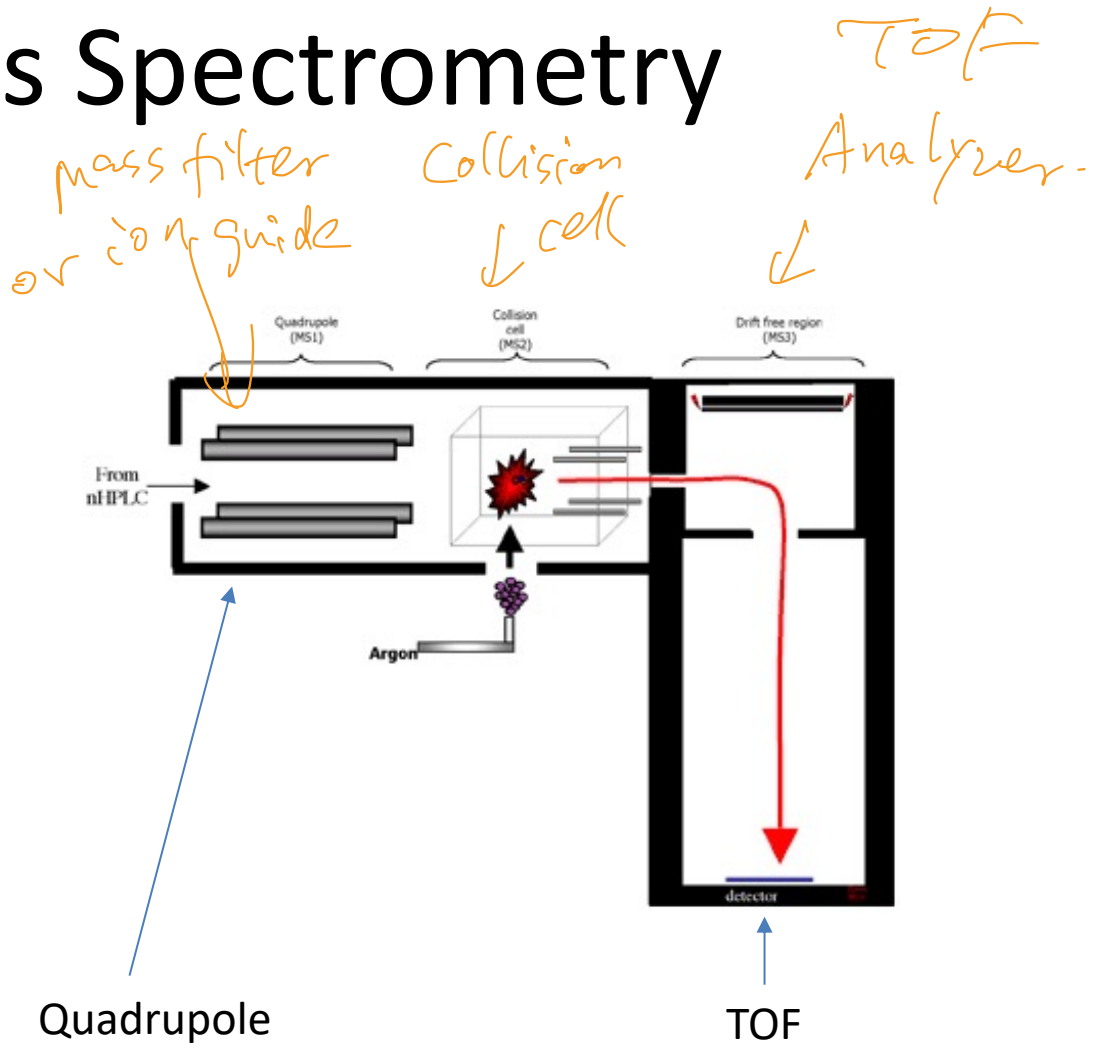
Database Search Method for Peptide Identification with MS/MS

Tandem Mass Spectrometry

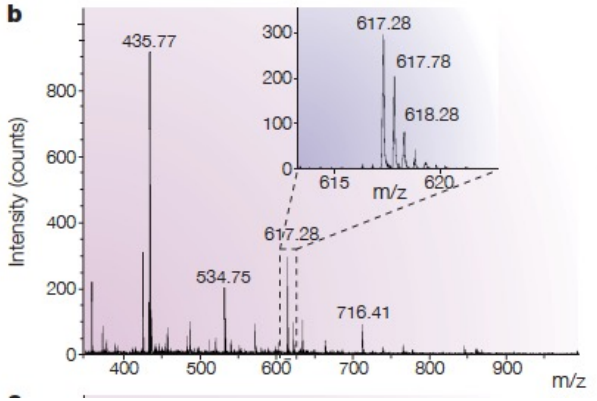
- Tandem MS combines different mass analyzers. E.g. Q-Tof.
- Quadrupole can run in either ion guide or ion filter modes.
- To measure the precursor ions
  - Quadrupole in ion guide mode
  - Collision off
- To measure the fragment ions of a precursor ion
  - Quadrupole to select the target m/z
  - Collision on



