

Metabolomics

Johannes Rainer

Eurac Research, Bolzano, Italy

johannes.rainer@eurac.edu - github/twitter: jotsetung

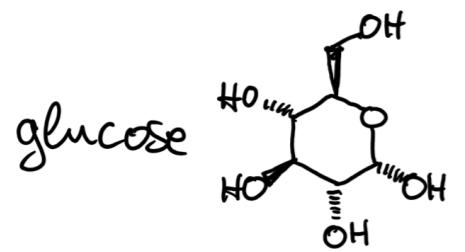
12 July 2018

Content

- Introduction to metabolomics
- Preprocessing of LC-MS data in Bioconductor
- Normalization
- Annotation/identification

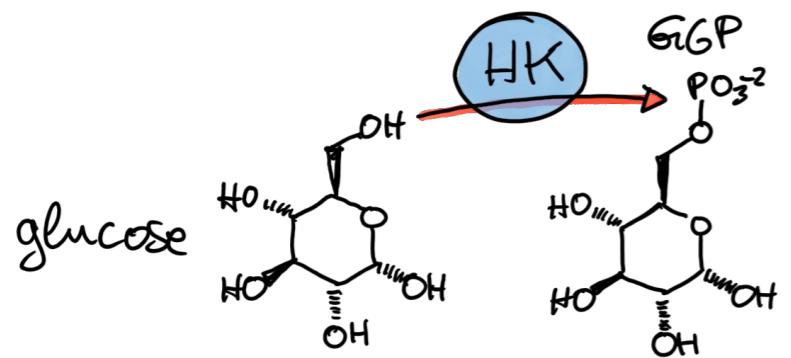
Metabolite? Metabolism?

- Glycolysis



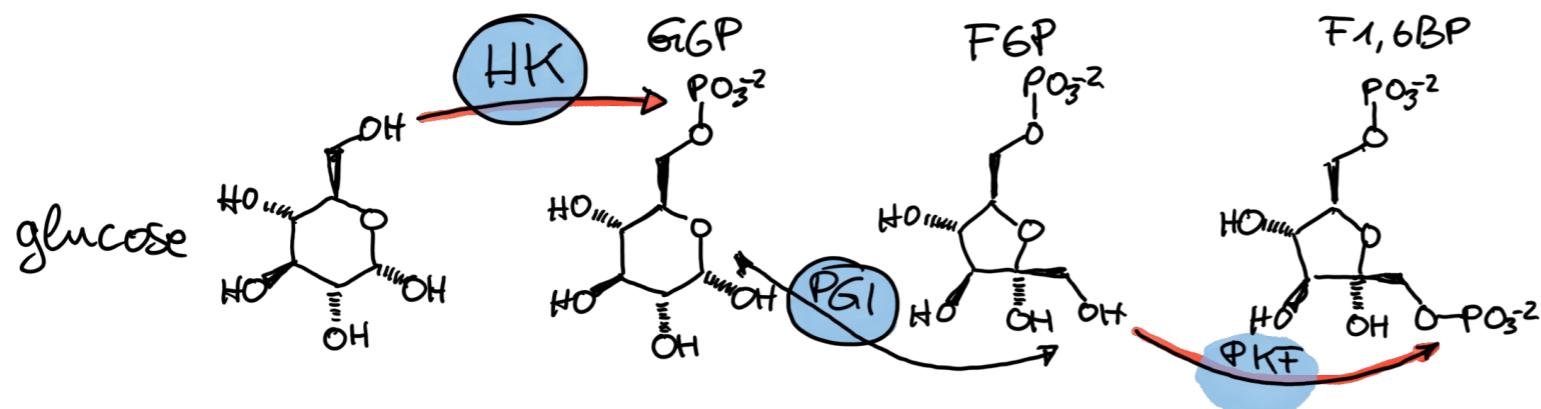
Metabolite? Metabolism?

- Glycolysis



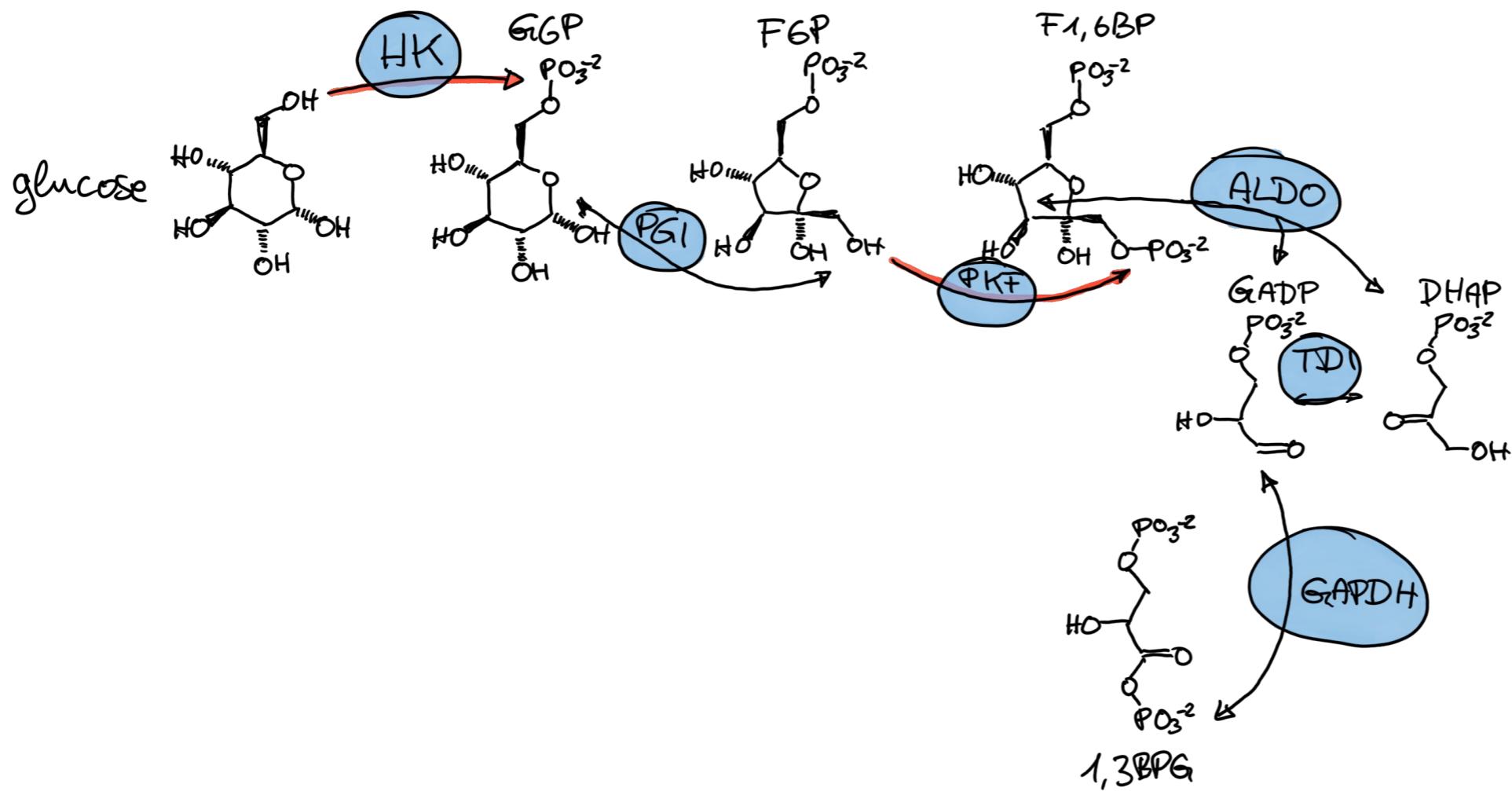
Metabolite? Metabolism?

- Glycolysis



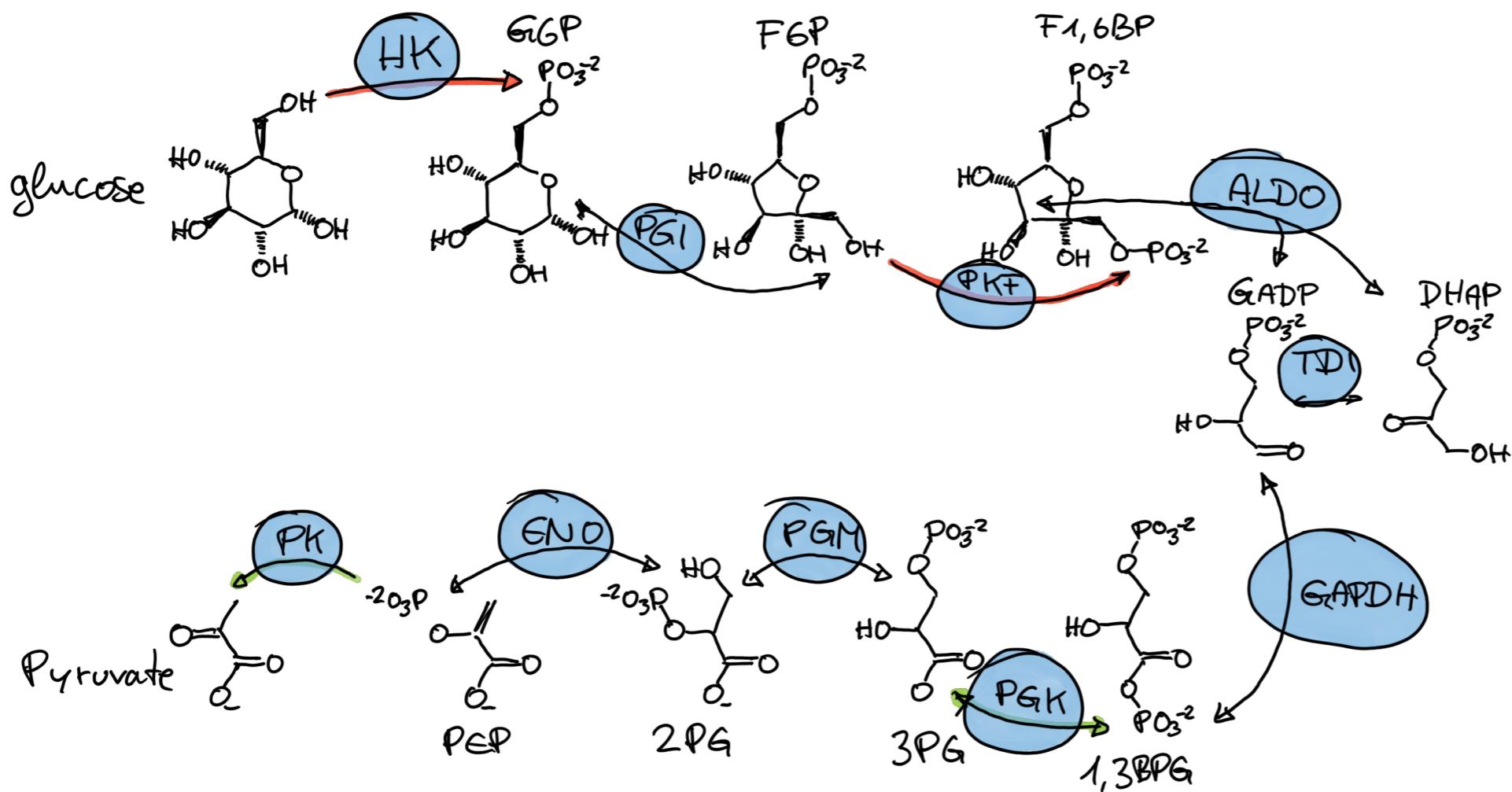
Metabolite? Metabolism?

- Glycolysis



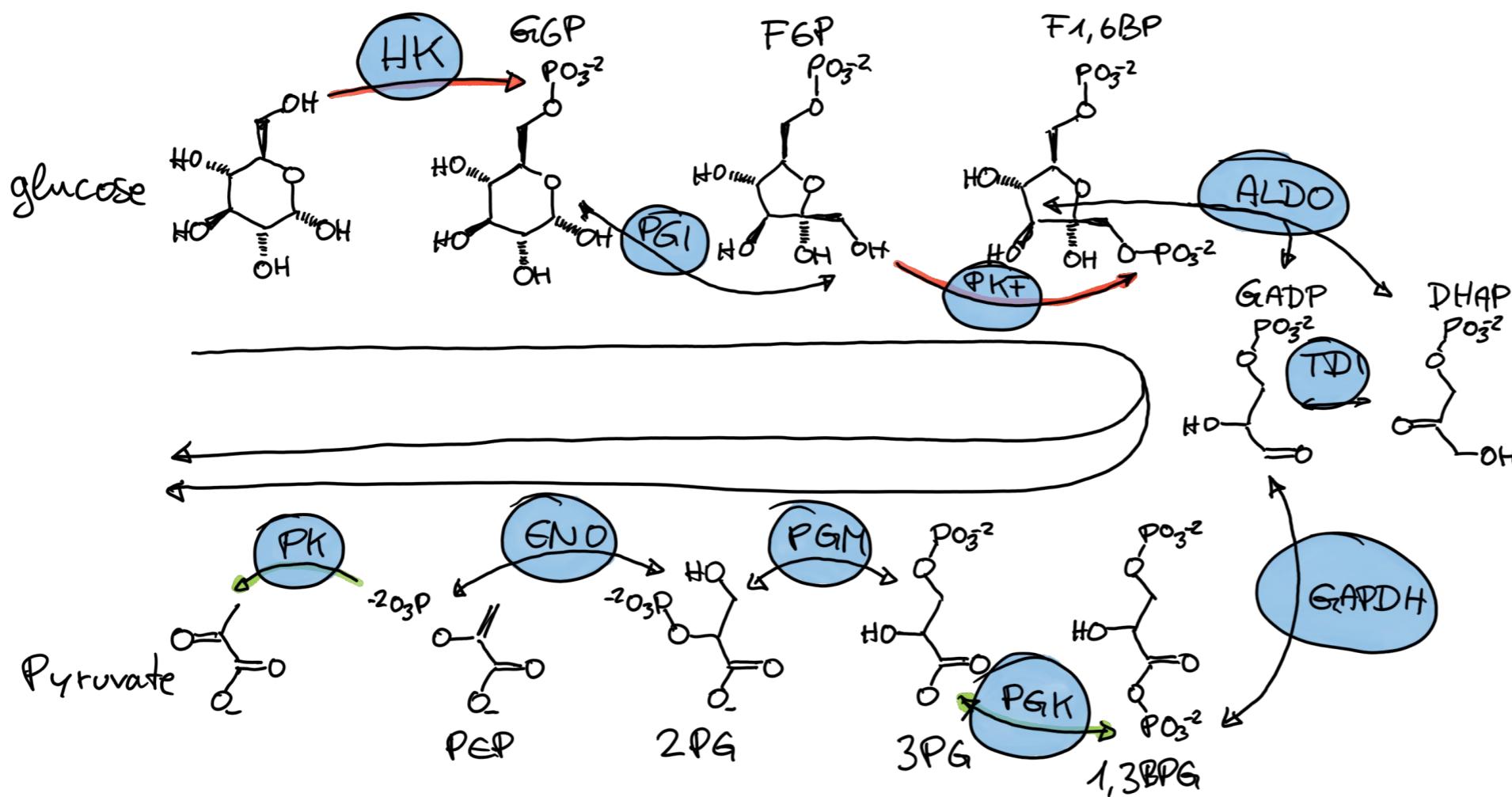
Metabolite? Metabolism?

