Mass Spectrometry – How it works, what it does, and how it can help you

Lisa Jenkins, PhD

National Cancer Institute

Proteomics

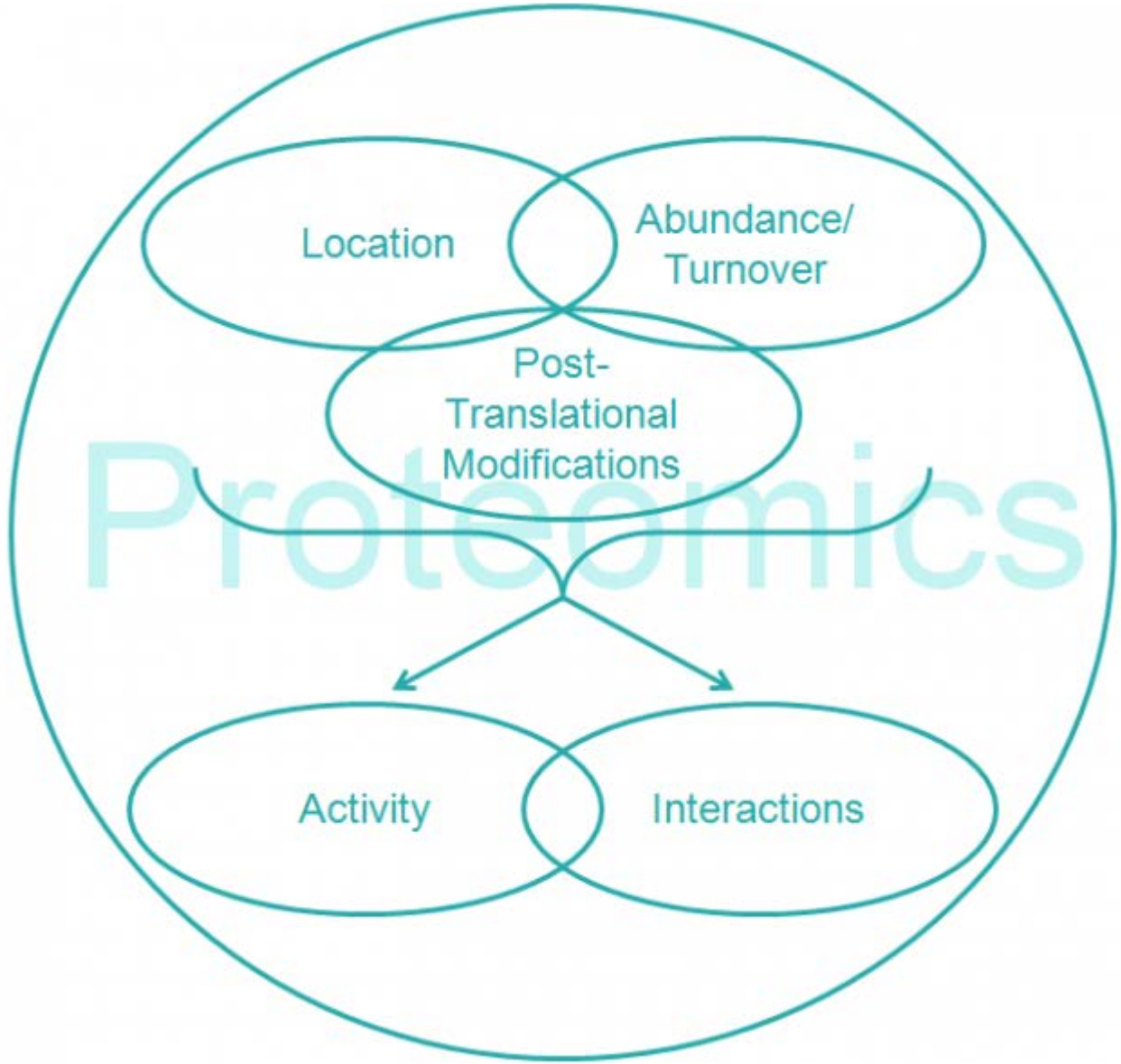
What is a proteome?

Entire set of proteins expressed by a genome, cell, tissue, or organism at a certain time

- *given type of cell or organism*
- *at a given time*
- *under defined conditions*

A proteome is a snapshot of a dynamic system.

The exact specifications of a “proteome” and “proteomics experiment” can change significantly depending on the question being investigated.

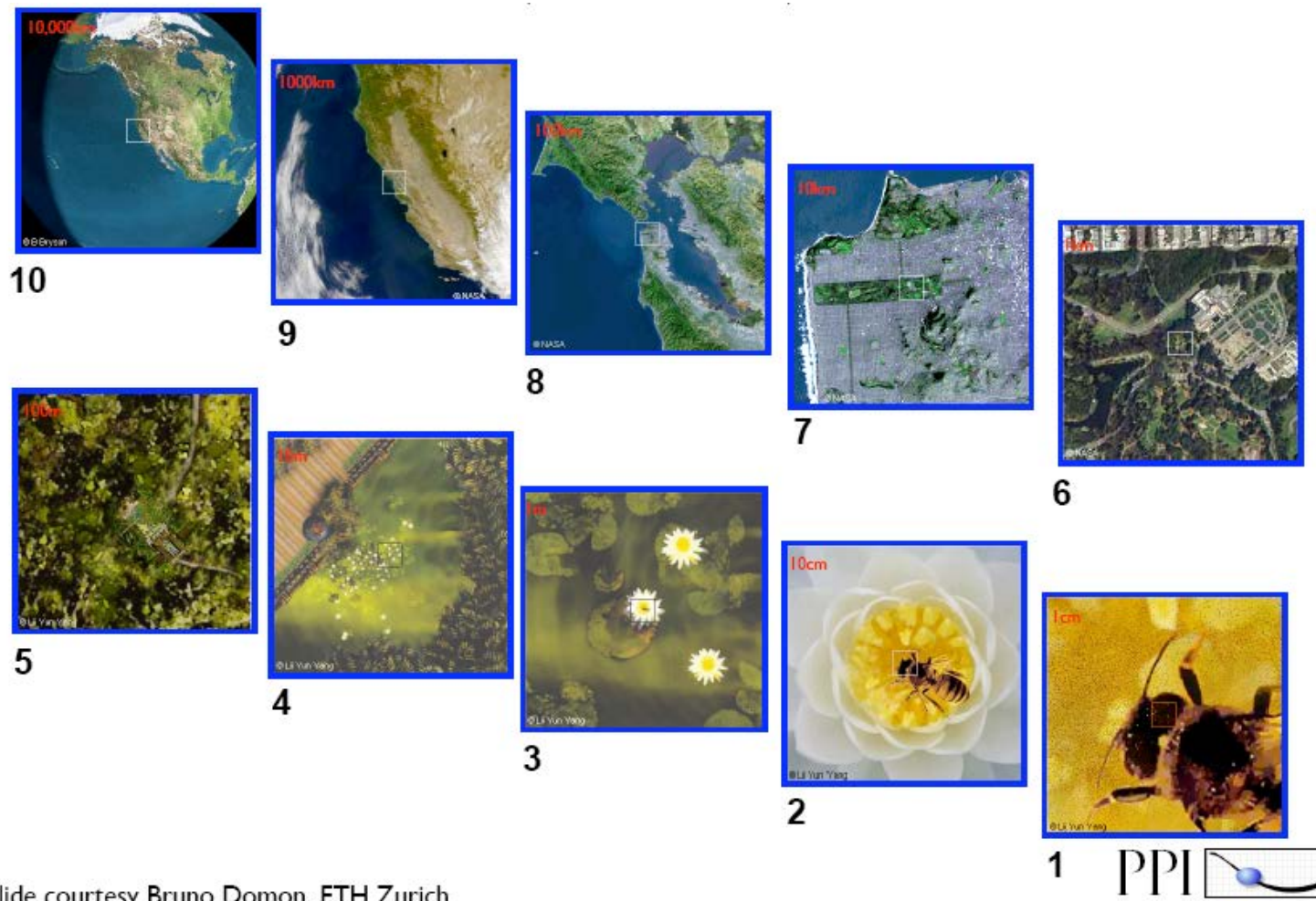


https://www.ebi.ac.uk/

Two critical issues in proteomics in general are the wide range of protein concentrations and the complexity of the proteome. PTMs further complicate both issues as only a fraction of the total protein may be modified and a single protein can be modified in multiple different ways.

In a single proteome, it is estimated that there are 10^6 different proteins.

In a cell, the protein concentration spans 10-12 orders of magnitude.



Slide courtesy Bruno Domon, ETH Zurich

PPI

Why mass spectrometry to study proteomes?

MS is an analytical technique that ionizes chemical species and sorts the ions based on their mass-to-charge ratio.

Allows the identification and quantitation of proteins in complex mixtures, as well as sites of post-translational modification.

Facilitated by publicly-available genome sequences, development of high-throughput methods and high resolution, accurate mass instruments, and improved bioinformatics.

Unable to characterize the full protein complement of a mammalian cell, MS can characterize ~5000 proteins in a single experiment.

