overlapping peaks



Conclusion: We cannot inject into the spectrometer all the peptides from a complex sample (e.g. cell lysate) <u>at the same time.</u> The instrument will not be able to measure anything.

We solve these problems with Liquid Chromatography (LC)

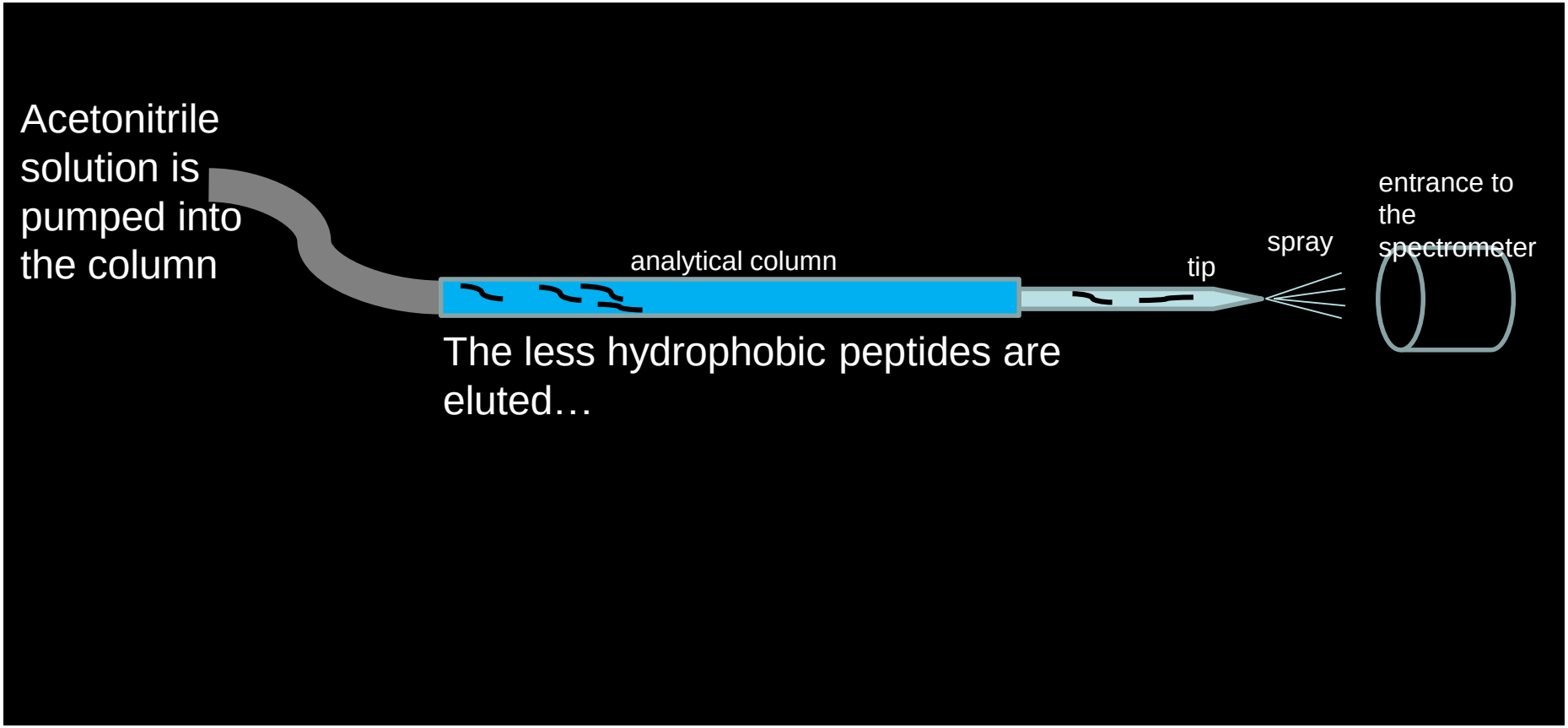


The peptide mixture is pumped into the column…

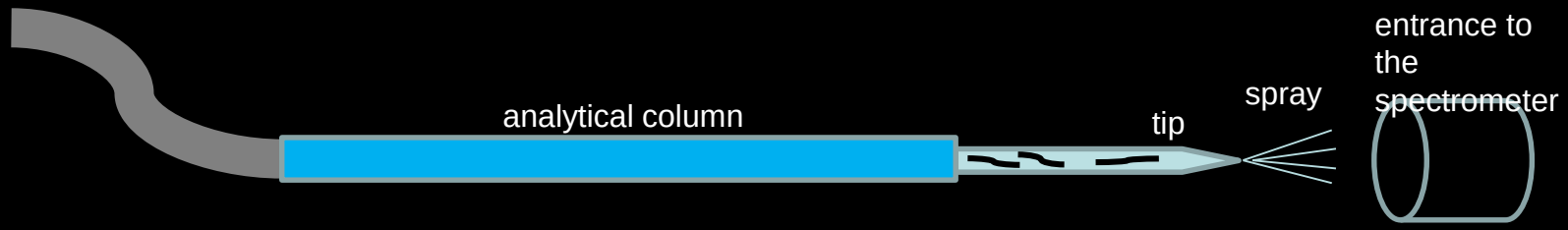
The analytical column is filled with C18, which is a hydrophobic resin. All the peptides initially stick to the resin, because they are slightly hydrophobic…

