SV PTECH PTP-PCF

Lab Methods (BIOENG-519) Methods: omics methods in biomedical research

Proteomics

Maria Pavlou, PhD

Head of Proteomics Core Facility

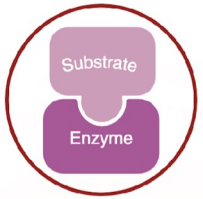
Name your favorite protein

Maria Pavlou

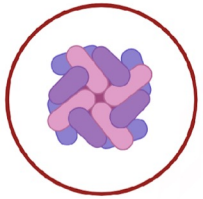
Maria Pavlou

Meet the most famous ones

Enzymes
(e.x. lipases, proteases)

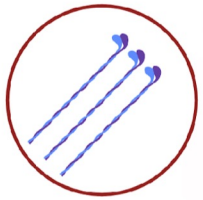


Storage
(e.x. ferritin, albumin)

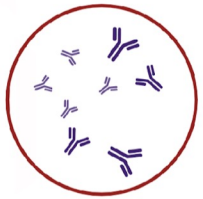


Hormones
(e.x. insulin, oxytocin)

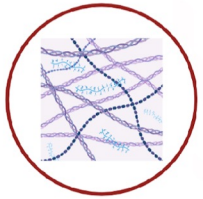
Motor
(e.x. actin, myosin)



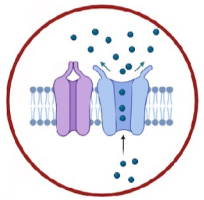
Defence
(e.x. antibodies)

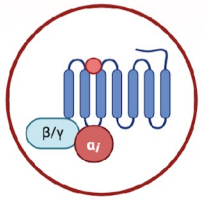


Structure
(e.x. collagen, keratins)



Transport
(e.x. channel proteins)





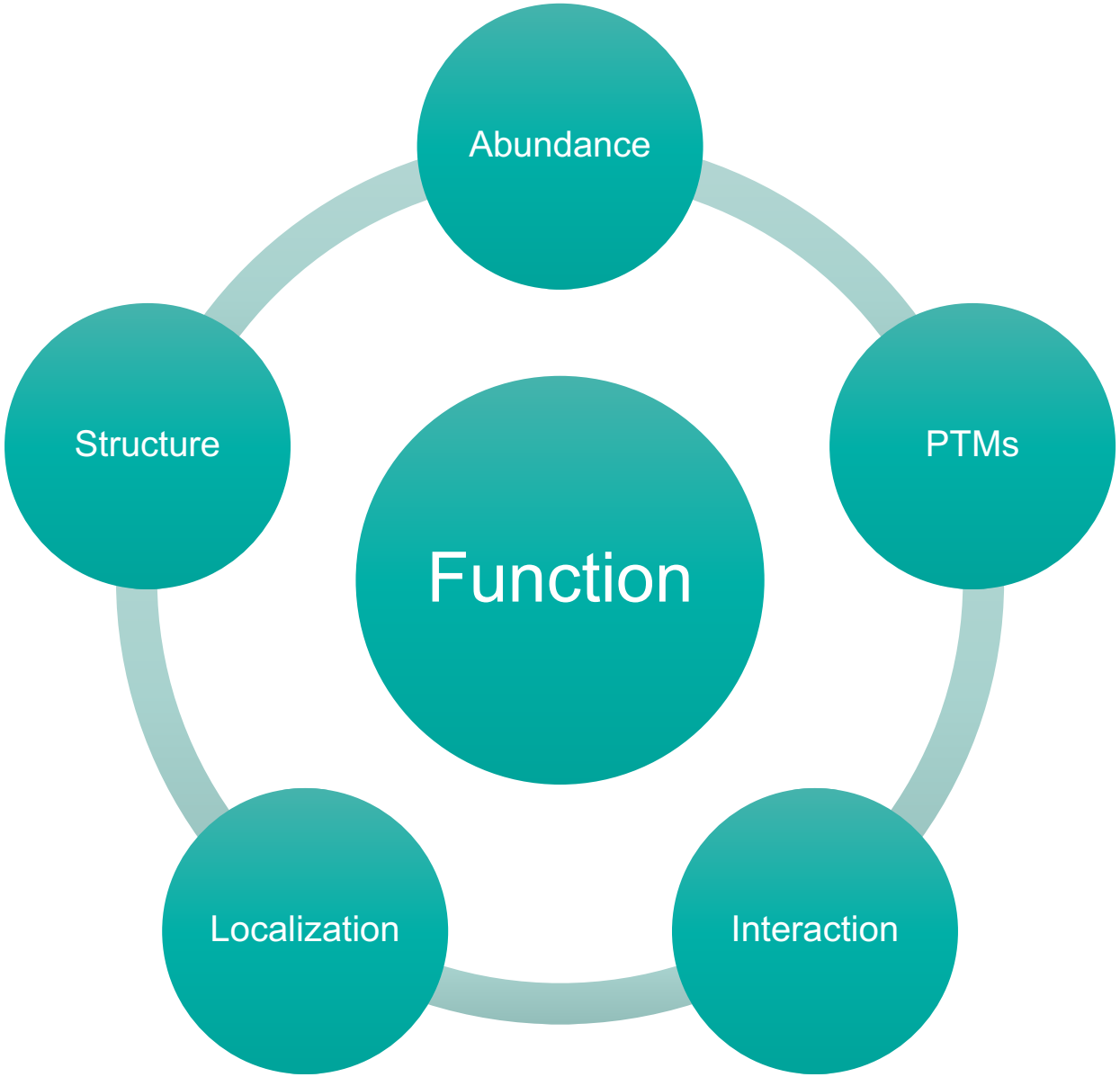
Receptors
(e.x. GPCRs)

Super dynamic molecules

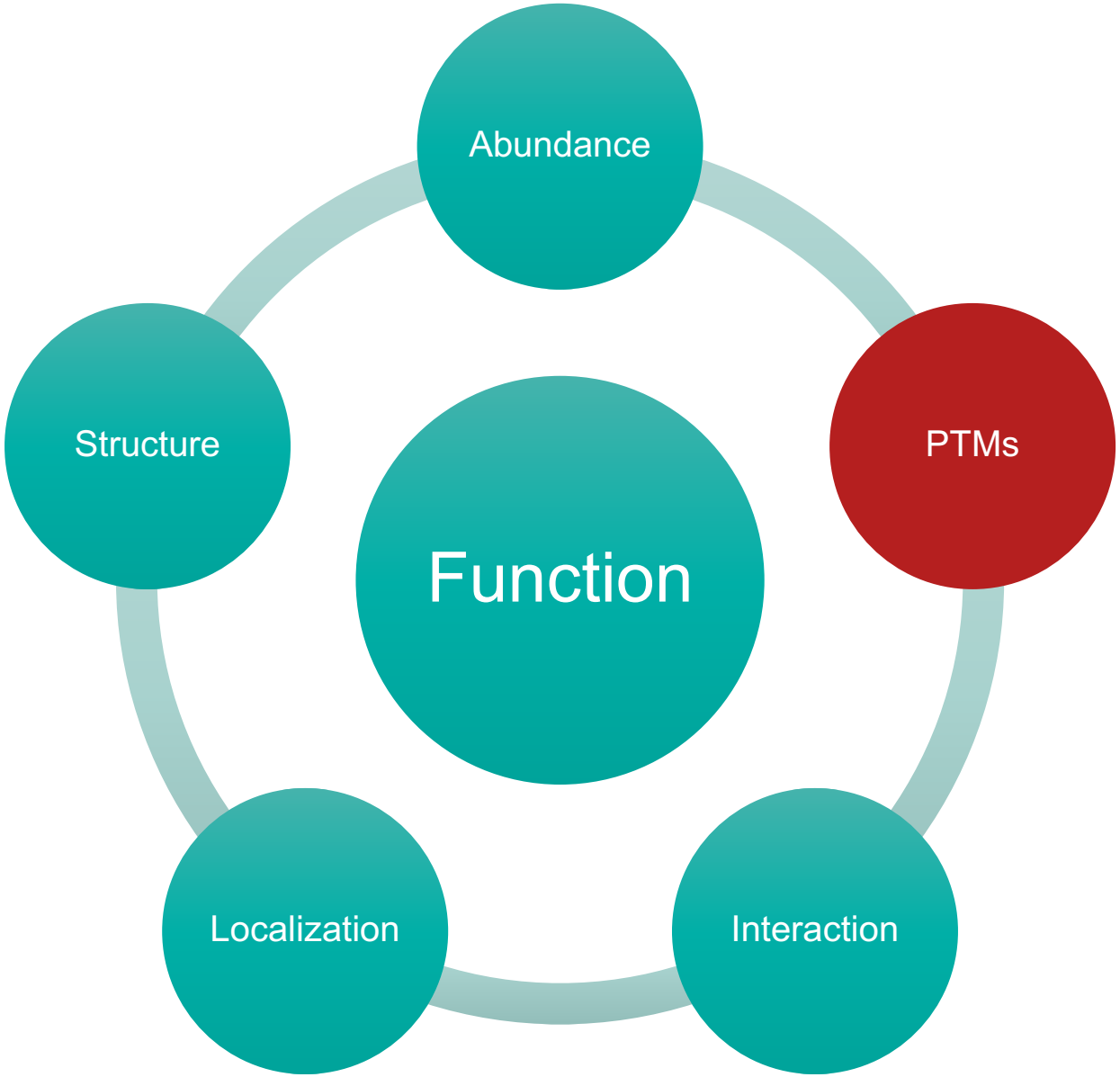
- https://youtu.be/MZ47-G4XKDw

Protein function is multi-faceted

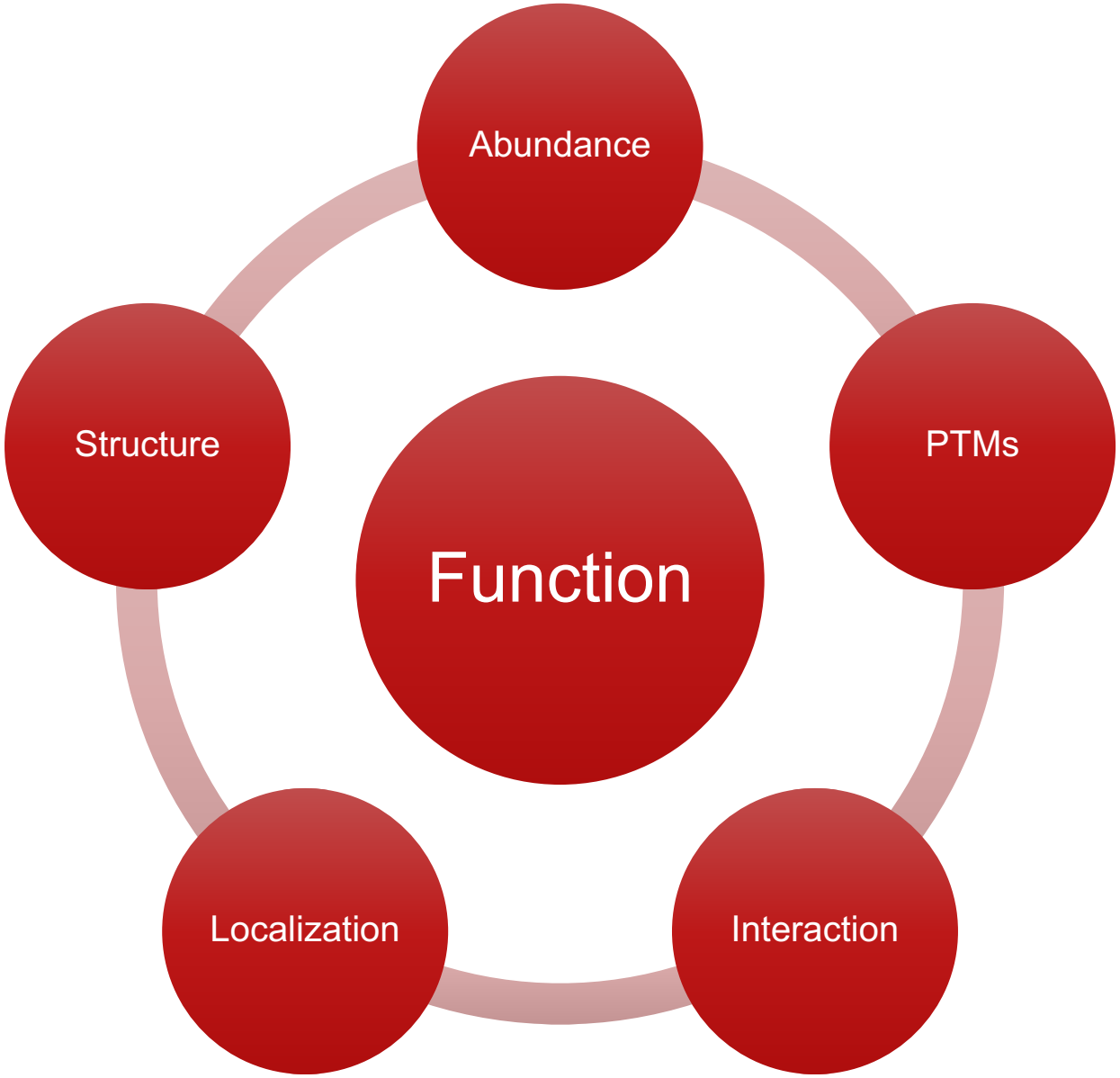
Protein function is multi-faceted



Protein function is multi-faceted



Protein function is multi-faceted



Proteomes and Proteomics

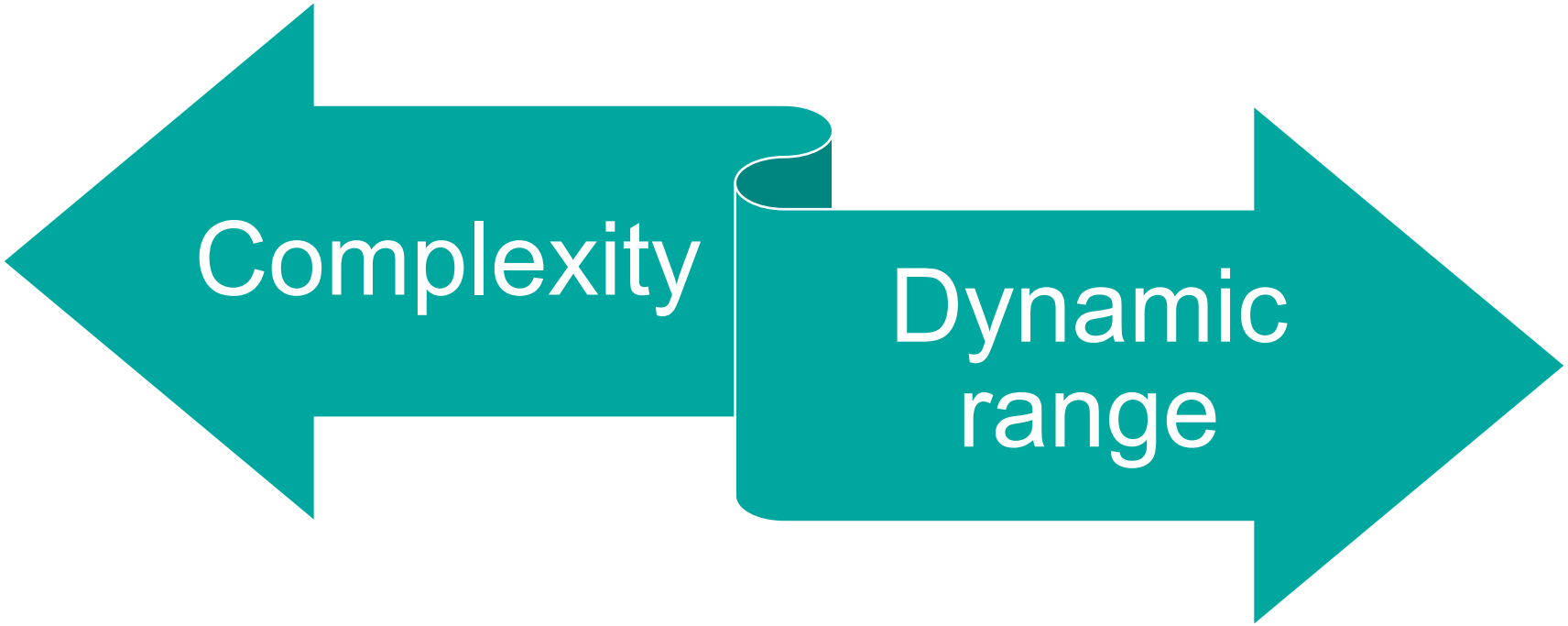
"**Proteome**": **PROTE**ins expressed by a gen**OME**

"**Proteomics**": methods (-omics) dedicated to the analysis of **proteomes**

Represents the effort to establish the *identities*, *quantities*, *structures*, and *biochemical and cellular functions* of all proteins in an organism, organ, or organelle, and how these properties vary in space, time, or physiological state.