- Glycolysis



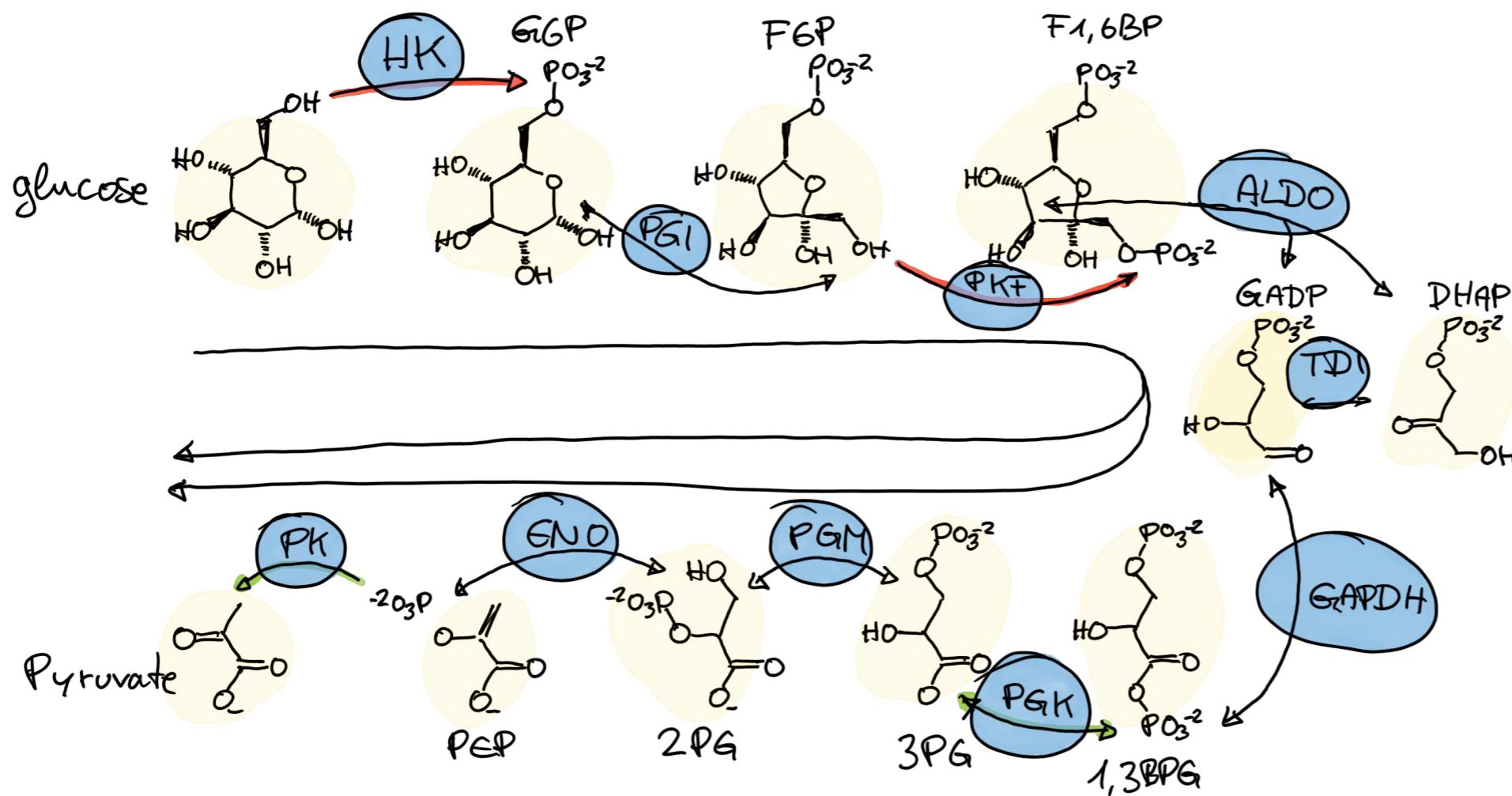
Metabolite? Metabolism?

- Glycolysis



Metabolite? Metabolism?

- Glycolysis



- Metabolites: intermediates and products of cellular processes.

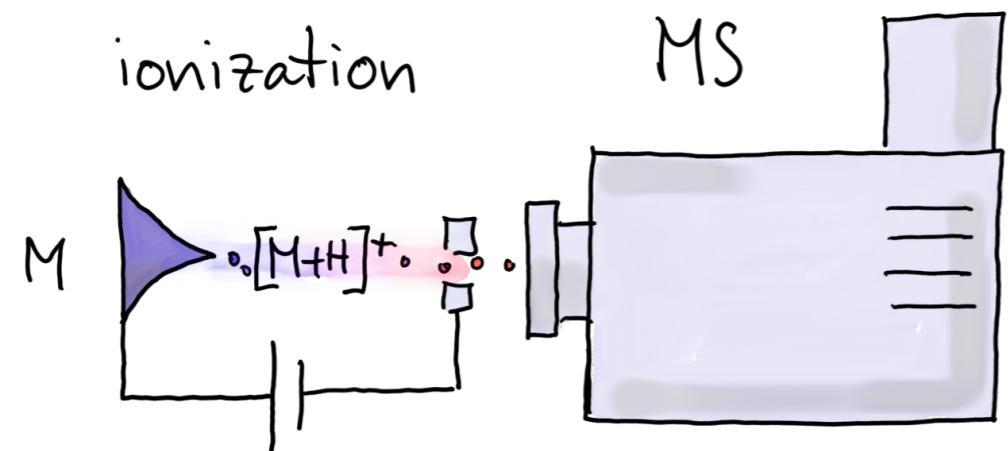
Metabolomics?

- Large-scale study of small molecules (metabolites) in a system (cell, tissue, organism).
- Comparison of the different -omes:
- **Genome**: what can happen.
- **Transcriptome**: what appears to be happening.
- **Proteome**: what makes it happen.
- **Metabolome**: what actually happened.
- Metabolome influenced by genetic **and** environmental factors.

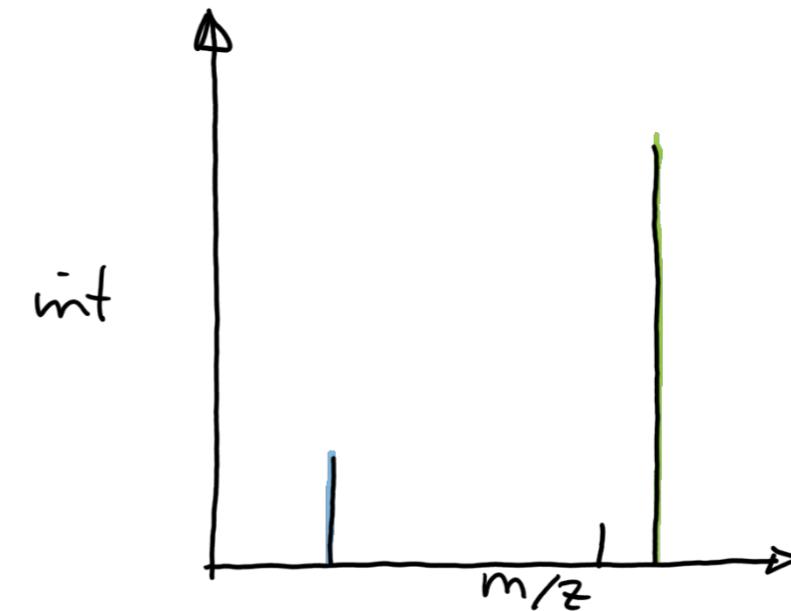
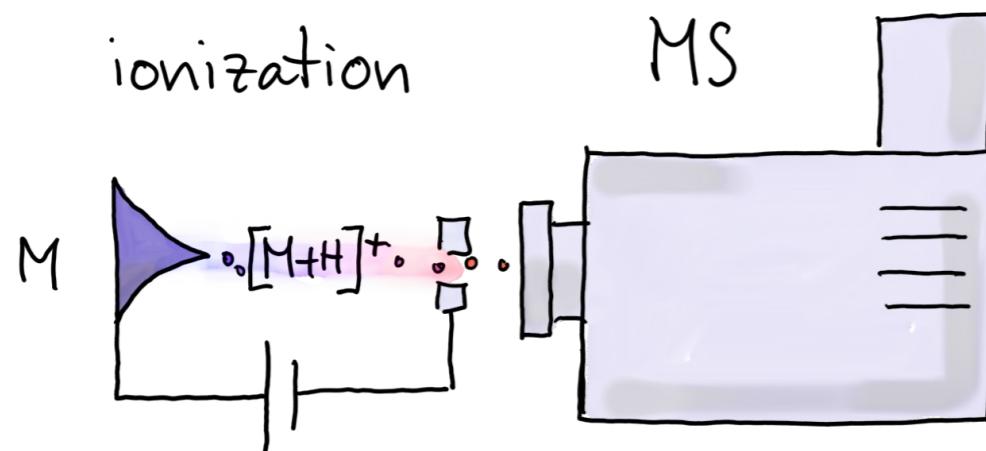
How can we measure metabolites?

- Nuclear Magnetic Resonance (NMR) - not covered here.
- Mass spectrometry (MS)-based metabolomics.
- Targeted/untargeted metabolomics:
 - **targeted**: quantitative measurement of selected metabolites.
 - **untargeted**: semi-quantitative measurement of all metabolites (detectable with the setup) in a sample.

Mass Spectrometry (MS)



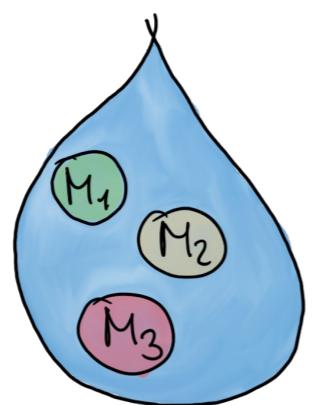
Mass Spectrometry (MS)



- **Problem:** unable to distinguish between metabolites with the same mass-to-charge ratio (m/z).
- **Solution:** separate metabolites prior to MS by another property.

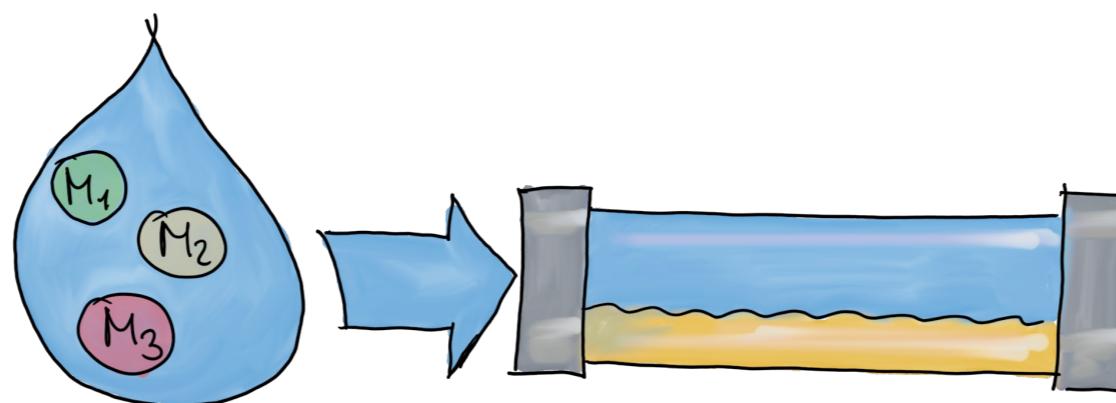
Liquid chromatography

- Sample is dissolved in a fluid (mobile phase).