This initial “loading” step is all done in weak acid in 100% water

LC (2)

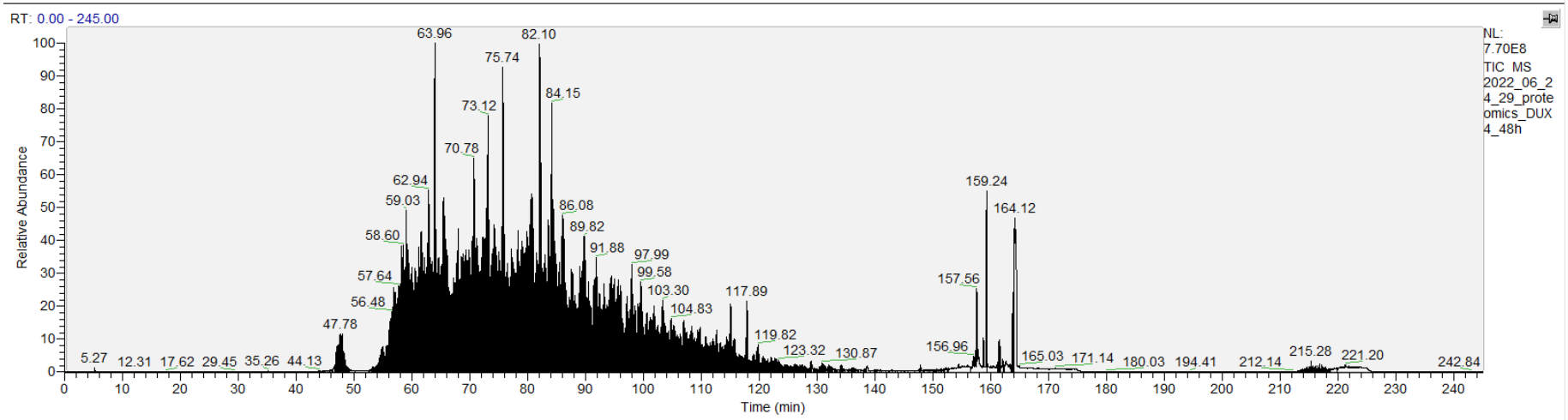

Acetonitrile solution is pumped into the column
analytical column
tip
spray
entrance to the spectrometer
The less hydrophobic peptides are eluted…

LC (3)



At higher acetonitrile concentrations (40% ACN in water), essentially all the peptides elute…

LC (4)



Loading

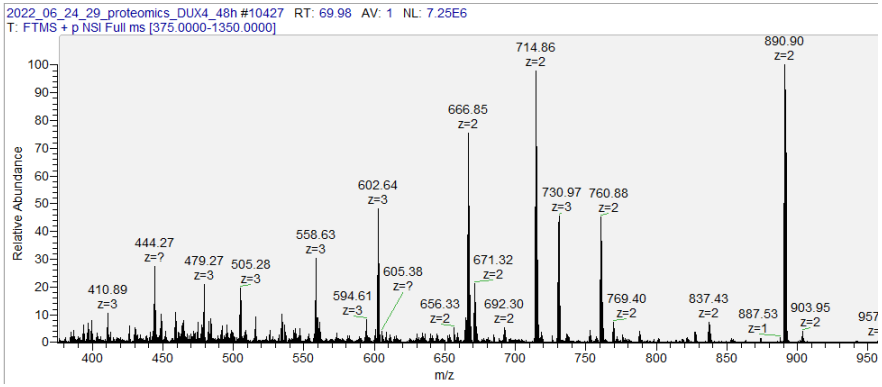
0% ACN

40% ACN

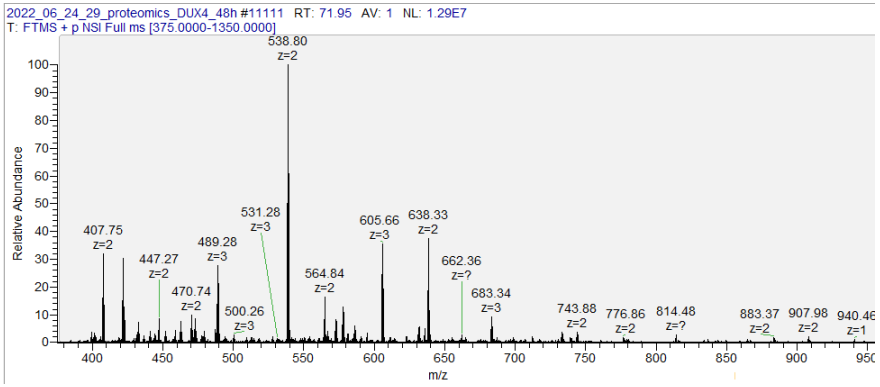
Column Wash

Acetonitrile gradient

LC (5)



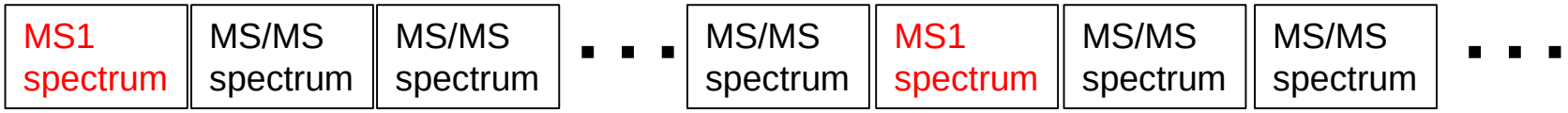
MS1 spectrum at 70 min



MS1 spectrum at 72 min

Completely different!

