



University of
Zurich^{UZH}

SIAF
Swiss Institute of Allergy &
Asthma Research

Proteomics

28.10.2025

PD Dr. Katja Bärenfaller
katja.baerenfaller@siaf.uzh.ch



Swiss Institute of
Bioinformatics

After this lecture, you will be able:

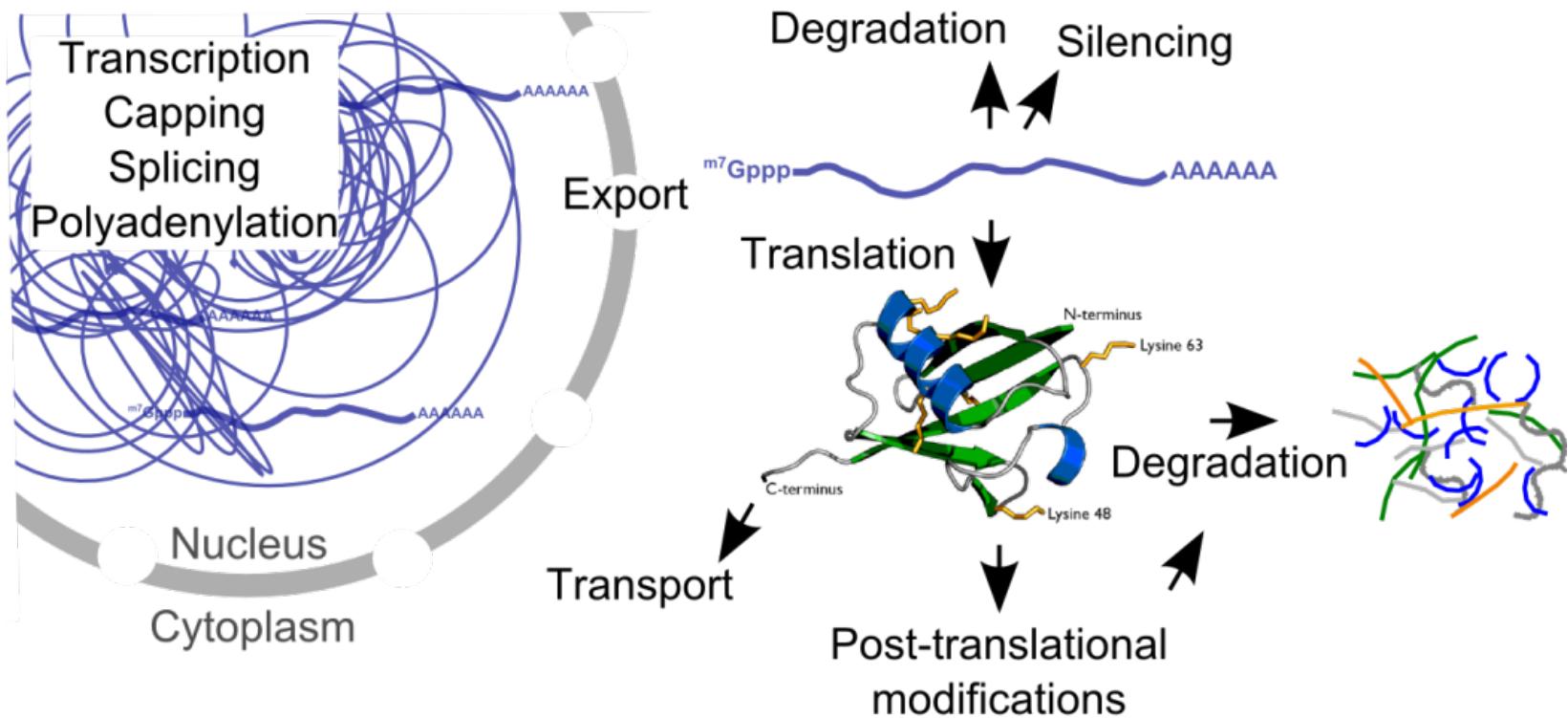
- to recognize and describe the significance of protein characterization
- to explain the principle of a mass spectrometry-based experiment
- to apply the principle of *de novo* peptide spectrum matching
- to describe the principle of database-dependent peptide spectrum matching
- to calculate the false discovery rate based on a target decoy approach
- to describe various application methods of proteomics

➤ Buzz group (5 min):

Discuss in groups of two or three why examining proteins is important and what information can be gained (only) through the direct characterization of proteins.