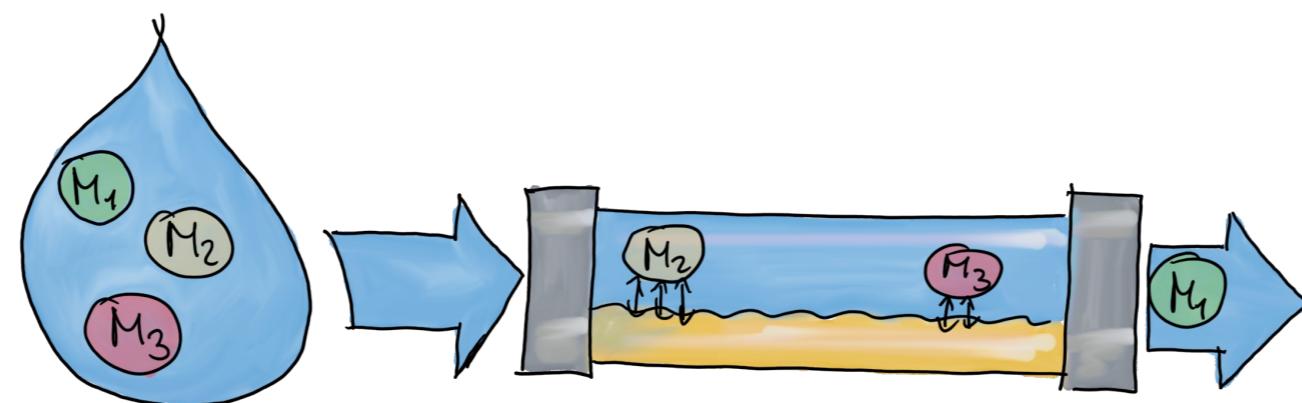
Liquid chromatography

- Sample is dissolved in a fluid (mobile phase).
- Mobile phase carries analytes through a column with a stationary phase.



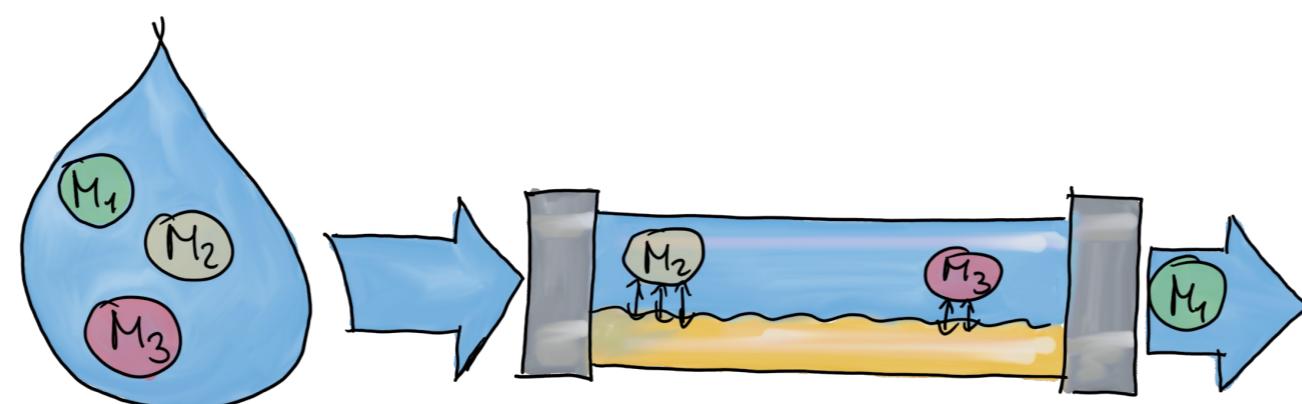
Liquid chromatography

- Sample is dissolved in a fluid (mobile phase).
- Mobile phase carries analytes through a column with a stationary phase.
- Separation based on affinity for the column's stationary phase.

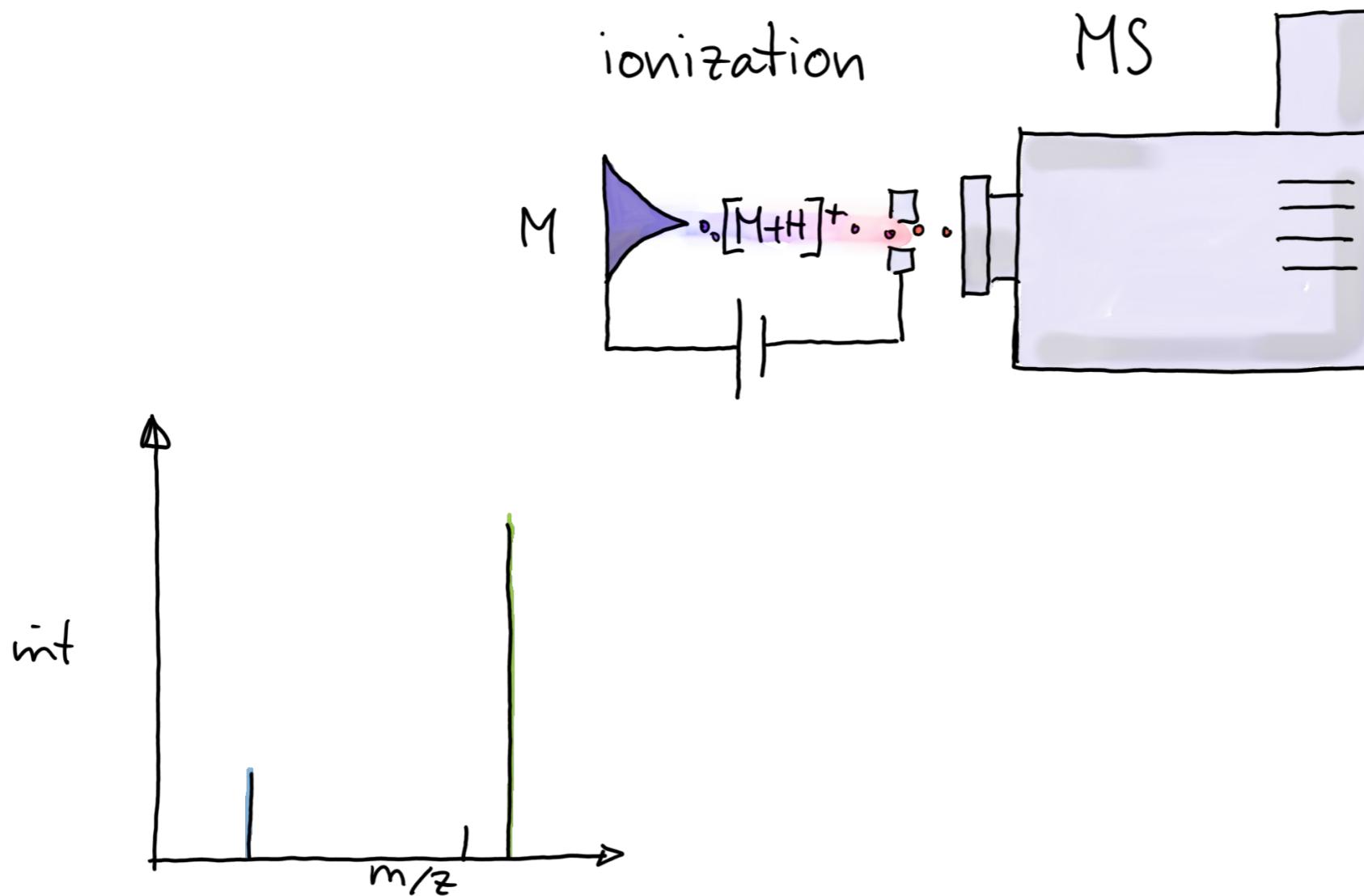


Liquid chromatography

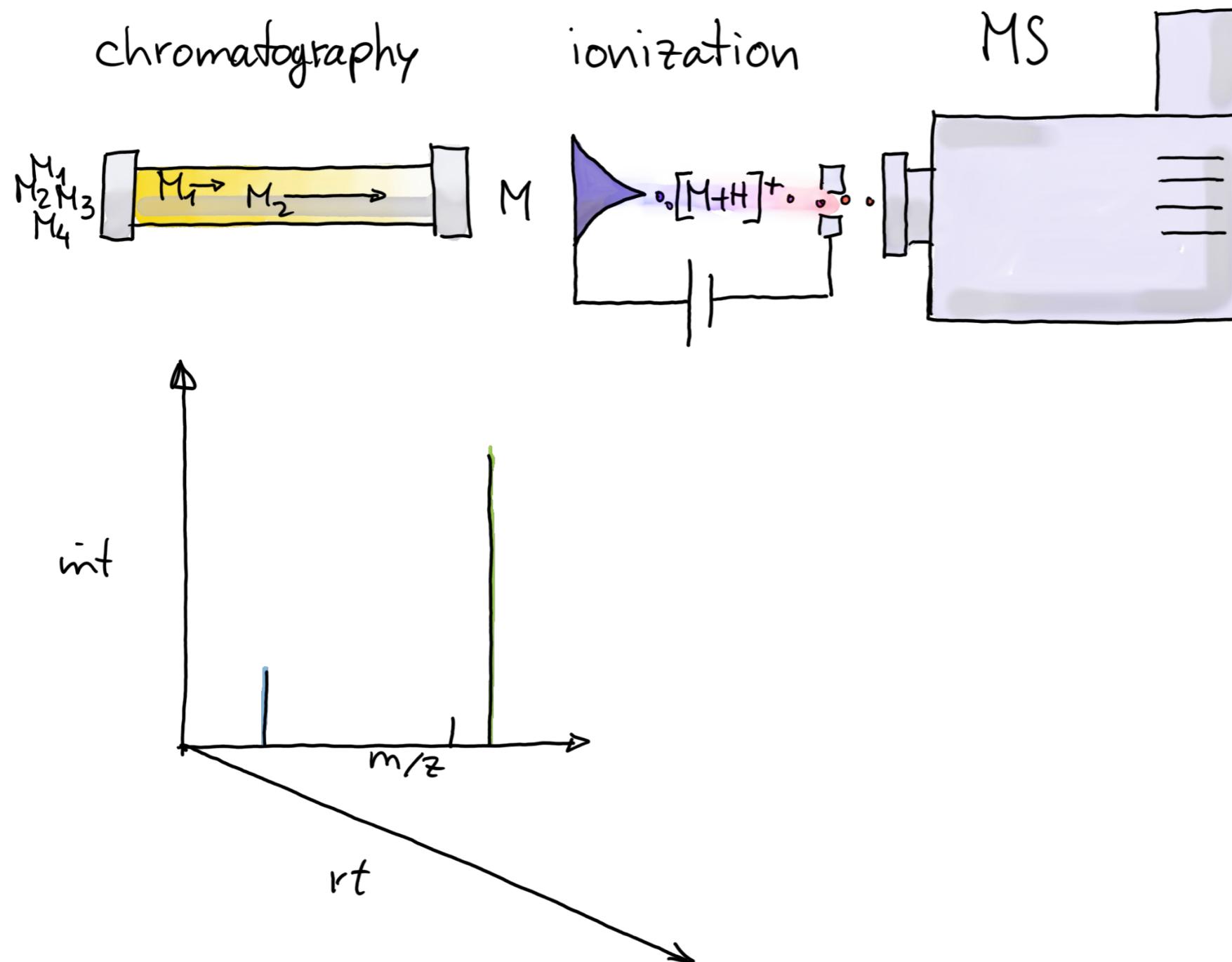
- Sample is dissolved in a fluid (mobile phase).
- Mobile phase carries analytes through a column with a stationary phase.
- Separation based on affinity for the column's stationary phase.
- HILIC (hyrophilic liquid interaction chromatography):
 - Hydrophilic, polar stationary phase.
 - Analytes solved in mobile phase.
 - Analytes separated by polarity: low polarity elute first, high polarity later.



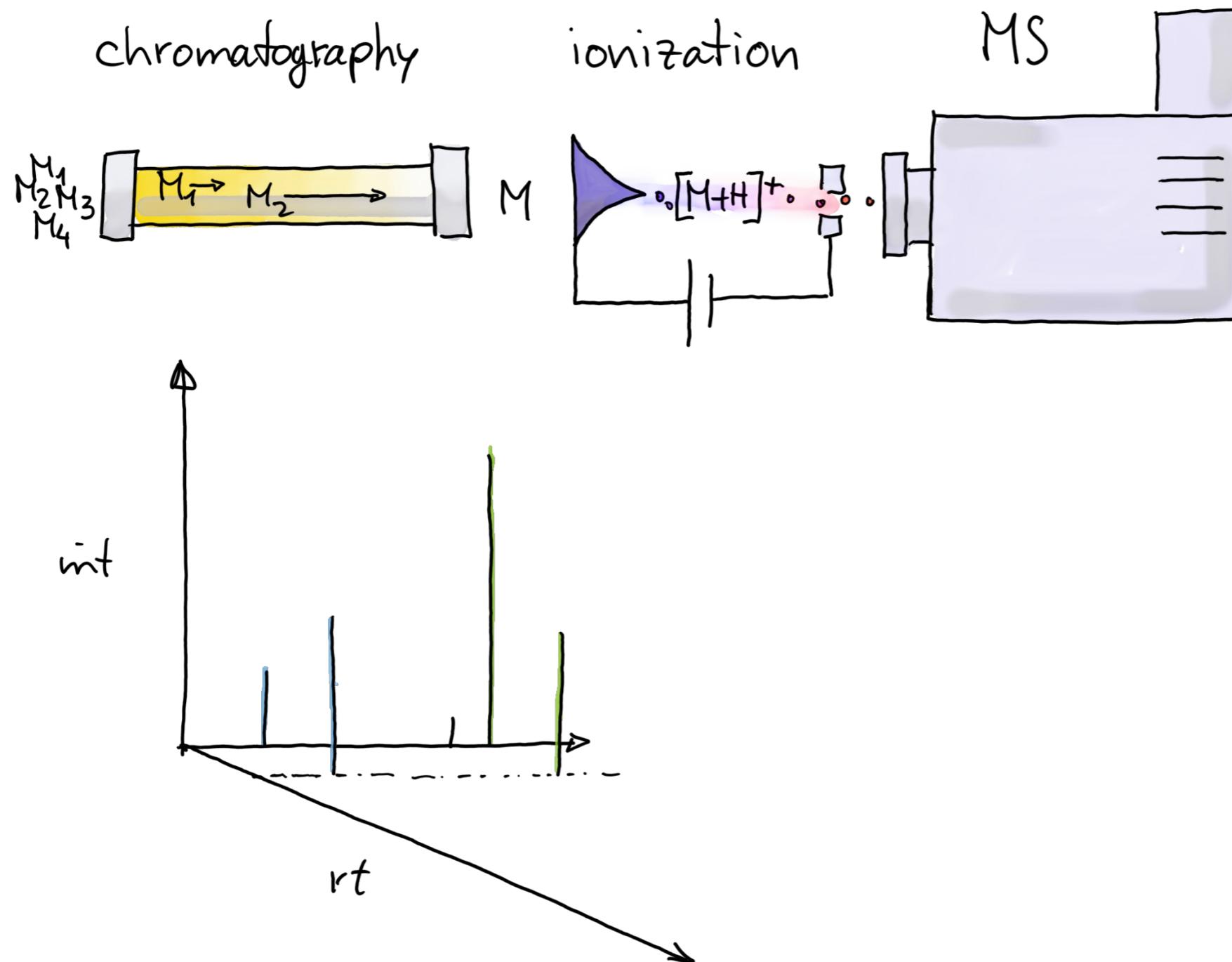
Liquid Chromatography Mass Spectrometry (LC-MS)