Types of mass spectrometers

Mass spectrometers are comprised of an ionization source and a mass detector:

MALDI-TOF – matrix assisted laser desorption ionization time-of-flight detection

ESI-trap – electrospray ionization ion trap detection

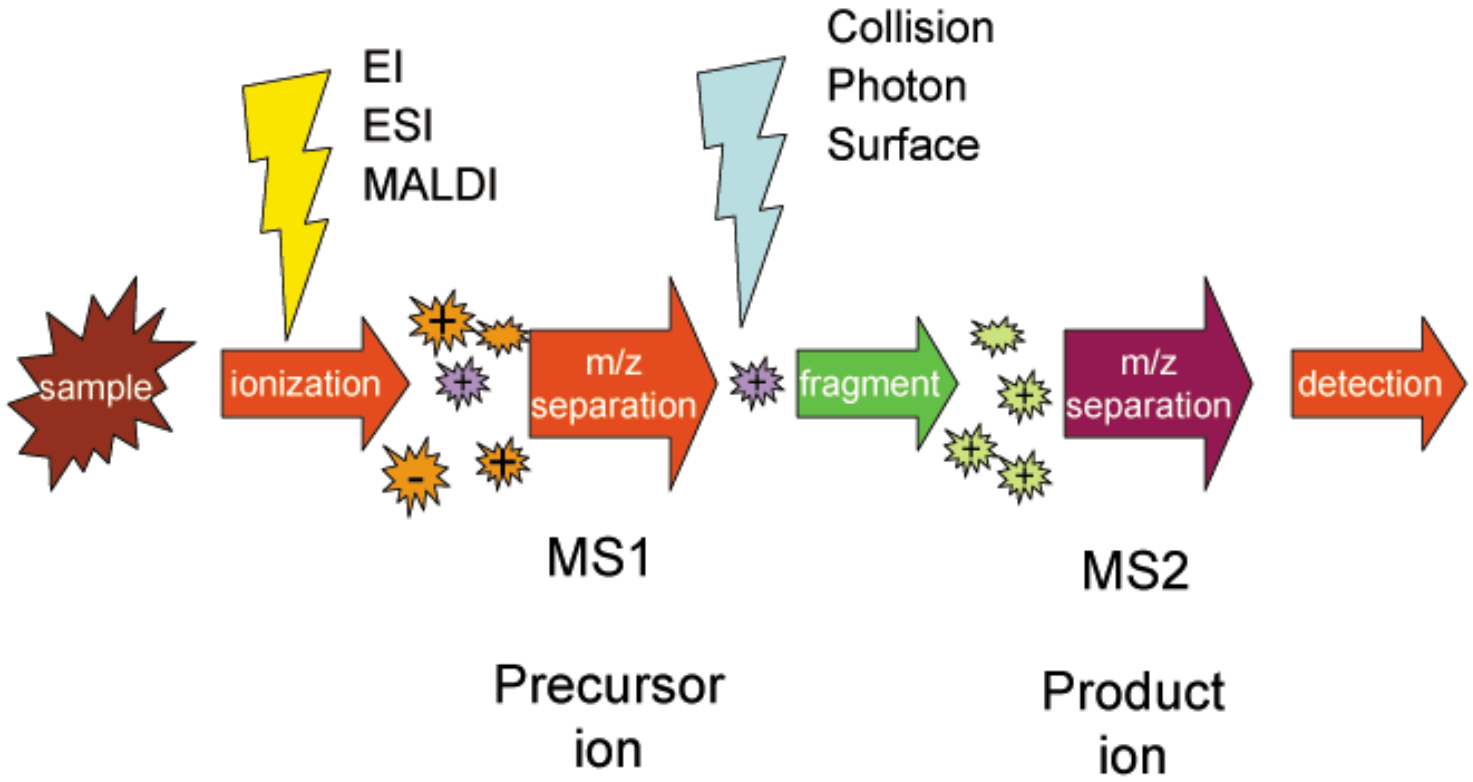
For proteomics, the most common methods of ionization are MALDI and ESI/nESI. The most common detectors are: TOF, trap, FT-ICR, and quadrupole.

Thermo
Fusion
Q Exactive

Waters
Synapt G2-Si
Xevo G2-XS

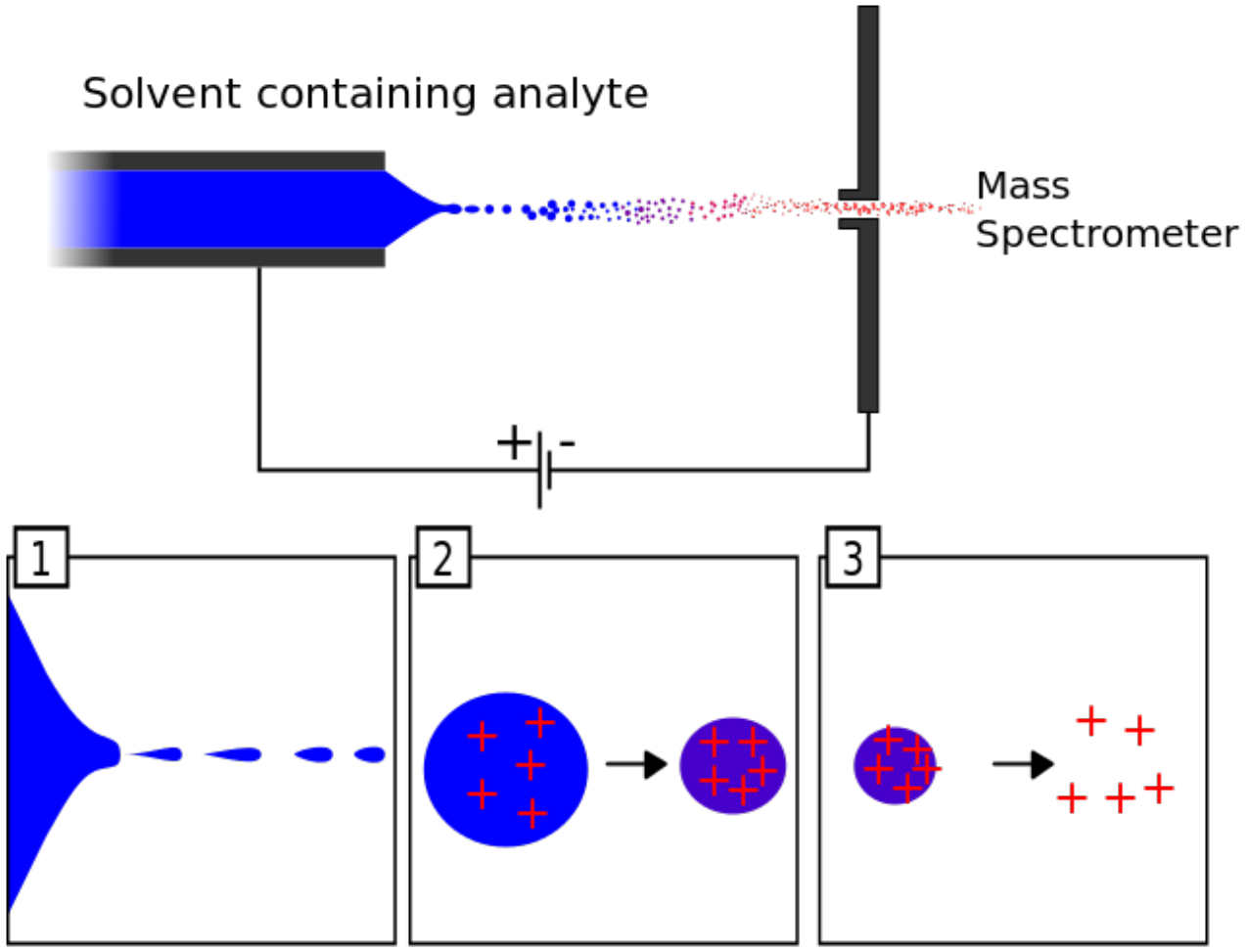
SCIEX
Triple TOF 6600

General schematic for analysis using a mass spectrometer



K. Murray (Kkmurray), https://commons.wikimedia.org/w/index.php?curid=1943319

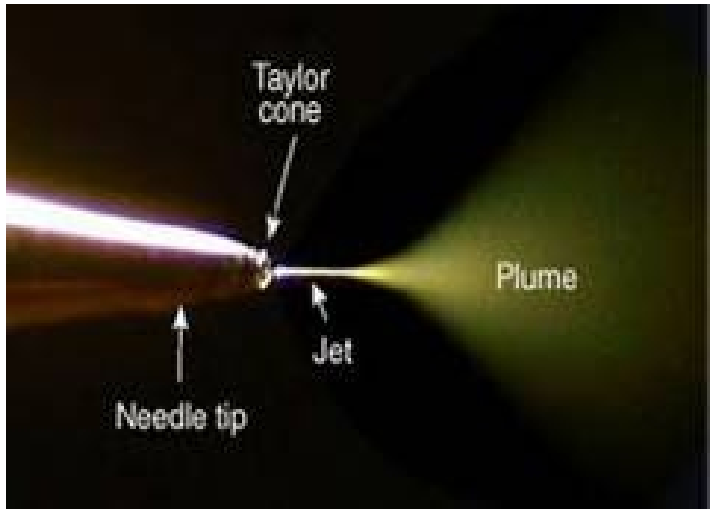
How ionization works (electrospray example)