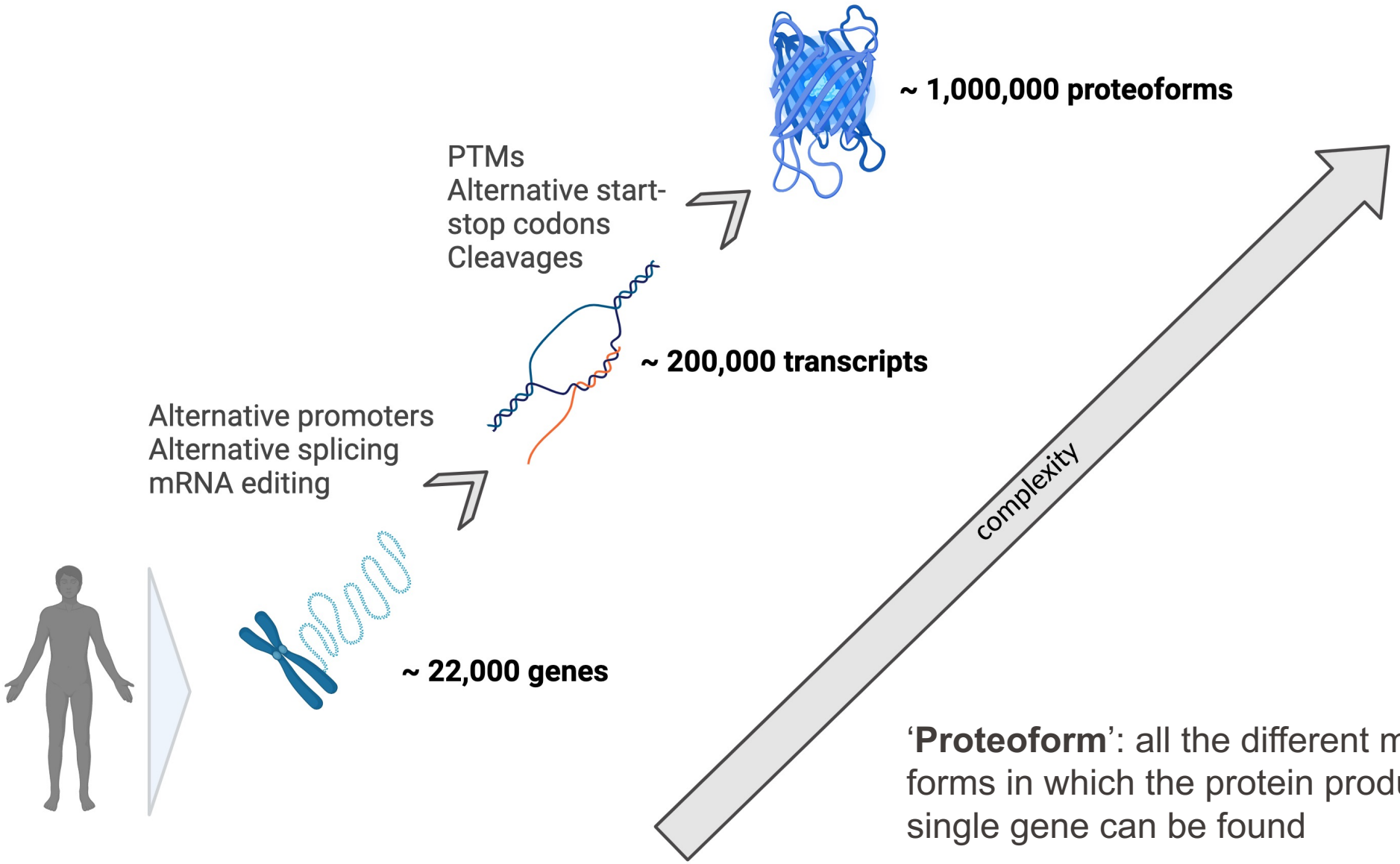
MCP 1.10 pg 675 National Research Council Steering committee

Main challenges of proteomics



Maria Pavlou

Increasing complexity

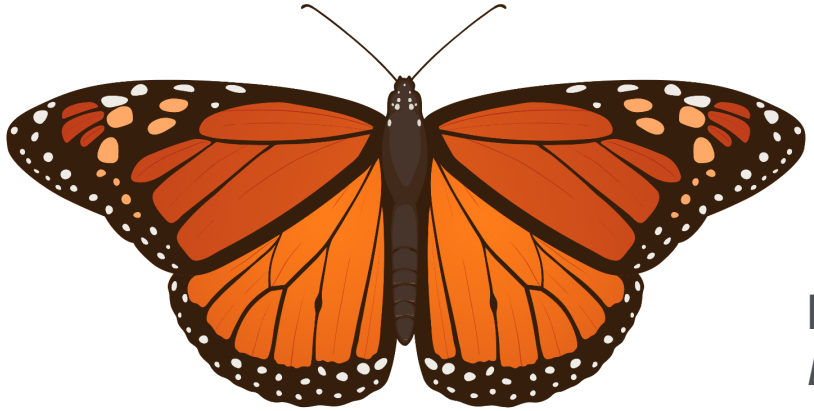


'**Proteoform**': all the different molecular forms in which the protein product of a single gene can be found

Nat Methods. 2013 Mar; 10(3): 186–187.

Same genome- different proteome- different phenotype

Maria Pavlou

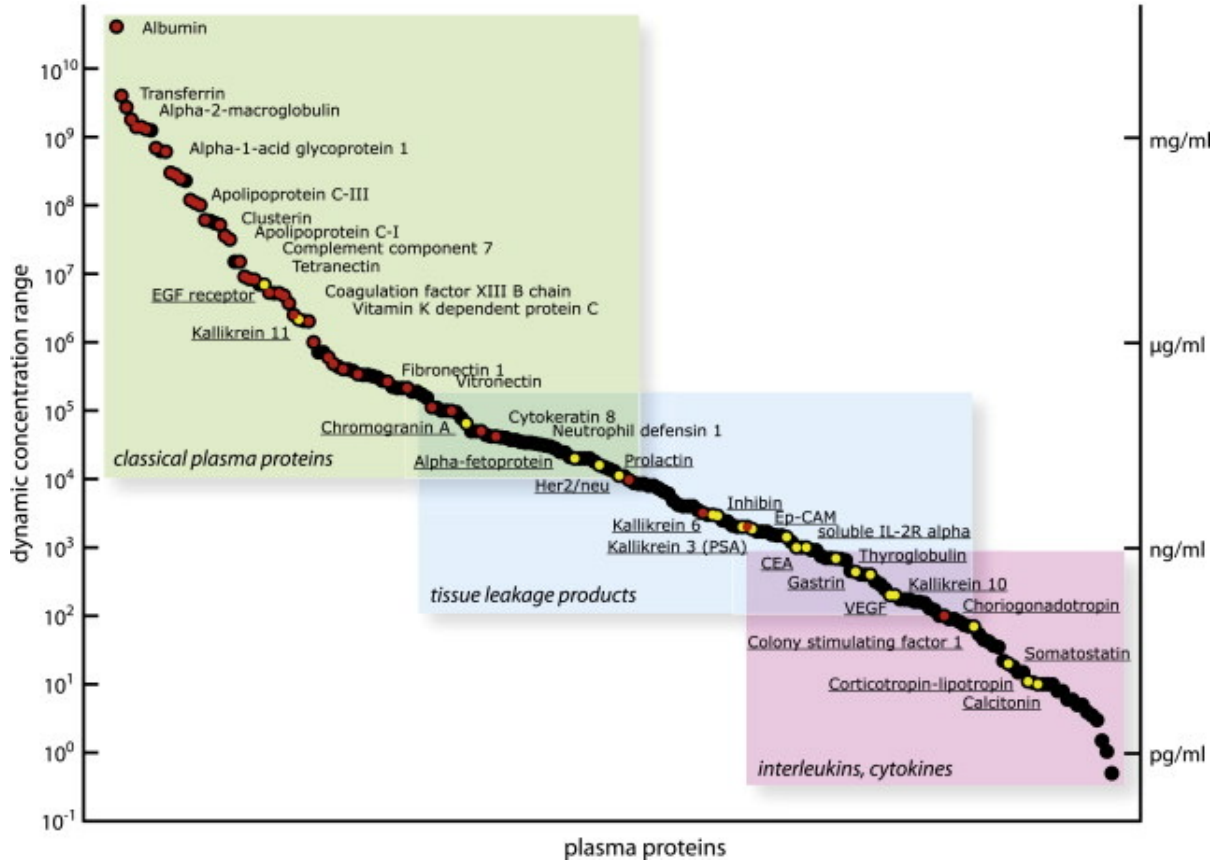


Monarch butterfly
Danaus plexippus

Maria Pavlou

Dynamic range

- Factors of 10^5 to 10^{10} between low and high-abundance proteins observed in biological samples (especially plasma)



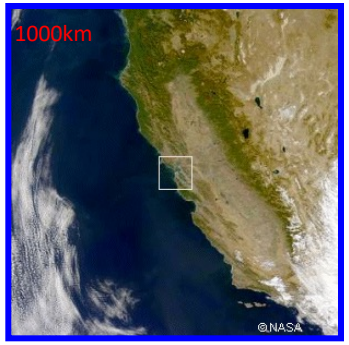
- https://doi.org/10.1016/j.molonc.2008.12.001

Maria Pavlou

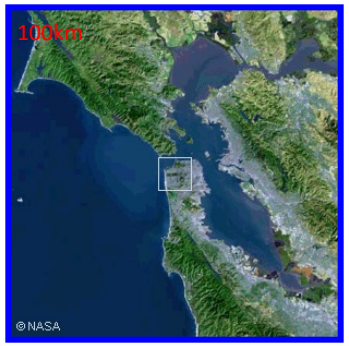
Can you see the bees on the flowers while you are landing?

10^{10} is pretty large dynamic range…!

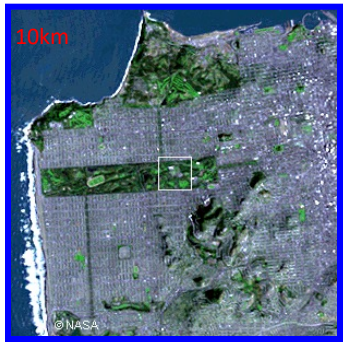
10



9



8

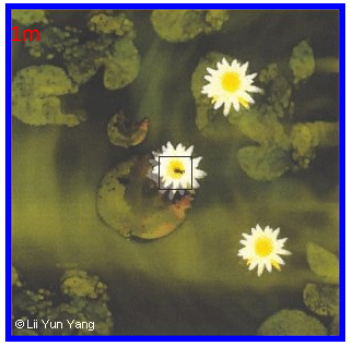


7

6

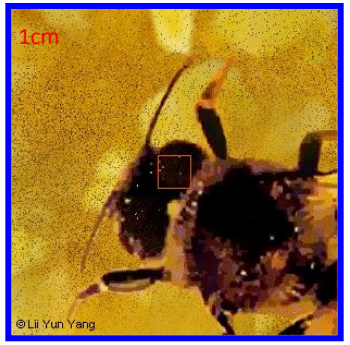
5

4



3

2

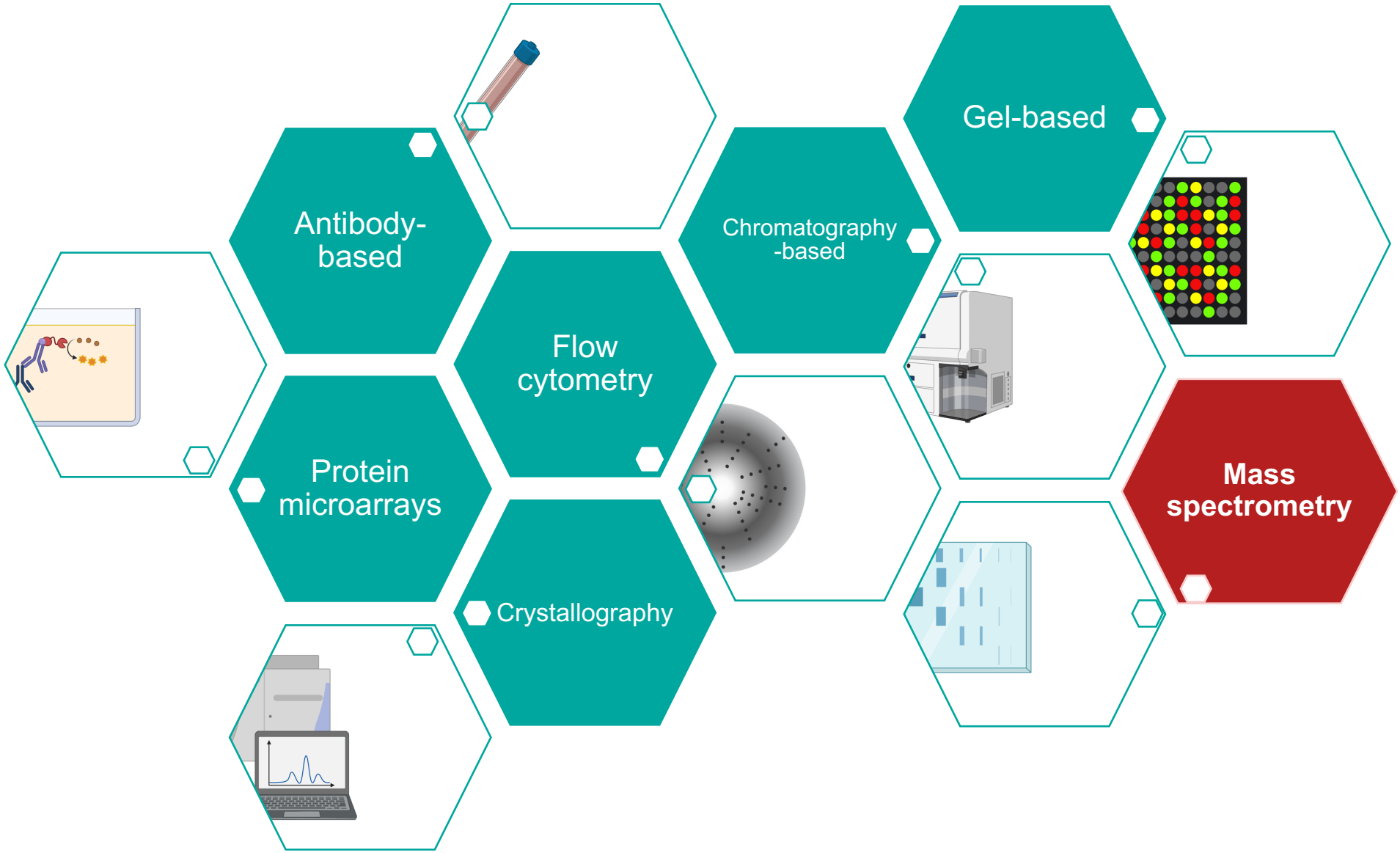


1

…and there is no PCR for proteins!

Slide courtesy Bruno Domon, ETH Zurich

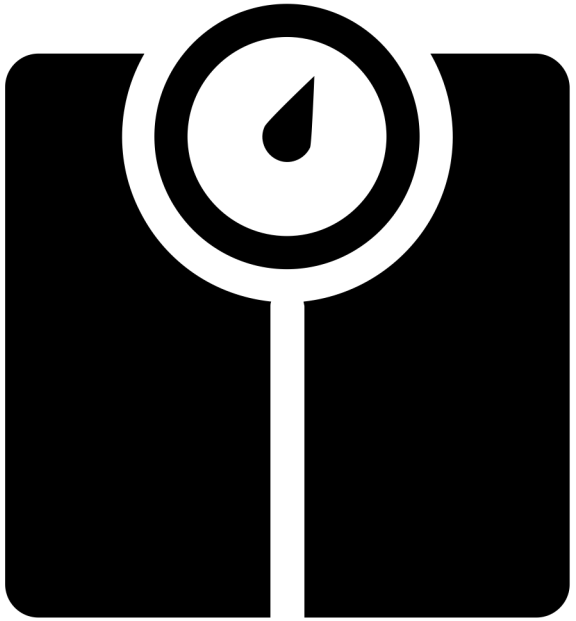
Proteomic tools



Maria Pavlou

What is a mass spectrometer?

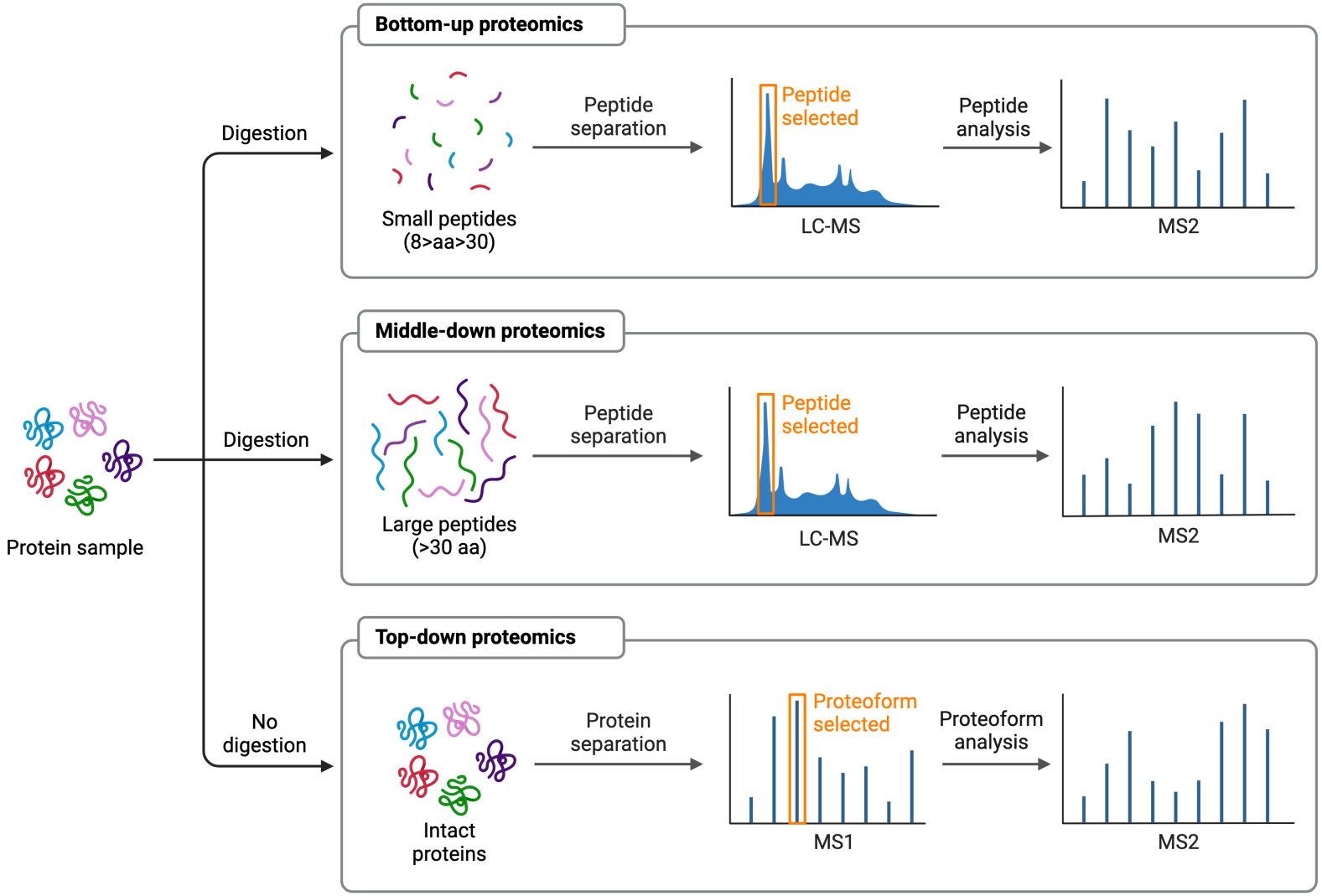
A Mass Spectrometer (MS) measures the mass-to-charge ratio (m/z) of ions



Molecular Scale

Maria Pavlou

MS-based proteomics workflows
Top or bottom? Up or down?