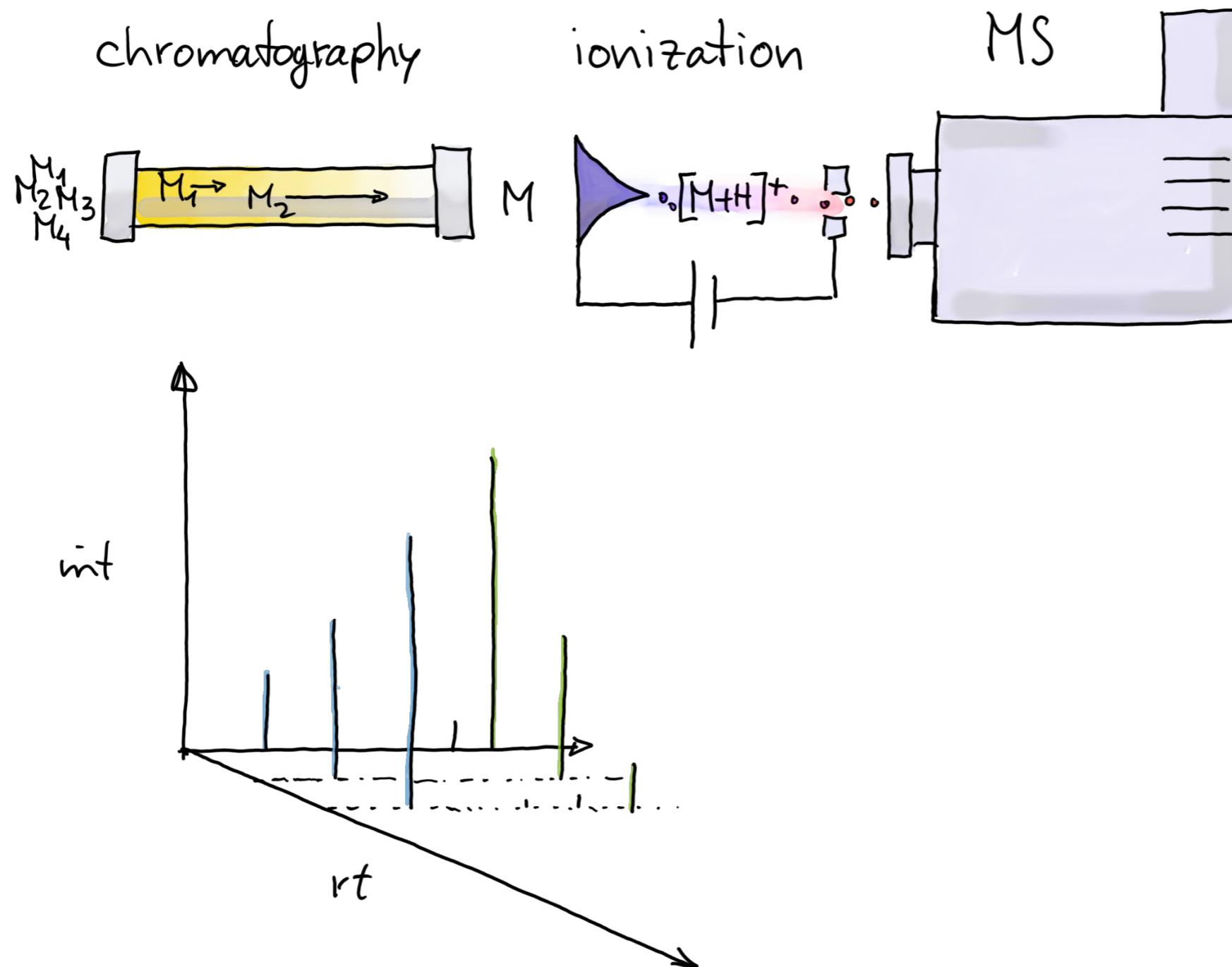
Liquid Chromatography Mass Spectrometry (LC-MS)



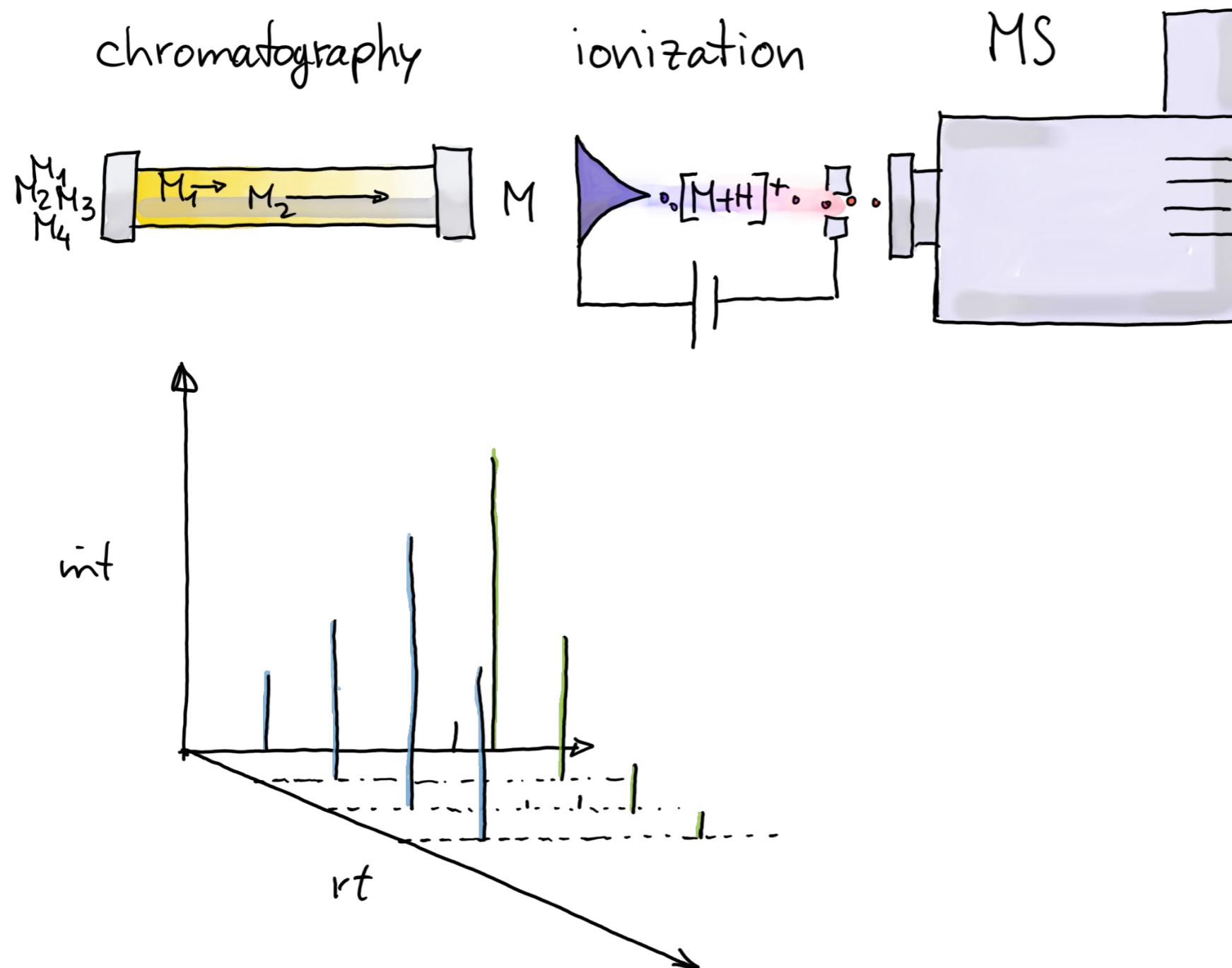
Liquid Chromatography Mass Spectrometry (LC-MS)



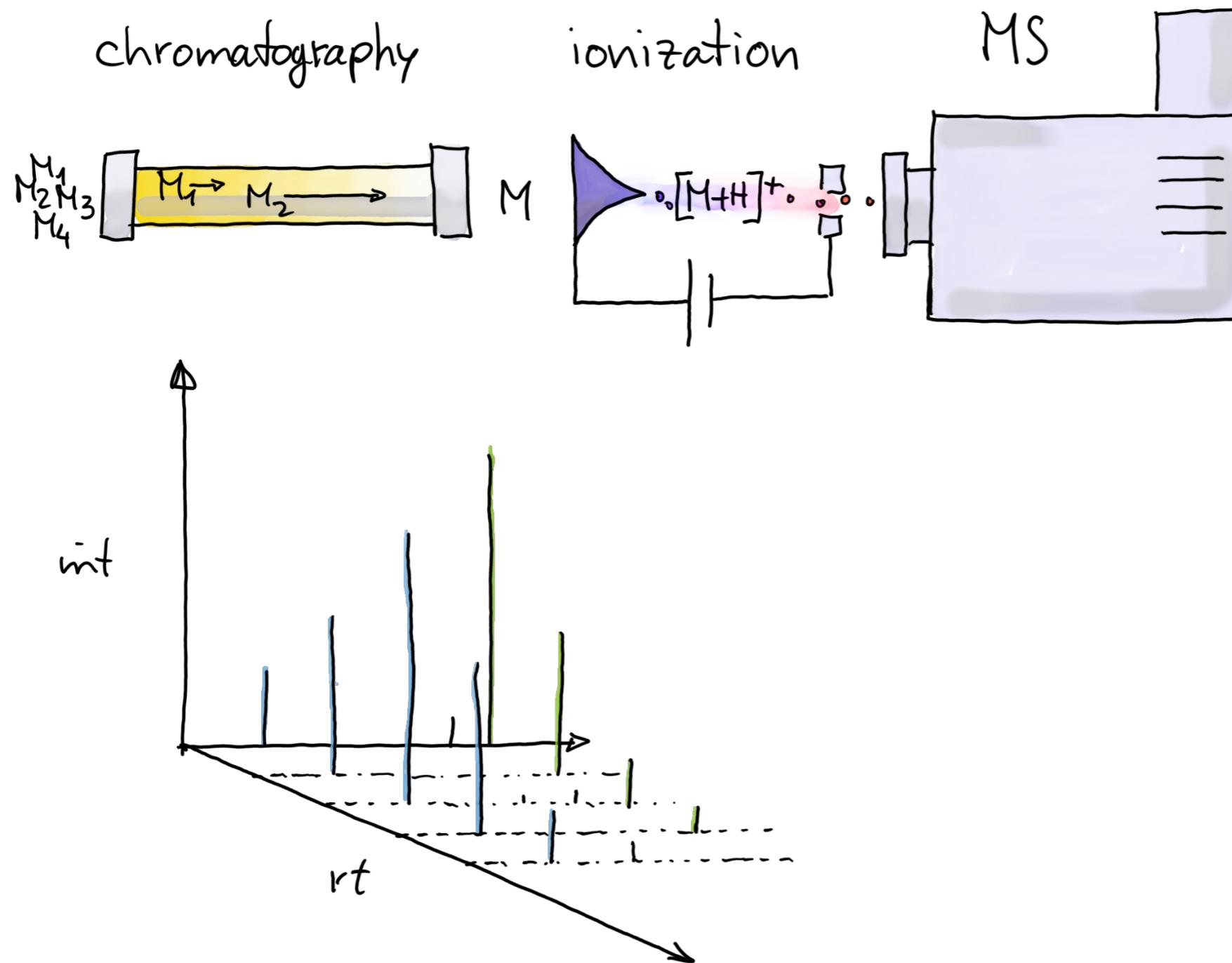
Liquid Chromatography Mass Spectrometry (LC-MS)