(1) Under high voltage, the Taylor Cone emits a jet of liquid drops

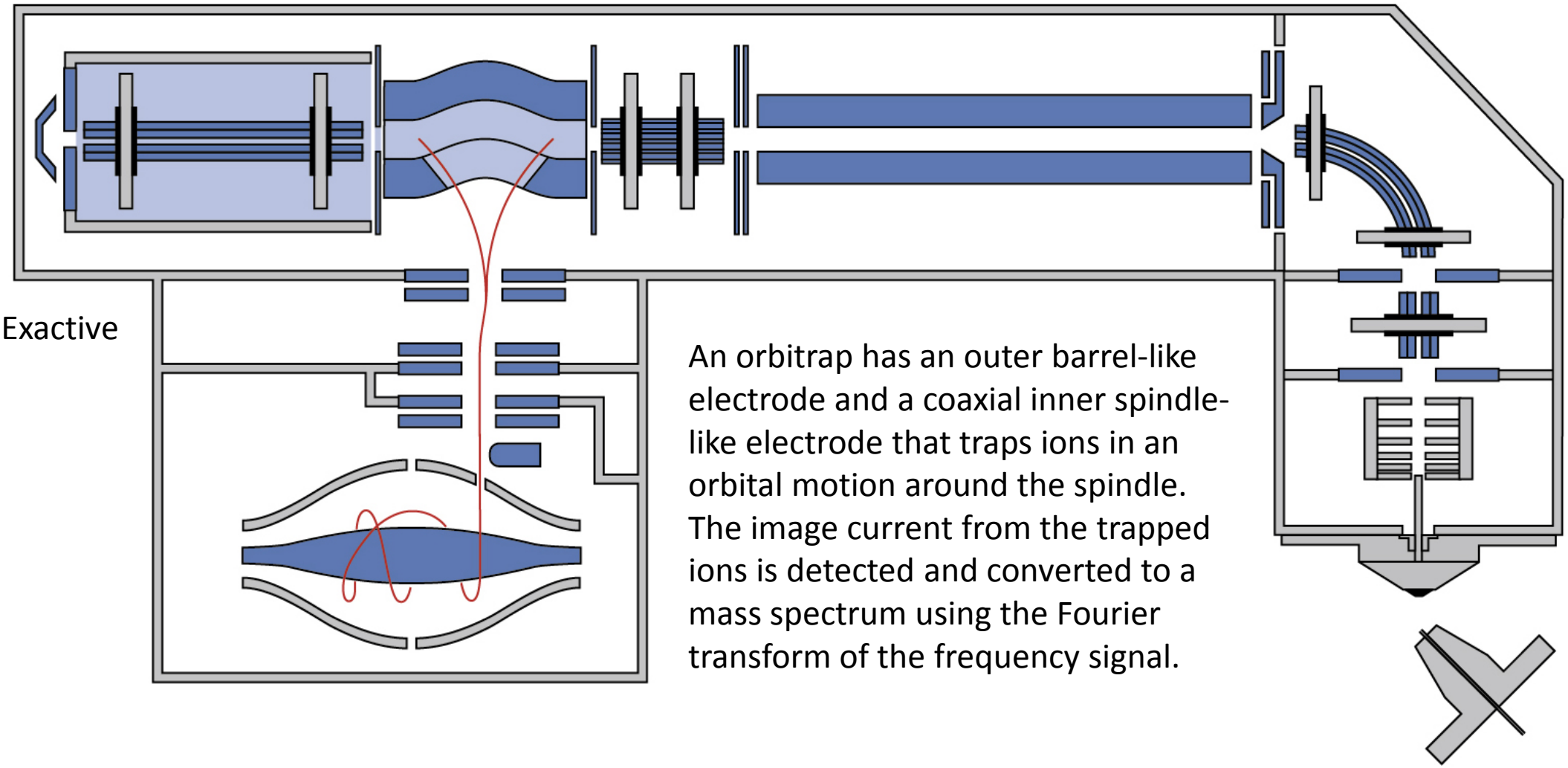
(2) The solvent from the droplets progressively evaporates, leaving them more and more charged

(3) When the charge exceeds the Rayleigh limit the droplet explosively dissociates, leaving a stream of charged ions



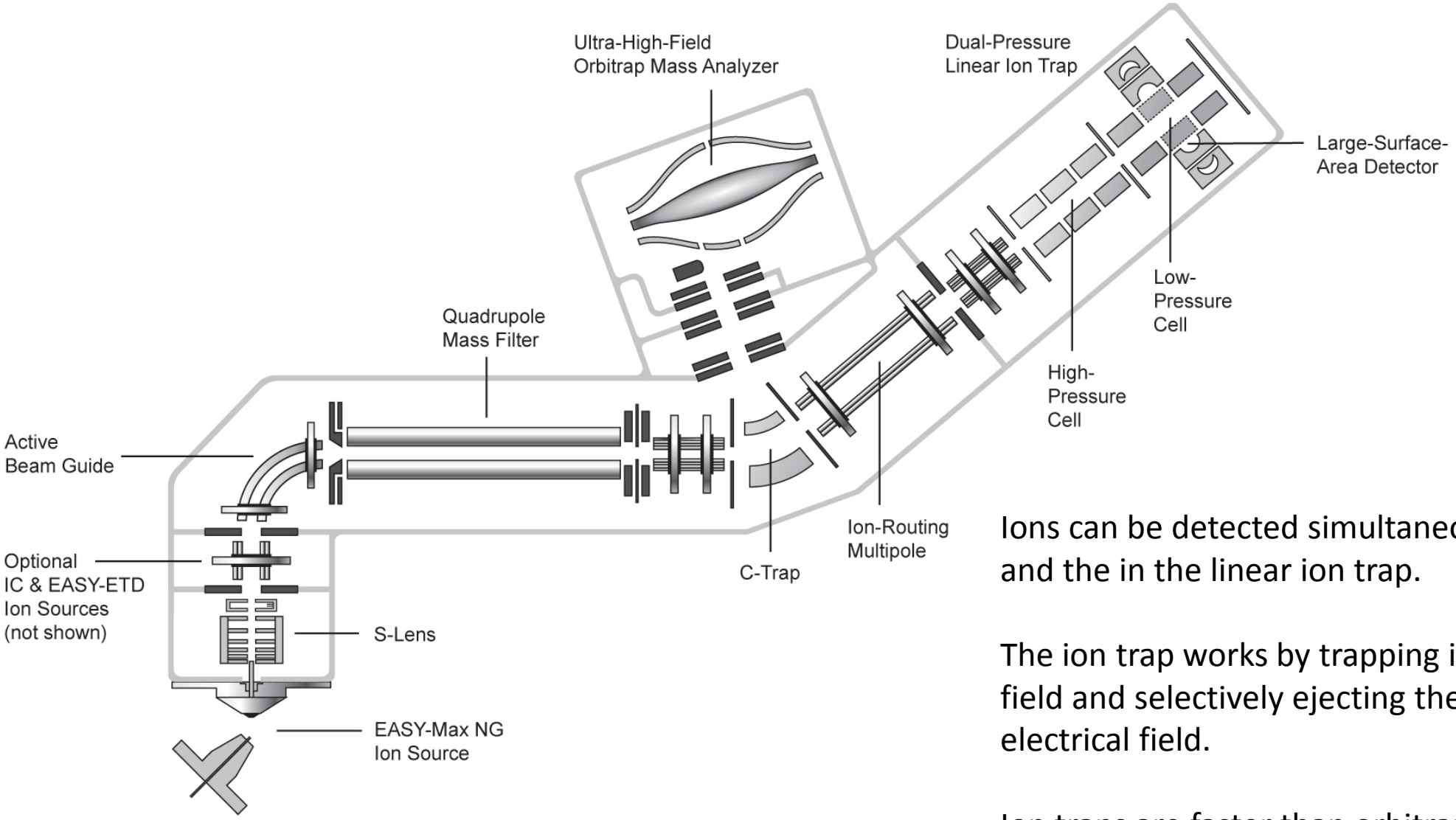
Internal schematic of an Orbitrap instrument

Thermo Q Exactive



An orbitrap has an outer barrel-like electrode and a coaxial inner spindle-like electrode that traps ions in an orbital motion around the spindle. The image current from the trapped ions is detected and converted to a mass spectrum using the Fourier transform of the frequency signal.

An orbitrap can be combined with a linear ion trap



Ions can be detected simultaneously in the orbitrap and the in the linear ion trap.

The ion trap works by trapping ions in an electrical field and selectively ejecting them by adjusting the electrical field.

Ion traps are faster than orbitraps, but do not have as high of mass accuracy.