SV
PCF

SB
MSEAP

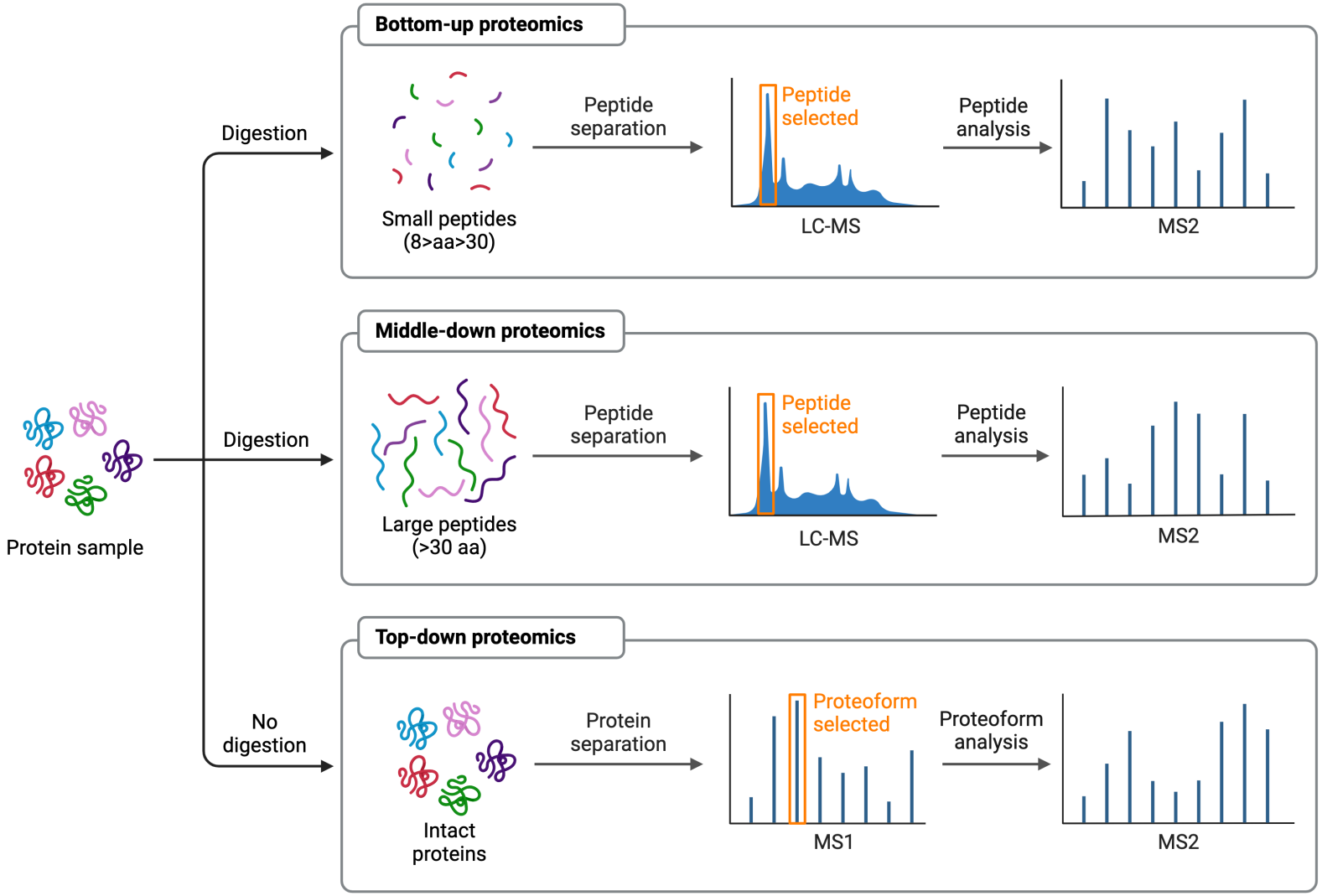
Top Down prons and cons

- Identification of proteoforms
- De novo sequencing
- Rich information, less false positives

- Limited sensitivity and throughput
- Pure samples required
- Insoluble proteins and big proteins difficult to be
- analysed
- Highly sophisticated instrumentation

Maria Pavlou

MS-based proteomics workflows
Top or bottom? Up or down?



SV
PCF

SB
MSEAP

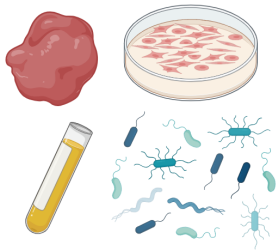
Bottom Up prons and cons

- Simpler
- Higher-throughput
- Less sophisticated instrumentation
- Applicable for “tough” proteins
- Peptide separation is easier

- PTM and isoform information is often lost
- Good “flying” peptides have to be generated
- Protein inference based on peptides can be tricky

Maria Pavlou

Typical bottom-up workflow



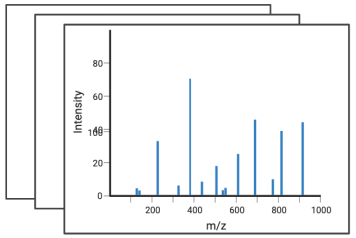
Protein extraction

Denatured proteins

Enzymatic digestion

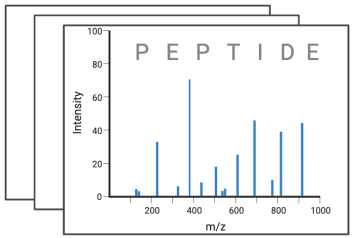
Peptides

Liquid chromatography tandem MS



raw data

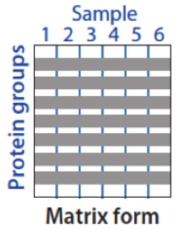
Database search



Peptide sequence

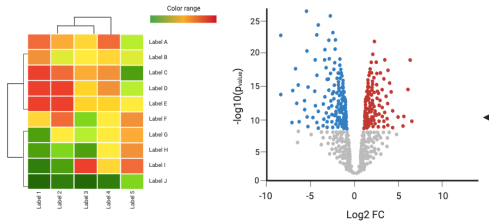
Protein inference

Quantitation

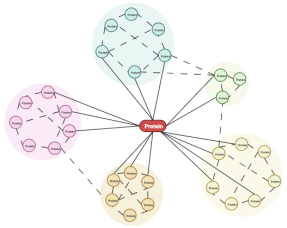


Matrix form

Statistical analysis



Biological interpretation



Maria Pavlou

Protein extraction

Physical disruption

- Sonication
- Bead-beating
- Freeze-thaw
- Grinding

Detergents and chaotropic substances

- Protein extraction
- Protein solubilisation

Detergents and MS

Common detergents are incompatible with

1. Reverse phase liquid chromatography (compromise fractionation)
2. Mass spectrometry (ion suppression)

Detergent removal

✓Dialysis

✓Filtration

✓Electrophoresis (eg. SDS)

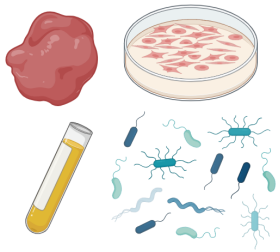
✓Protein precipitation

✓Dilution (eg. Urea)

MS-compatible detergents

Maria Pavlou

Typical bottom-up workflow



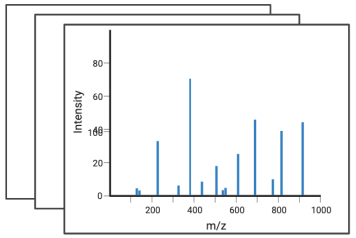
Protein extraction

Denatured proteins

Enzymatic digestion

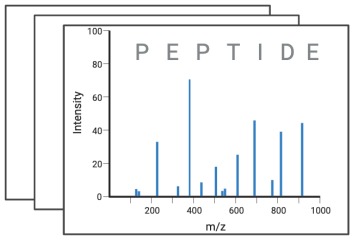
Peptides

Liquid chromatography tandem MS



raw data

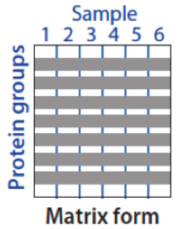
Database search



Peptide sequence

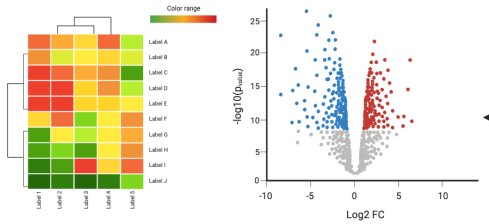
Protein inference

Quantitation

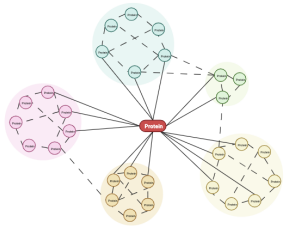


Matrix form

Statistical analysis



Biological interpretation



Trypsin- the star of proteases

- ✓Highly specific and efficient; C-term of the basic residues Lysine and Arginine (except when followed by Proline)
- ✓Lys and Arg are relatively abundant and usually well distributed throughout a protein; many peptides of MS-reasonable size
- ✓Relatively cheap
- ✓Produces peptides with at least two charges (important for ionization)

Maria Pavlou

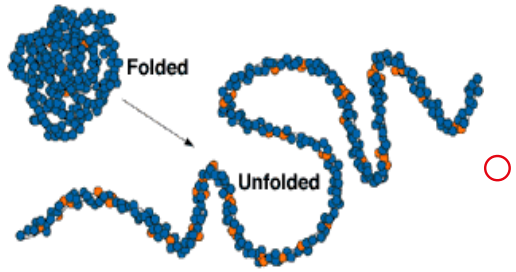
Protein digestion: why not trypsin?

Protease	Organism	Enzyme family	Specificity	pH range
Arg-C	Clostridium histolyticum	Cysteine-protease	R’	7.2-8
Asp-N	Pseudomonas fragi	Metallo-protease	‘D	7-8
Glu-C	Staphylococcus aureus	Serine-protease	E’	4-7.8
Lys-C	Lysobacter enzymogenes	Serine-protease	K’	8.5-8.8
Lys-N	Lysobacter enzymogenes	Metallo-protease	‘K	8
Trypsin	Bos taurus	Serine-protease	K’, R’	7.5-9
Chymotrypsin	Bos taurus	Serine-protease	F’, W’, Y’	7-9

Maria Pavlou

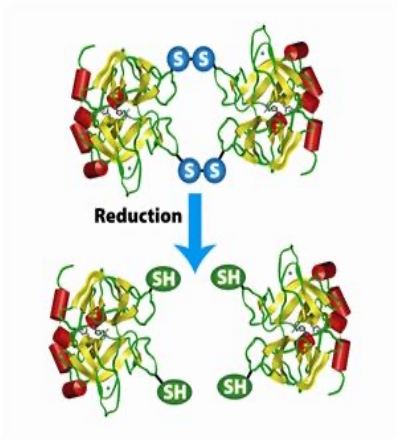
Digestion of proteins

1. Denaturation



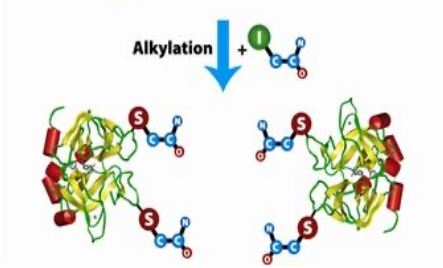
- Denaturing agents
  - Urea, guanidinium chloride, SDS, Rapigest...
  - ✧Think about the way how to remove the detergent afterwards!
  - ✧Don’t forget to dilute denaturing agents before adding the digestion enzyme (***why?***)

2. Reduction



- Buffers
  - Tris, HEPES, Ammonium bicarbonate
  - ✧ Be aware of the optimal pH of your digestion enzyme

3. Alkylation



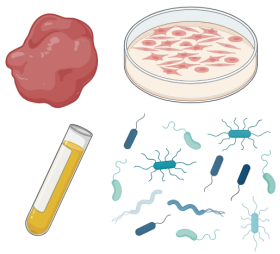
Peptide fractionation or enrichment

a) Off-gel electrophoresis : pI
b) Cation exchange: charge state in solution
c) Affinity chromatography : special groups
d) Reversed phase: hydrophobicity

- ✓ Isolation of subset of peptides / reduction of sample complexity
- ✓ Off-line or on-line

