Biopolymers: Protein Analysis by Mass Spectrometry

Ruedi Aebersold

Methoden der Biologischen Analytik

December 8th, 2017

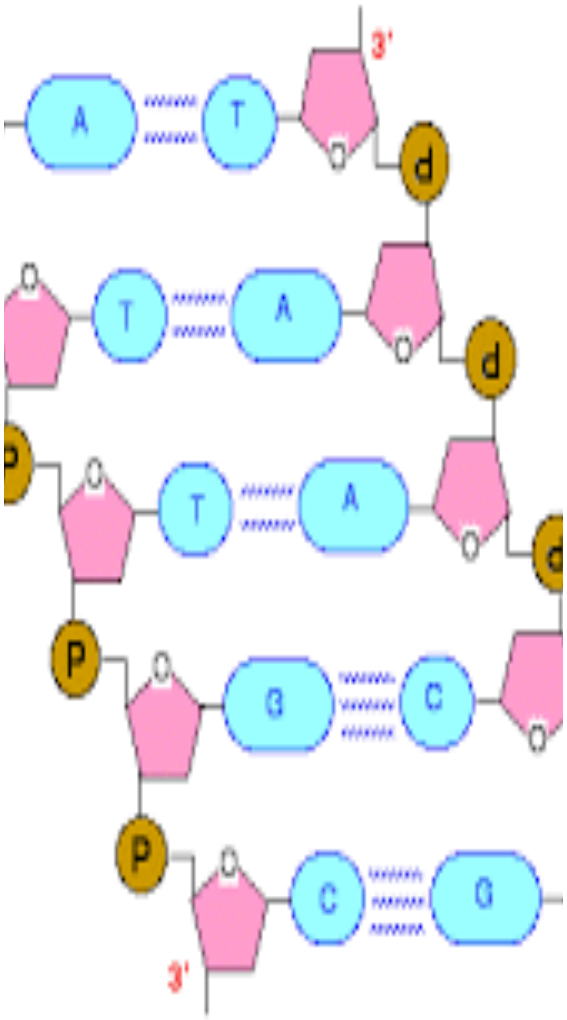
Disclaimer

I assume that you have studied and understood the tutorial related to mass spectrometry that was posted on the Moodle.

Outline

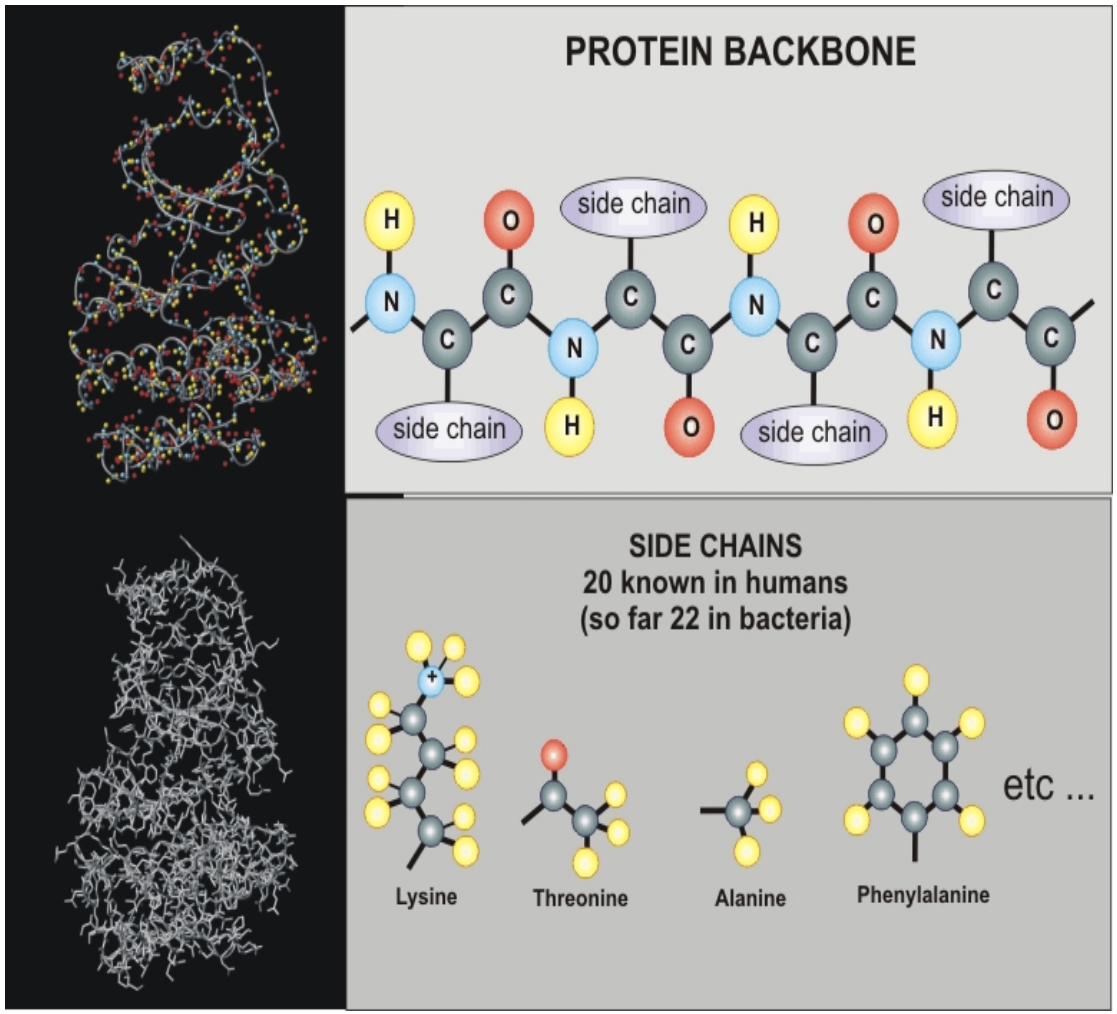
- About DNA and proteins
- Proteins and proteomes
- Likely muddiest points
  - Why does a pure peptide generate more than one signal in a mass spectrometer and why can the same signal represent different peptides?
  - How do peptides fragment and how do we read the sequence out of fragment ion spectra

Nucleic acids



- 4 types of monomers
- Complementarity through base pairing
- Enzymatic synthesis of complementary strand from template
- Amplification of sequences
- Modified bases??