The mode of operation



typically 15 MS/MS

New MS1 spectrum is taken every ~2 second

Some numbers…

- ~30,000 peptides from ~4,000 proteins can be identified in a typical LC-MS run of lysate from human cells (HeLa), using a 100-minute ACN gradient.
- The minimal number of molecules of a specific peptide that are required for detection is 1,000 - 10,000.
- Further separation steps prior to the LC, can lead to the final detection of ~8,000 proteins (combining several runs together). This is called “deep proteomics”.
- Very good results can be measured from 100 ng of digest.
- Price of a complete LC-MS system: 750K– 2.2M$. Maintenance: 50,000$/year.
- Price of a single run: ~50$.

Issues of protein identification

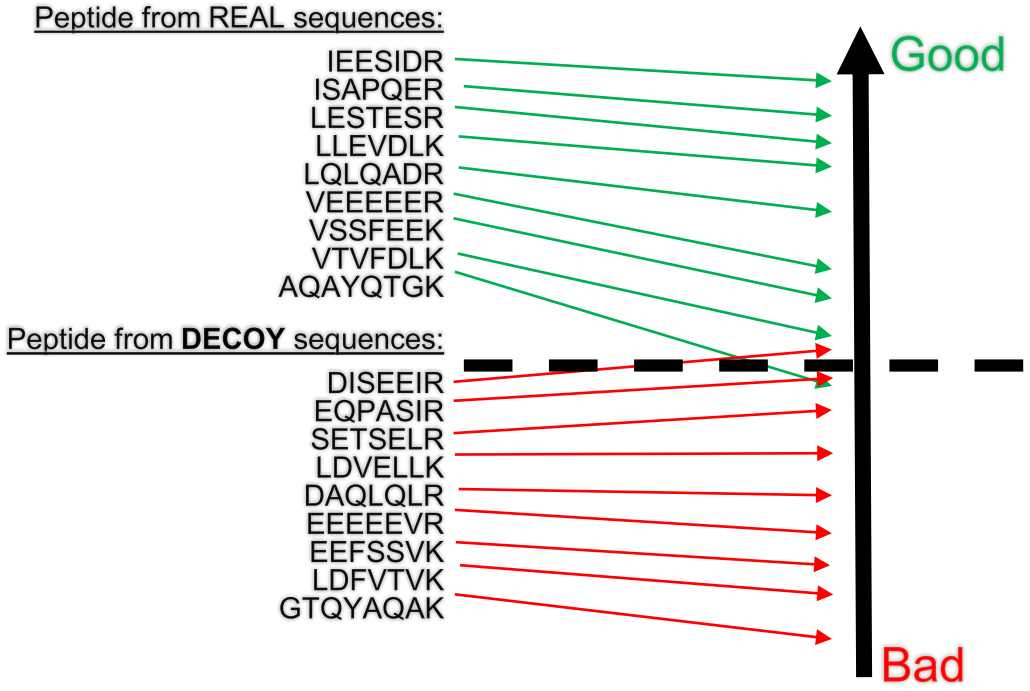
In Eukaryotes it is very common for multiple copies of (almost) the same protein to occur in the genome:

A1 and A2 are indistinguishable by mass spectrometry.

You can only determine that A3 occurred in the sample if you are “lucky” to measure peptides that span the red sections. This may not always occur, especially if the sequence similarity between A3 and A1 is very high (>95%). Hence, they will be classified to the same “protein group”.

False Discovery Rate (FDR)