Maria Pavlou

Typical bottom-up workflow



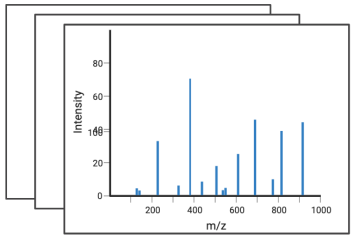
Protein extraction

Denatured proteins

Enzymatic digestion

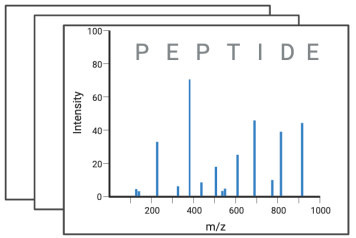
Peptides

Liquid chromatography tandem MS



raw data

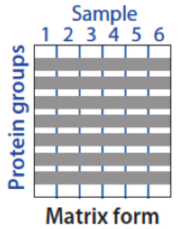
Database search



Peptide sequence

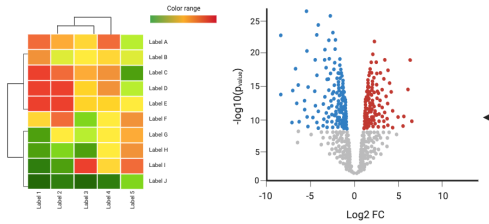
Protein inference

Quantitation

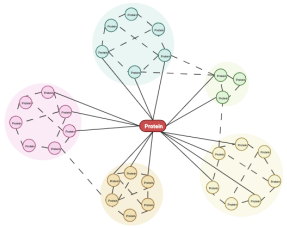


Matrix form

Statistical analysis

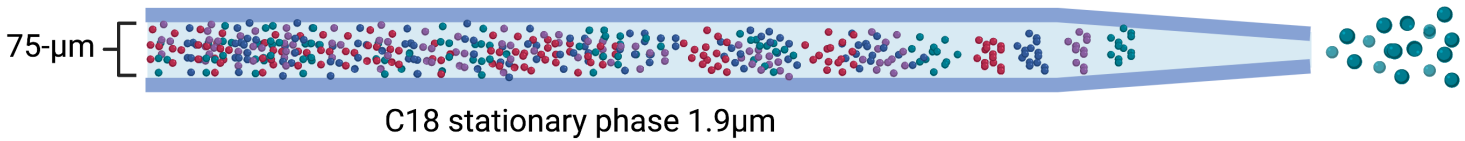


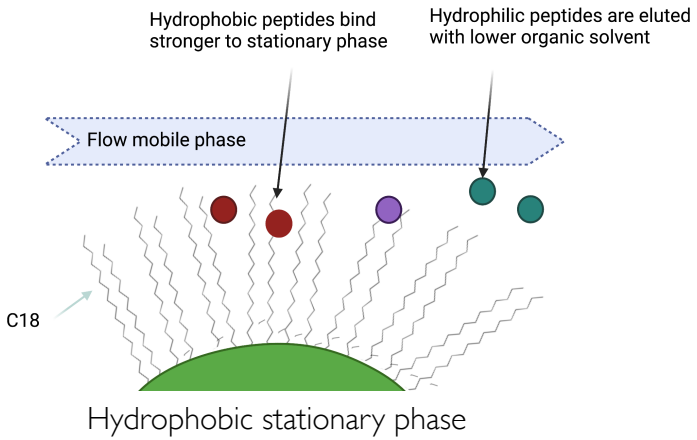
Biological interpretation



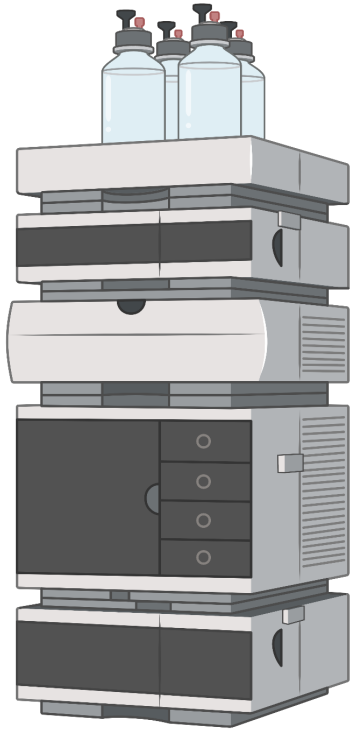
Reverse Phase (RP) chromatography

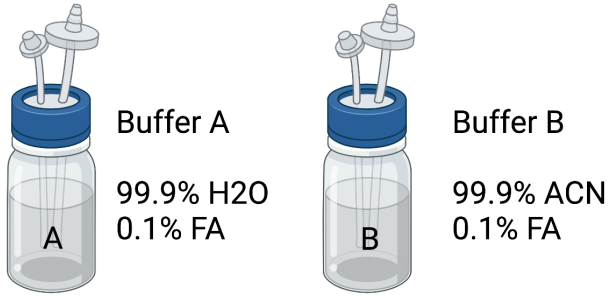
Maria Pavlou

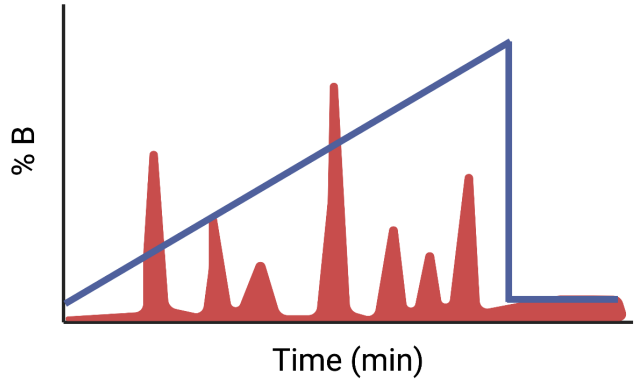




Gradient: 3-90% organic solvent

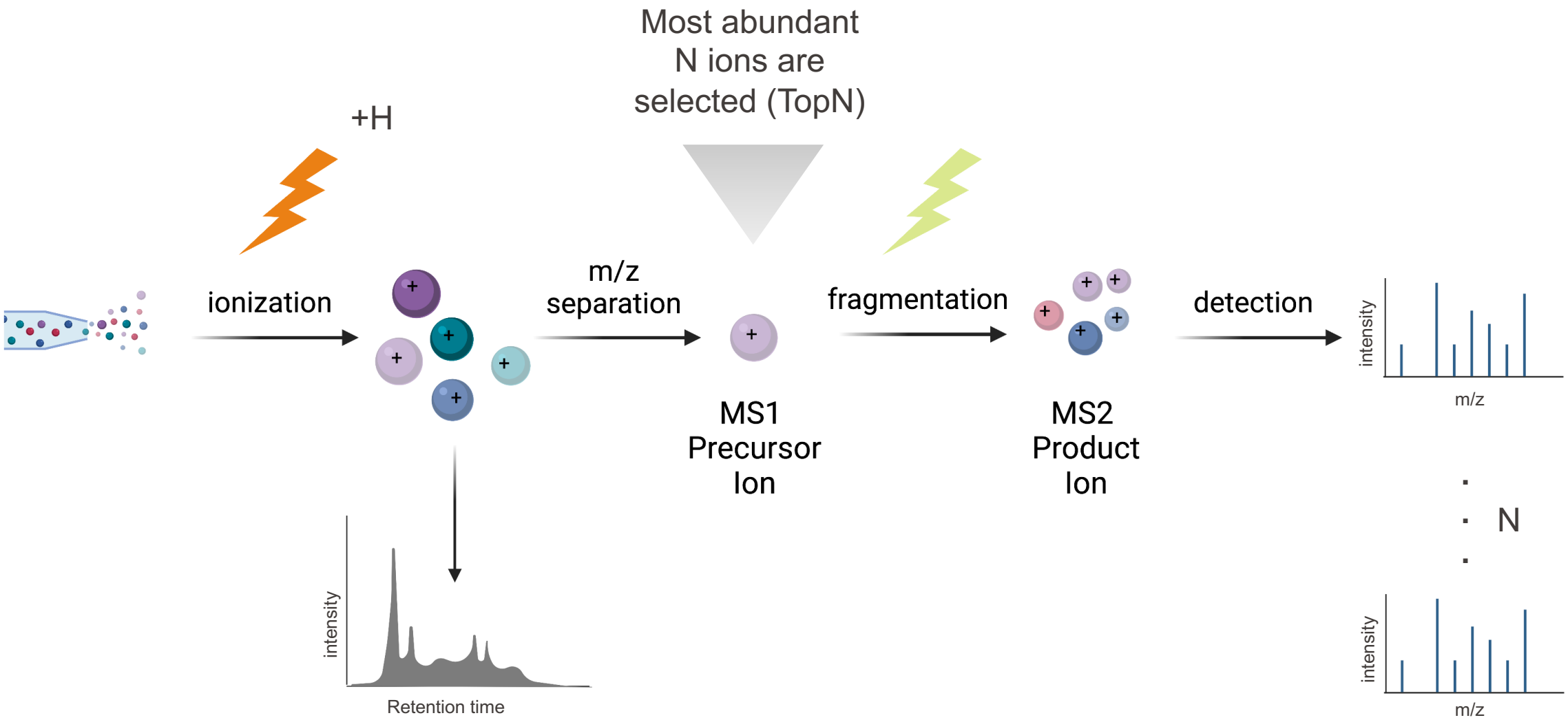




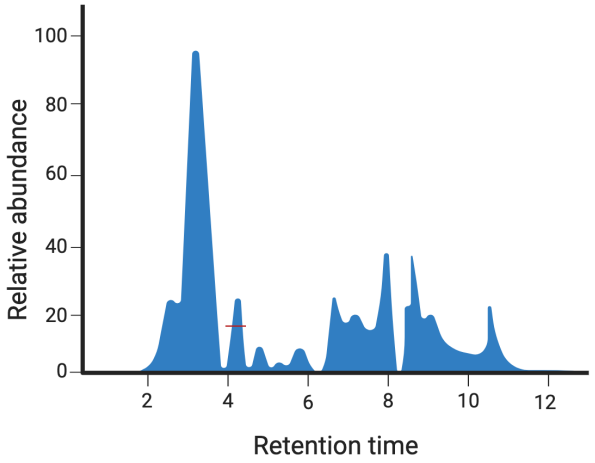


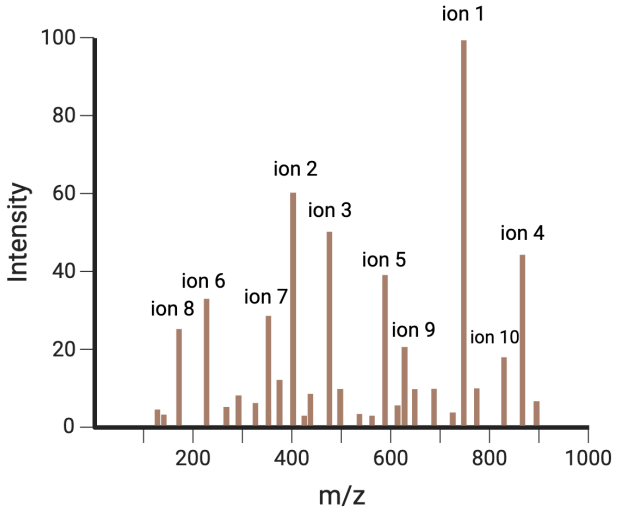
Maria Pavlou

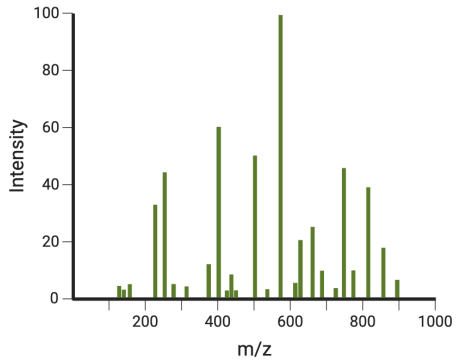
Tandem MS (MS/MS)
Data Dependent Acquisition (DDA)

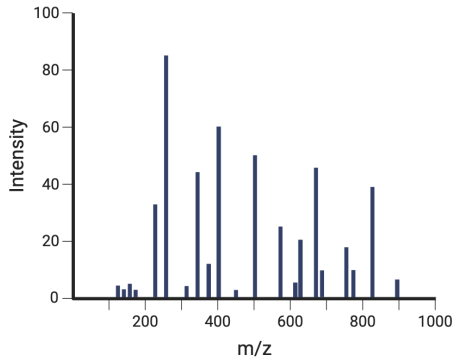


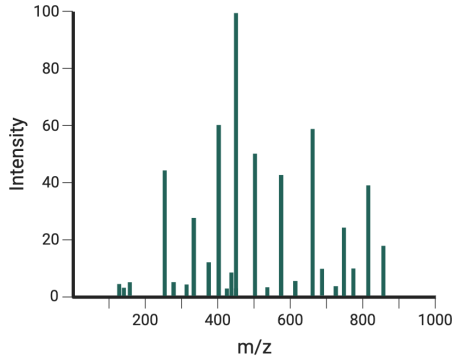
Mass analyser @ 4min







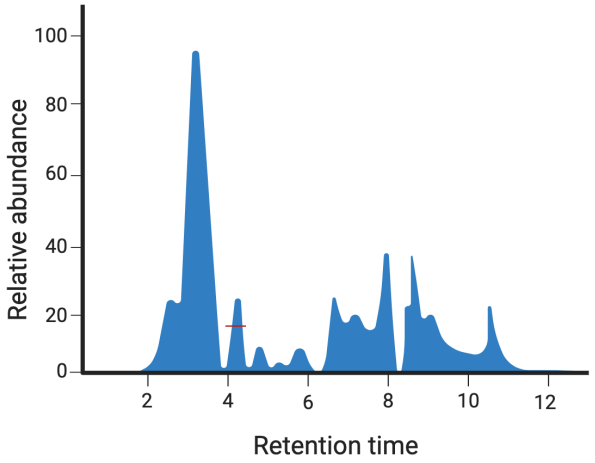


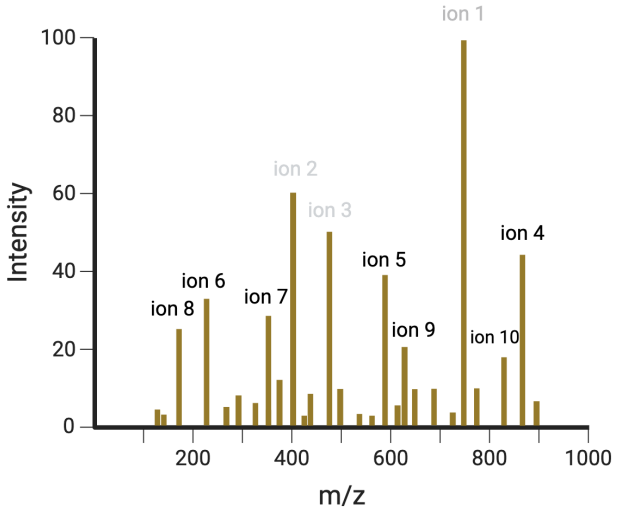


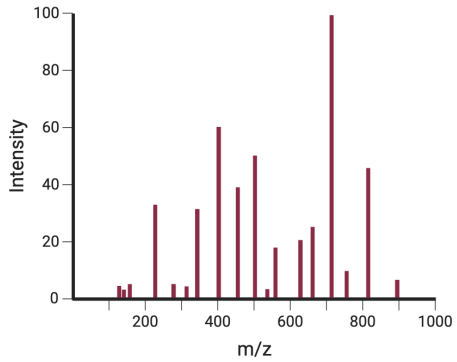
Analysis time (sec)

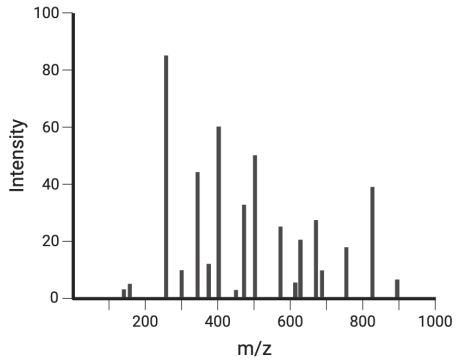
MS1	Peptide ions
MS2 ion 1	Fragments of selected ion
MS2 ion 2	Fragments of selected ion
MS2 ion 3	Fragments of selected ion

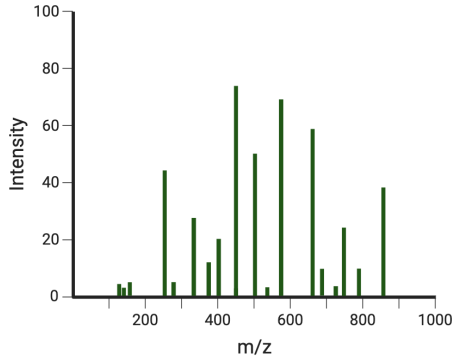
Few seconds later…











Analysis time (sec)

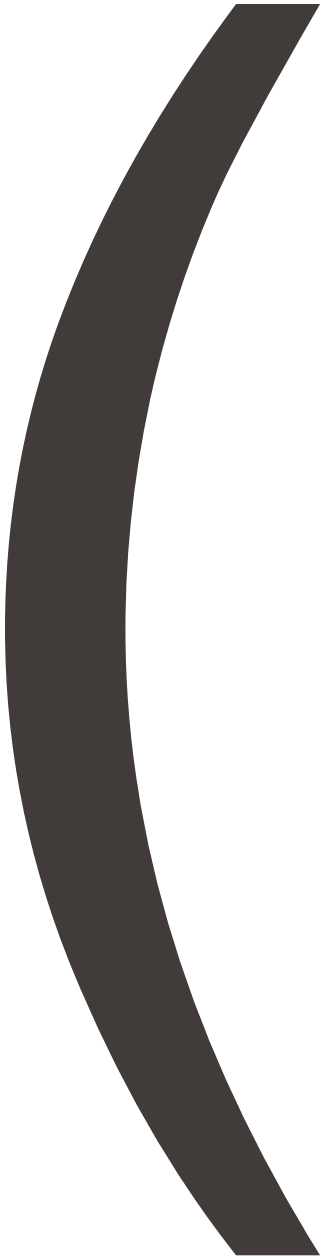
MS1	Peptide ions
MS2 ion 1	Fragments of selected ion
MS2 ion 2	Fragments of selected ion
MS2 ion 3	Fragments of selected ion
MS1	**Peptide ions**
MS2 ion 4	Fragments of selected ion
MS2 ion 5	Fragments of selected ion
MS2 ion 6	Fragments of selected ion
MS1	**Peptide ions**
……	

Thermo video MS/MS

https://youtu.be/zJagpUbnv-Y

Recorded information

- Retention time
- **Peptide ion m/z**
- Peptide spectrum
- Intensity of ions



Maria Pavlou

Natural isotopic distribution: relative abundance of isotopes

- Most elements occur in nature as a mixture of isotopes
- Isotopes are atom species of the same chemical element that have different masses
- They have the same number of protons and electrons, but a different number of neutrons (1Da)
- The main elements occurring in proteins are CHNOPS

element (symbol)	isotope	abundance %
hydrogen (H)	^{1}H	99.988 %
	^{2}H	0.012 %
carbon (C)	^{12}C	98.93 %
	^{13}C	1.07 %
nitrogen (N)	^{14}N	99.636 %
	^{15}N	0.364 %
oxygen (O)	^{16}O	99.757 %
	^{17}O	0.038 %
	^{18}O	0.205 %
phosphor (P)	^{31}P	100 %
sulfur (S)	^{32}S	94.99 %
	^{33}S	0.75 %
	^{34}S	4.25 %
	^{36}S	0.01 %

Mass can be many things

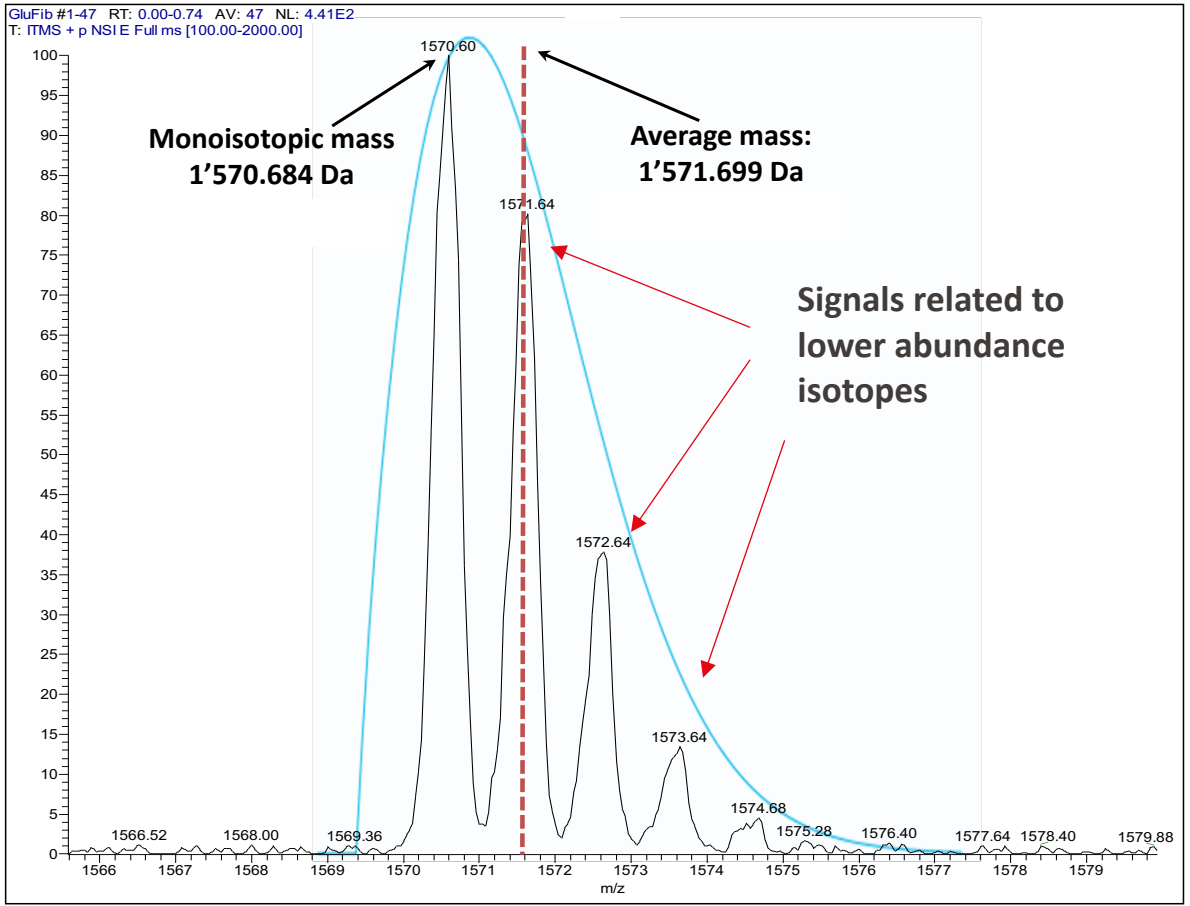
- Nominal: sum of integer atomic weights
- Average: the centroid of the complete isotopic envelope
- Monoisotopic: the mass of the first peak of the isotope distribution (most abundant elements)

GluFib: ***EGVNDNEEGFFSAR***

Nominal mass 1569 Da

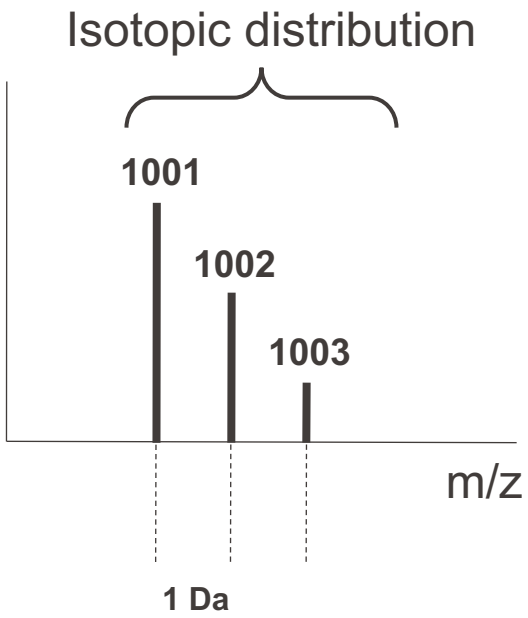
Chemical Formula:

- $C_{66}H_{95}N_{19}O_{26}$



The distances between isotopic peaks reveal charge state

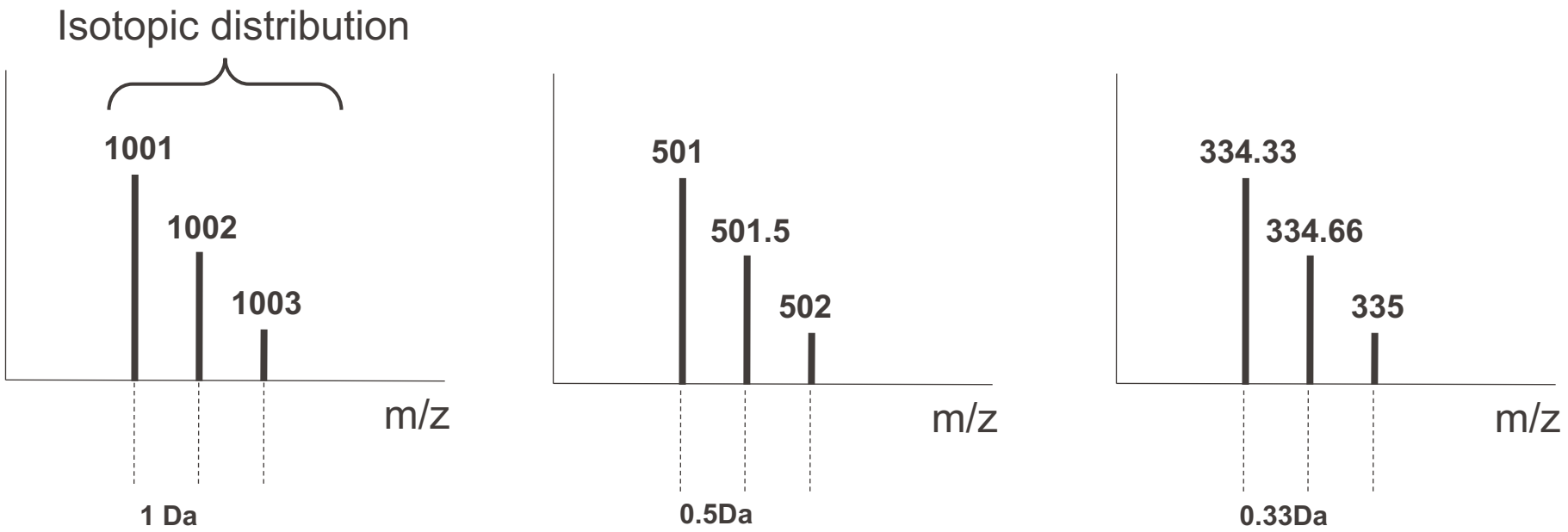
- M = 1000 Da (not charged)
- $[M+1H]^{1+} \longrightarrow$ z = 1
- m/z = 1001 Da (Monoiso.)