Proteins



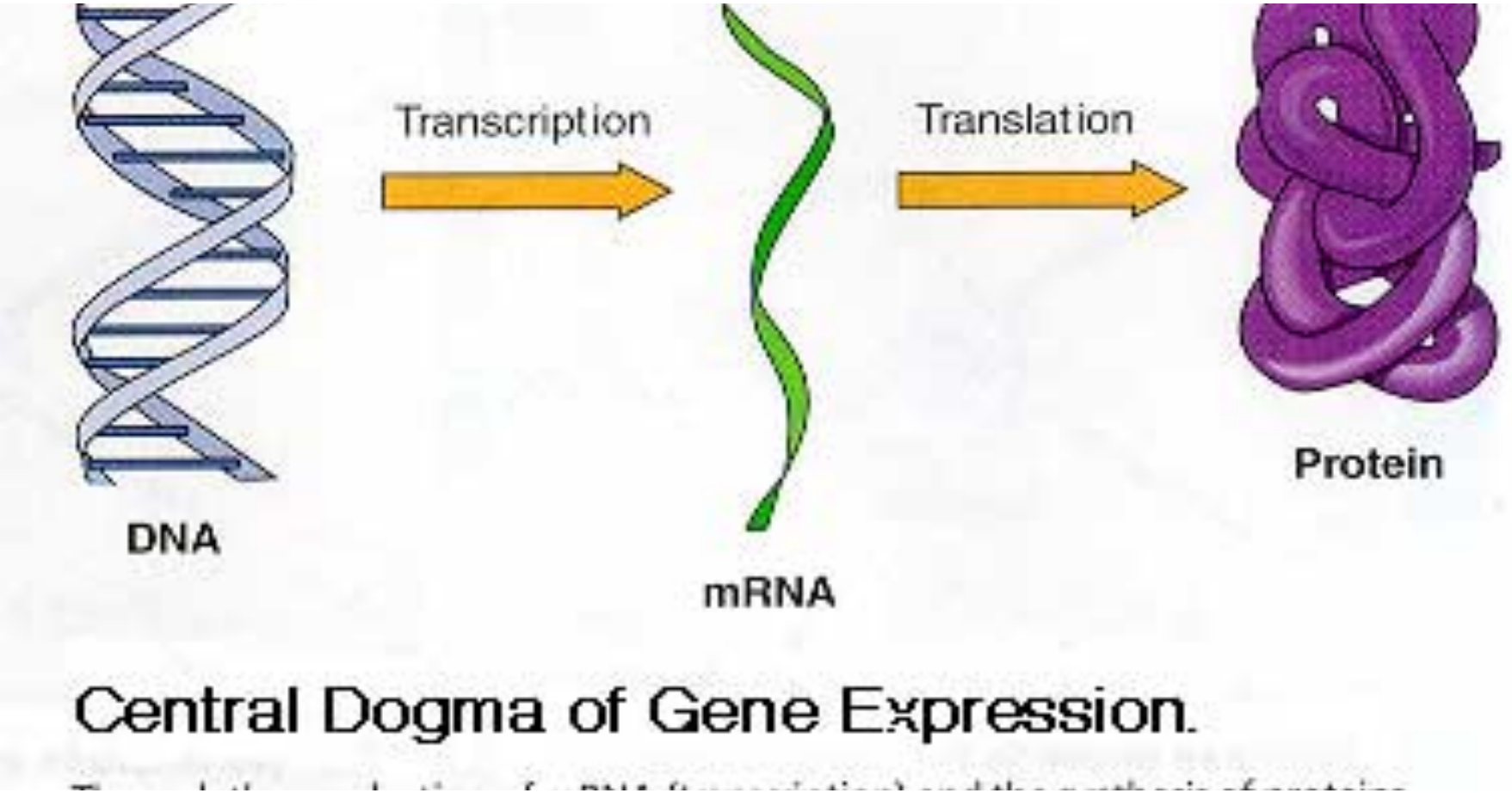
- 20 types of monomers (amino acids)
- Functional only if folded
- No amplification method known
- Amino acids are frequently modified (post translational modification, PTM)

Nucleic acids vs. proteins

- **Nucleic Acids**
- We sequence a synthetic complement to a template sequence
- The synthetic template can be extensively amplified
- We can sequence minute amounts of nucleic acids. E.g. single cells

- **Proteins**
- We directly sequence the proteins extracted from biological samples
- No amplification
- We **cannot** sequence minute amounts of protein e.g. from single cells
- It is a lot easier to sequence a gene than a protein

The Central Dogma of Biology



„once (sequential) information has passed into protein it cannot get out again“, Crick (1958)

So why do we bother analyzing proteins?

- If:
  - We can predict the sequence of proteins from DNA/RNA
  - Protein analysis is harder than DNA analysis
- Because:
  - Protein quantities cannot be predicted from DNA/RNA
  - Proteins are frequently modified and modifications are important for the function of a protein

Outline

- About DNA and proteins
- Proteins and proteomes
- Likely muddiest points
  - Why does a pure peptide generate more than one signal in a mass spectrometer and why can the same signal represent different peptides?
  - How do peptides fragment and how do we read the sequence out of fragment ion spectra

Classical biopolymer analysis (single protein)