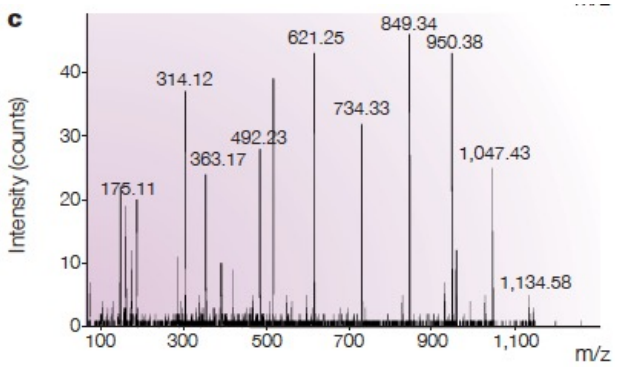
Tandem Mass Spectrometry Procedure

survey

MS



MS/MS

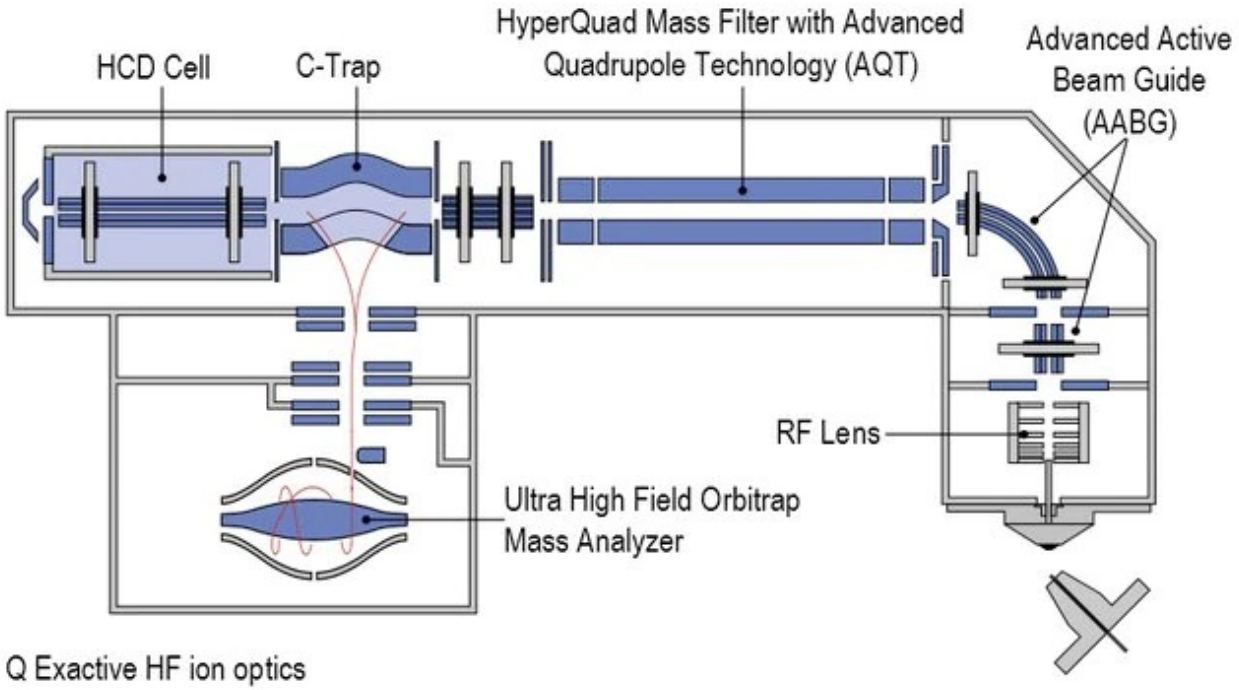


- Step 1. All precursor ions are measured to produce the survey scan.
- Step 2. A precursor ion is selected (by m/z) and fragmented. All fragment ions are measured to produce the tandem MS scan (also called as MS/MS or MS2 scan).
- Repeat Step 2 a few times. Then go back to Step 1.

DDA

Note: For each MS2 spectrum, we additional know the precursor m/z.