Maria Pavlou

The distances between isotopic peaks reveal charge state

- M = 1000 Da (not charged)
- $[M+1H]^{1+} \longrightarrow$ z = 1
- m/z = 1001 Da (Monoiso.)

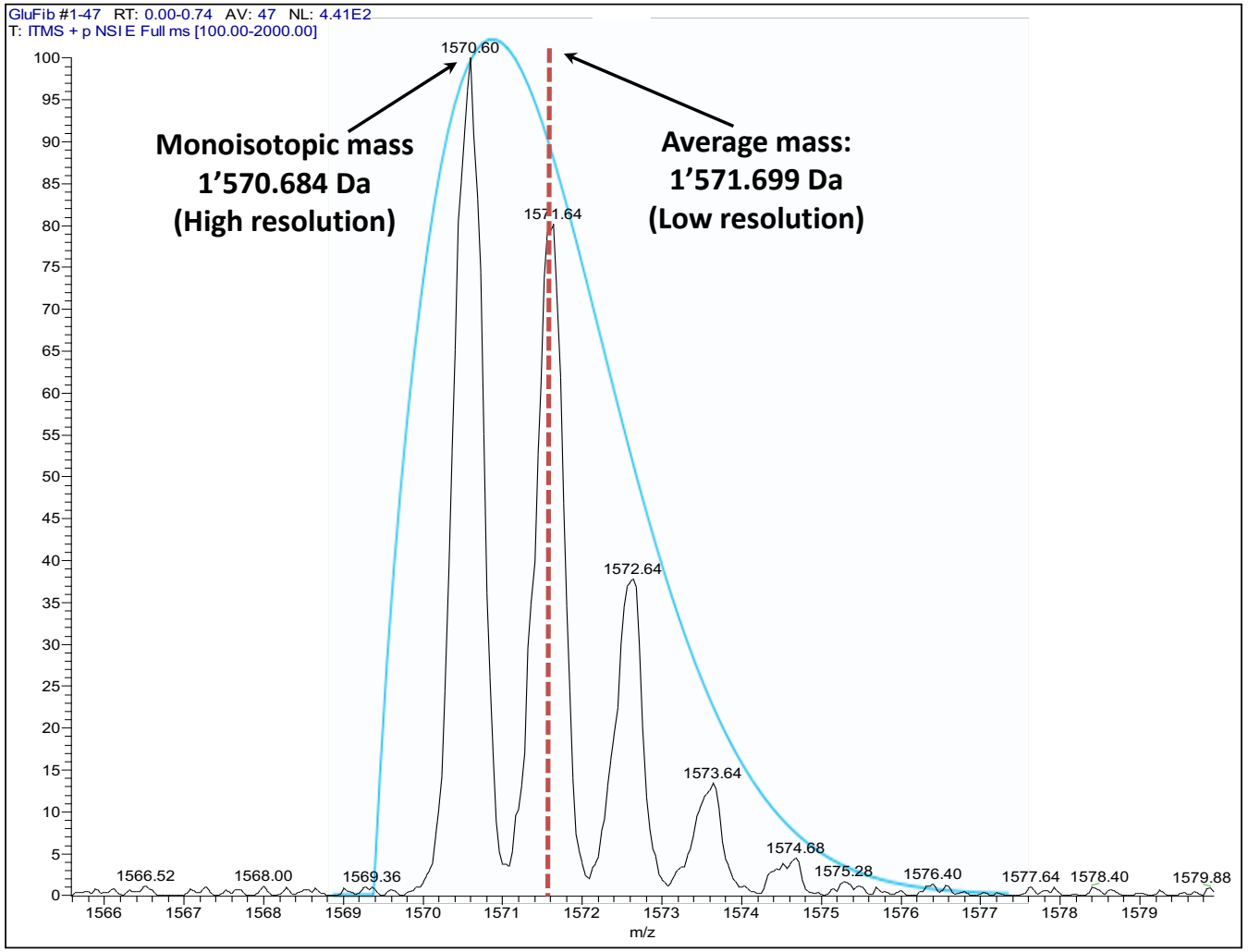


Low vs high resolution

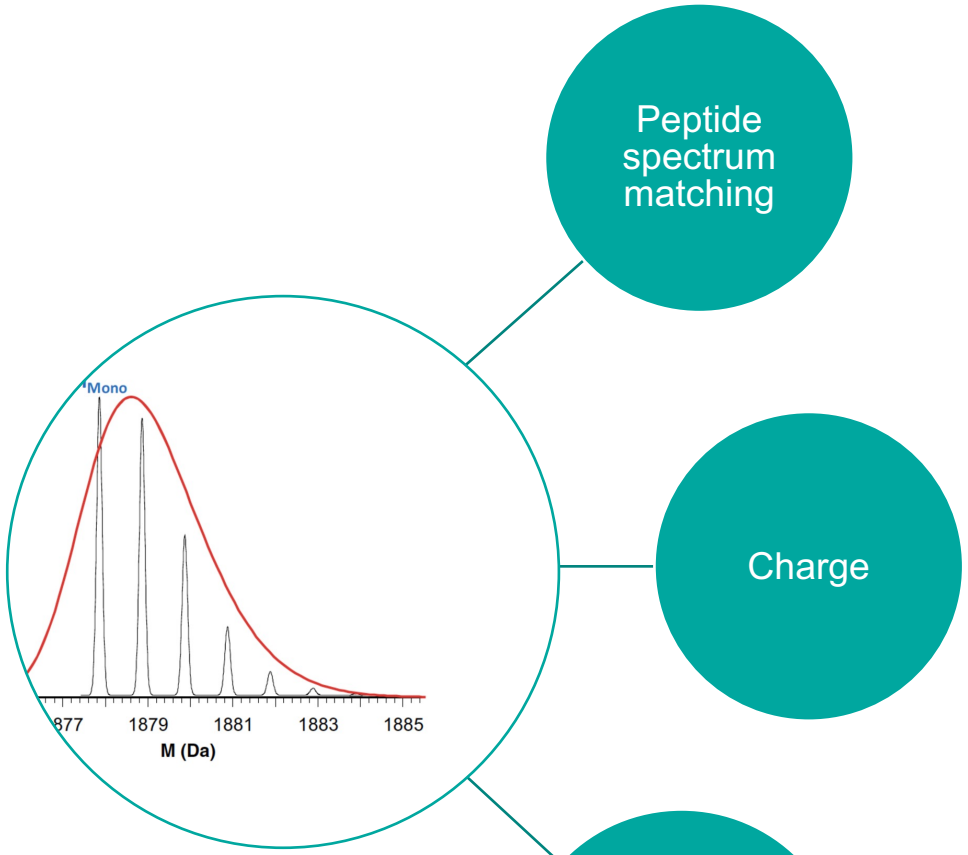
GluFib: ***EGVNDNEEGFFSAR***
MW: 1569.6696 Da

Chemical Formula:

- $C_{66}H_{95}N_{19}O_{26}$



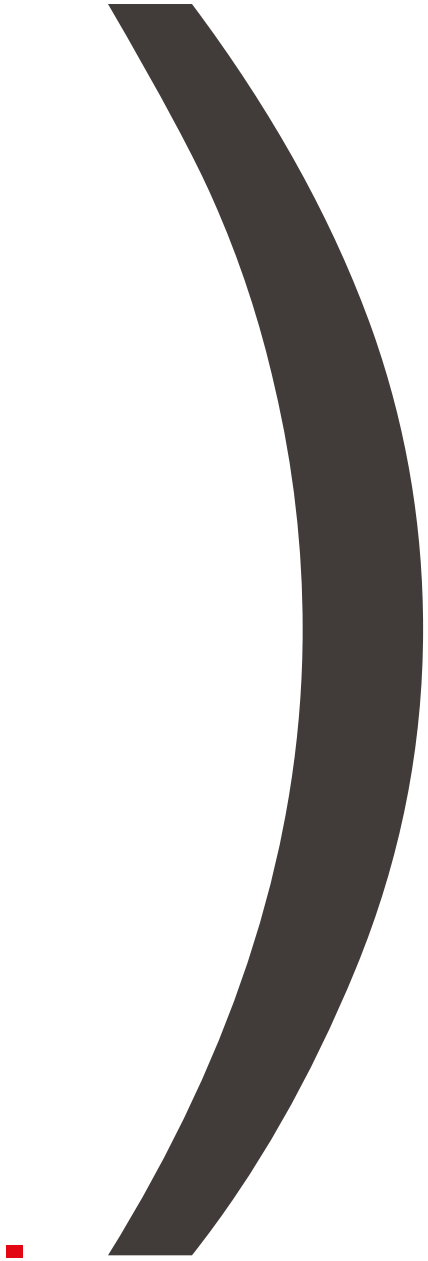
The importance of resolution and isotopic envelope