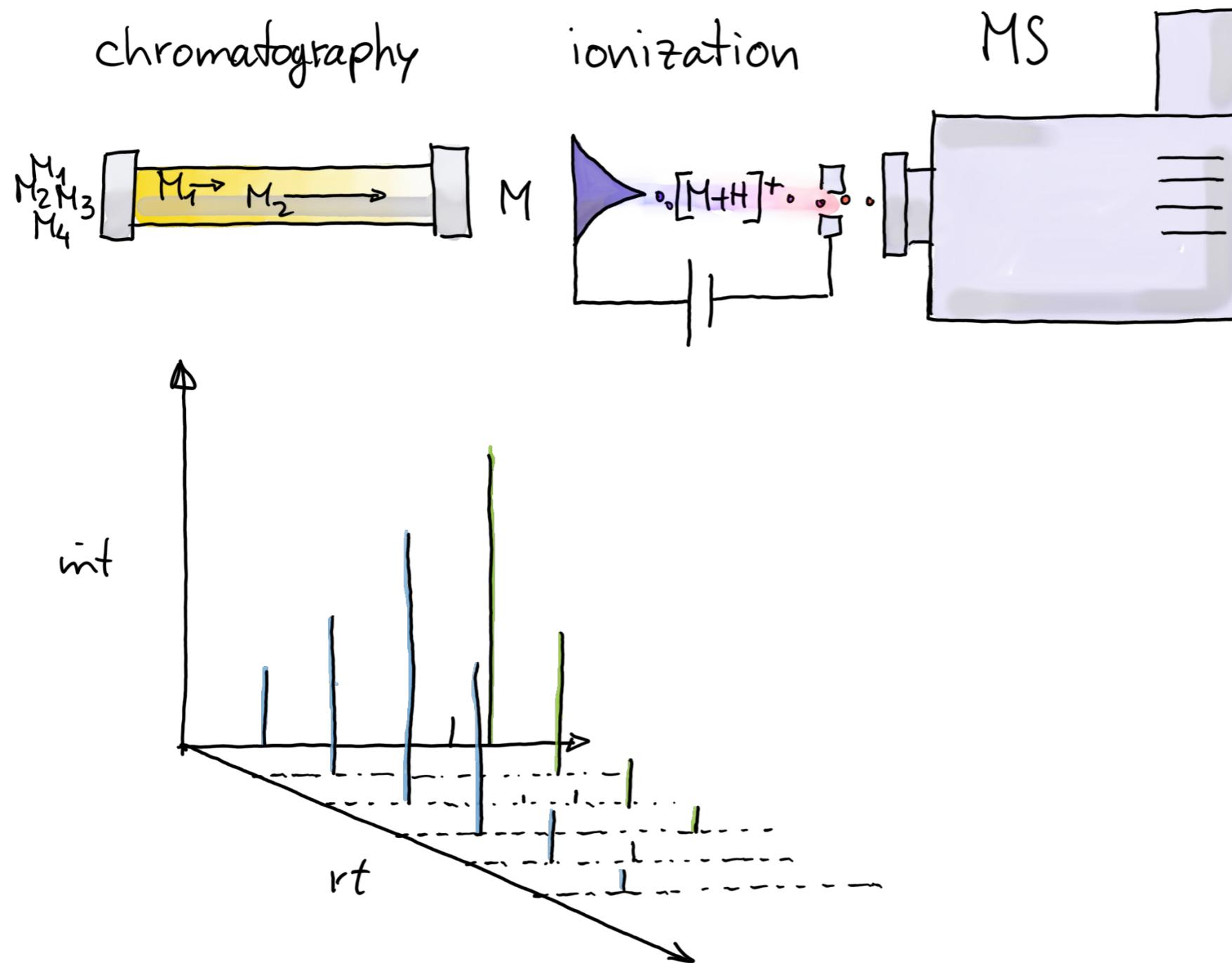
Liquid Chromatography Mass Spectrometry (LC-MS)



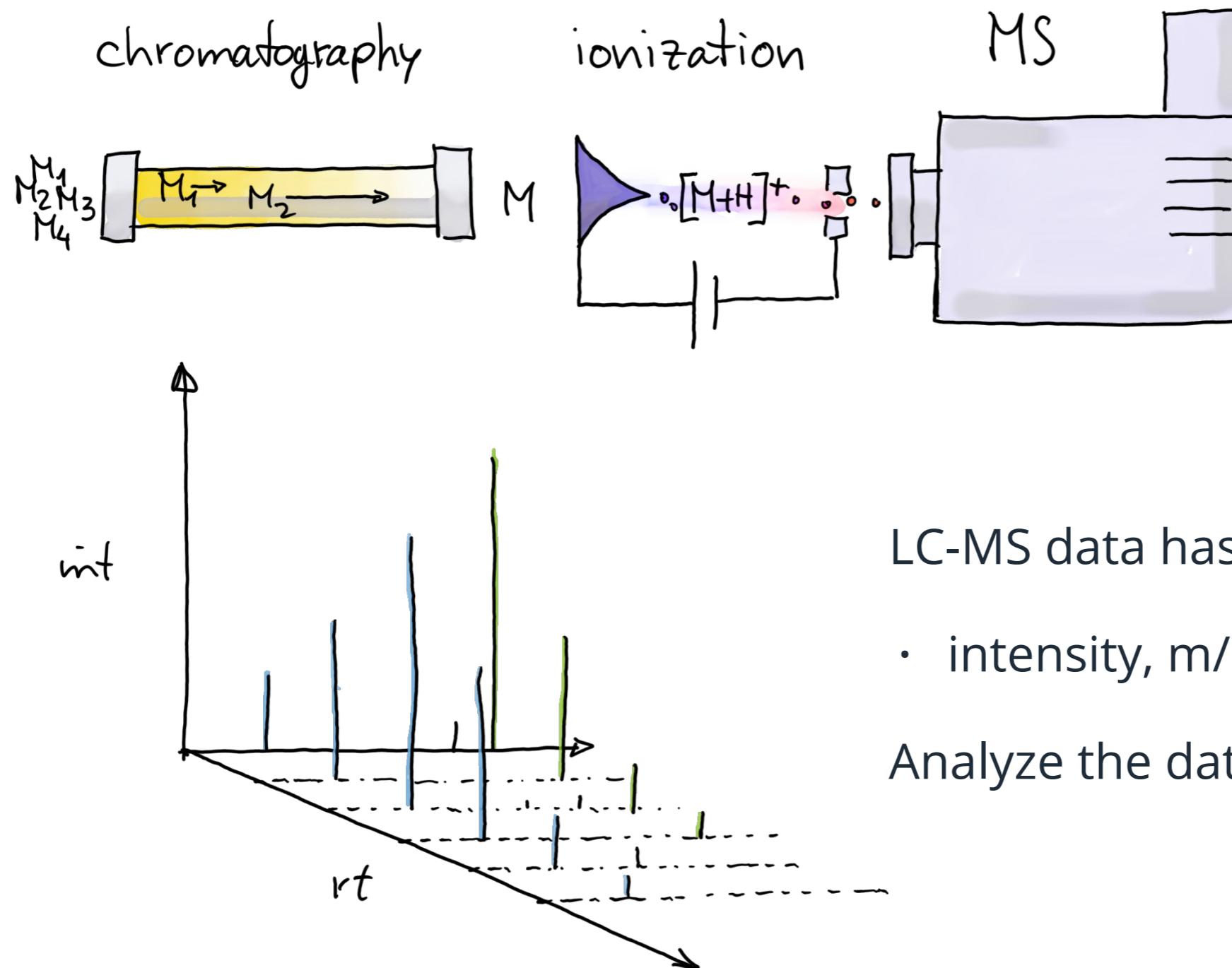
Liquid Chromatography Mass Spectrometry (LC-MS)



Liquid Chromatography Mass Spectrometry (LC-MS)



Liquid Chromatography Mass Spectrometry (LC-MS)



LC-MS data has thus 3 dimensions:

- intensity, m/z , retention time.

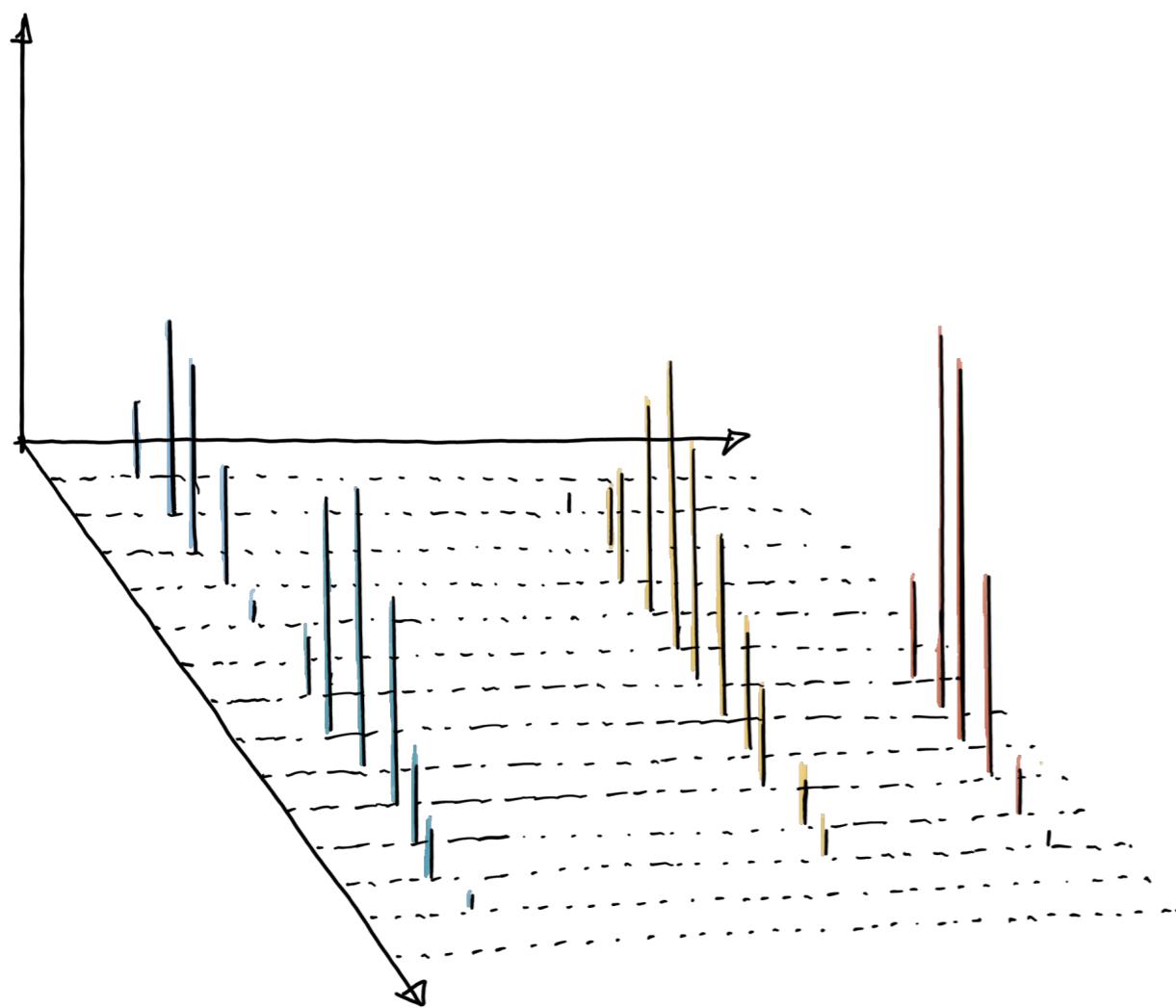
Analyze the data along retention time.

LC-MS data preprocessing

- Chromatographic peak detection
- Alignment
- Correspondence

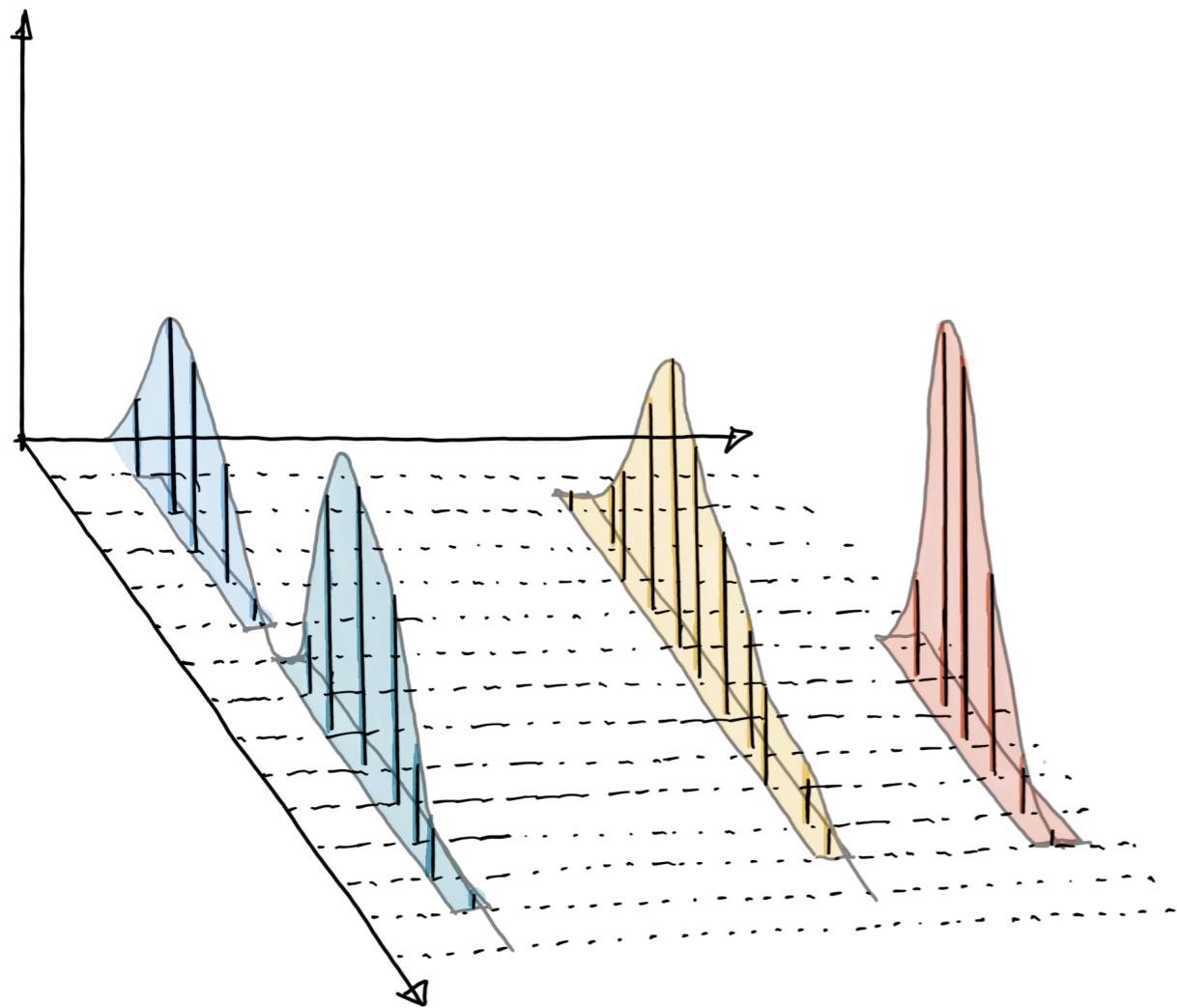
Chromatographic peak detection

- Aim: identify chromatographic peaks in the data.