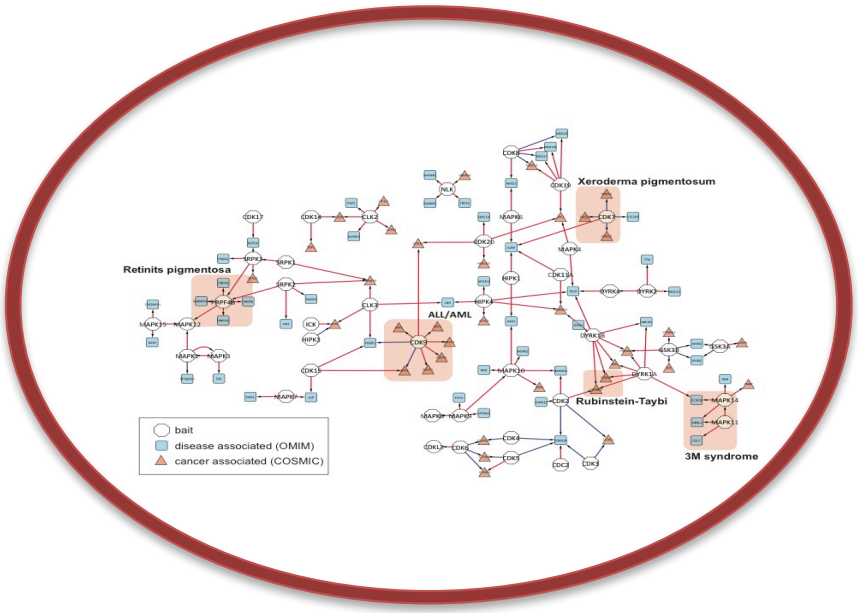
- We need a method to purify a specific protein activity to purity
- We need to sequence the corresponding protein, at least partially
- We then generate a probe to isolate the corresponding gene and sequence the gene

Classical biopolymer analysis (single, purified protein)

- We need a method to purify a specific protein activity
- We need to sequence the corresponding protein, at least partially
- We then generate a probe to isolate the corresponding gene and sequence the gene

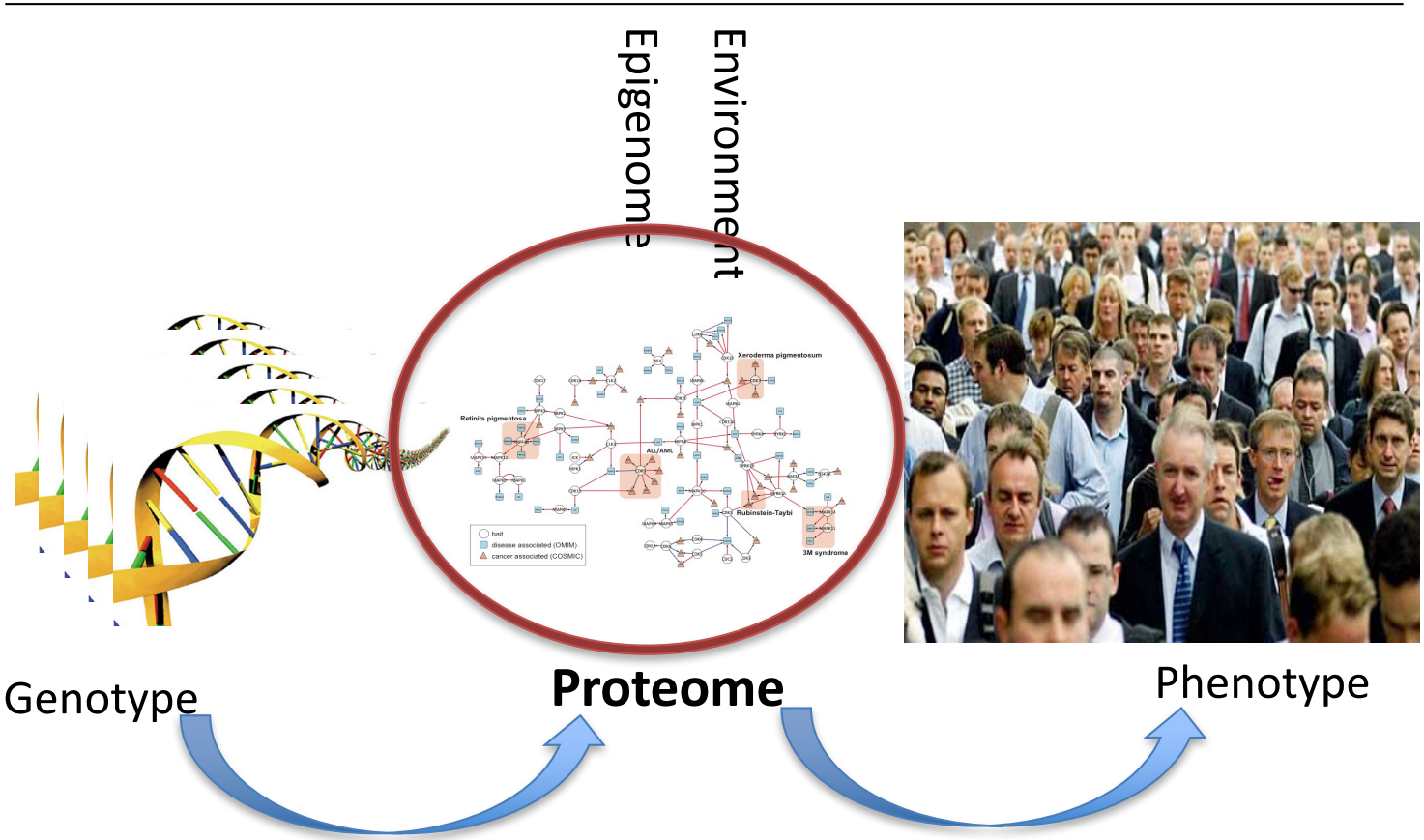
What would you consider the most significant challenges with this strategy?

Proteome: The ensemble of all proteins of a cell or organism



Human: ~20.000 ORF's; potentially a million proteoforms (splicing, modifications)

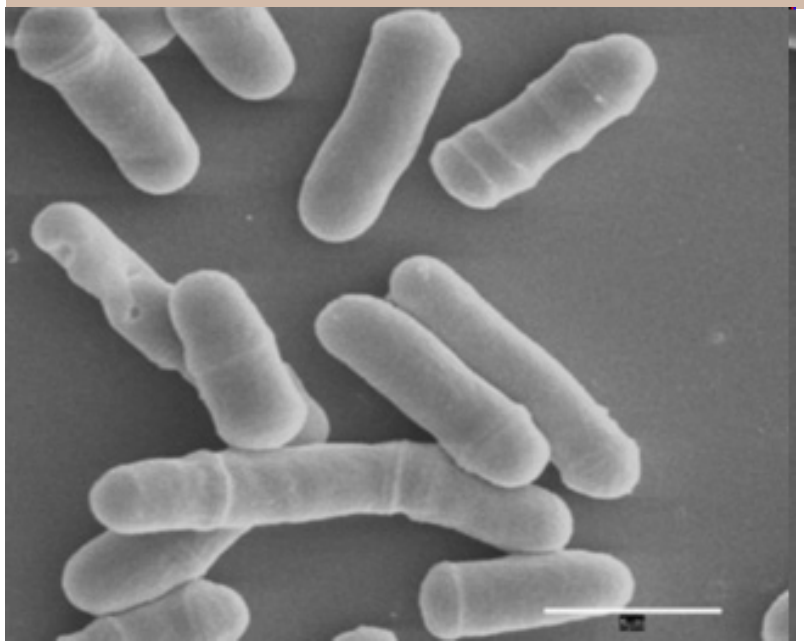
The position of the proteome in experimental biology