Peptide spectrum matching

Charge

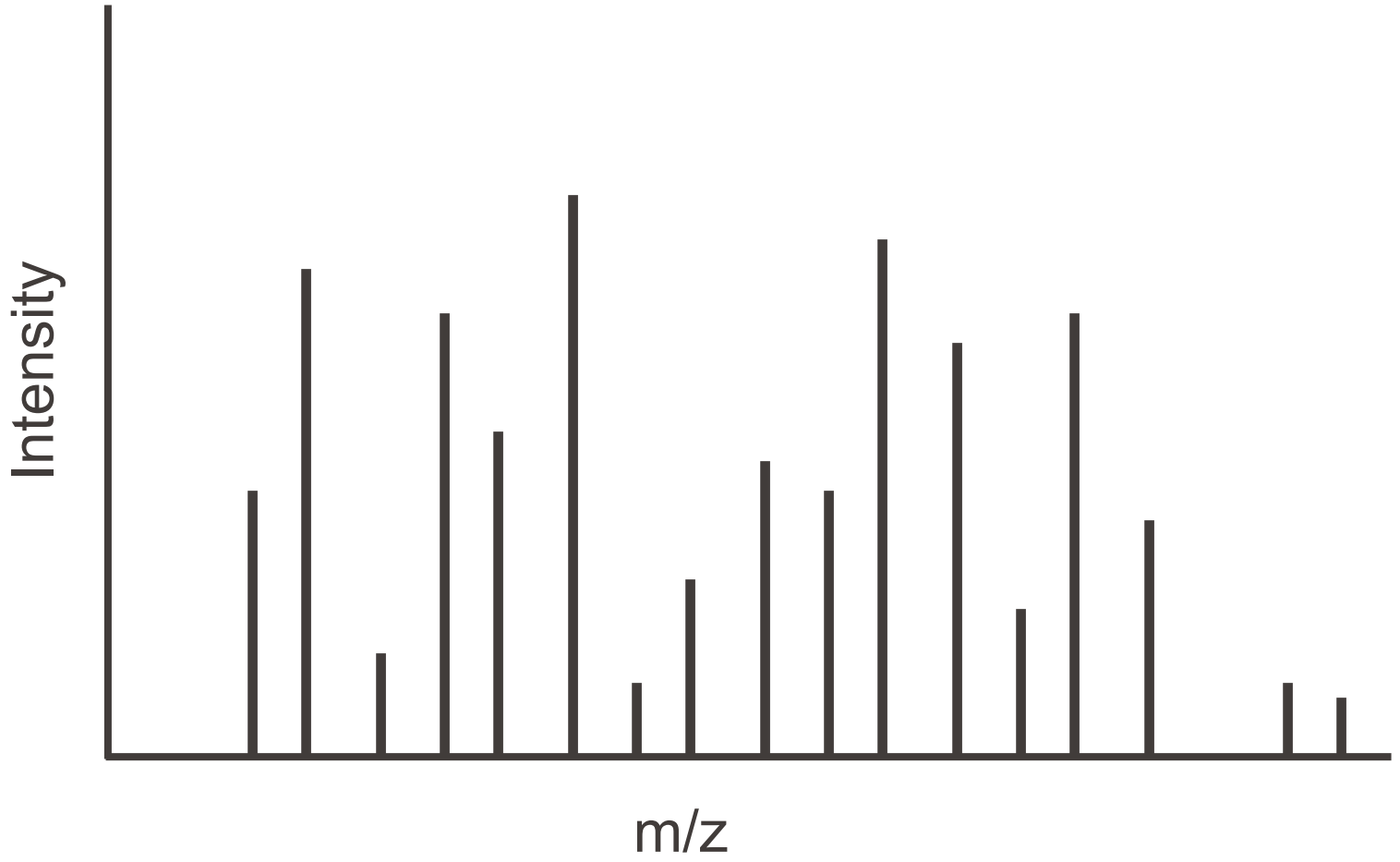
Monoisotopic mass



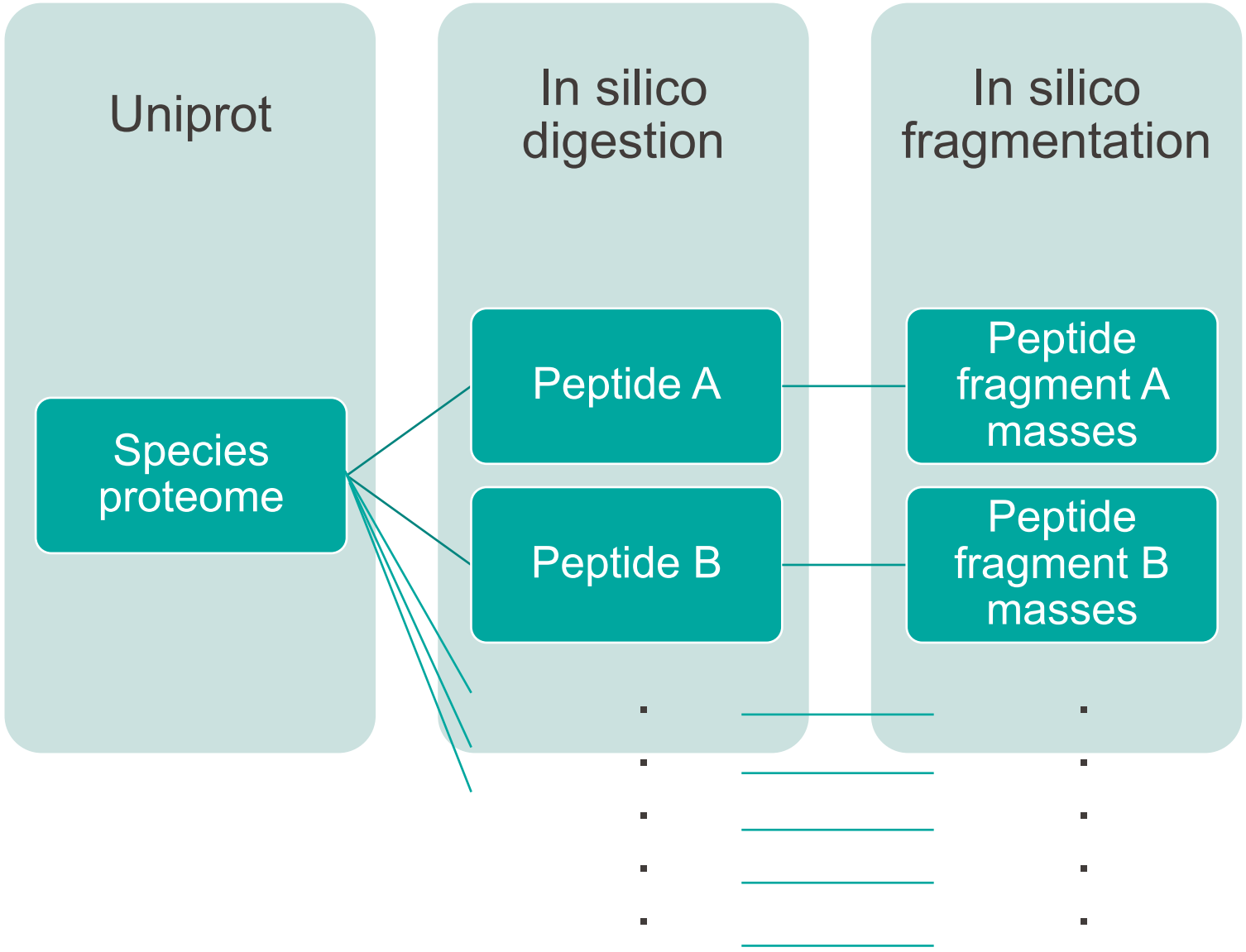
Recorded information

- Retention time
- Peptide ion m/z
- **Peptide spectrum**
- Intensity of ions

MS2 spectrum in 2D



Database-based peptide sequence identification



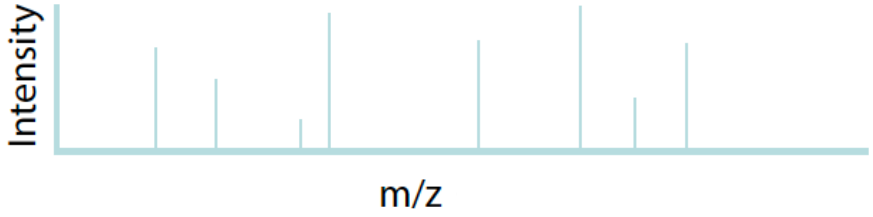
Maria Pavlou

Database search

1. MS1 filter
2. MS2 scoring
3. Probabilistic analysis

Observed Mass
1000 ± 0.010 Da

Corresponding MS^2 data



Peptide A Mass
999.980

Peptide B Mass
999.993

Peptide C Mass
1000.005

Peptide D Mass
1000.010

Peptide E Mass
1000.025

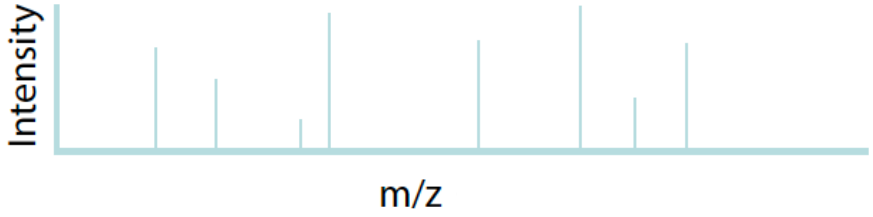
Maria Pavlou

Database search

1. MS1 filter
2. MS2 scoring
3. Probabilistic analysis

Observed Mass
1000 ± 0.010 Da

Corresponding MS^2 data



~~Peptide A Mass 999.980~~

Peptide B Mass 999.993

Peptide C Mass 1000.005

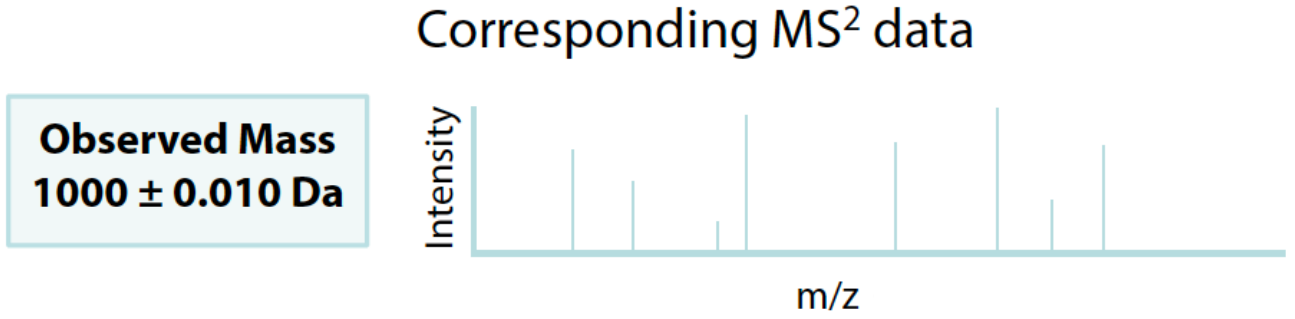
Peptide D Mass 1000.010

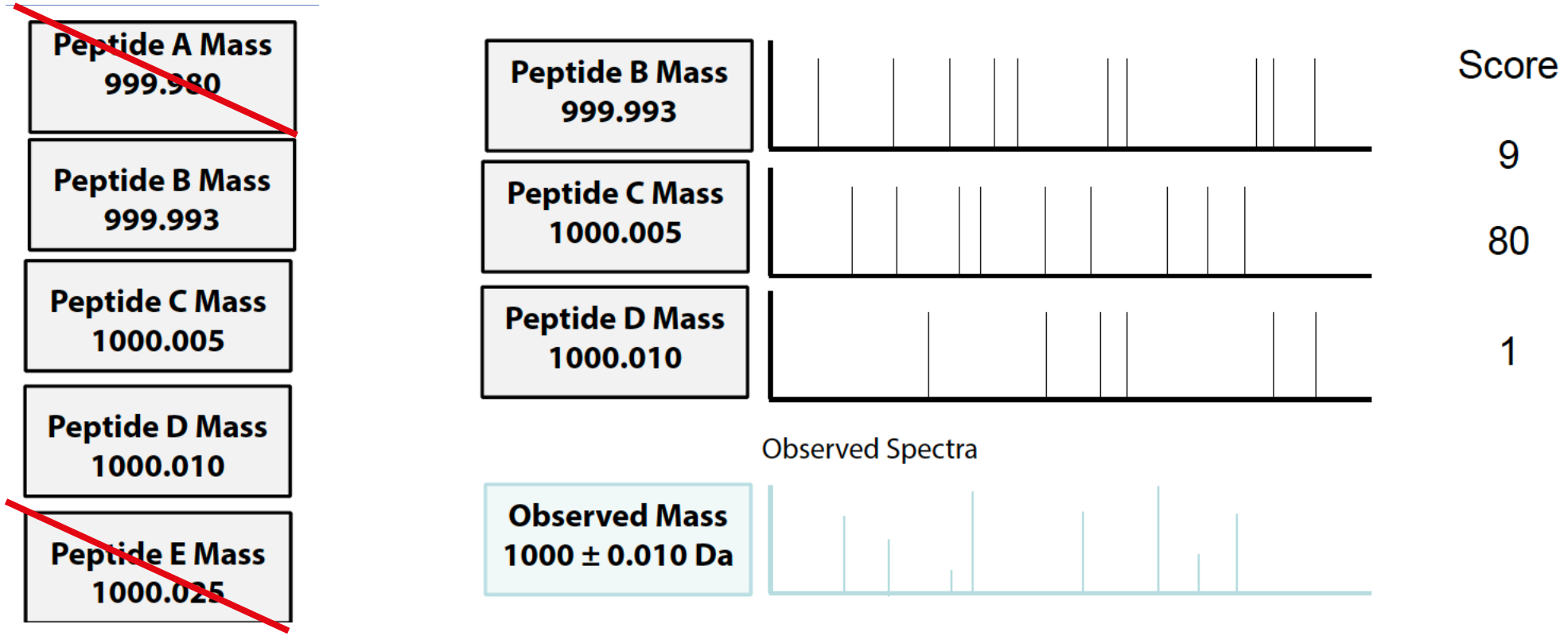
~~Peptide E Mass 1000.025~~

Maria Pavlou

Database search

1. MS1 filter
2. MS2 scoring
3. Probabilistic analysis

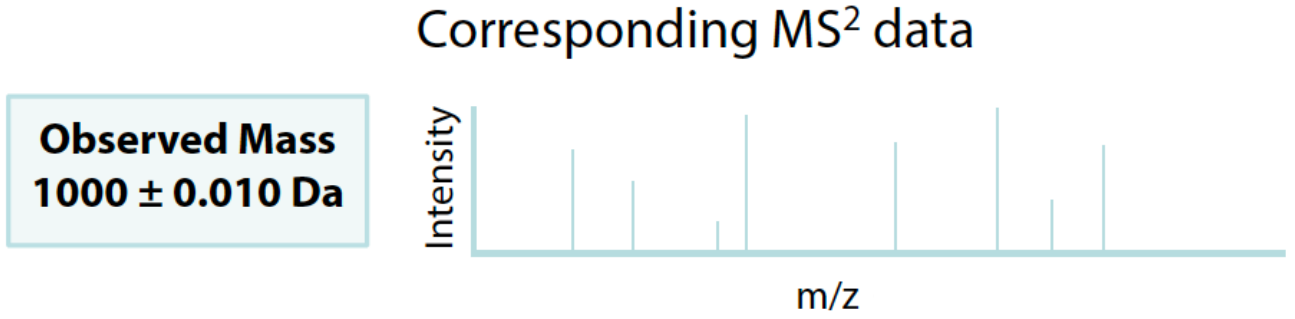


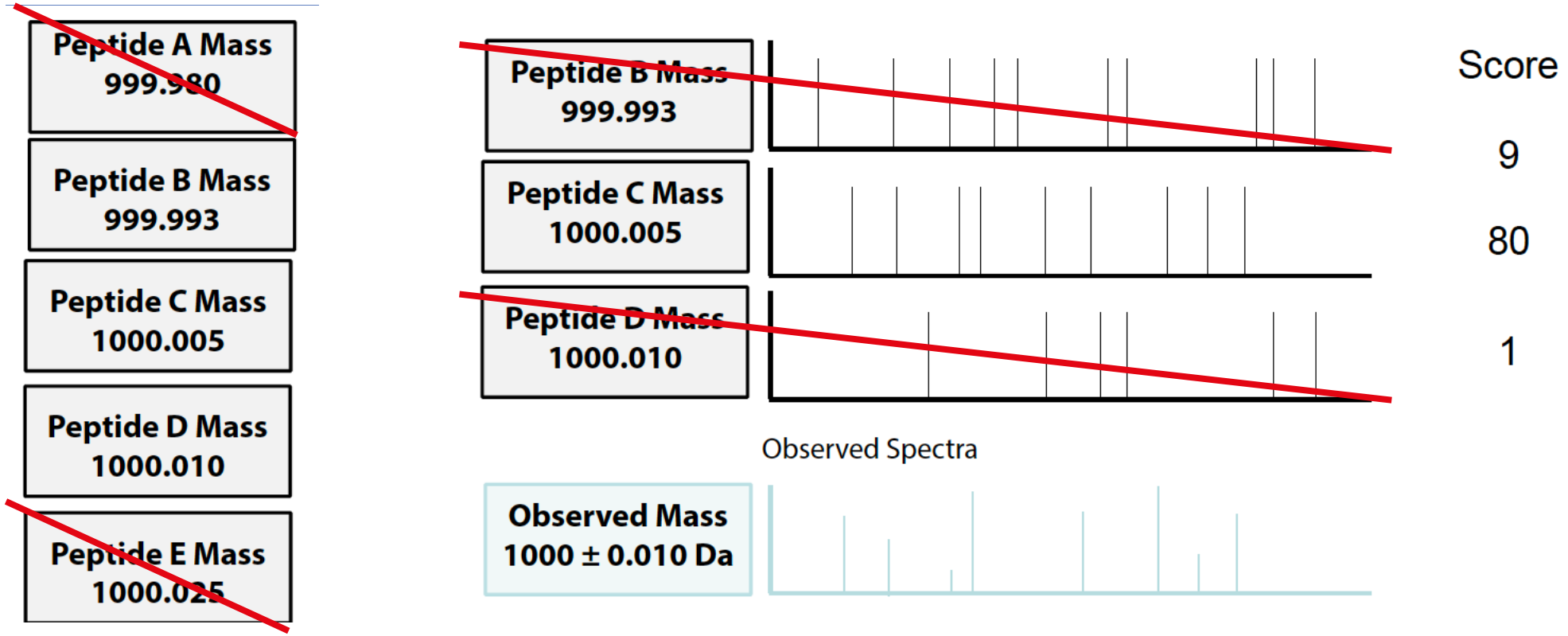


Maria Pavlou

Database search

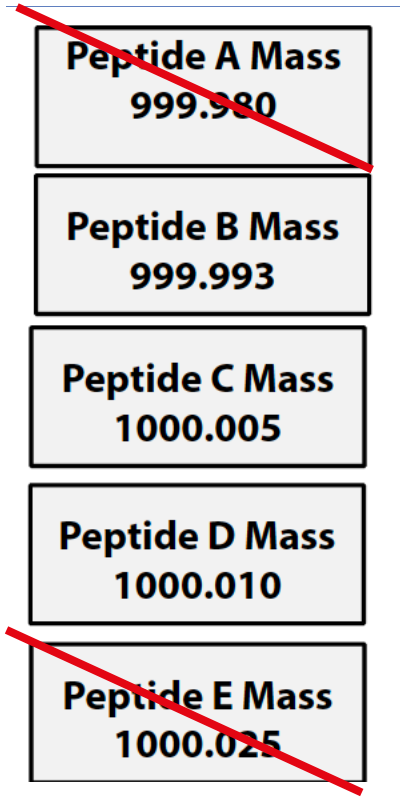
1. MS1 filter
2. MS2 scoring
3. Probabilistic analysis

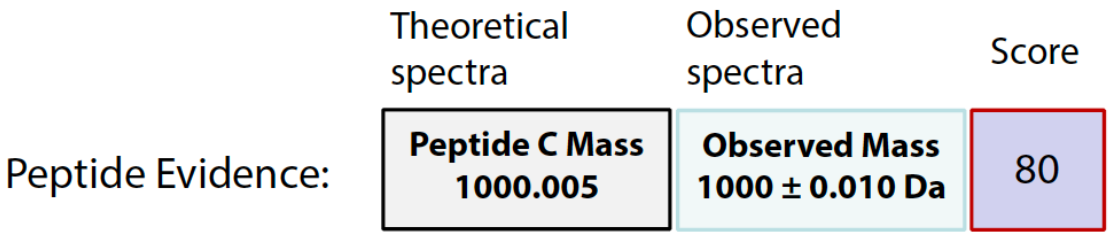




Database search

1. MS1 filter
2. MS2 scoring
3. Probabilistic analysis





Decoy/target strategy to determine FDR

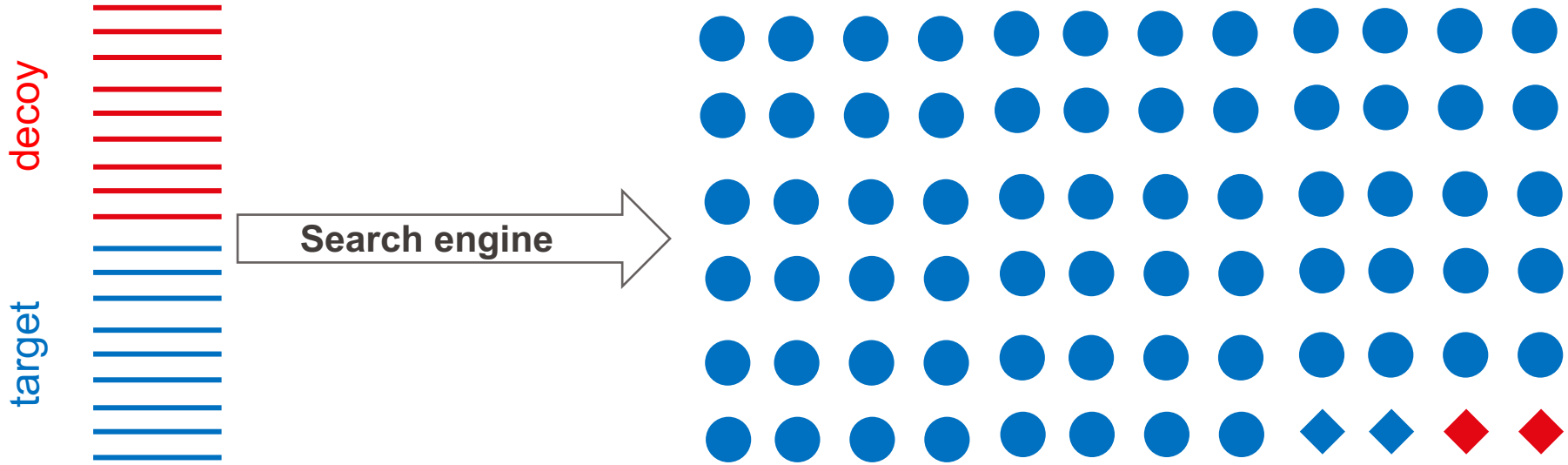
Maria Pavlou

AEPTIR
target

ITPEAR
decoy

Concatenated search

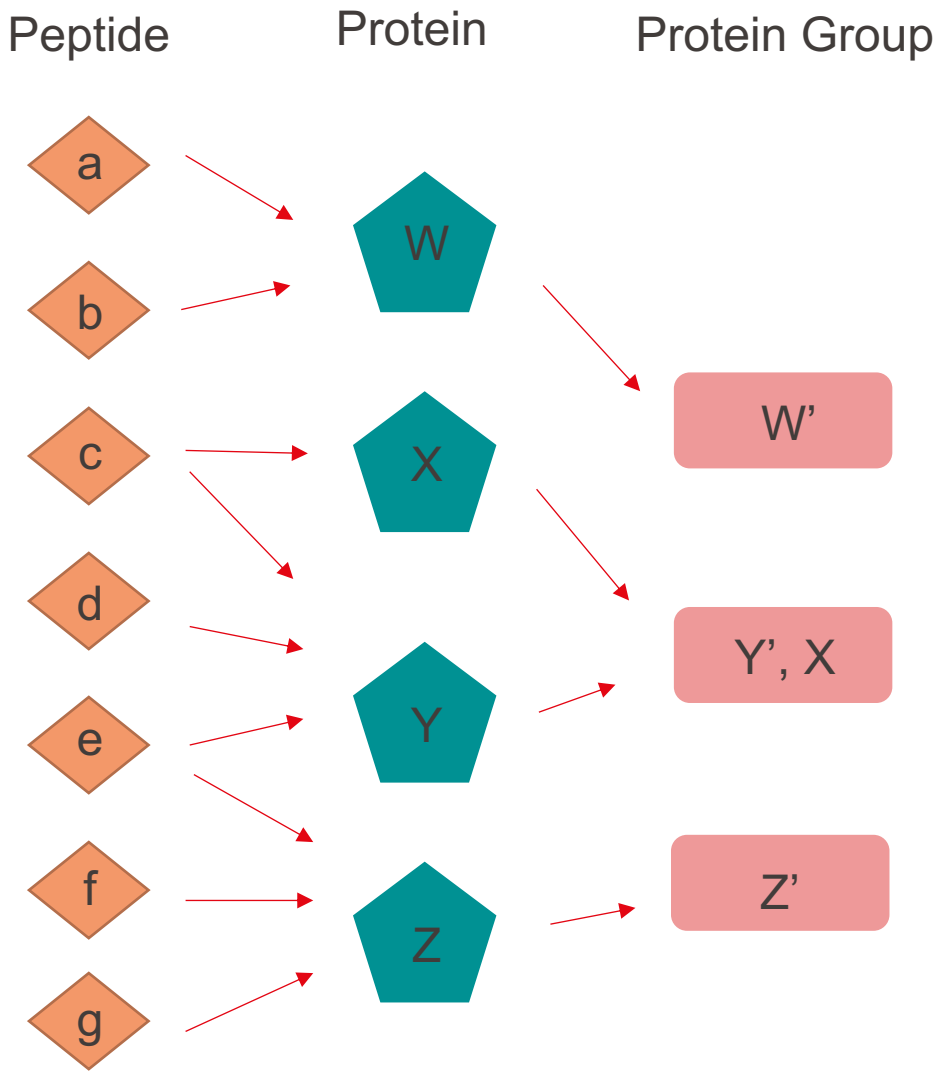
Commonly used is 1% FDR:
1 Decoy hit is accepted among
99 real hits