➤ If you are wondering what proteins are, spend some time looking at the Glossary instead.

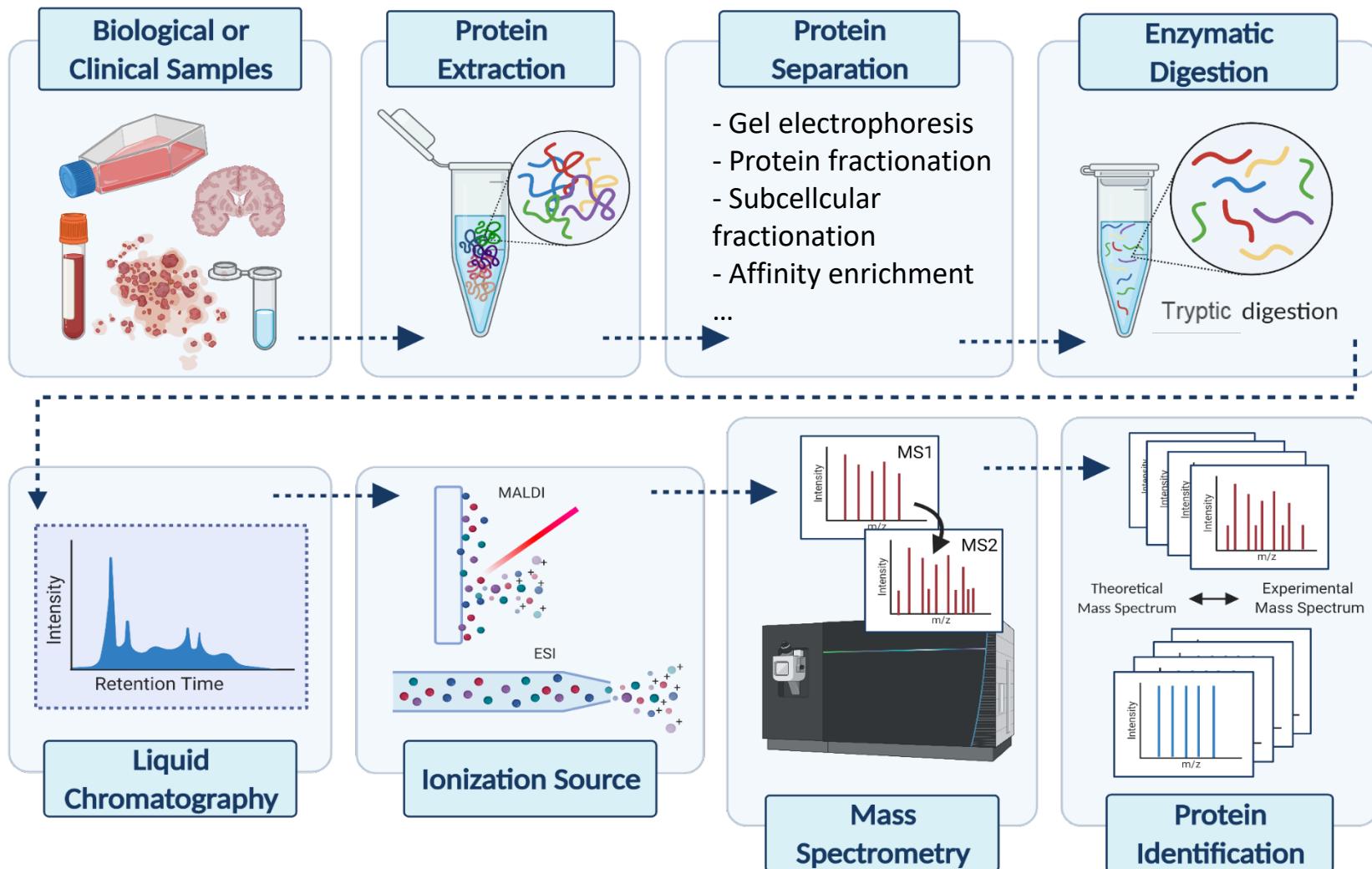
Why proteome research?