Score = Goodness of fragmentation fit Measurement vs. Theoretical



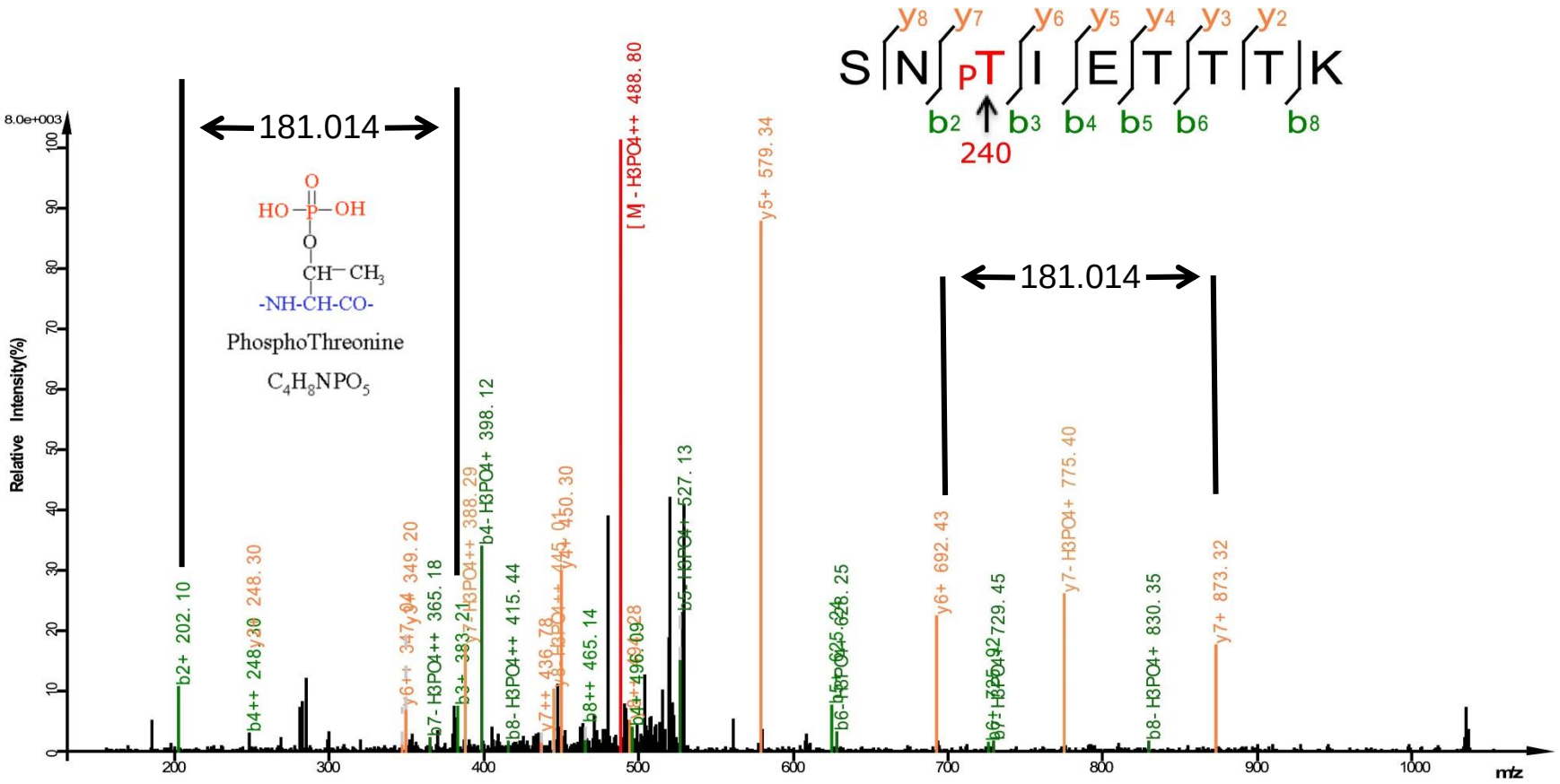
We will want to put the threshold for the score about here. The FDR would be:

FDR = (Number of REDs **above** threshold)/(Total number of REDs and GREENs **above** Threshold)

Common application:

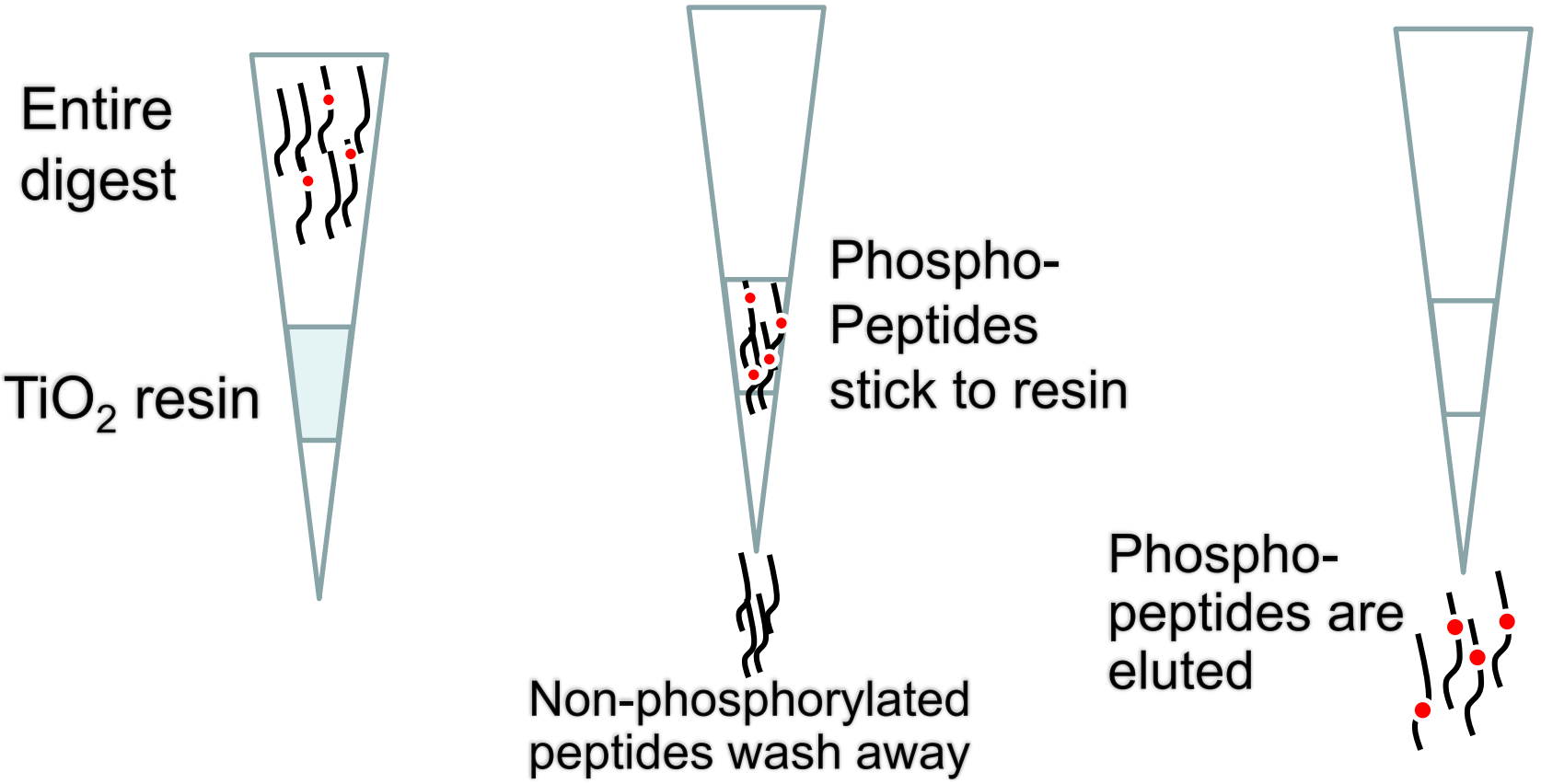
Post-Translational Modifications (PTMs)

PTM detection– MS/MS



Enrichment for phospho-peptides

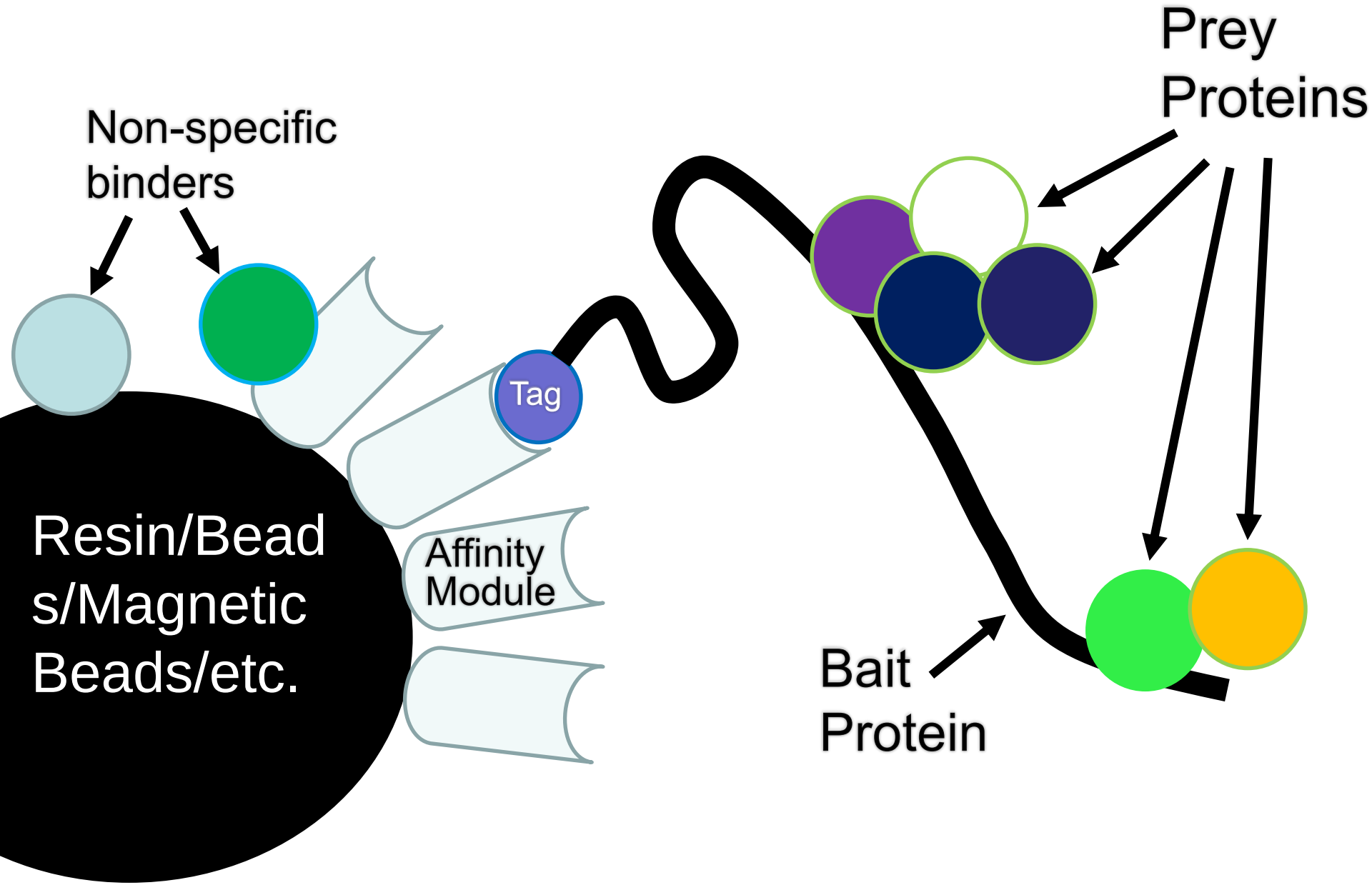
Phospho-peptides have a strong affinity towards TiO_2