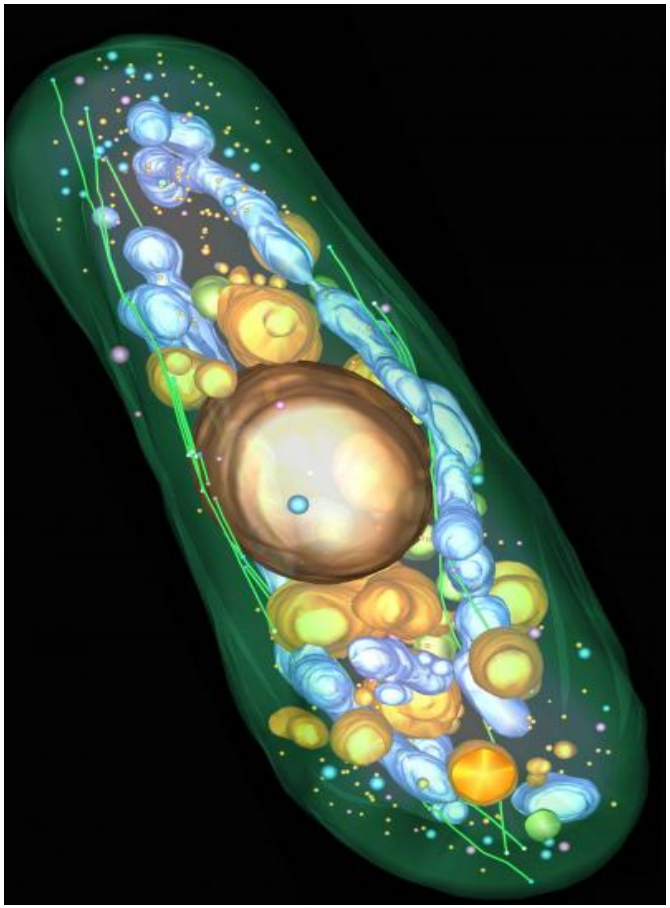
Consequences for biopolymer analysis
Proteome scale

- To know all proteins means to know all functions?
- The technology needs to be able to identify hundreds of thousands of peptides reliably
- High throughput and high precision
- Shotgun mass spectrometry for proteins
- The data from mass spectrometric analysis are searched against nucleic acid sequences

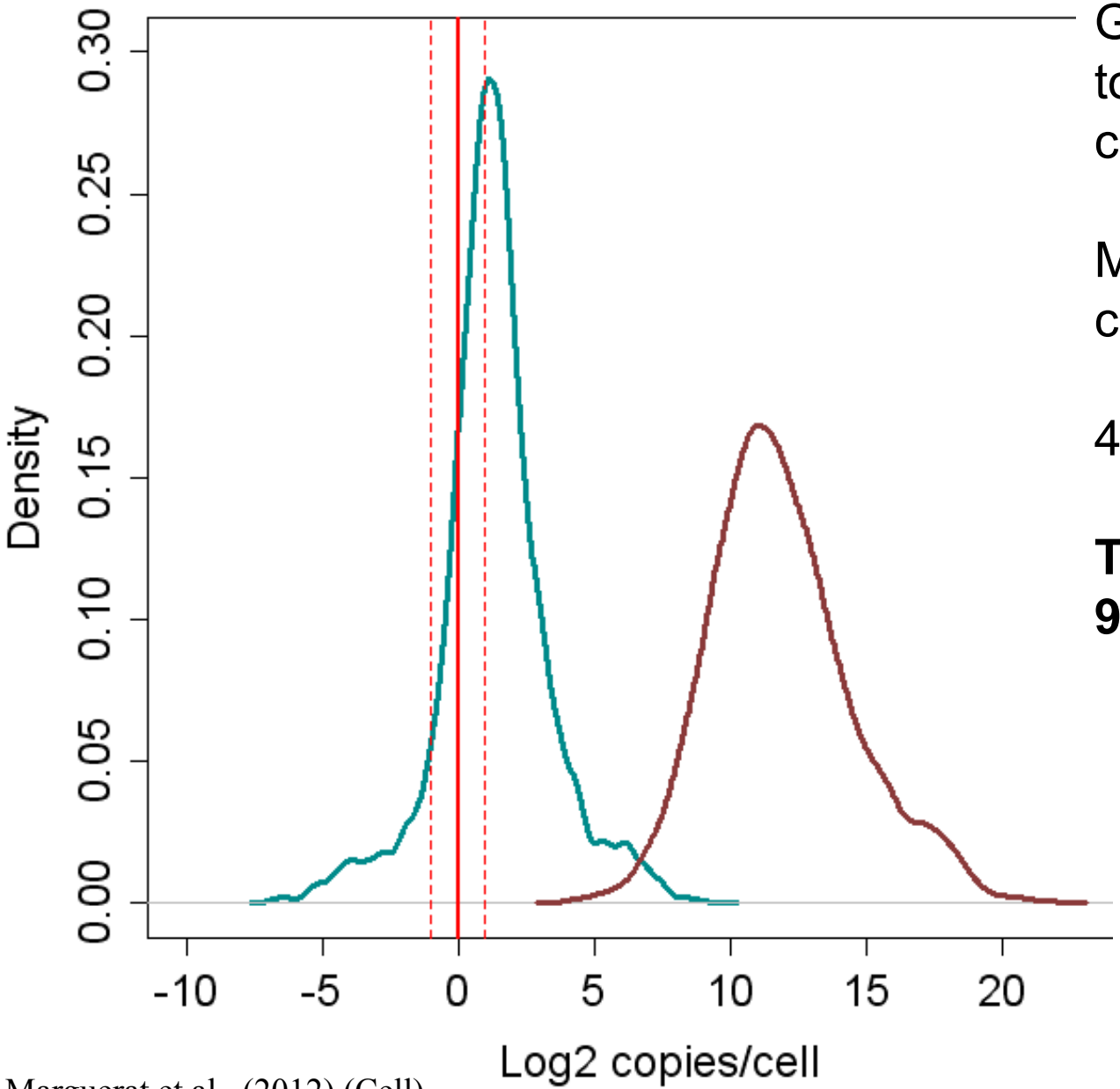
The amazing complexity of the proteome: fission yeast (*Schizosaccharomyces pombe*)