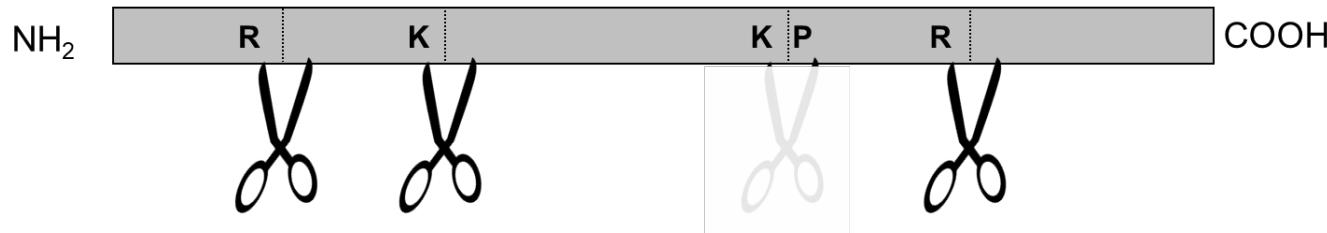
- The phenotypic profiles are determined by the molecular profiles of the individual cellular components.
- It is primarily the proteins, their activities, modifications, subcellular localization, and interactions that are responsible for determining the appearance and state of a biological organism.
- Various processes determine protein levels and activities
- Proteomics is the large-scale study of proteins and the proteome, which is the entire set of proteins produced or modified by an organism or system.