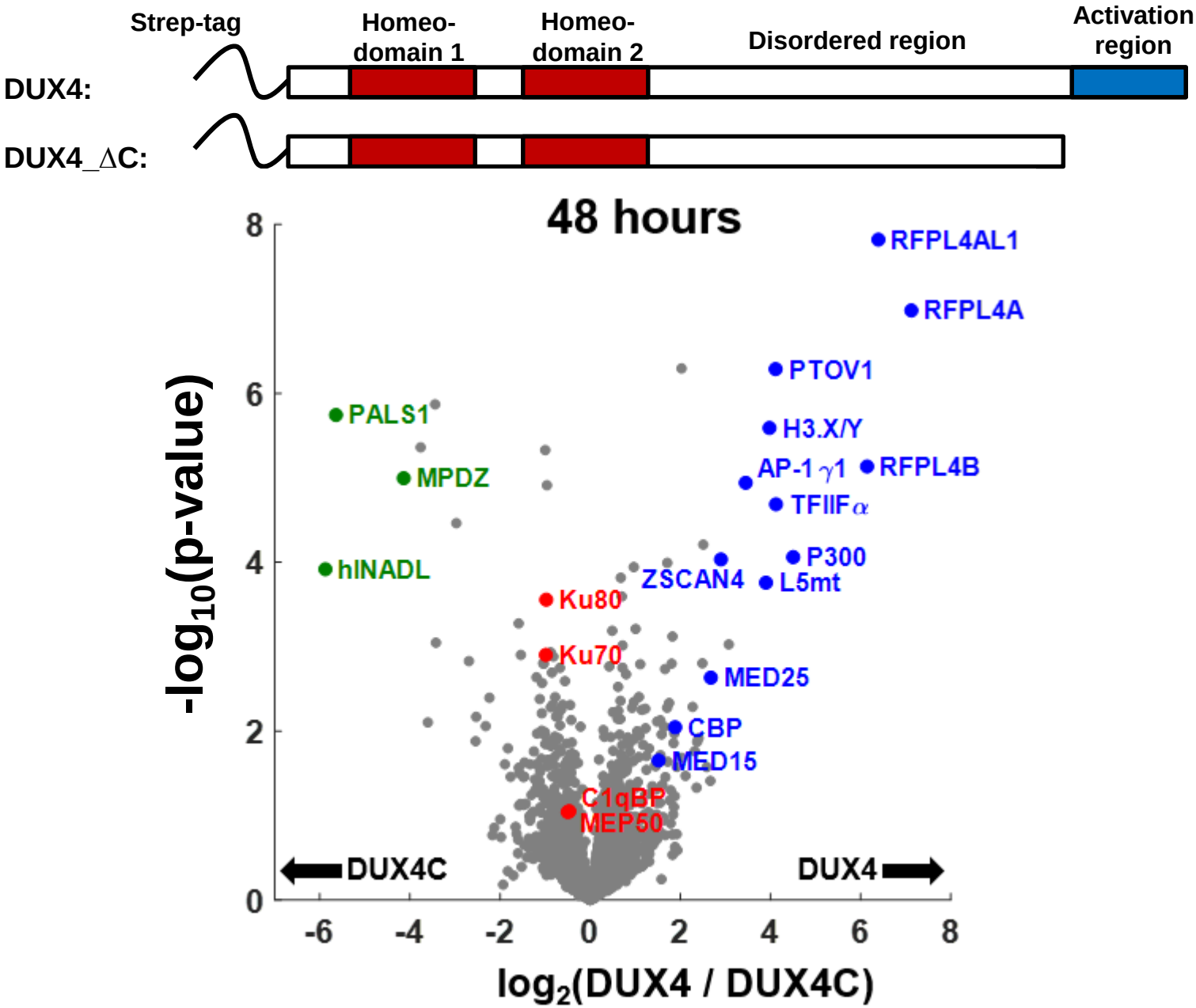
Common application:

Protein-Protein Interactions (PPIs)