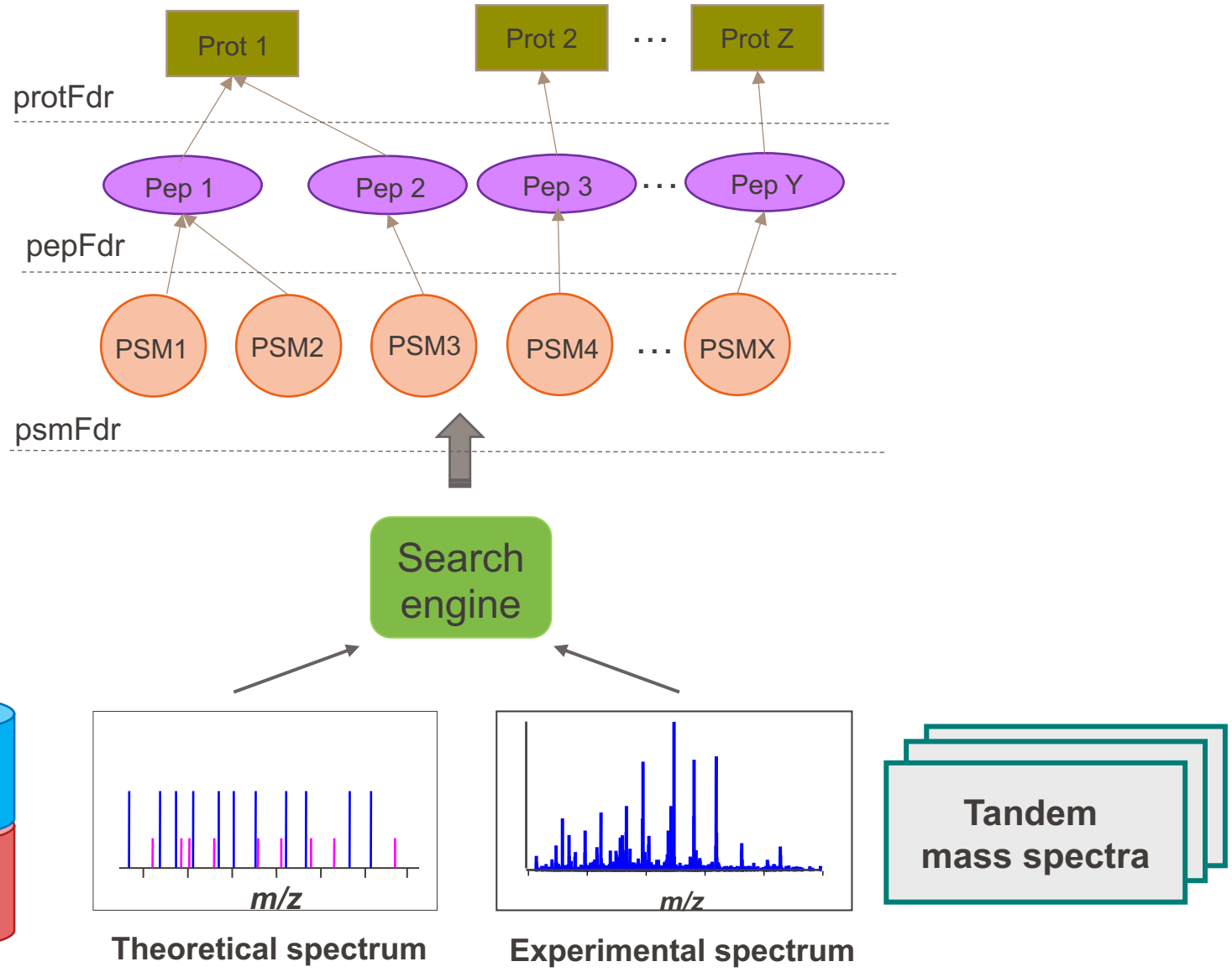
$$FDR = \frac{\#\ decoy}{\#\ target}$$

Maria Pavlou

Protein inference



From spectra to protein identifications



Maria Pavlou

Databases and search constrains

- Publicly available (Uniprot) or custom made
- Not too large, not too small
- Include a common set of contaminants (keratins, BSA…)

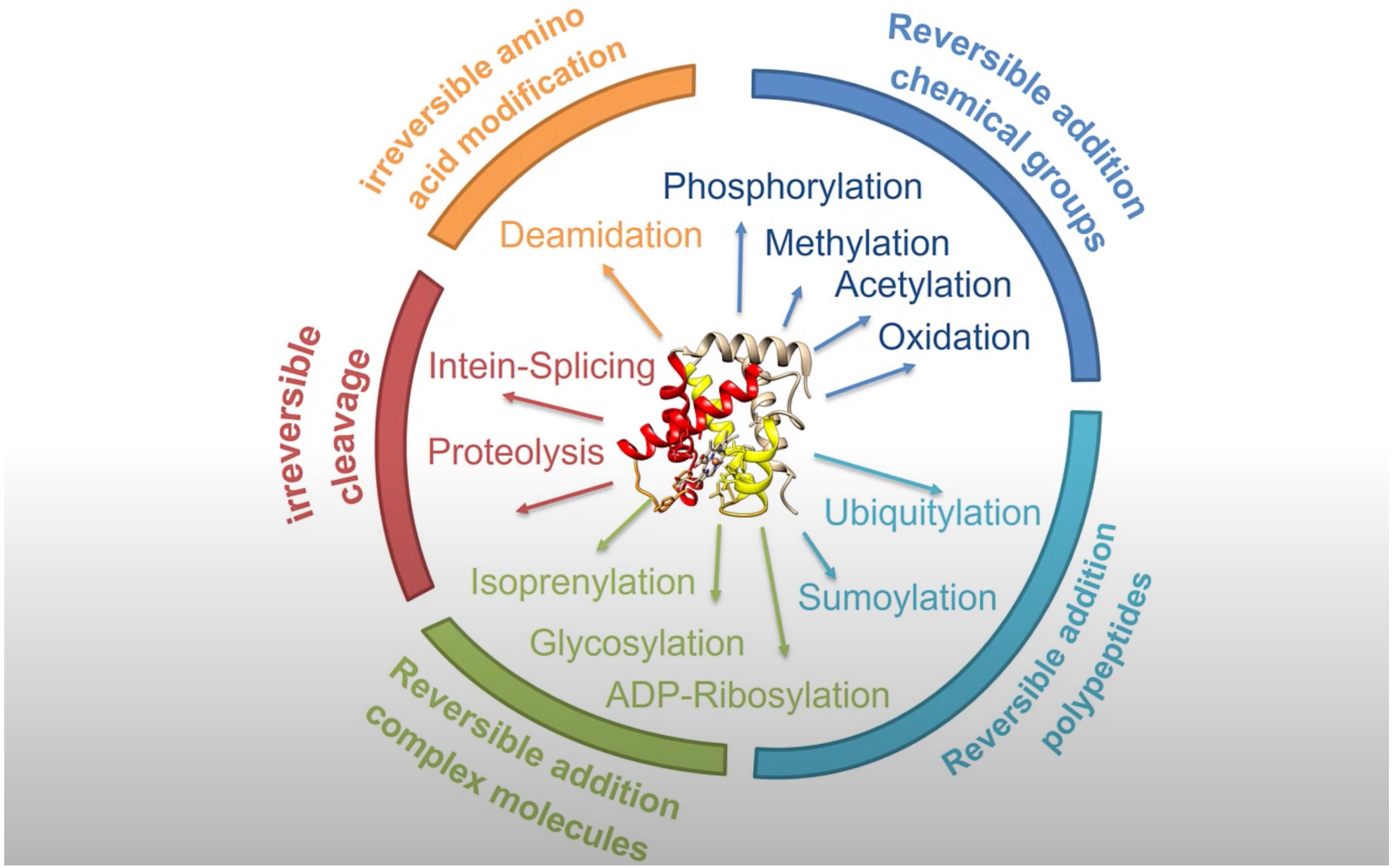
Commonly used search engines

- Mascot
- Sequest
- X!Tandem
- Andromeda
- Comet
- …

- All search engines use different criteria, producing different scores
- Using multiple search engines simultaneously yields better results

What about PTMs?

But what is a PTM Maria???

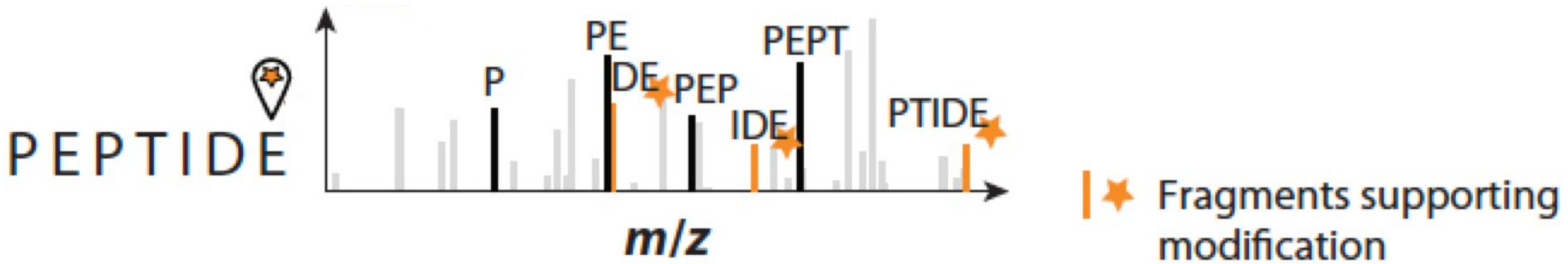


What about PTMs?

Introducing a mass shift

All protein molecules? -> fixed (in-silico spectra taking into account the mass difference)

Some protein molecules? -> variable (two forms of in-silico spectra: with and without)



How many peptides can you “see”?

human mitogen-activated protein kinase-8 (MAPK8)

a MS-compatible peptides

MSRSKRDNNF YSVEIGDSTF TVLKRYQNLK PIGSGAQGIV CAAYDAILER NVAIKKLSRP FQNQTHAKRA YRELVLMKCV
NHKNIIGLLN VFTPQKSLEE FQDVYIVMEL MDANLCQVIQ MELDHERMSY LLYQMLCGIK HLHSAGIIHR DLKPSNIVVK
SDCTLKILDF GLARTAGTSF MMTPYVVTRY YRAPEVILGM GYKENVDLWS VGCIMGEMVC HKILFPGRDY IDQWNKVIEQ
LGTPCPEFMK KLQPTVRTYV ENRPKYAGYS FEKLFPDVLF PADSEHNKLK ASQARDLLSK MLVIDASKRI SVDEALQHPY
INVWYDPSEA EAPPPKIPDK QLDEREHTIE EWKELIYKEV MDLEERTKNG VIRGQPSPLG AAVINGSQHP SSSSSVNDVS
SMSTDPTLAS DTDSSLEAAA GPLGCCR

b Observed peptides

MSRSKRDNNF YSVEIGDSTF TVLKRYQNLK PIGSGAQGIV CAAYDAILER NVAIKKLSRP FQNQTHAKRA YRELVLMKCV
NHKNIIGLLN VFTPQKSLEE FQDVYIVMEL MDANLCQVIQ MELDHERMSY LLYQMLCGIK HLHSAGIIHR DLKPSNIVVK
SDCTLKILDF GLARTAGTSF MMTPYVVTRY YRAPEVILGM GYKENVDLWS VGCIMGEMVC HKILFPGRDY IDQWNKVIEQ
LGTPCPEFMK KLQPTVRTYV ENRPKYAGYS FEKLFPDVLF PADSEHNKLK ASQARDLLSK MLVIDASKRI SVDEALQHPY
INVWYDPSEA EAPPPKIPDK QLDEREHTIE EWKELIYKEV MDLEERTKNG VIRGQPSPLG AAVINGSQHP SSSSSVNDVS
SMSTDPTLAS DTDSSLEAAA GPLGCCR

Proteotypic: experimentally observable peptides that can be used to uniquely identify a protein

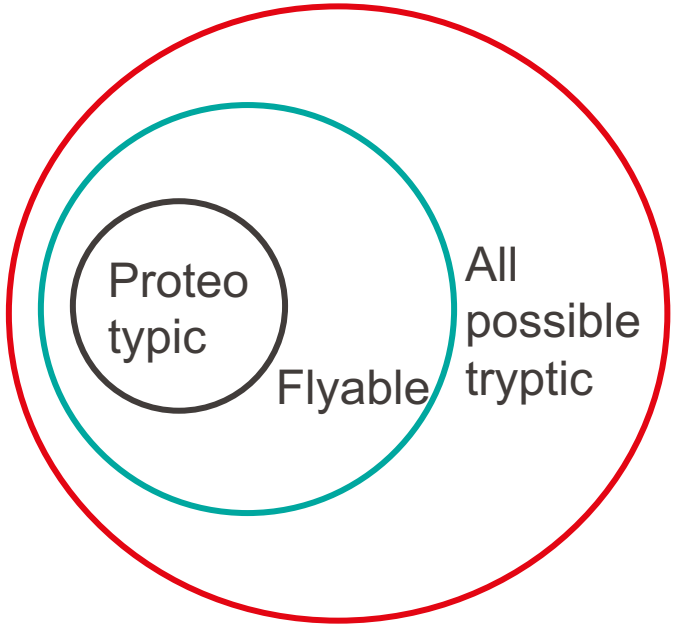
c Proteotypic peptides

MSRSKRDNNF YSVEIGDSTF TVLKRYQNLK PIGSGAQGIV CAAYDAILER NVAIKKLSRP FQNQTHAKRA YRELVLMKCV
NHKNIIGLLN VFTPQKSLEE FQDVYIVMEL MDANLCQVIQ MELDHERMSY LLYQMLCGIK HLHSAGIIHR DLKPSNIVVK
SDCTLKILDF GLARTAGTSF MMTPYVVTRY YRAPEVILGM GYKENVDLWS VGCIMGEMVC HKILFPGRDY IDQWNKVIEQ
LGTPCPEFMK KLQPTVRTYV ENRPKYAGYS FEKLFPDVLF PADSEHNKLK ASQARDLLSK MLVIDASKRI SVDEALQHPY
INVWYDPSEA EAPPPKIPDK QLDEREHTIE EWKELIYKEV MDLEERTKNG VIRGQPSPLG AAVINGSQHP SSSSSVNDVS
SMSTDPTLAS DTDSSLEAAA GPLGCCR

Flyable and proteotypic peptides

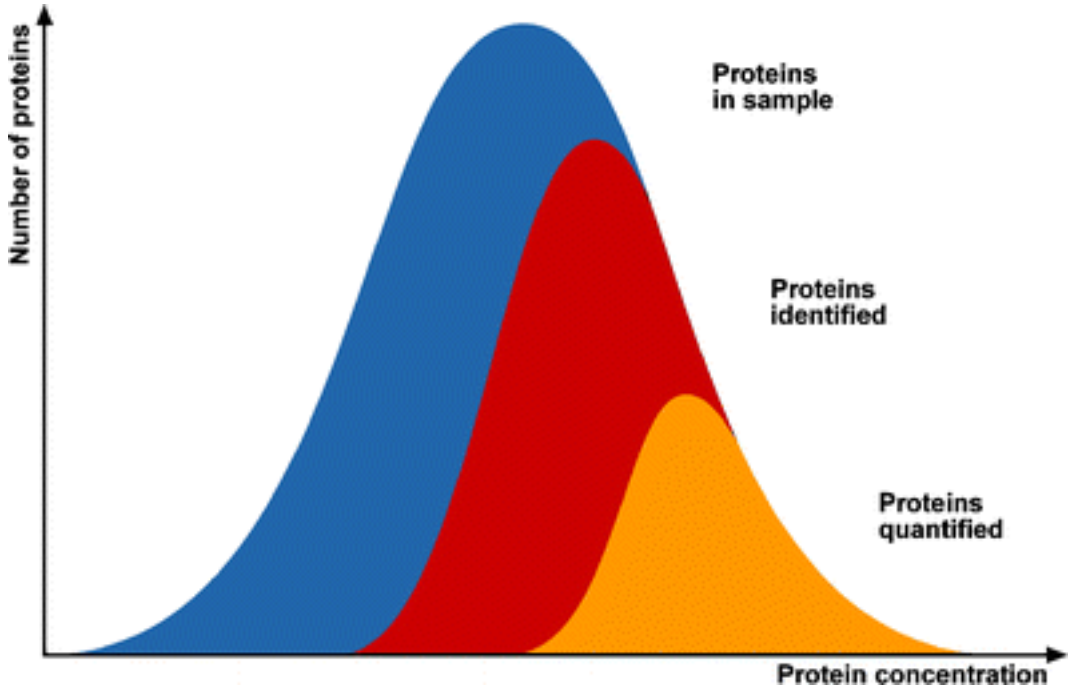
- Flyable: all peptides experimentally observed
- Proteotypic: experimentally observable peptides that can be used to uniquely identify a protein

  At least one proteotypic peptide is required for protein identification

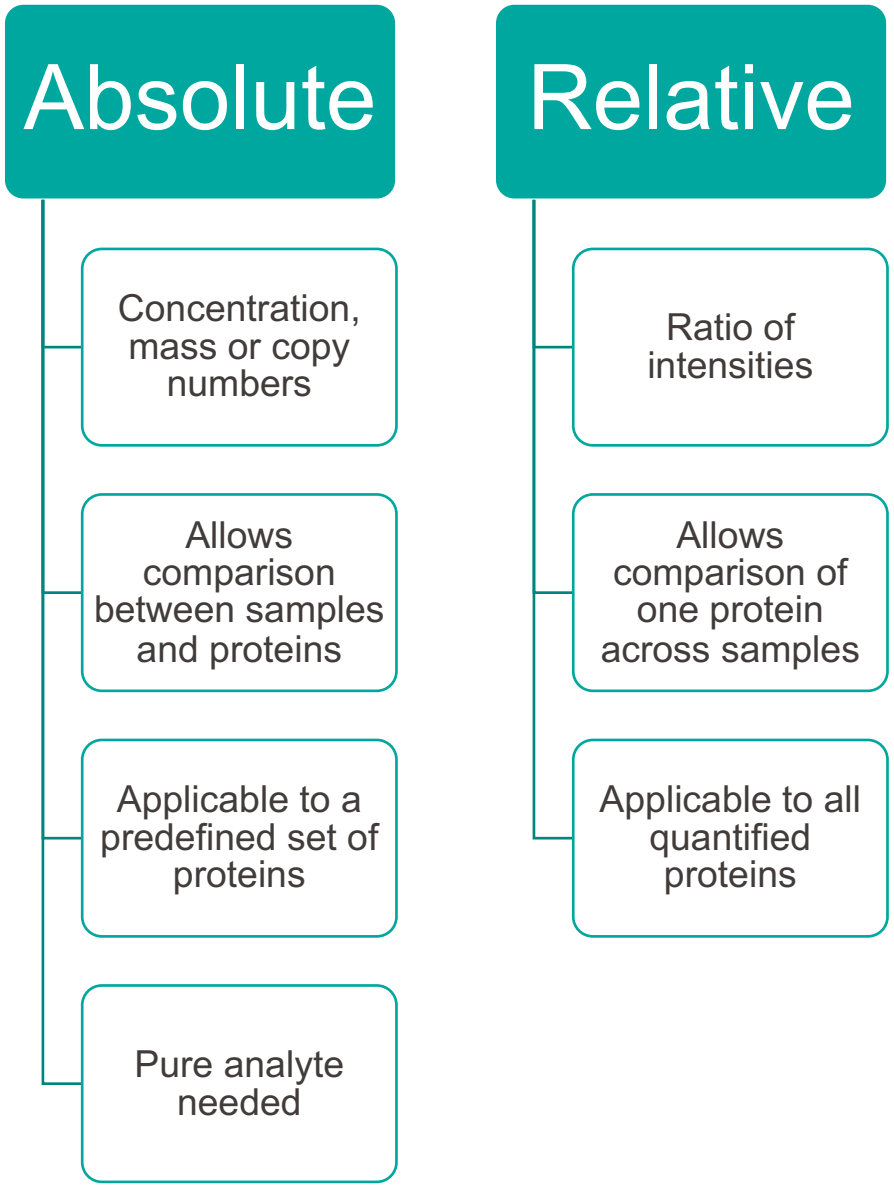


Q: I submited a gel band containing my favorite protein but you could not identify it. I know it is there because I also did a western blot.
Why???