Orbitrap QE-HF

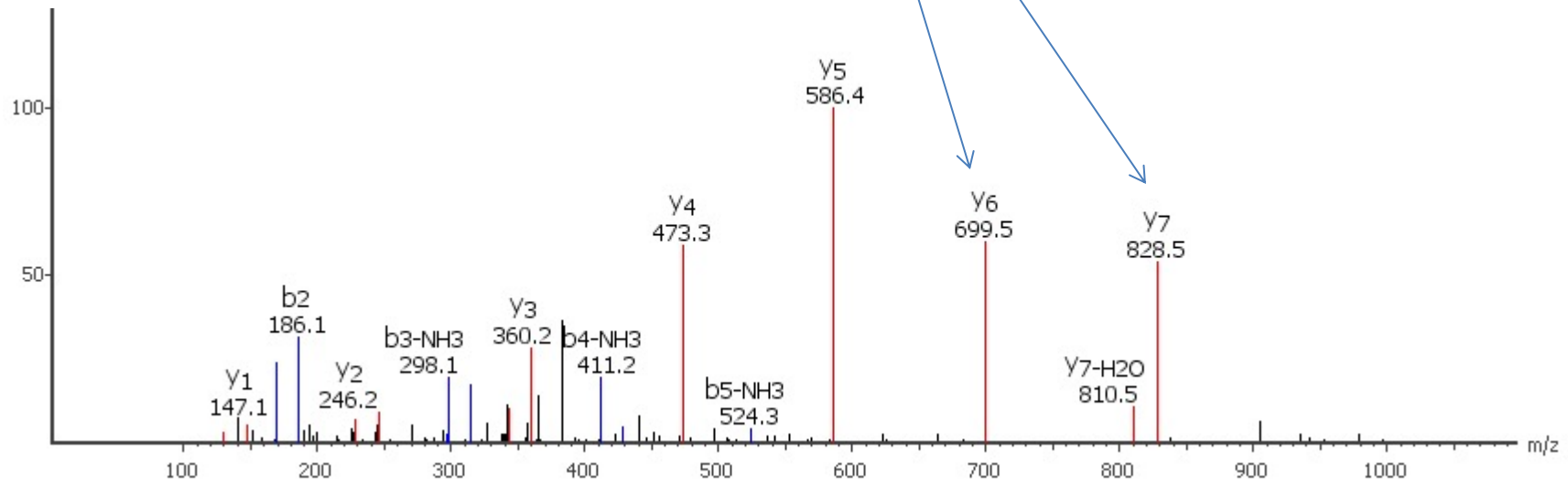


An actual instrument may consist many components for better sensitivity, accuracy, throughput and robustness. The figure illustrate the main components of an Orbitrap QE-HF instrument made by ThermoFisher Scientific.

Peptide-Spectrum Match

b_1	**A NELLLNVK**	y_8
b_2	**AN ELLLNVK**	y_7
b_3	**ANE LLLNVK**	y_6
b_4	**ANEL LLNVK**	y_5
b_5	**ANELL LNVK**	y_4
b_6	**ANELLL NVK**	y_3
b_7	**ANELLLN VK**	y_2
b_8	**ANELLLNV K**	y_1

828.499 (assuming 1+ charge)



y-ion m/z = (total of amino acid residue mass + 18.011 + z * 1.007) / z

b-ion m/z = (total of amino acid residue mass + z * 1.007) / z

H2O

proton

charge state

Mass Error Tolerance

- Peak matching allows certain mass error tolerance (due to instrument measurement errors).
- Error can be specified either in Da or in ppm (part-per-million).
- ppm error = 1e6 * (observed mass – theoretical mass)/theoretical mass
- Different instrument has different error tolerance:
  - Low resolution: often 0.5-1 Da
  - High resolution: often 1-20 ppm
- Precursor ions and fragment ions often have slightly different error tolerances.

Search Through Database

>sp|P02769|ALBU_BOVIN Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHRFKDLGEEHFKGLVLIAFSQYLQQCPFDEH
VKLVNELTEFAKTCVADESHAGCEKSLHTLFGDELCKVASLRETYGDMADCCEKQEPERNEC
FLSHKDDSPDLPKLKPDPNTLCDEFKADEKKFWGKYLYEIARRHPYFYAPELLYYANKYNGVF
QECCQAEDKGACLLPKIETMREKVLASSARQRLRCASIQKFGERALKAWSVARLSQKFPKA
EFVEVTKLVTDLTKVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKECCDKPLLEKSHC
IAEVEKDAIPENLPPLTADFAEDKDVCKNYQEAKDAFLGSFLYEYSRRHPEYAVSVLLRLAKEY
EATLEECCAKDDPHACYSTVFDKLKHLVDEPQNLIKQNCDQFEKLGEYGFQNALIVRYTRKV
PQVSTPTLVEVSRSLGKVGTRCCTKPESERMPCTEDYLSLILNRLCVLHEKTPVSEKVTKCCTE
SLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLPDTEKQIKKQTALVELLKHKPKATEEQL
KTVMENFVAFVDKCCAADDKEACFAVEGPKLVVSTQTALA

1. Use the enzyme digestion rule to cut each protein into peptides

MK|WVTFISLLLLFSSAYSR|GVFR|R|DTHK|SEIAHR|FK|DLGEEHFK|GLVLIAFSQYLQQCPFDEHVK|L

2. For each peptide, compare with the spectrum to see how well they match.

An Empirical Score

- y-ion m/z at charge one = total residue mass + 19.0178.
- Find approximate matching peak. Assume relative intensity = x.
- Relative intensity = current_peak_intensity / max_peak_intensity.
- Score contribution = $\max \begin{cases} \log_{10} 100 \cdot x \,, if\ x > 0.01 \\ 0, otherwise \end{cases}$
- Add up all score contributions of all y-ions.
- Better score functions will be discussed later.

Database Search

- Input:
  - A list of MS/MS spectra
  - A protein sequence database
- Algorithm 1:

For each MS/MS spectrum

For each protein in the database

In-silico digest the protein into peptides

For each peptide

Evaluate the peptide-spectrum match

Assign the highest-scoring peptide to the spectrum