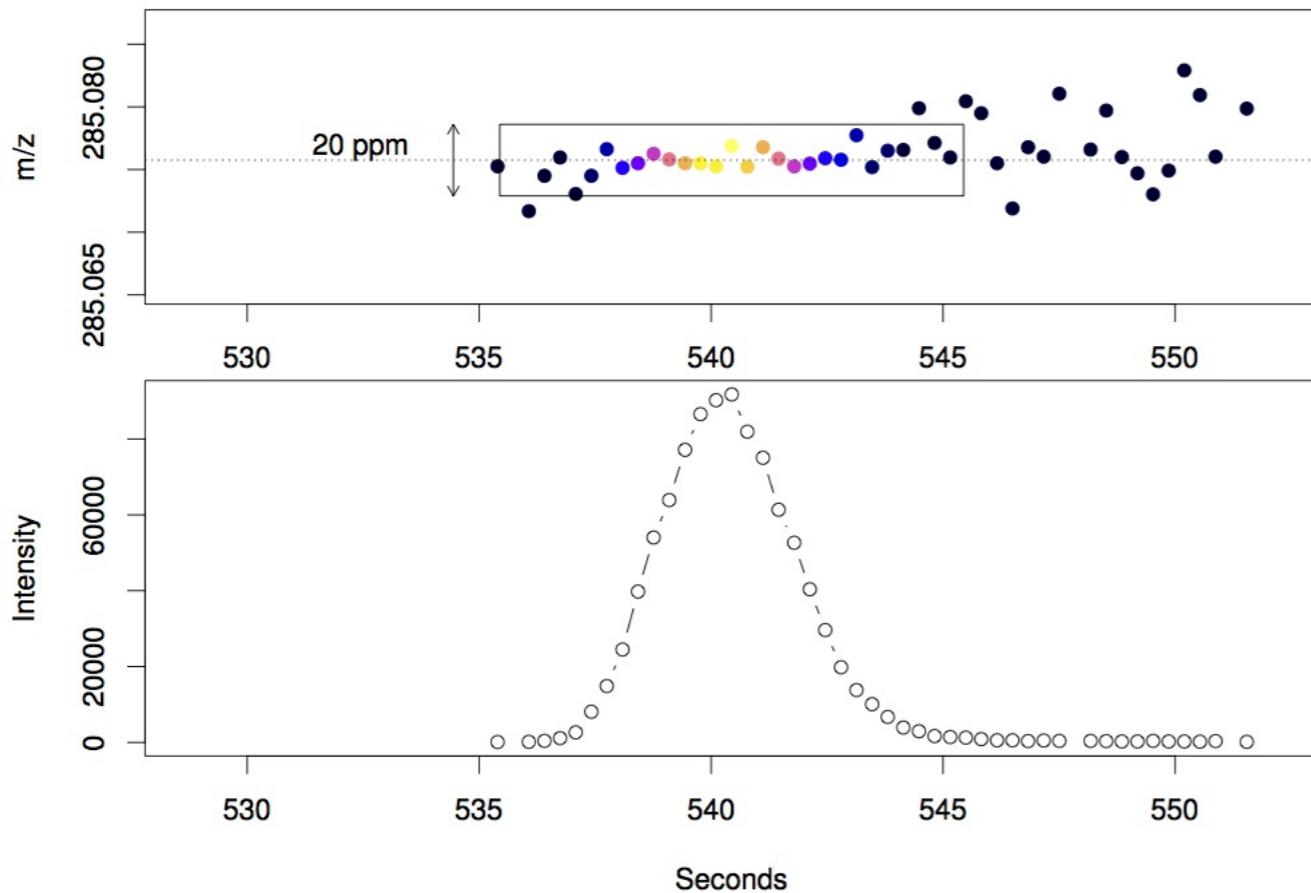
Chromatographic peak detection

- Aim: identify chromatographic peaks in the data.



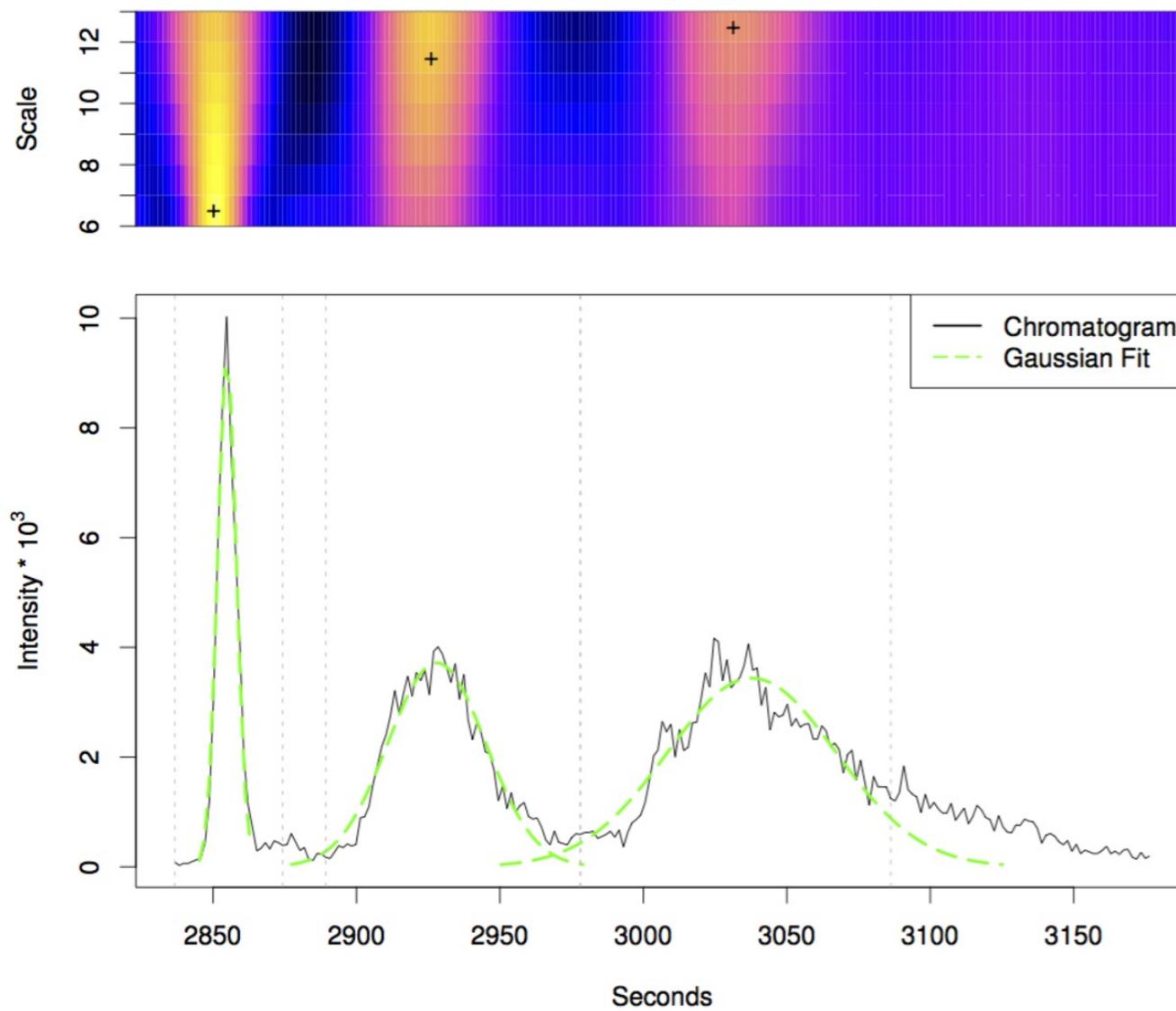
Chromatographic peak detection

- **centWave** [Tautenhahn et al. BMC Bioinformatics, 2008]:
- Step 1: identify regions of interest.



Chromatographic peak detection

- Step 2: peak detection using continuous wavelet transform.
- Allows detection of peaks with different widths.



Chromatographic peak detection

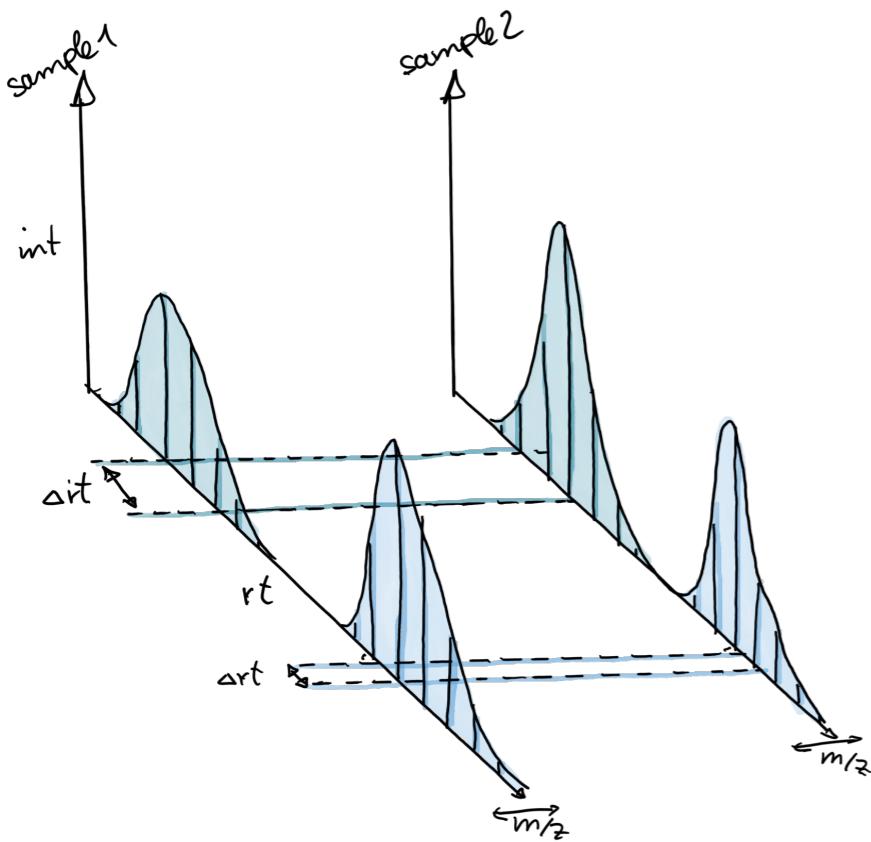
- After reading the data with `readMSData` (`MSnbase` package):
- `xcms`: `findChromPeaks` function, passing settings along with an algorithm-specific parameter object.

```
cwp <- CentWaveParam(peakwidth = c(2, 10), snthresh = 5)
data <- findChromPeaks(data, param = cwp)
head(chromPeaks(data), n = 3)
```

```
##          mz    mzmin    mzmax      rt    rtmin    rtmax     into     intb
## CP001 114.0907 114.0899 114.0929  1.954  0.280  3.907 1559.829 1555.923
## CP002 114.0913 114.0884 114.0929  5.860  4.465  8.650 1890.221 1885.757
## CP003 114.0914 114.0899 114.0929 10.882  8.650 13.114 1950.953 1946.210
##          maxo    sn sample is_filled
## CP001 584.9510 584       1       0
## CP002 601.8881 601       1       0
## CP003 691.9580 691       1       0
```

Alignment

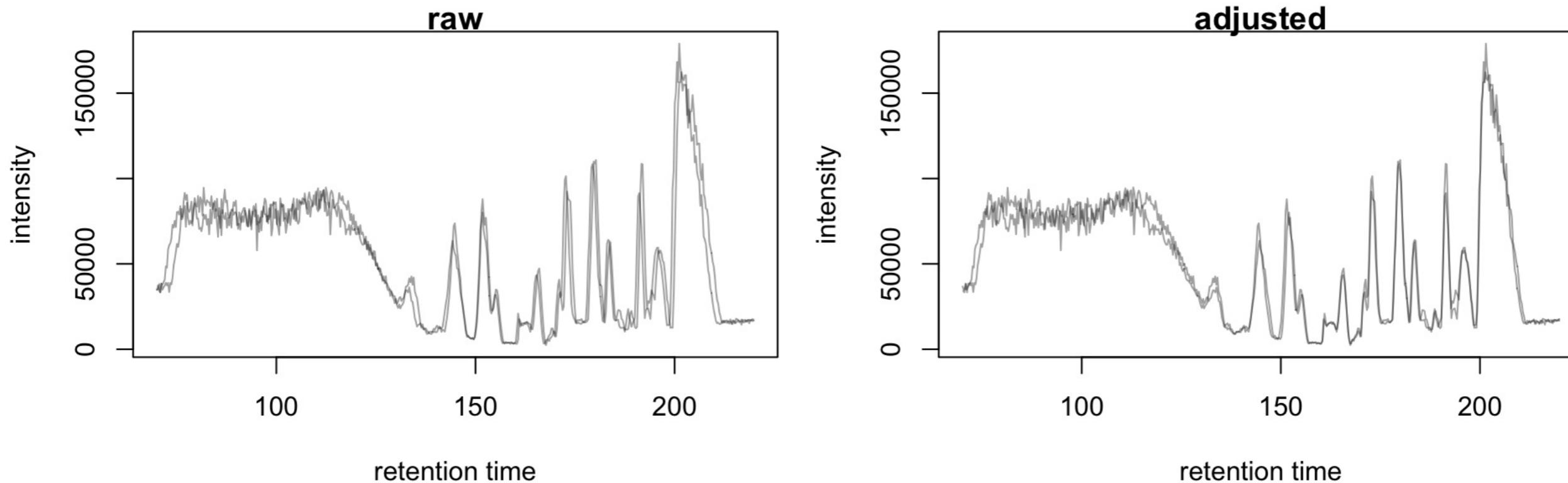
- Alignment: adjust differences in retention times between samples.
- Same analyte may elute at different time between measurements.