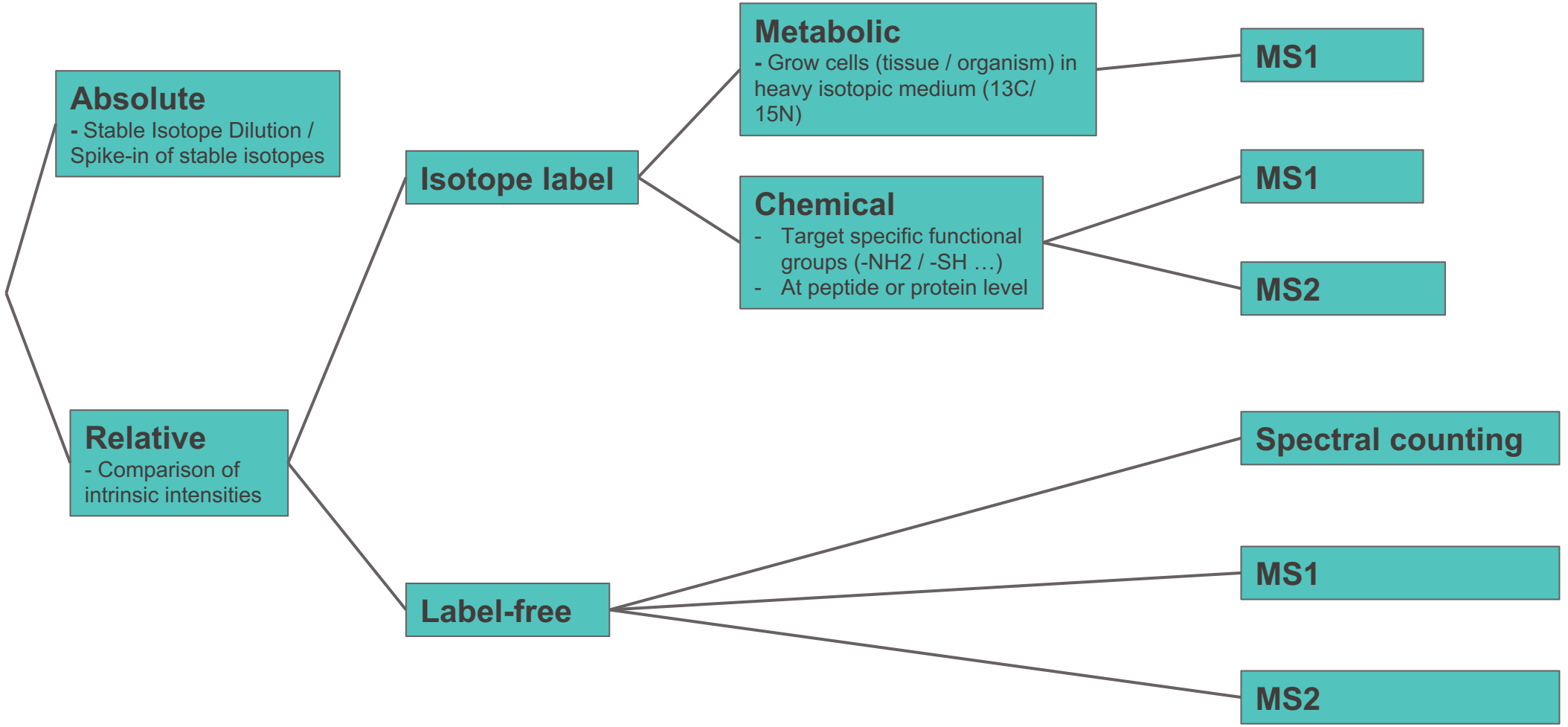
Quantitative proteomics



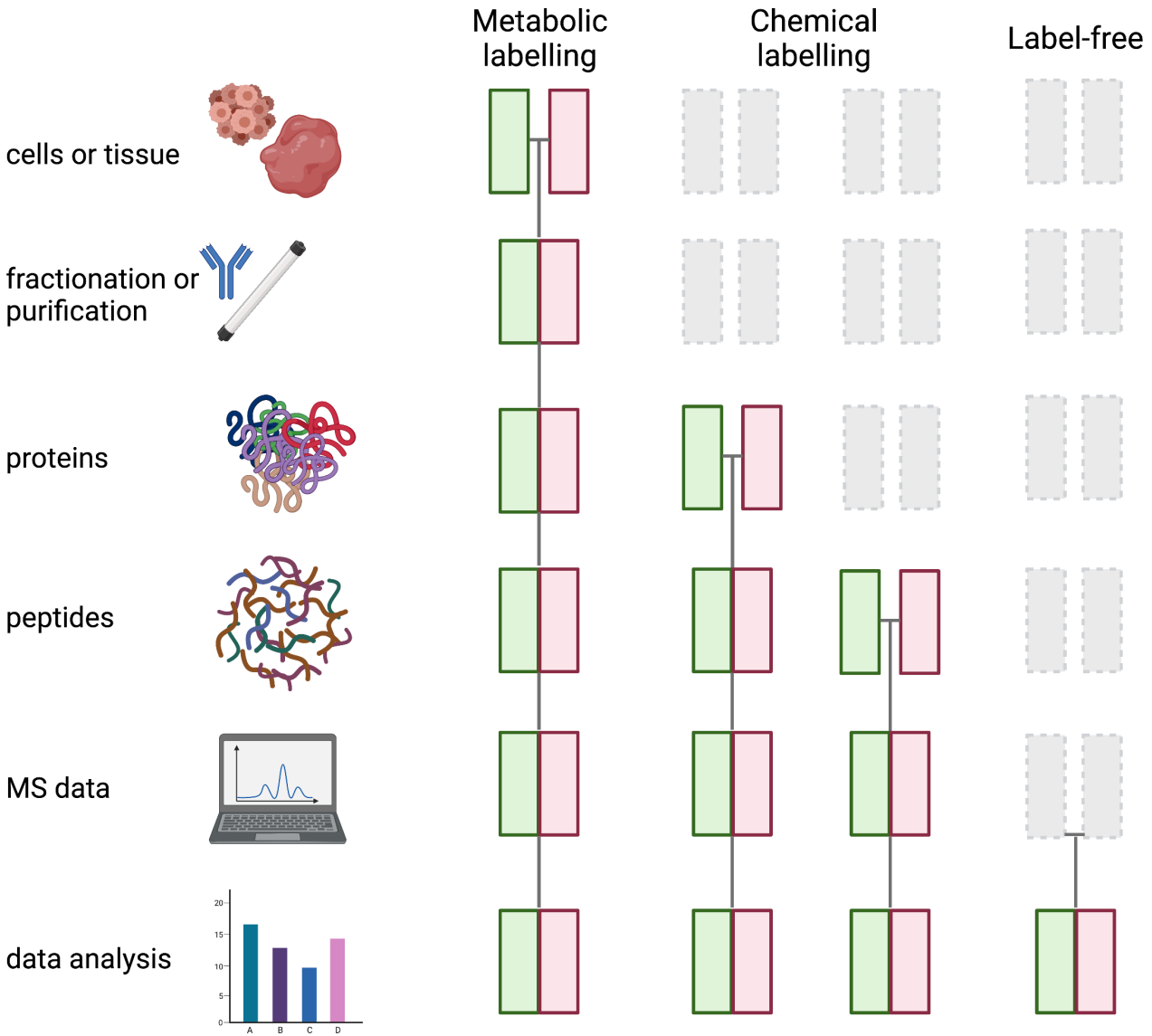
Absolute vs Relative



Quantitative proteomics strategies



Experimental design of different startegies



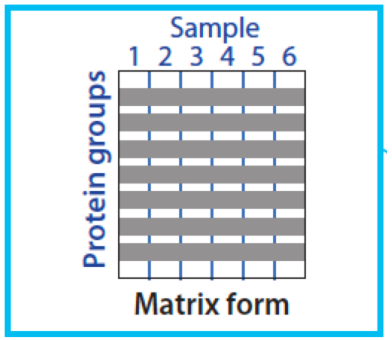
Maria Pavlou

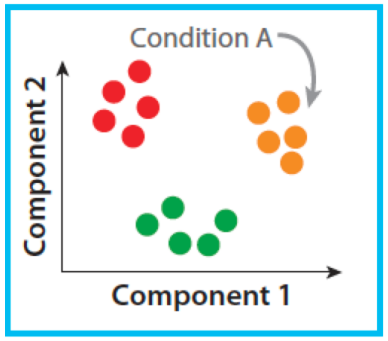
Comparison of different strategies

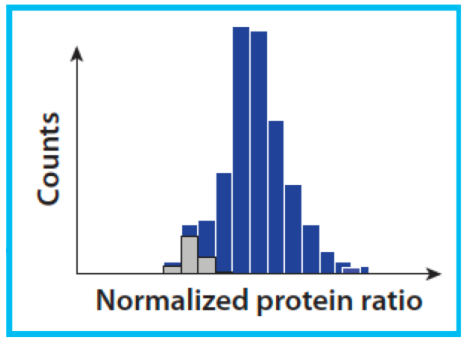
Method	Multiplexing	Dynamic range	Strong points	Weak points
Metabolic	2-3	1-2 logs	High accuracy with global labelling (early mixing of samples)	Requires actively growing cells (5 doublings)
Chemical (TMT)	Up to 18	2 logs	Highest multiplexing capability and low % of missing values	Expensive; ratio compression
Label-free	1	2-3 logs	No labelling requirements	Long processing, variabilty, missing values due to inherent stochasticity

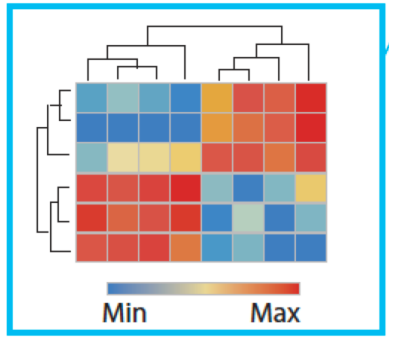
Downstream analysis

Maria Pavlou

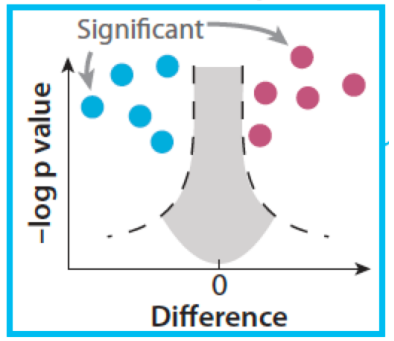


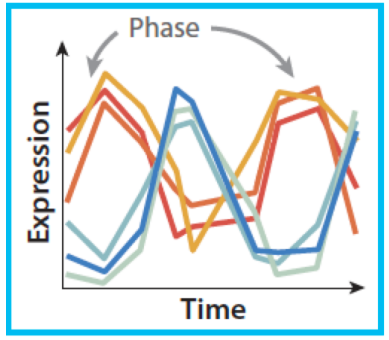




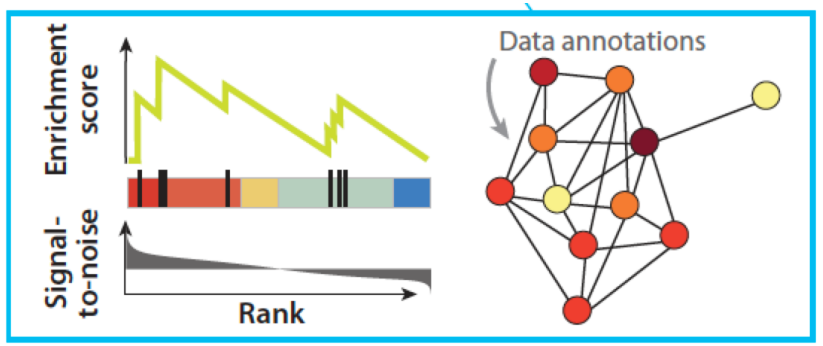


Quality Control



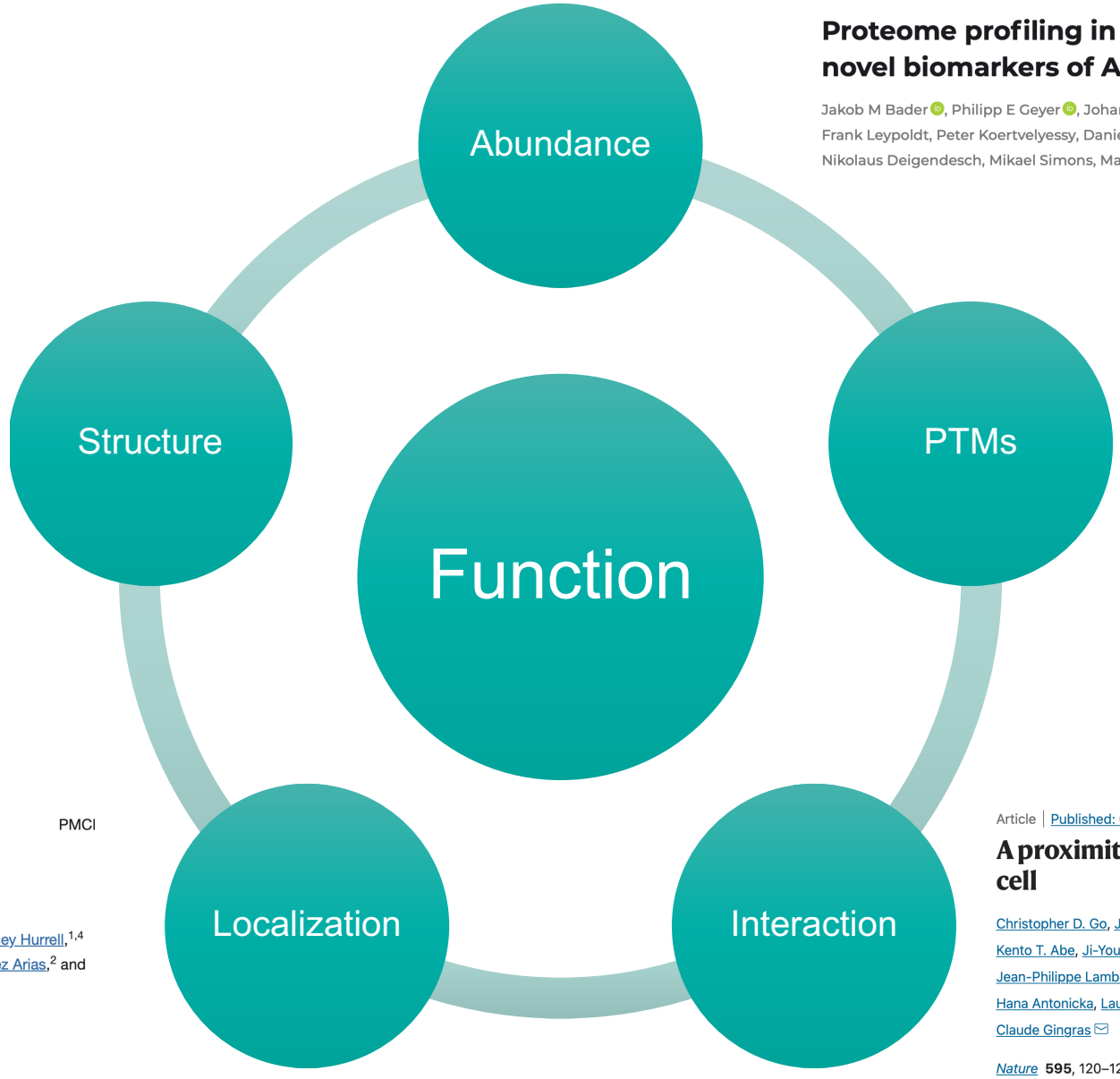


Statistical analysis



Interpretation

MS-based proteomic studies



Article | 2 June 2020 | OPEN ACCESS | TRANSPARENT PROCESS

Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease

Jakob M Bader, Philipp E Geyer, Johannes B Müller, Maximilian T Strauss, Manja Koch, Frank Leypoldt, Peter Koertvelyessy, Daniel Bittner, Carola G Schipke, Enise I Incesoy, Oliver Peters, Nikolaus Deigendesch, Mikael Simons, Majken K Jensen, Henrik Zetterberg, Matthias Mann

Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry

Simone Schopper, Abdullah Kahraman, Pascal Leuenberger, Yuehan Feng, Ilaria Piazza, Oliver Müller, Paul J Boersema & Paola Picotti

Nature Protocols **12**, 2391–2410 (2017) | Cite this article

Article | Open access | Published: 07 December 2022

Phosphoproteomic analysis of neoadjuvant breast cancer suggests that increased sensitivity to paclitaxel is driven by CDK4 and filamin A

S. Mouron, M. J. Bueno, A. Lluch, L. Manso, I. Calvo, J. Cortes, J. A. Garcia-Saenz, M. Gil-Gil, N. Martinez-Janez, J. V. Apala, E. Caleiras, Pilar Ximénez-Embún, J. Muñoz, L. Gonzalez-Cortijo, R. Murillo, R. Sánchez-Bayona, J. M. Cejalvo, G. Gómez-López, C. Fustero-Torre, S. Sabroso-Lasa, N. Malats, M. Martinez, A. Moreno, D. Megias, ... M. Quintela-Fandino + Show authors

Nature Communications **13**, Article number: 7529 (2022) | Cite this article

Nat Commun. 2016; 7: 9992. PMCI
Published online 2016 Jan 12. doi: 10.1038/ncomms9992

A draft map of the mouse pluripotent stem cell spatial proteome

Andy Christoforou,[1,2] Claire M. Mulvey,[1,2] Lisa M. Breckels,[1,3] Aikaterini Geladaki,[1,2] Tracey Hurrell,[1,4] Penelope C. Hayward,[2] Thomas Naake,[1,3] Laurent Gatto,[1,3] Rosa Viner,[5] Alfonso Martinez Arias,[2] and Kathryn S. Lilley[a,1]

Article | Published: 02 June 2021

A proximity-dependent biotinylation map of a human cell

Christopher D. Go, James D. R. Knight, Archita Rajasekharan, Bhavisha Rathod, Geoffrey G. Hesketh, Kento T. Abe, Ji-Young Youn, Payman Samavarchi-Tehrani, Hui Zhang, Lucie Y. Zhu, Evelyn Popiel, Jean-Philippe Lambert, Étienne Coyaud, Sally W. T. Cheung, Dushyandi Rajendran, Cassandra J. Wong, Hana Antonicka, Laurence Pelletier, Alexander F. Palazzo, Eric A. Shoubridge, Brian Raught & Anne-Claude Gingras

Nature **595**, 120–124 (2021) | Cite this article