MS-based proteomics experiments

1) **“top-down”**

Intact proteins are analyzed directly

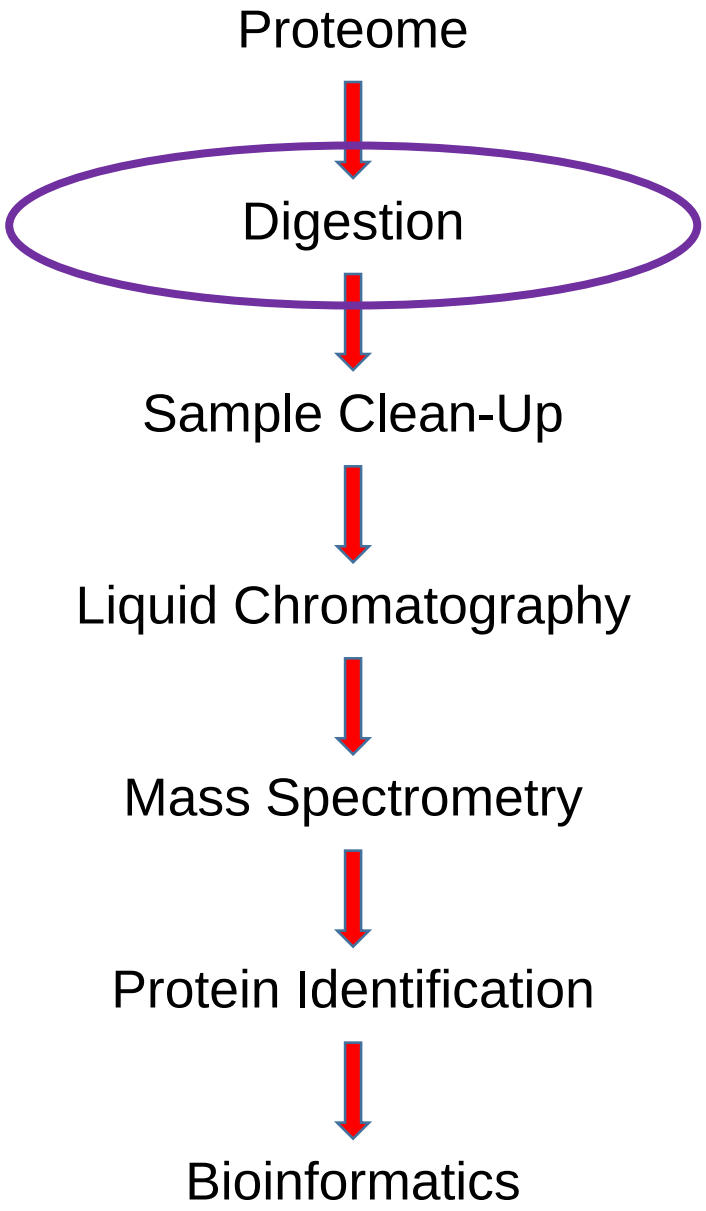
Can be performed under native or denaturing conditions

2) **“bottom-up” or “shotgun sequencing”**

Proteins are digested into peptides with either enzymatic or chemical methods and the peptides are analyzed

Always a denaturing method

Standard bottom-up workflow



Sample preparation for bottom-up proteomics

1) Gel electrophoresis followed by in-gel trypsin digestion

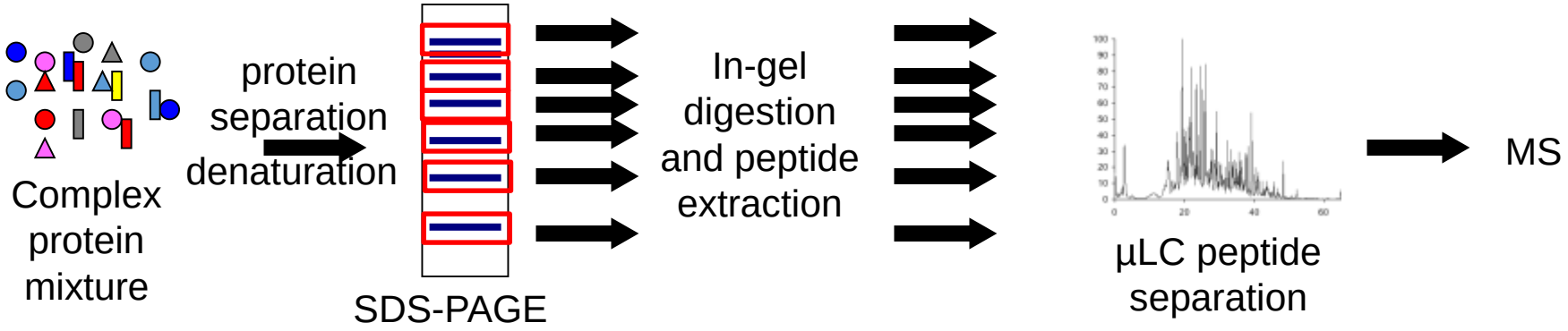
Shevchenko et al. Nature Protocols 2007 1: 2856-60

2) In-solution digestion followed by peptide fractionation

Wiśniewski Methods Enzymol. 2017 585:15-27

Separation at either the protein or peptide level facilities depth of coverage of identifications.

Sample preparation using 1D gels



Why is this one of the most common ways to analyze samples?

- Gel removes any possible contaminants from sample
- Gel staining allows for visual confirmation of sample
- Separation of proteins by MW allows for more identifications

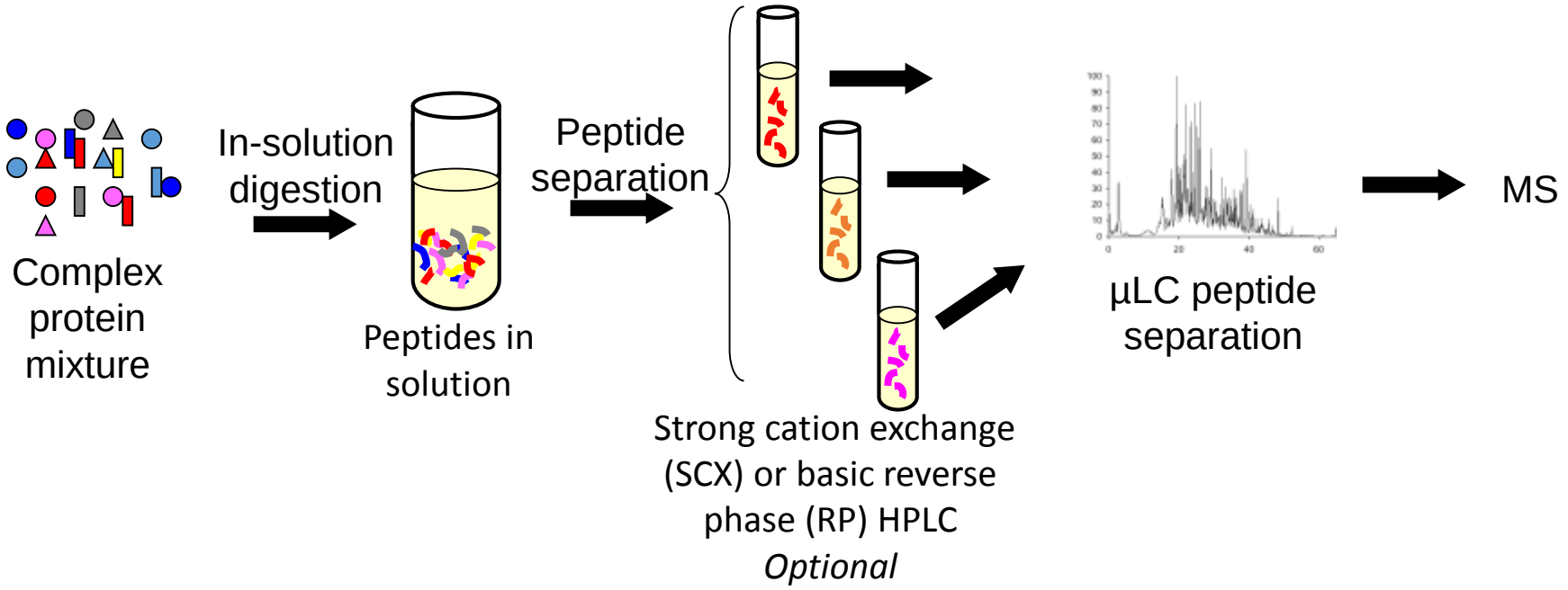
What are the problems with this method

- Peptide extraction is variable and inefficient
- Can be very time consuming
- Hydrophobic peptides don't extract well
- Keratin contamination can be a problem

Standard Protocol

http://www.abrf.org/archives/hmail/97028/0004.html

In-solution digestion and separation



Pros:

- Reduces sample loss
- Identify more proteins than gel systems
- Easily coupled to quantitation

Cons:

- Reagents necessary for solubility interfere with digestion
- Some proteins may elute with difficulty from SCX or RP

Proteolytic digestion

Trypsin is the most common enzyme used to digest proteins

- Cleaves at arginine (R) and lysine (K), except when followed by proline
- Leaves a basic charge on the peptide, which improves ionization into the mass spectrometer
- Most proteins have a good proportion of K and R residues

However, there are some limitations:

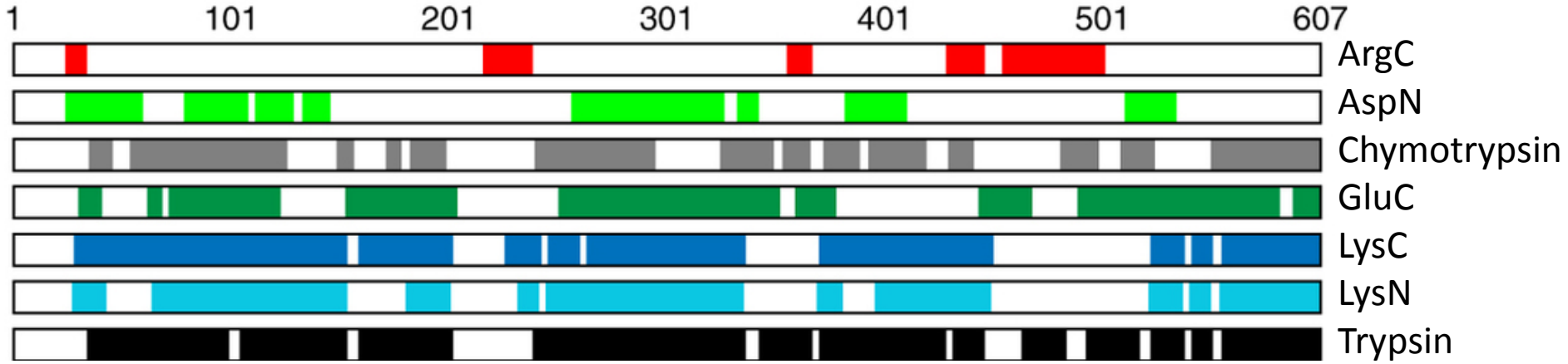
- Tryptic peptides are usually short, so trypsin alone covers only a restricted portion of the proteome
- Presence of negatively-charged amino acids (D, E, pS, pT) in close proximity to R or K prevents tryptic cleavage and leads to missed cleavages and longer peptides
- Trypsin exhibits a lower cleavage efficiency toward K than R
- Often unable to produce MS-identifiable peptides derived from the C-termini of proteins