- Why? Chromatography setup, age of the column, column gets dirty...

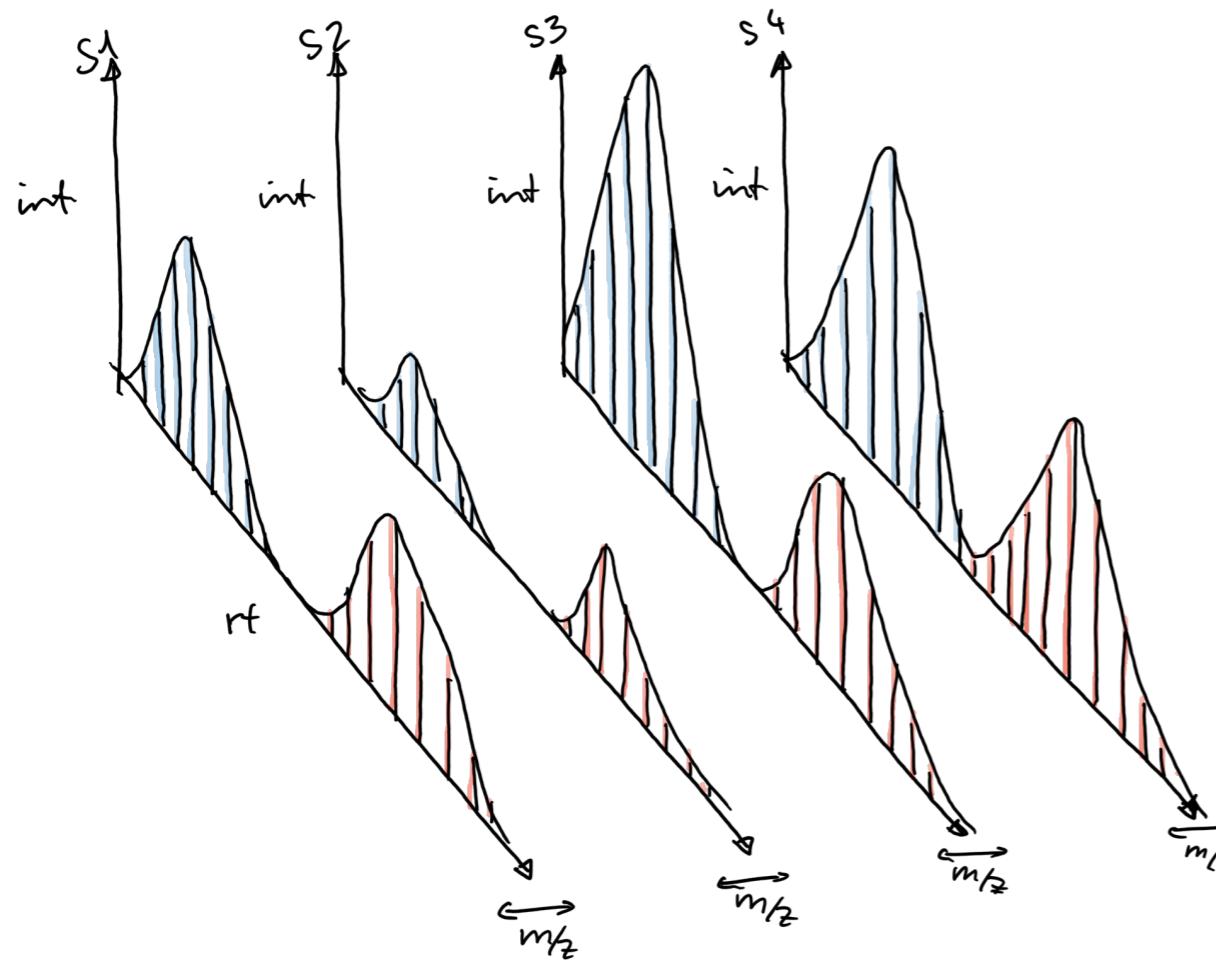
Alignment

- Many algorithms available [Smith et al. Brief Bioinformatics 2013]
- `xcms`: `adjustRtime` function with `PeakDensityParam` [Smith et al. Anal. chem. 2006] or `ObiwarppParam` [Prince et al. Anal. chem. 2006].



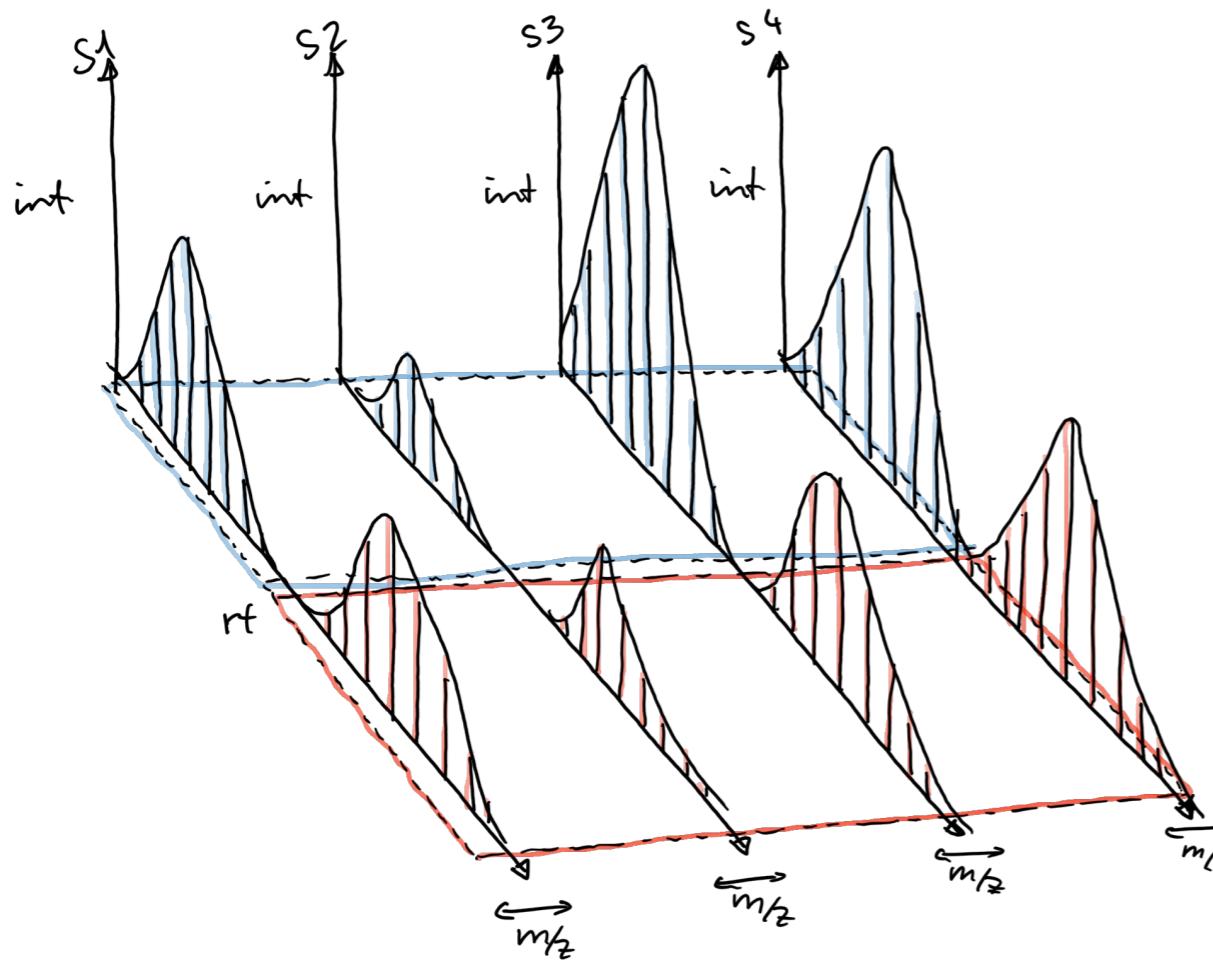
Correspondence

- Aim: group peaks across samples, assuming they represent the same ion.
- Result: matrix of abundances, rows features, columns samples.
- Depends on proper alignment.



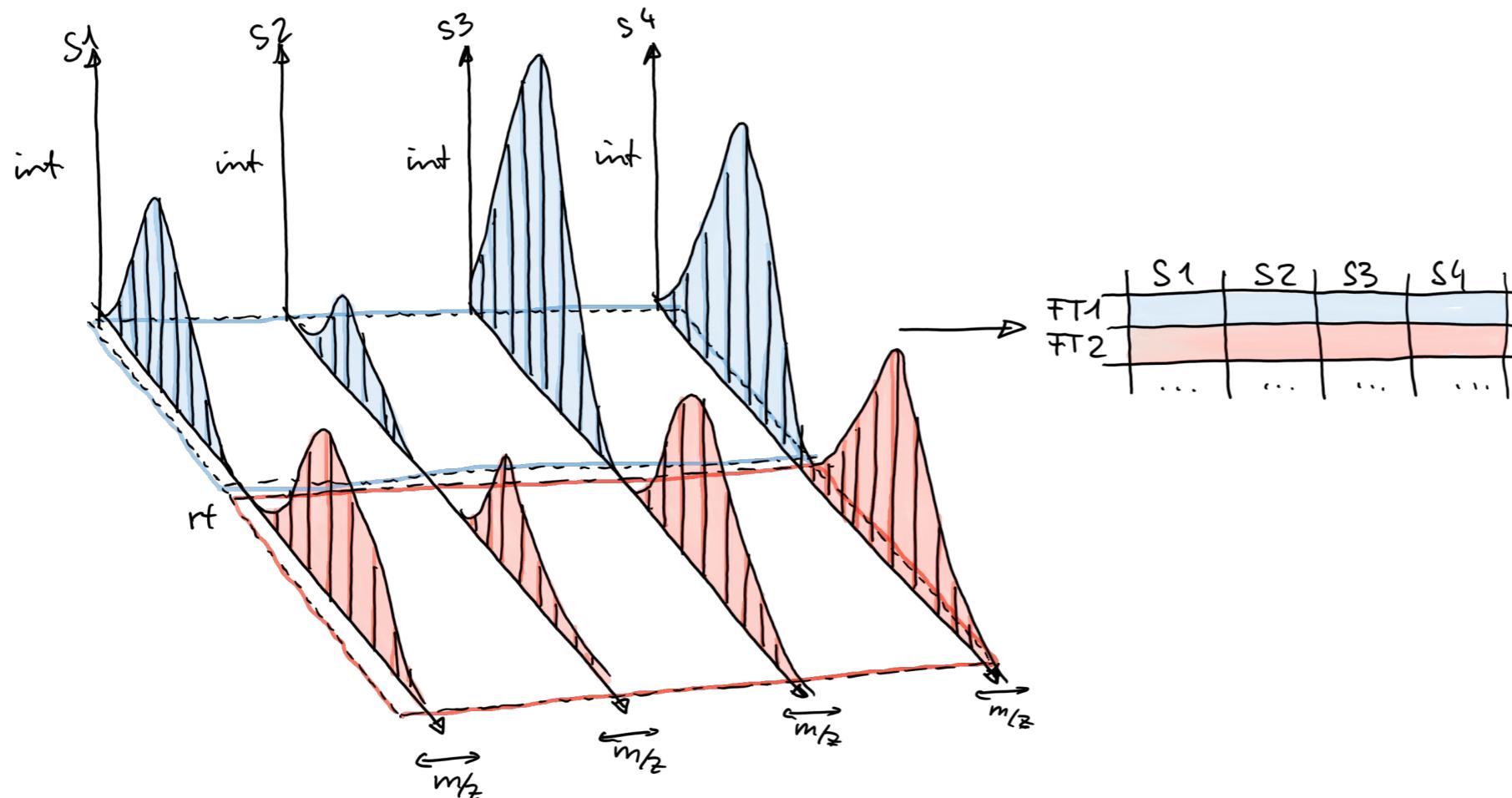
Correspondence

- Aim: group peaks across samples, assuming they represent the same ion.
- Result: matrix of abundances, rows features, columns samples.
- Depends on proper alignment.