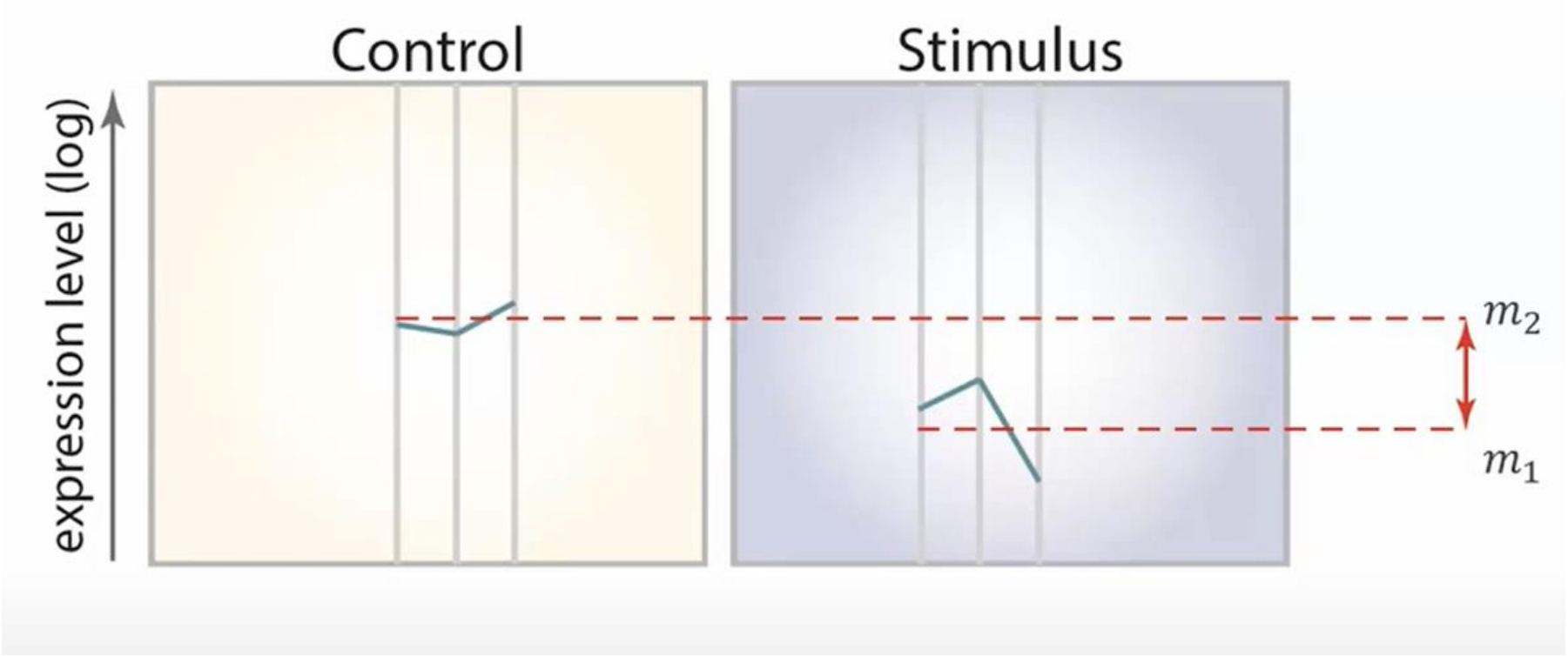
Affinity Purification (AP-MS)