- unicellular eukaryote (fungus)
- genome : 14 Mb, ~5000 genes
- 7-14 micrometer in length



The amazing complexity of the proteome



Genes produce between 30 to 1.1 million protein copies/cell

Median protein copies/cell: 3919

42,000 mRNAs

Total number of proteins: 95 Mio per cell

Much larger dynamic range than for mRNAs

Total protein concentration ~300mg/ml

Marguerat et al , (2012) (Cell)

Consequences for biopolymer analysis
Proteome scale

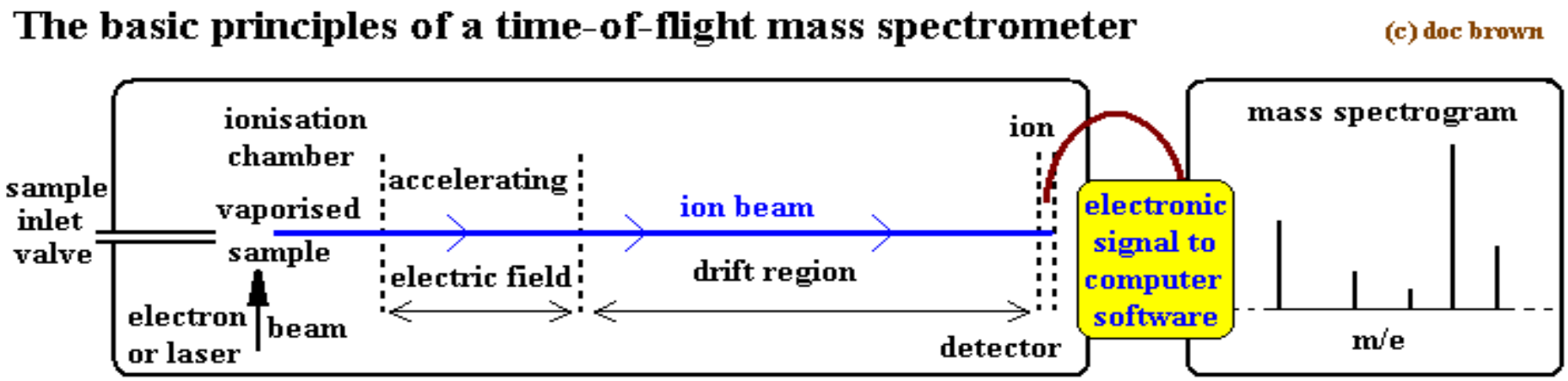
- To know all proteins means to know all functions?
- The technology needs to be able to identify hundreds of thousands of peptides reliably
- High throughput and high precision
- Shotgun mass spectrometry for proteins
- The data from mass spectrometric analysis are searched against nucleic acid sequences

- Do we now all functions if we know all proteins?
- Why do we not purify all proteins and sequence each one separately?
- Why do we need nucleic acid sequences to search fragment ion spectra?
- Why do we sequence peptides and not proteins?

Outline

- About DNA and proteins
- Proteins and proteomes
- Likely muddiest points
  - Why does a pure peptide generate more than one signal in a mass spectrometer and why can the same signal represent different peptides?
  - How do peptides fragment and how do we read the sequence out of fragment ion spectra

Single stage mass spectrometer



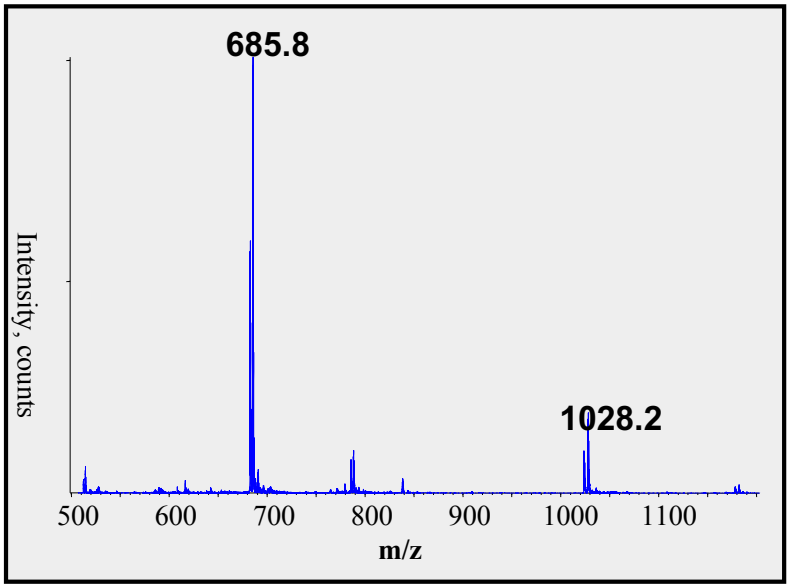
- The single stage mass spectrometer generates a mass spectrum of molecular ions

Questions around peptide mass spectra

- How would a mass spectrum look like if you analyzed a digest of a single, purified protein?
- How would a mass spectrum look like if you analyzed a digest of a proteome?
- Is the mass of a peptide sufficient to determine the sequence of a peptide?
- Is the mass of several peptides of a protein sufficient to identify the protein?