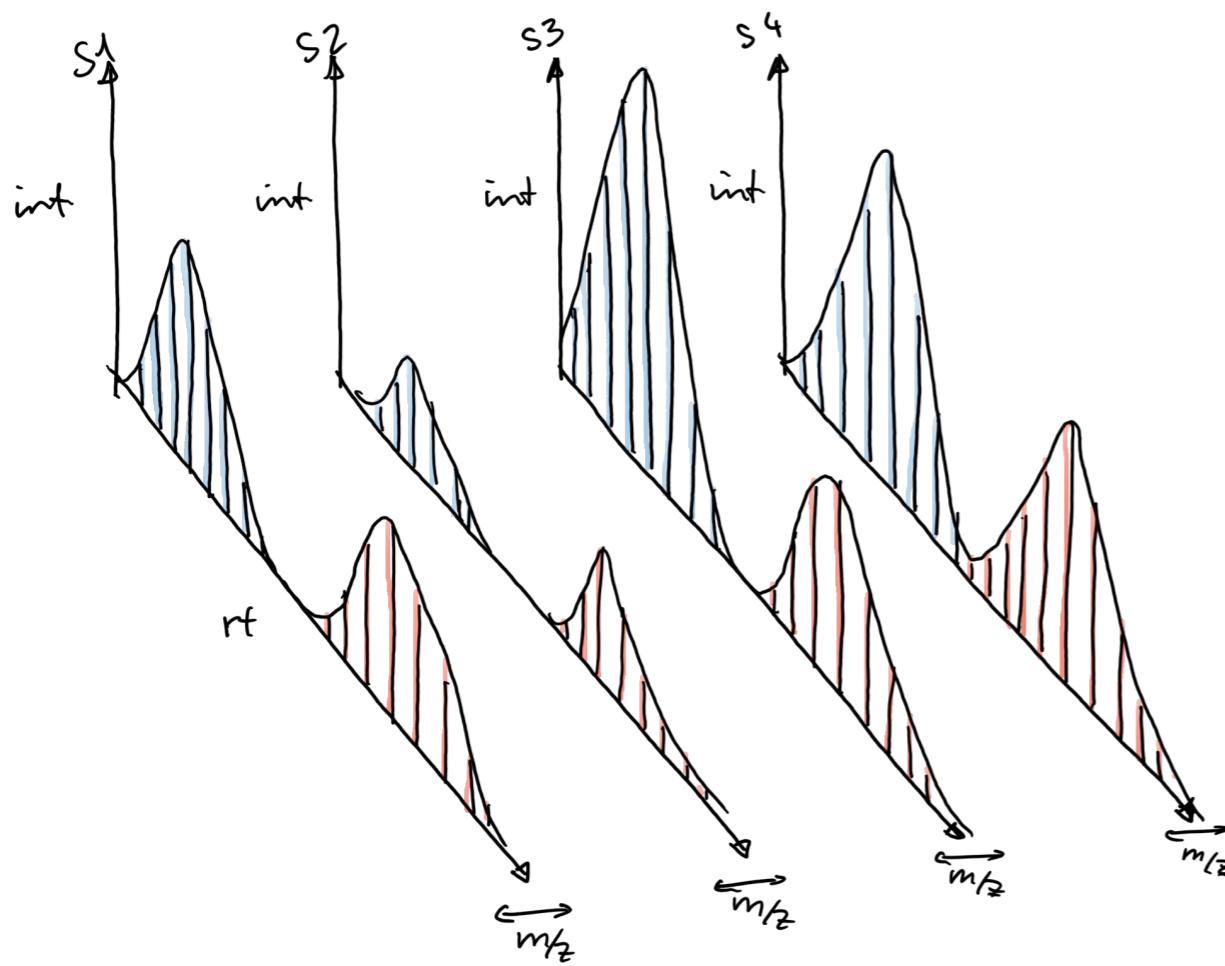
Correspondence

- Aim: group peaks across samples, assuming they represent the same ion.
- Result: matrix of abundances, rows features, columns samples.
- Depends on proper alignment.



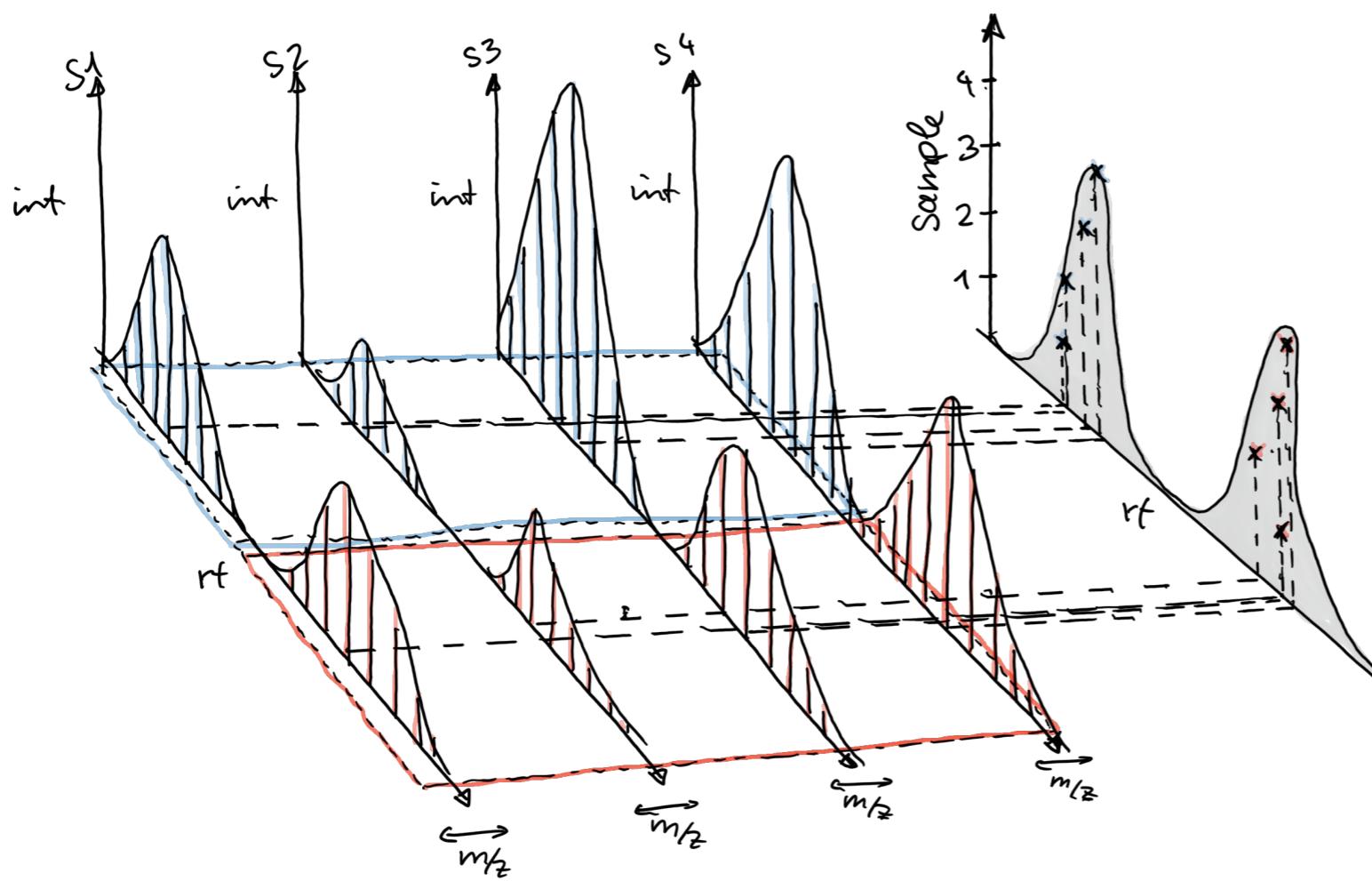
Correspondence

- `xcms: groupChromPeaks` with `NearestPeaksParam` [Katajamaa et al. Bioinformatics 2006] and `PeakDensityParam` [Smith et al. Anal. chem. 2006].
- peak density approach (for a given m/z slice):



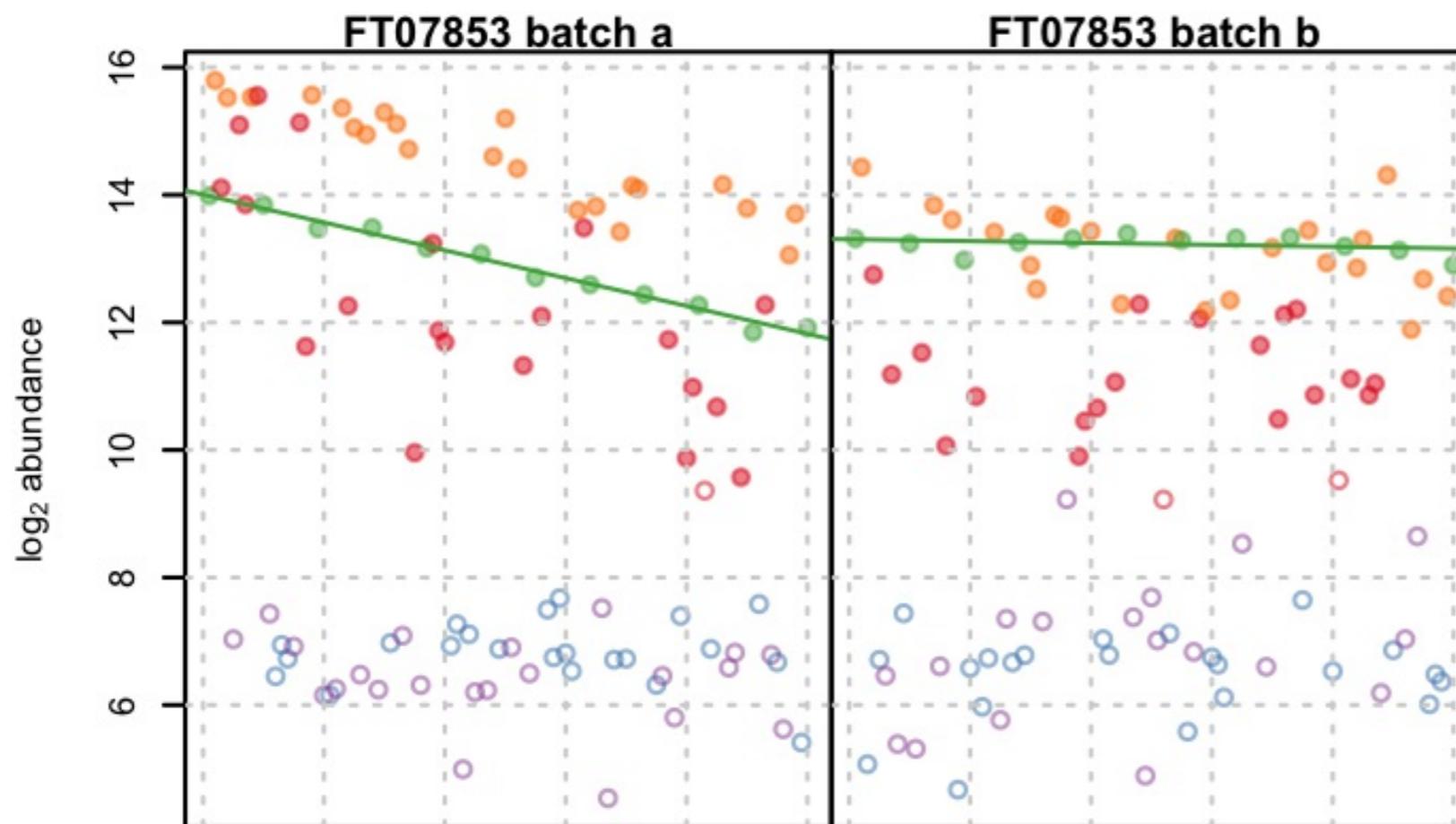
Correspondence

- `xcms: groupChromPeaks` with `NearestPeaksParam` [Katajamaa et al. Bioinformatics 2006] and `PeakDensityParam` [Smith et al. Anal. chem. 2006].
- peak density approach (for a given m/z slice):



Normalization

- Effects related to batch/measurement run.
- LC-dependent effects: seem to affect each metabolite in a different way.



Normalization

- Good practice:
 - QC control samples measured repeatedly.
 - Internal standards.
 - Replicates.
 - Measurement of samples in randomized order.
- Popular methods:
 - RUV [De Livera et al. Anal. Chem. 2015]
 - linear models [Wehrens et al. Metabolomics 2016]
 - linear and higher order models [Brunius et al. Metabolomics 2016].

Identification

- Features have to be annotated to metabolites.

```
## DataFrame with 4 rows and 4 columns
##           mzmed      rtmed      POOL_1      POOL_2
##           <numeric>  <numeric>  <numeric>  <numeric>
## FT001 105.041814839707 167.961095453642 229.490739260736 3093.75184315684
## FT002 105.041653033614 157.083057856508 4762.39872227772 6601.45091358641
## FT003 105.069636149683 31.8108067962868 699.723986763235 1033.23232267732
## FT004 105.11027064078 63.7513630255991 20211.2633706294 15839.5504368189
```

- Match compounds based on features' m/z.
- Will result in an 1:n mapping.

Identification based on m/z matching

- Databases containing masses of (synthetic and biological) compounds.
 - The Human Metabolome Database (HMDB): <https://hmdb.ca>
 - Chemical Entities of Biological Interest: <https://www.ebi.ac.uk/chebi>
 - PubChem (human and other animals) <https://pubchem.ncbi.nlm.nih.gov>
 - ...
- m/z is **not** the mass.
- Mass of an $[M+H]^+$ ion: m/z - mass of 1 hydrogen.
- Different ions from the same compound: $[M+H]^+$, $[M+2H]^{++}$, $[M+Na]^+$, ...

Improved identification

Identify compounds based on m/z **and**:

- **retention time**: requires lab-internal database with approximate retention times for specific compounds (manually defined).
- **MS2 spectrum**:
 - Needs MS2 data of the compounds (e.g. SWATH data).
 - Only works if reference spectrum available in database.
- Up and coming: [CompoundDb](#) package (similar to [ensemblDb](#) and alike): contains compound data + spectra.

Afternoon metabolomics lab

- LC-MS data handling ([MSnbase](#)).
- LC-MS data preprocessing using [xcms](#).