Other proteolytic enzymes

Protease	Cleavage site	Advantages	Limitation
ArgC	C-terminal of R	ArgC is mostly combined with other proteases to investigate PTMs and to increase the proteome coverage qualitatively	ArgC peptides are generally longer than tryptic peptides
AspN	N-terminal of D	AspN can perform hydrolysis of peptide bonds at the amine side of D residues. It also functions within a pH range of 4–9	If detergents are present in the digestion buffer, Asp-N can cleave at the amine side of E residues AspN cleaves more efficiently at the N termini of D than E residues, resulting in many missed cleavages AspN peptides are generally longer than tryptic peptides
Chymotrypsin	C-terminal of F, Y, L, W and M	Chymotryptic peptides cover a proteome space that is most orthogonal to that of trypsin, in both a qualitative and quantitative manner. Chymotrypsin is particularly useful for covering transmembrane regions of membrane proteins	The efficiency of chymotrypsin toward different hydrophobic amino acid residues varies and results in quite a few missed cleavages
GluC	C-terminal of D	GluC can be combined with other proteases for the study of PTMs and to increase proteome coverage qualitatively	Specificity of GluC depends on the pH and the buffer composition. At pH 4, the enzyme preferentially cleaves at the C terminus of E, whereas at pH 8 it additionally cleaves at D residues GluC peptides are generally longer than tryptic peptides
LysC	C-terminal of K	Often used to complement trypsin in a serial LysC > trypsin digestion protocol to complement the somewhat lower efficiency of trypsin toward K residues LysC is resistant to denaturants (such as 8 M urea). This allows proteins to be digested in their optimal denatured state which enhances digestion efficiency	Peptides generated by LysC alone overlap significantly with tryptic peptides, and therefore sequence coverage of proteins may not increase significantly LysC peptides are generally longer than tryptic peptides
LysN	N-terminal of K	LysN is more resistant to denaturants than trypsin. It may also be heated to 70 °C The combination of LysN with ETD peptide fragmentation provides unique and straightforward sequence interpretation, and it allows facile de novo sequencing	Occasionally, LysN also cleaves N-terminally to A, S and R

Giansanti *Nature Protocols* 2016 11:993-1006

Comparison of enzyme cleavages for BSA digest

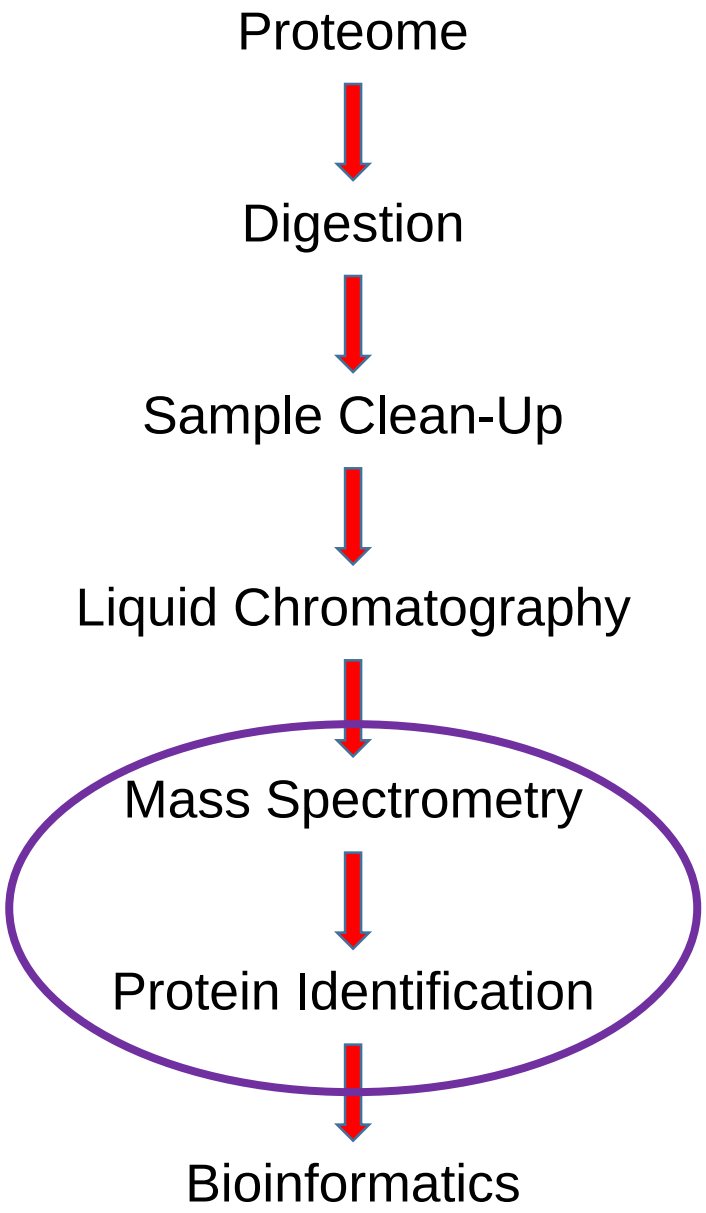


Trypsin shows the best overall coverage, followed by chymotrypsin. There are area that only one enzyme can resolve, and areas that are unmappable.

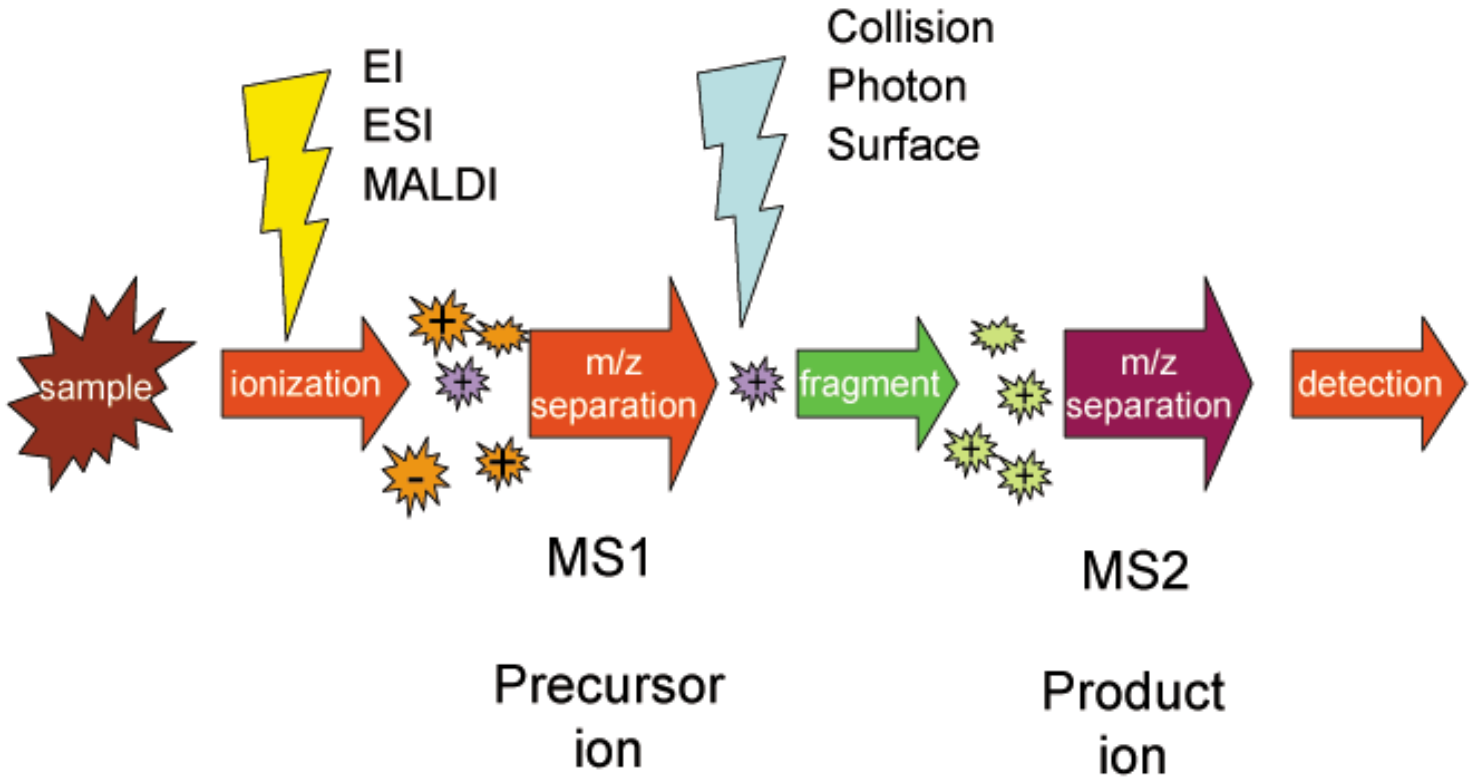
Sometimes combination of enzymes, either on separate samples or combined in the same tube, is needed for optimal coverage.