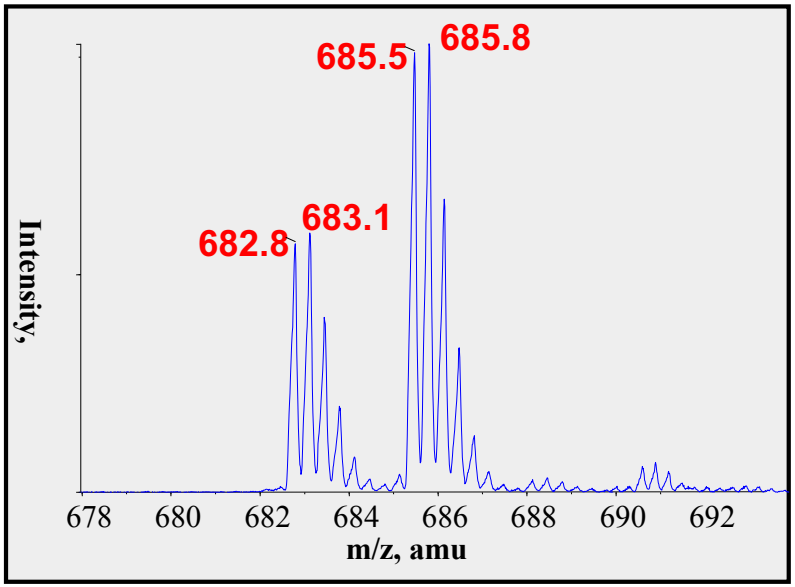
The relationship between peaks detected in a mass spectrum and the corresponding analyte peptide

We inject a pure peptide into the mass spectrometer. Why do we see more than one peak?



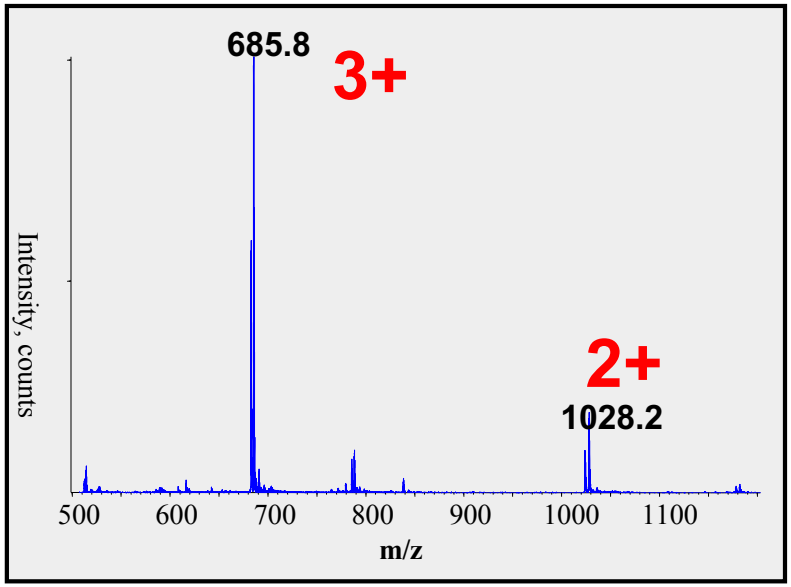
The relationship between peaks detected in a mass spectrum and the corresponding analyte peptide

We zoom a specific peak with a high resolution mass spectrometer. Why does the peak now appear serrated?