AP-MS of DUX4 C-term

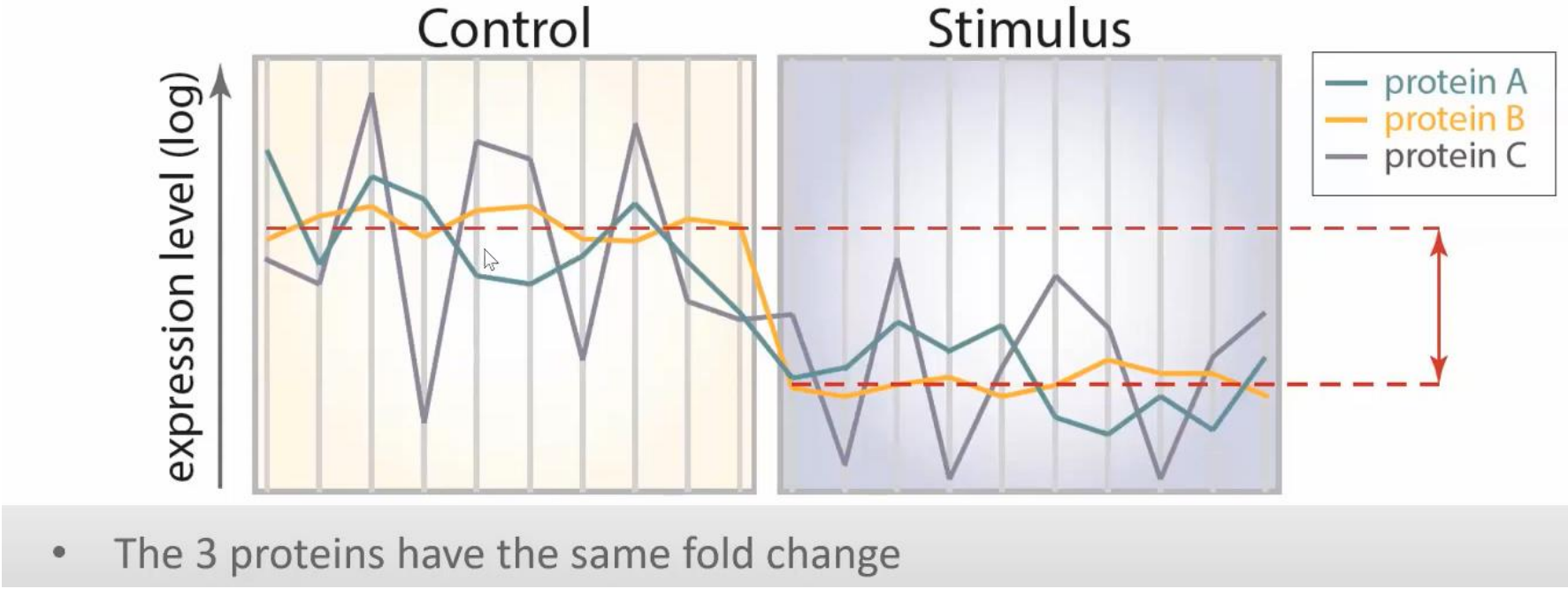


The basic question in quantification



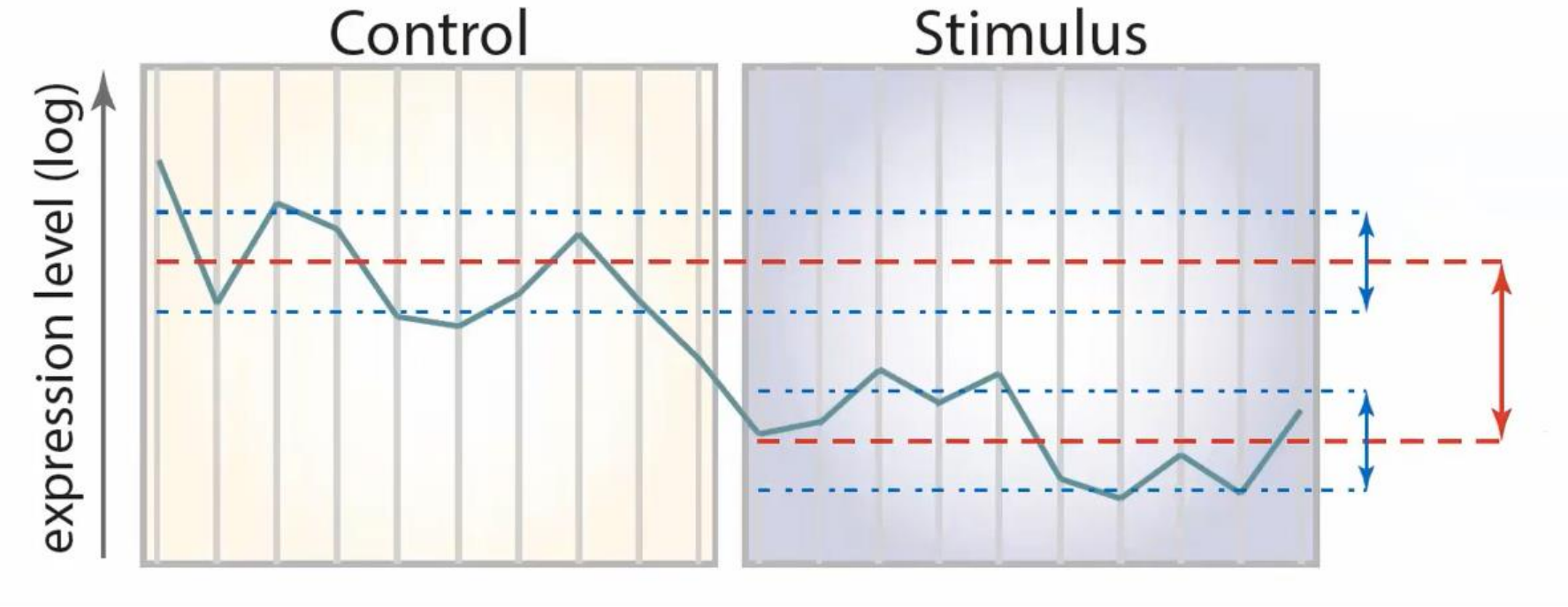
Credit: Juergen Cox

The basic question in quantification (2)



- The 3 proteins have the same fold change

The t-test assign a confidence value to the difference in mean



$$t = \frac{variance\ between\ groups}{variance\ within\ groups}$$

ספקטרומטריית מסות - סיכום
(פרוטיאומיקה מול גנומיקה)

- חקירה של תופעות רלוונטיות לרמת החלבון (ביטוי, מודיפיקציות, אינטרקציות, מבנה, ליגנדים).
- יש לספק DB עם כל הרצפים האפשריים.
- רגישה מאוד – אבל לא ברמת מולקולה בודדת (אי אפשר להגביר חלבון)