Generic mass spectrometry-based proteomics experiment

Elena Barletta



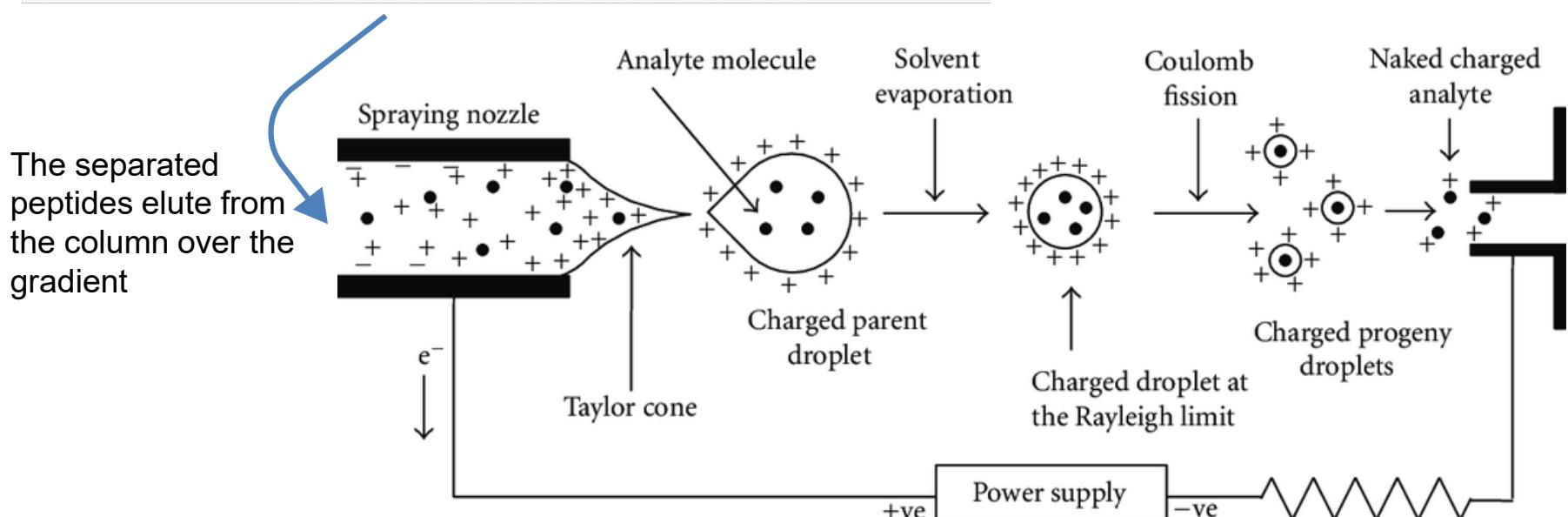
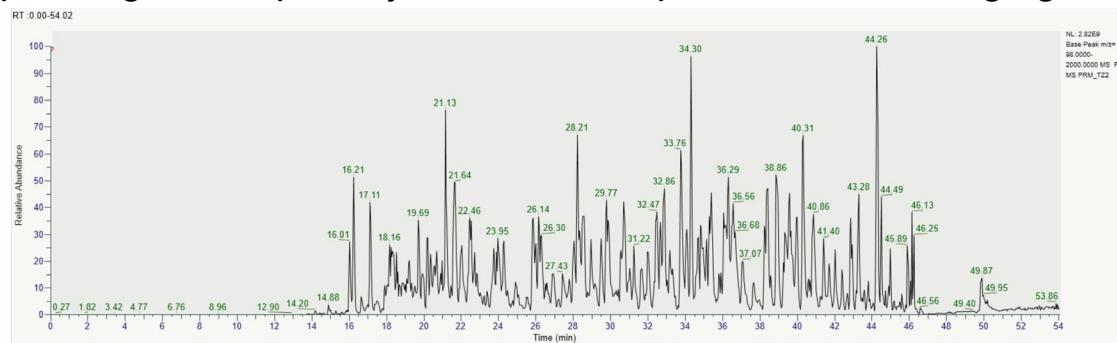
- Before analysis, the proteins are typically digested with a site-specific protease, most of the time with trypsin.
- Trypsin cuts after arginine or lysine, except when the cutting site is followed by proline, which leads to limited cleavage.