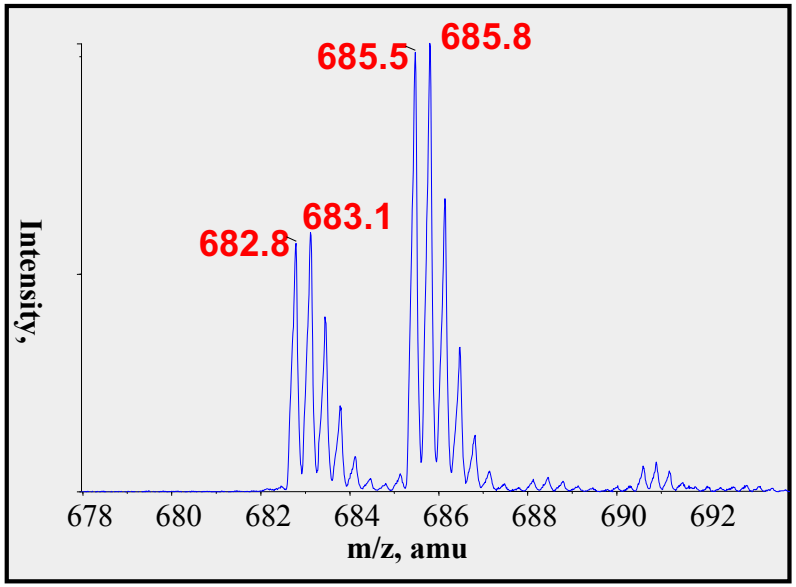


Basics of mass spectrometry

Multiply Charged Ions



Isotope Distribution

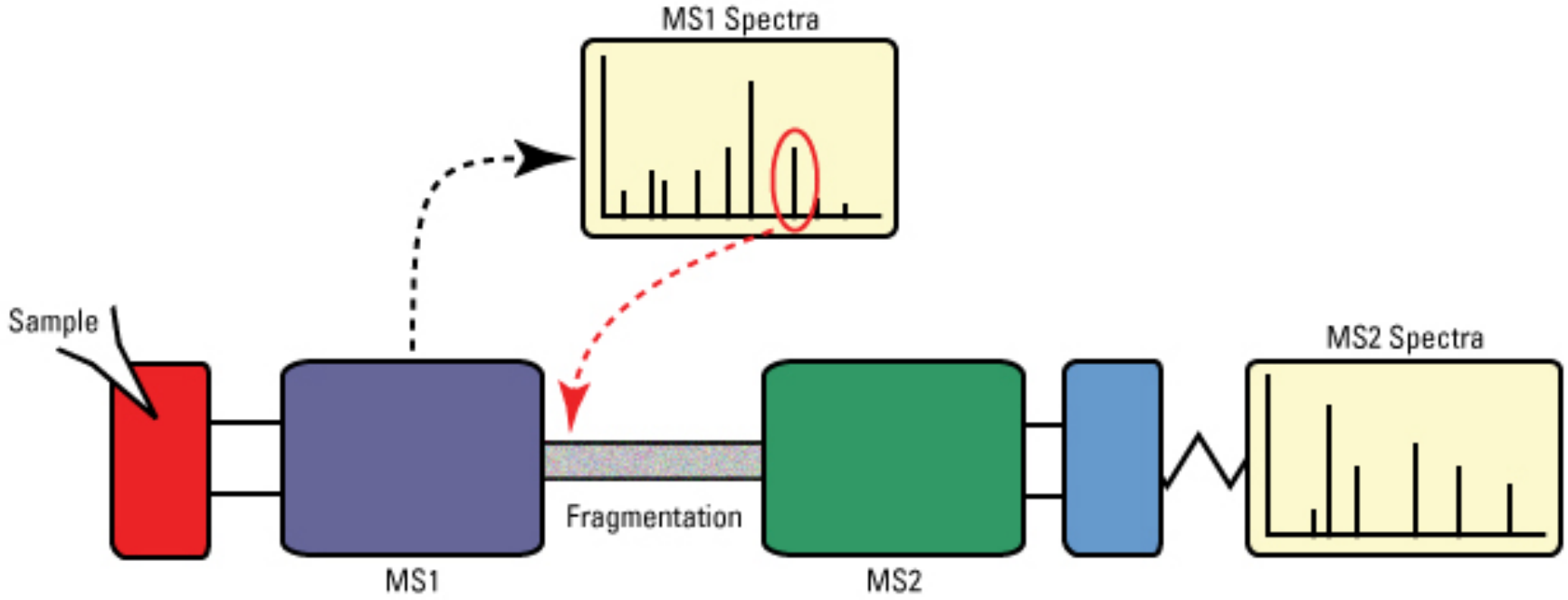


Atom	Mass	Rel. Abund.
Hydrogen	1.008	99.985
	2.001	0.015
Carbon	12.000	98.90
	13.003	1.10
Nitrogen	14.003	99.63
	15.000	0.37
Oxygen	15.995	99.76
	17.999	0.20
Sulfur	31.972	95.02
	33.968	4.21

Outline

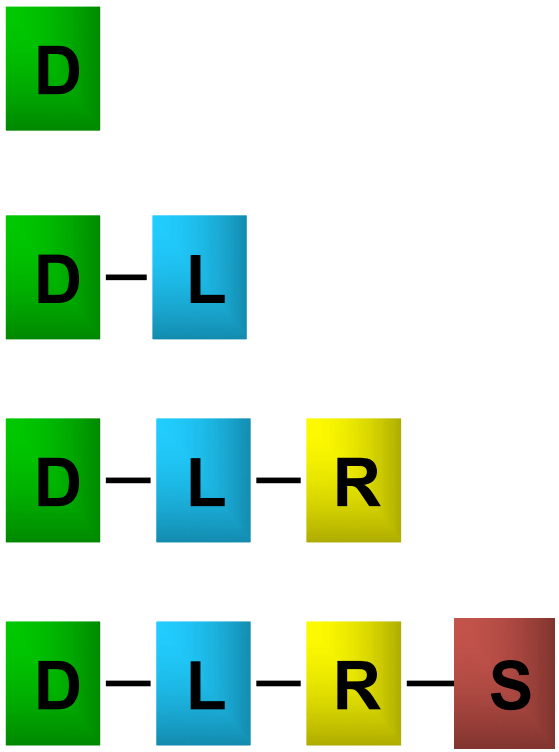
- About DNA and proteins
- Proteins and proteomes
- Likely muddiest points
  - Why does a pure peptide generate more than one signal in a mass spectrometer and why can the same signal represent different peptides?
  - How do peptides fragment and how do we read the sequence out of fragment ion spectra

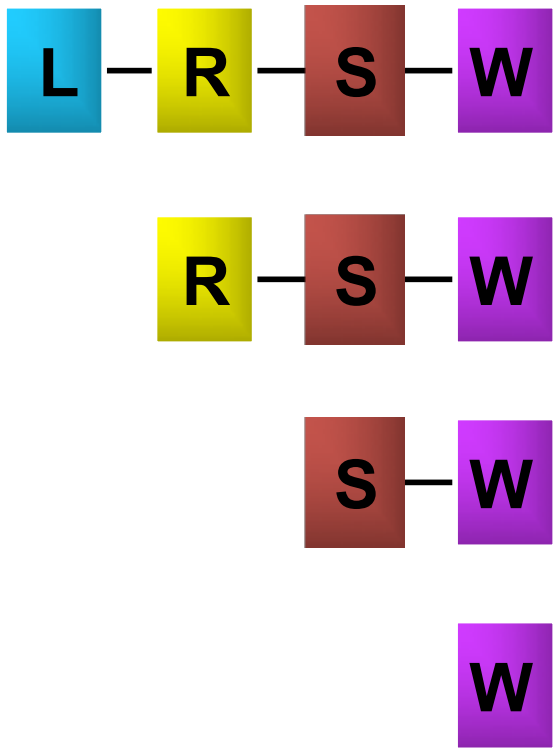
Tandem mass spectrometry: From mass to sequence