Giansanti *Nature Protocols* 2016 11:993-1006

Standard bottom-up workflow

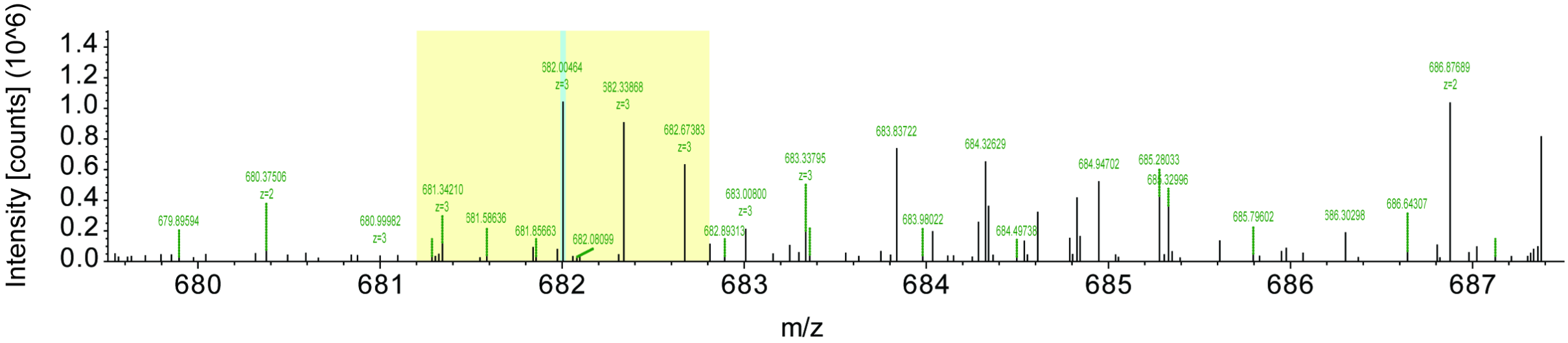


General schematic for analysis using a mass spectrometer



K. Murray (Kkmurray), https://commons.wikimedia.org/w/index.php?curid=1943319

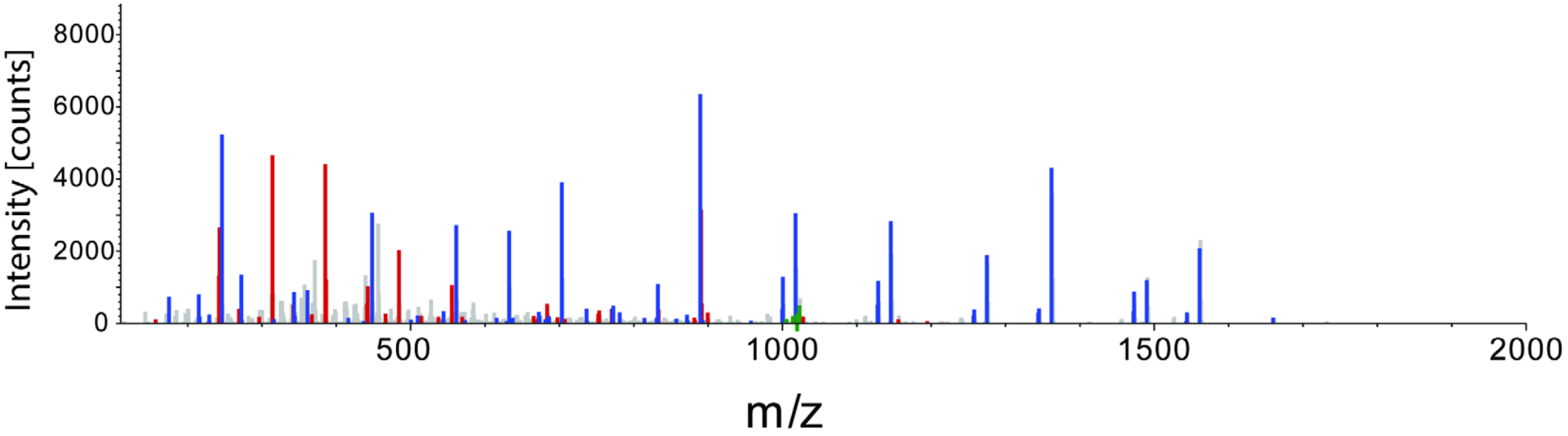
Example mass spectrum



In a mass spectrum, the x axis is always m/z (mass/charge). Depending on the size of the molecule and the type of ionization, the charge state can vary from 1+ to 8+ or higher. Higher charge state molecules can fold into a smaller m/z range (for example: 1000 Da peptide at z=2+ will be observed at 500 m/z).

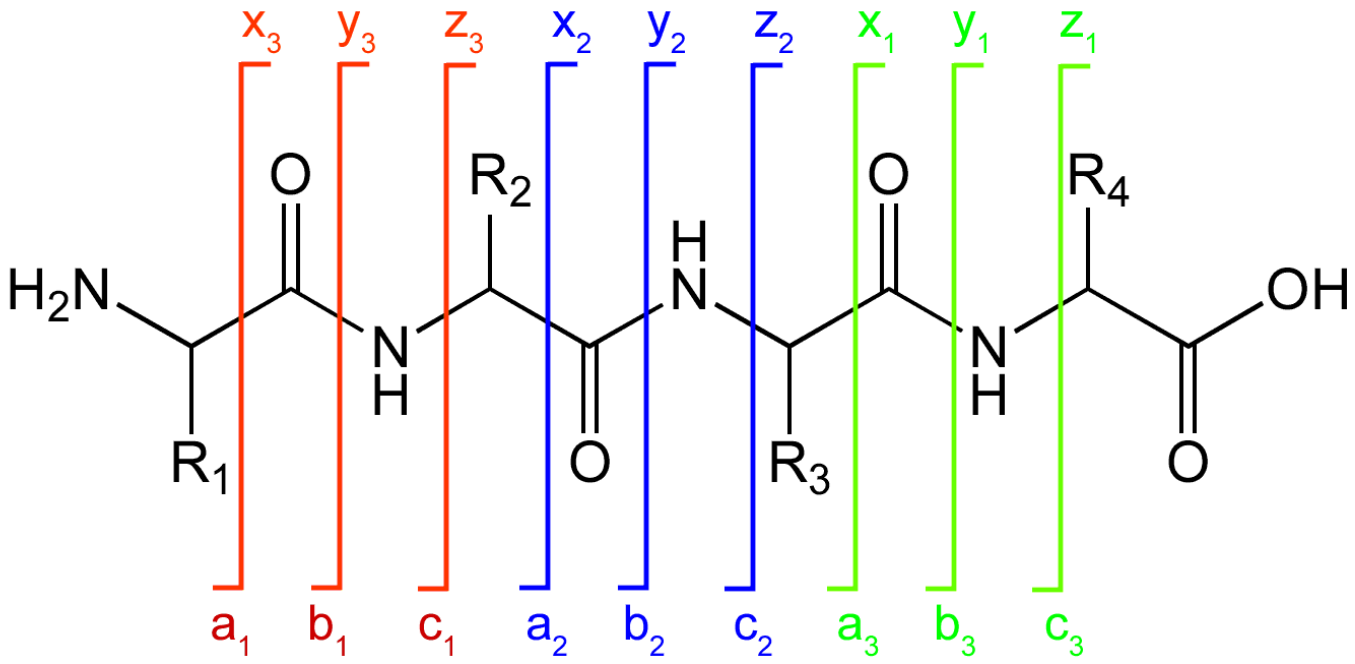
The y axis can be % abundance, scaled such that the most intense peak is 100%, or intensity (ion counts).

Example MS/MS spectrum



So, how do we go from an MS/MS to a peptide sequence?

Peptide sequencing by tandem MS



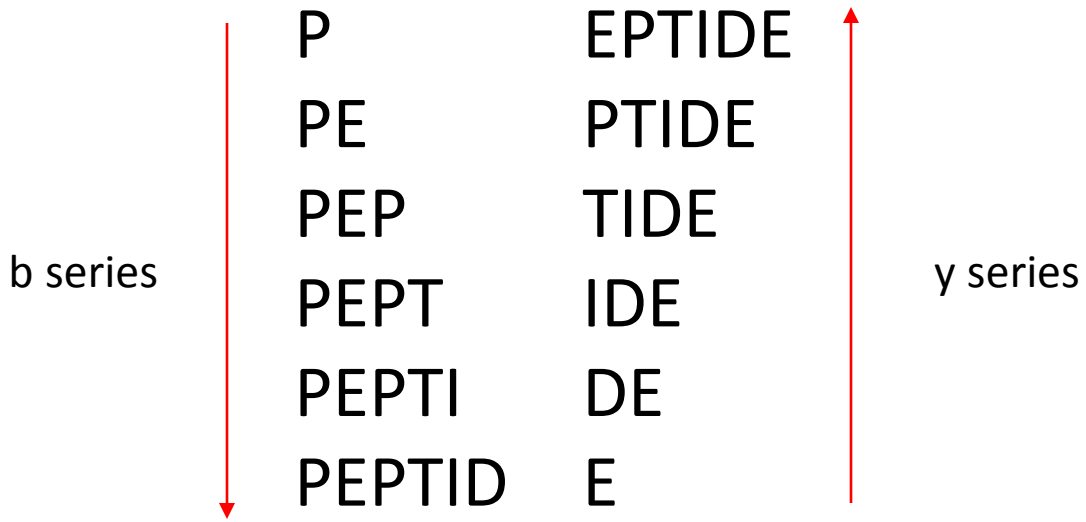
The fragment ions that are formed in the mass spectrometer depends on the method of fragmentation:

Collision with He or N_2 gas (CID/HCD): b/y ions

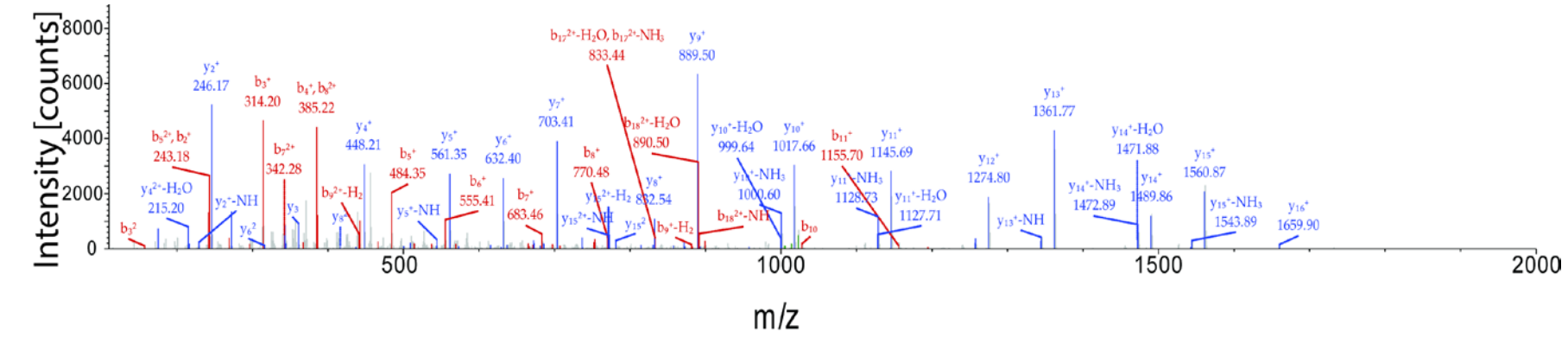
Electron-transfer dissociation (ETD): c/z ions

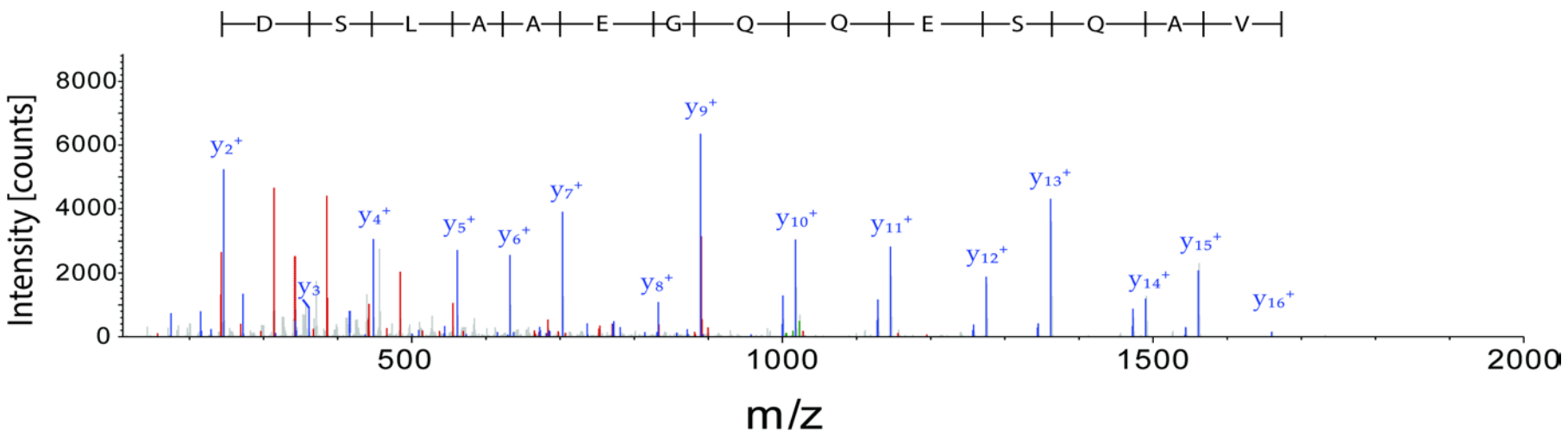
Photodissociation (UVPD): a, b, c, x, y, z ions

P E P T I D E


b series
P
PE
PEP
PEPT
PEPTI
PEPTID
EPTIDE
PTIDE
TIDE
IDE
DE
E
y series

Revisiting the MS/MS spectrum





b^+	b^{2+}	Seq.	y^+	y^{2+}
114.09	57.55	L		
243.13	122.07	E	1930.91	965.96
314.17	**157.59**	A	1801.87	901.44
385.21	193.11	A	1730.83	865.92
484.28	**242.64**	V	**1659.79**	**830.40**
555.31	278.16	A	**1560.72**	**780.87**
683.37	**342.19**	Q	**1489.69**	745.35
770.40	**385.71**	S	**1361.63**	**681.32**
899.45	450.23	E	**1274.60**	**637.80**
1027.51	**514.26**	Q	**1145.55**	**573.28**
1155.56	578.29	Q	**1017.50**	**509.25**
1212.59	606.80	G	**889.44**	445.22
1341.63	**671.32**	E	**832.42**	**416.71**
1412.67	**706.84**	A	**703.37**	352.19
1483.70	742.35	A	**632.34**	**316.67**
1596.79	798.90	L	**561.30**	281.15
1683.82	842.41	S	**448.22**	224.61
1798.85	**899.93**	D	**361.18**	181.10
1869.88	935.44	A	**246.16**	123.58
		R	**175.12**	88.06

Automated analysis of MS data

- There are several algorithms that are used to search MS-based proteomics data:
  - Sequest
  - Mascot
  - XTandem! (free)
  - Andromeda (free)
  - MSAmanda (free)
  - OMSSA (free)

- These algorithms are packaged into various software programs:
  - Proteome Discoverer
  - MaxQuant (free)
  - Trans-Proteome Pipeline
  - Progenesis
  - Scaffold

- They share a common basis:
  - match the experimental spectra to predicted spectra based upon an in silico digest of the proteins, with scoring for how well the experimental and predicted spectra match

Applications of bottom-up analysis

Identification of proteins

- What binds to my protein/nucleic acid of interest?

Characterization of sites of post-translational modification

- What sites on my protein are phosphorylated/acetylated/methylated/ ubiquitinated/react with a small molecule?

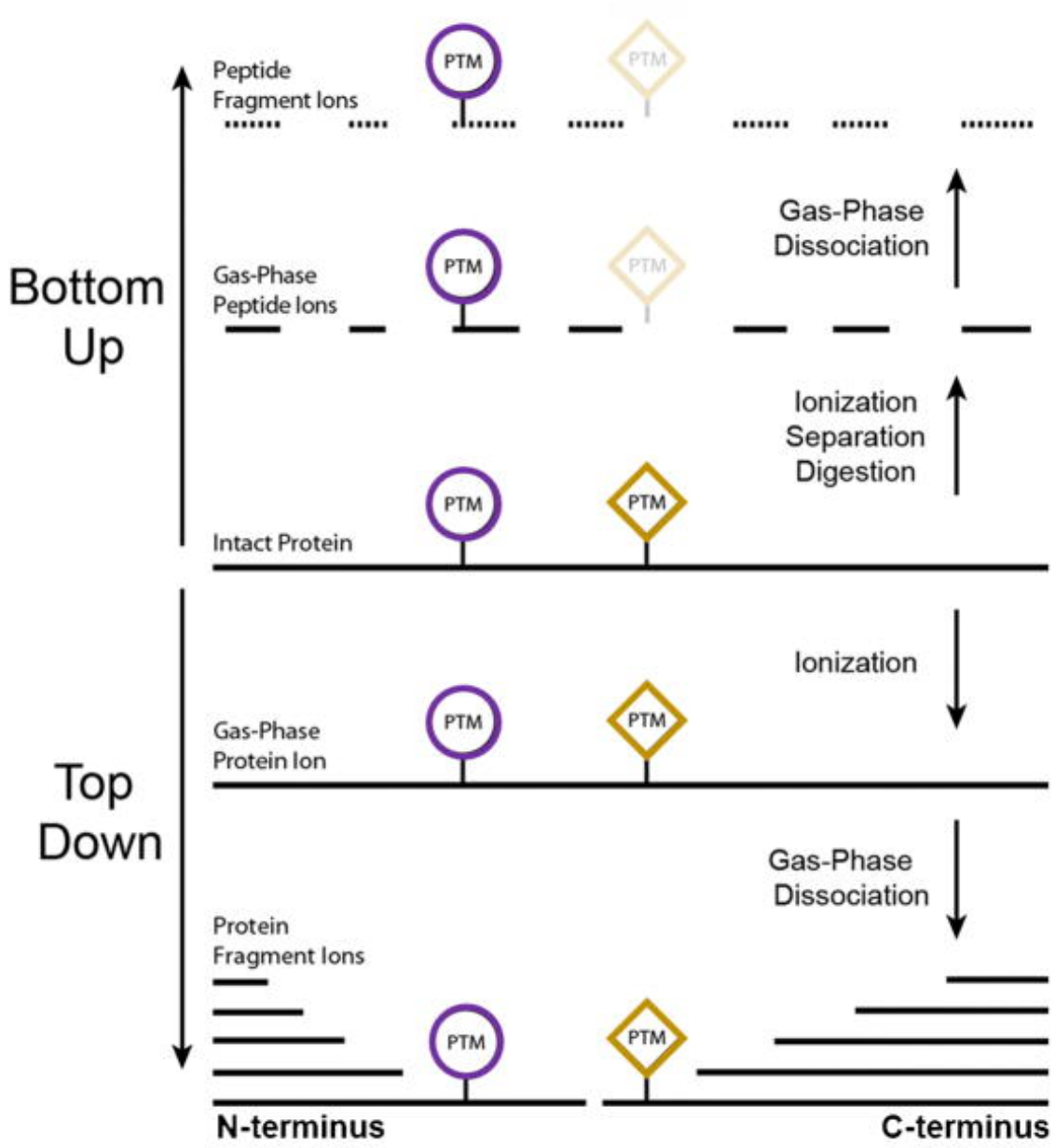
Quantitation of protein level changes

- What proteins are affected by knockdown/overexpression/inhibition of my protein?