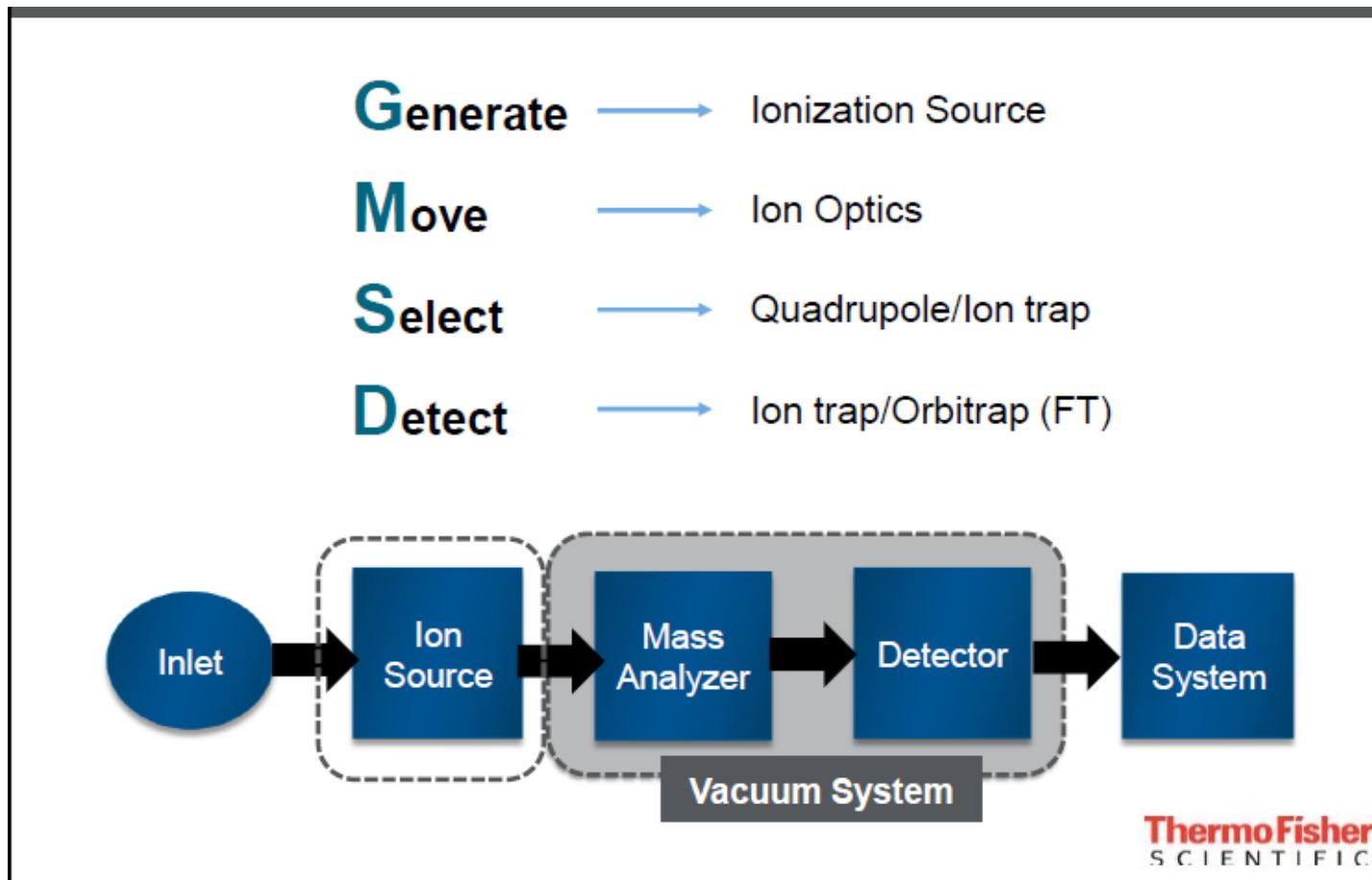


- After protein cleavage, the peptides need to be purified.
- The peptides are often fractionated or enriched for specific features such as specific post-translational modifications, internal cleavage sites, etc.

Liquid Chromatography and Ion Generation with Electrospray Ionization (ESI)

In reverse phase (RP) chromatography, the peptides are separated on a column filled with C18 beads; the components are separated from each other based on their affinity for the stationary phase depending on the polarity of the mobile phase that is changing over time.





Mass spectrometer types:

TOF = Time of Flight

Quadrupole (Q)