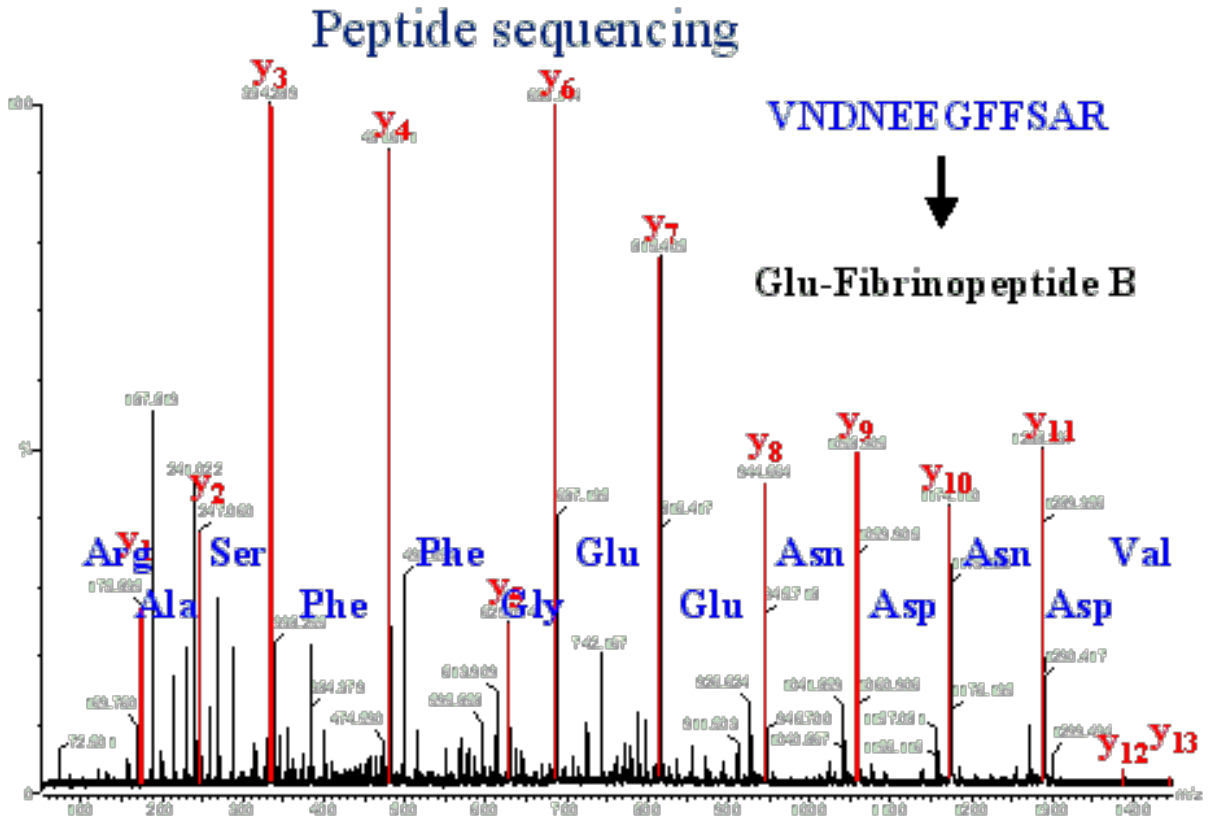
- Molecular ions of a peptide are isolated and further fragmented to generate a fragment ion or MS2 spectrum.
- The fragment ion spectrum records the mass and intensity of the fragment ions generated from the isolated molecular ion (precursor ion)

Fragmenting a peptide

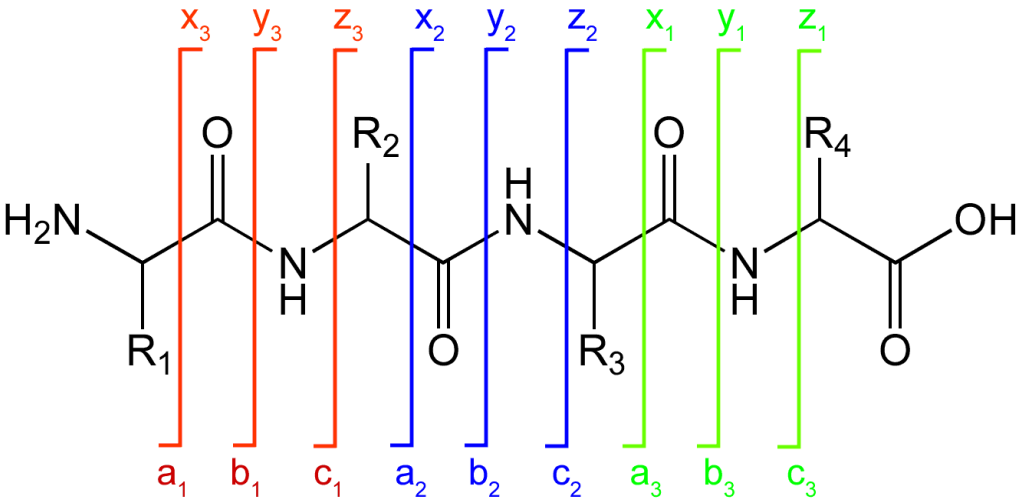


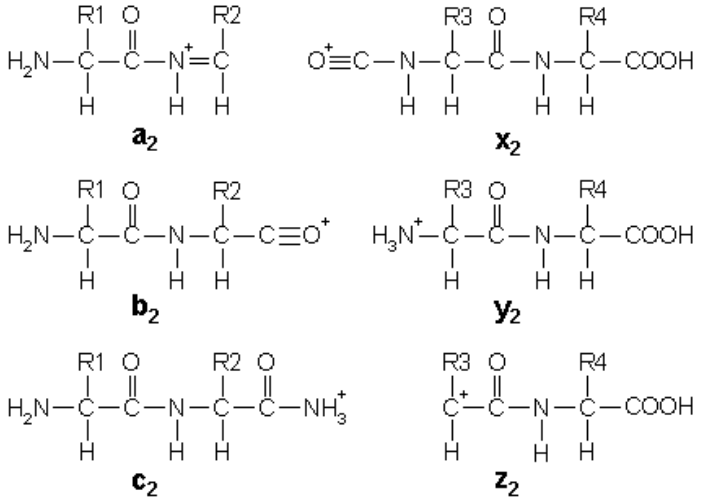


Peptide sequencing by MS/MS



Nomenclature of fragment ions





b- and y-ions are typically formed using collision-induced dissociation, the most common fragmentation technique

- Typically, only one bond is cleaved in a given molecule

- For an ensemble of molecules, different ions of a series are formed and generate the MS/MS spectrum

Questions around fragment ion mass spectra

- Is a single sequenced peptide sufficient to identify a protein?
- How long would a peptide need to be to be unique to a proteome?
- Are all fragment ion peaks of the same intensity? If not how does this affect your ability to read a sequence?
- If you detect two ion series, how do you know which one is the b- and which one the y-ion series?
- What happens to a tandem mass spectrometer if we inject the digest of a single, purified protein (assume 20 peptides)?
- What happens to a tandem mass spectrometer if we inject the digest of a proteome (assume 500.000 peptides)?
-