Crosslinking analysis/structural mass spectrometry

- Where are the disulfide bonds in my protein?
- Where does my protein bind another protein?

Top-down proteomics



In top-down proteomics experiments, intact proteins are ionized into the mass spectrometer.

Dissociation of the protein is performed in the gas phase to produce fragment ions that can be used for sequencing of the protein.

Benefits:

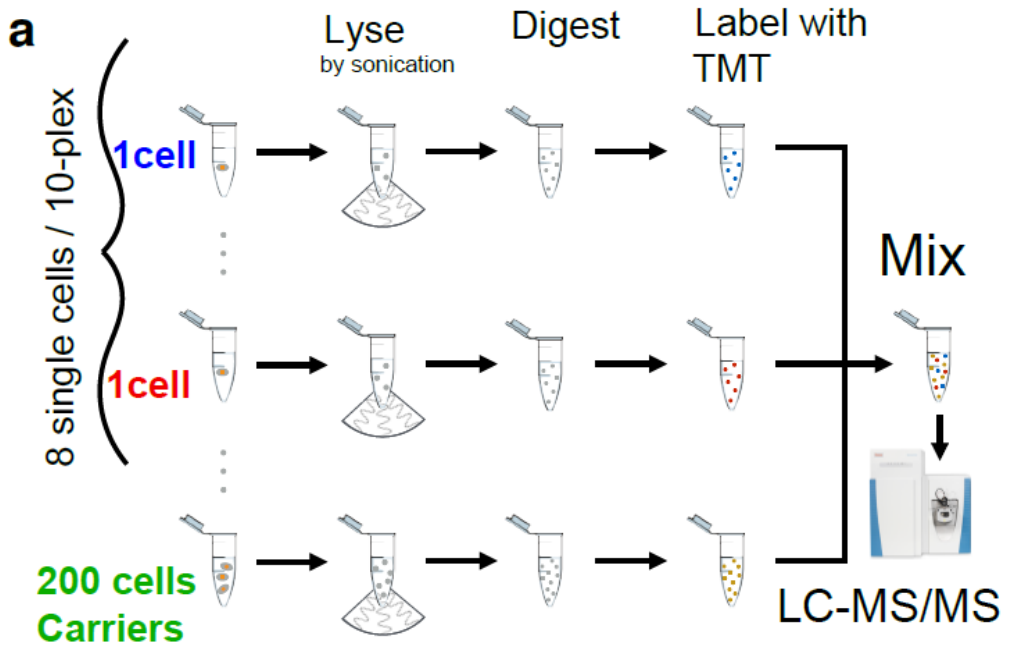
- Information about concurrent PTMs
- Amenable to studying splice variants, native forms of proteins (for example, metal-bound forms)

Limitations:

- Depending on proteins size, it may not be fully fragmented, so there may only be sequence information on the N- and C-termini
- Typically only analyzing 10-100s of proteins at once

Catherman Biochem Biophys Res Commun 2014 445:683–693

Approaching single-cell mass spectrometry



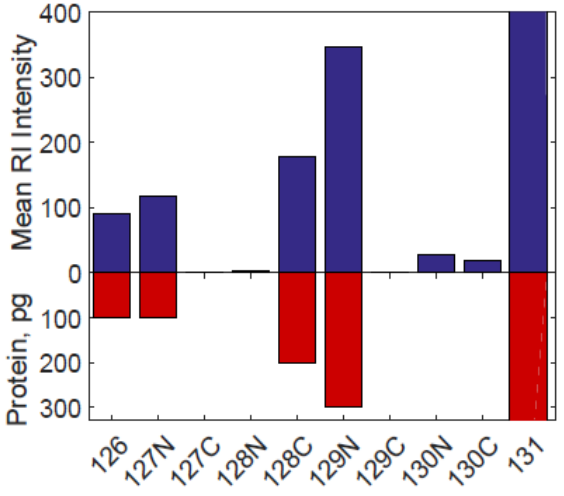
In the past, most single-cell mass spectrometry experiments have used large cells - Xenopus embryos.

Recently, the use of carrier cells combined with isobaric labeling has been suggested

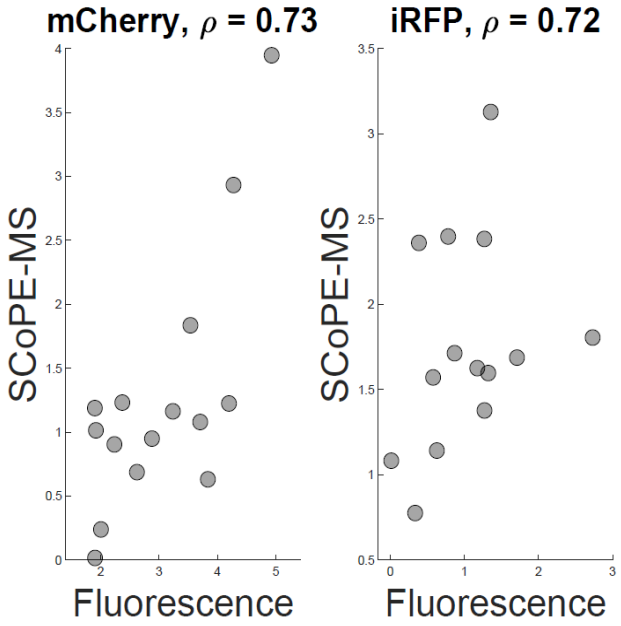
1) Carrier proteins provide peptide sequence identification
2) Isobaric label provides relative quantitation in the single cell relative to carrier population

Budnik bioRxiv 2017 http://dx.doi.org/10.1101/102681

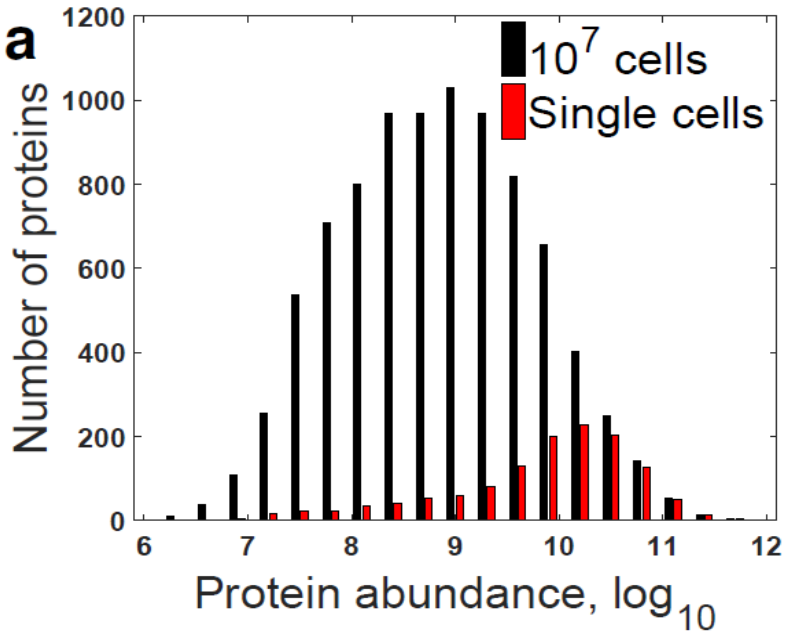
Approaching single-cell mass spectrometry



Signals for low protein amounts were above the background noise, suggesting that they could be reliably used for quantitation.



Using this method, 767 proteins could be quantified, with levels similar to that determined by FACS



Quantified proteins did skew towards higher cellular abundance.

Budnik bioRxiv 2017 http://dx.doi.org/10.1101/102681

Tips for real life experiments

- Ask a question and decide what type of results you need – the specific question determines sample preparation
  - Site of PTM?
  - Integral membrane proteins?

- Test trials on practice sample
  - Different digestion, fractionation strategies

- Discuss with your mass spectrometrist

Common contaminants and considerations

- Keratins (from skin, hair, clothes) are everywhere
- Most noticeable with small amounts of sample
- Best practices:
  - always wear gloves (nitrile)
  - use HPLC grade or filtered reagents
  - For in-gel digestions, keep separate staining trays
  - For the uber-strict, only open tubes in a laminar flow hood

For specific questions or projects, please feel free to contact me:

Building 37, Room 2140

jenkinsl@mail.nih.gov

240-760-7209

NATIONAL CANCER INSTITUTE
Center for Cancer Research

NCI Core Open House

March 7, 2018 from 2:30-4:00 PM

Bldg. 35 Atrium

Light refreshments provided

Meet with Core managers from Bldg. 37, Bldg. 41 and Frederick Cores to discuss projects and view posters summarizing capabilities and resources available to all CCR investigators.