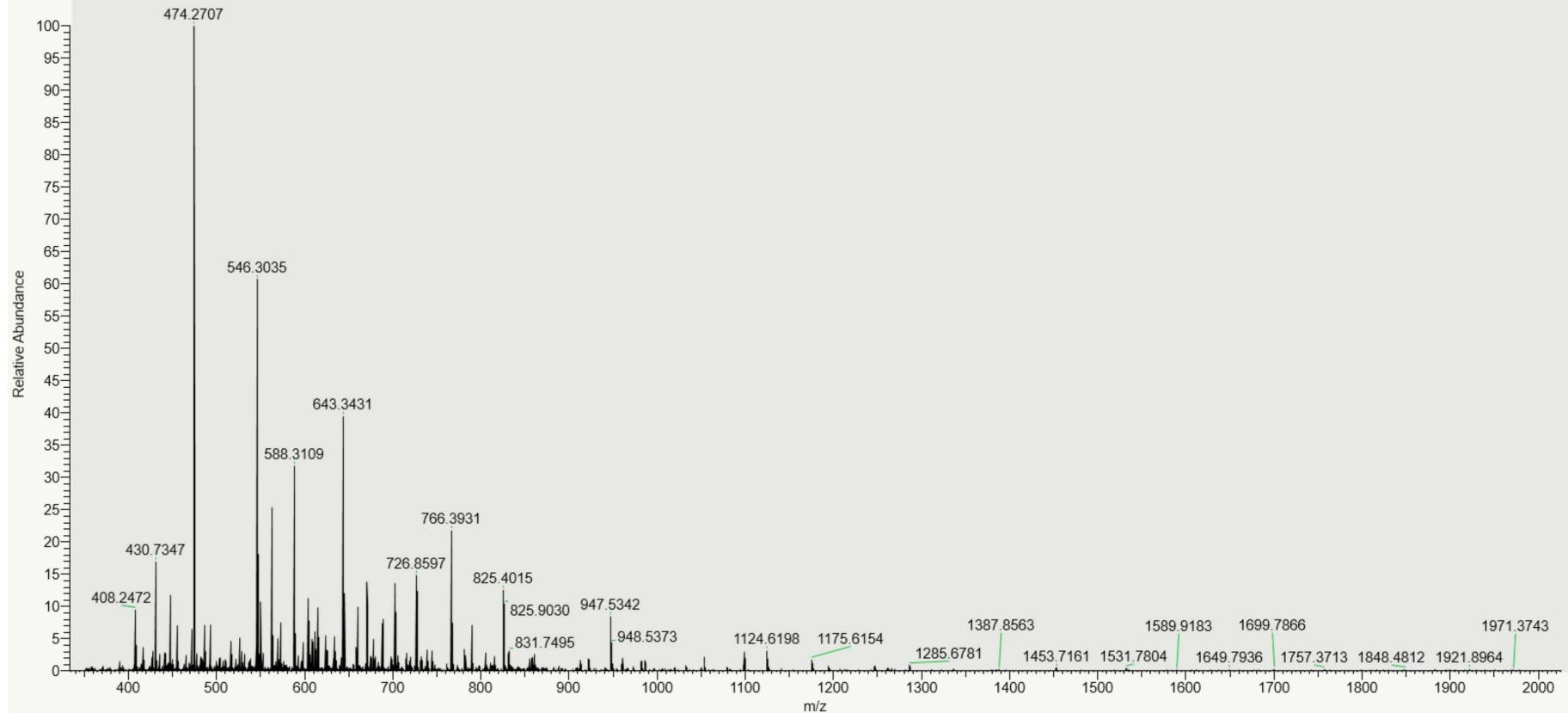
Ion Trap

Orbitrap

FT-ICR = Fourier Transform Ion Cyclotron Resonance

MS Spectrum

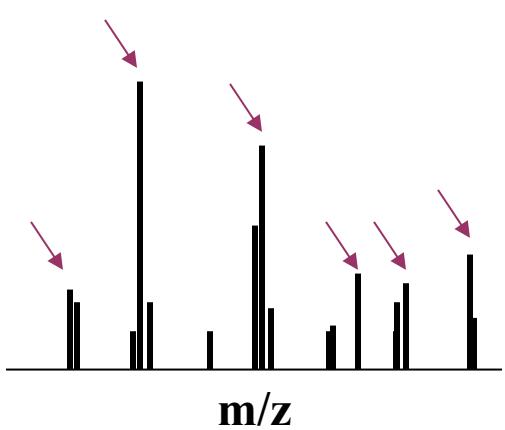
PRM_TZ2 #43407 RT: 32.27 AV: 1 NL: 2.22E+008
T: FTMS + p NSI Full ms [350.0000-2000.0000]



Peptide Mass Fingerprint

Identifying peptides using an MS spectrum:

List of peptide masses from MS scan



Search algorithm

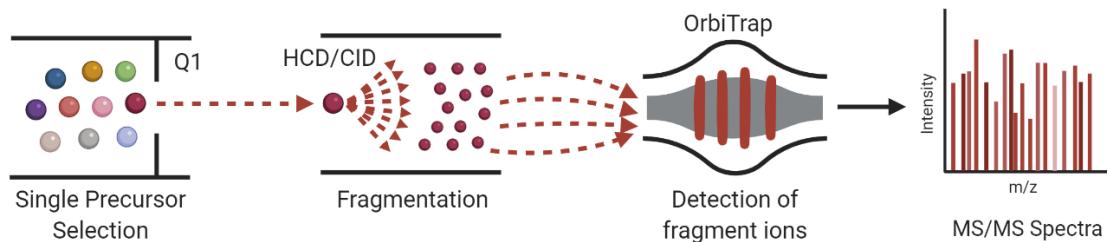
Identified peptide/protein

Sequence database

- Peptide spectrum assignment with Peptide Mass Fingerprinting is only advisable with samples of low complexity and small sequence databases, as the number of all possible peptides with a given mass over charge is huge in large sequence databases.

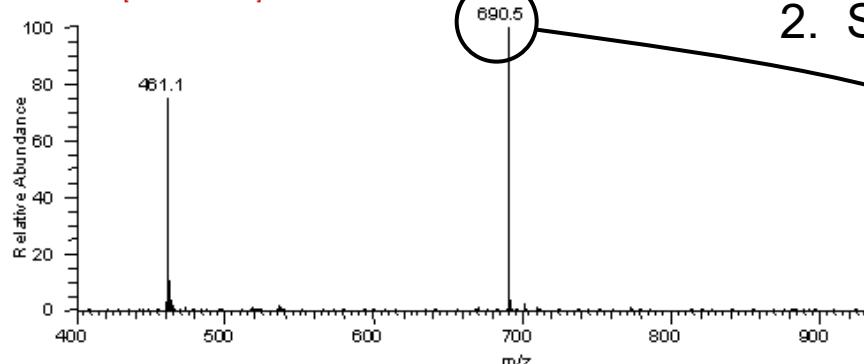
Tandem Mass Spectrometry (MS/MS)

- Obtaining sequence information for a peptide ion:



1. Acquire full (MS) scan

F: + c Full ms [400.00-1000.00]



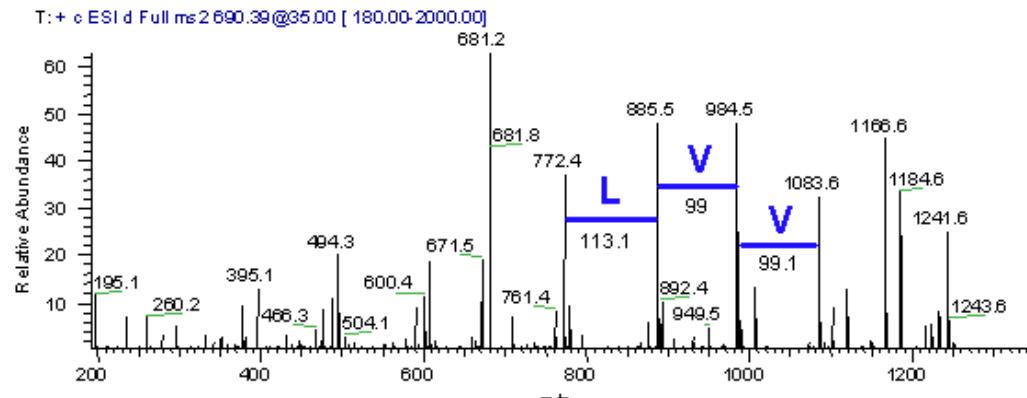
2. Select an ion

Radzikowska et al., *Omics technologies in allergy and asthma research: An EAACI position paper*, Allergy, 2022

3. Isolate the ion

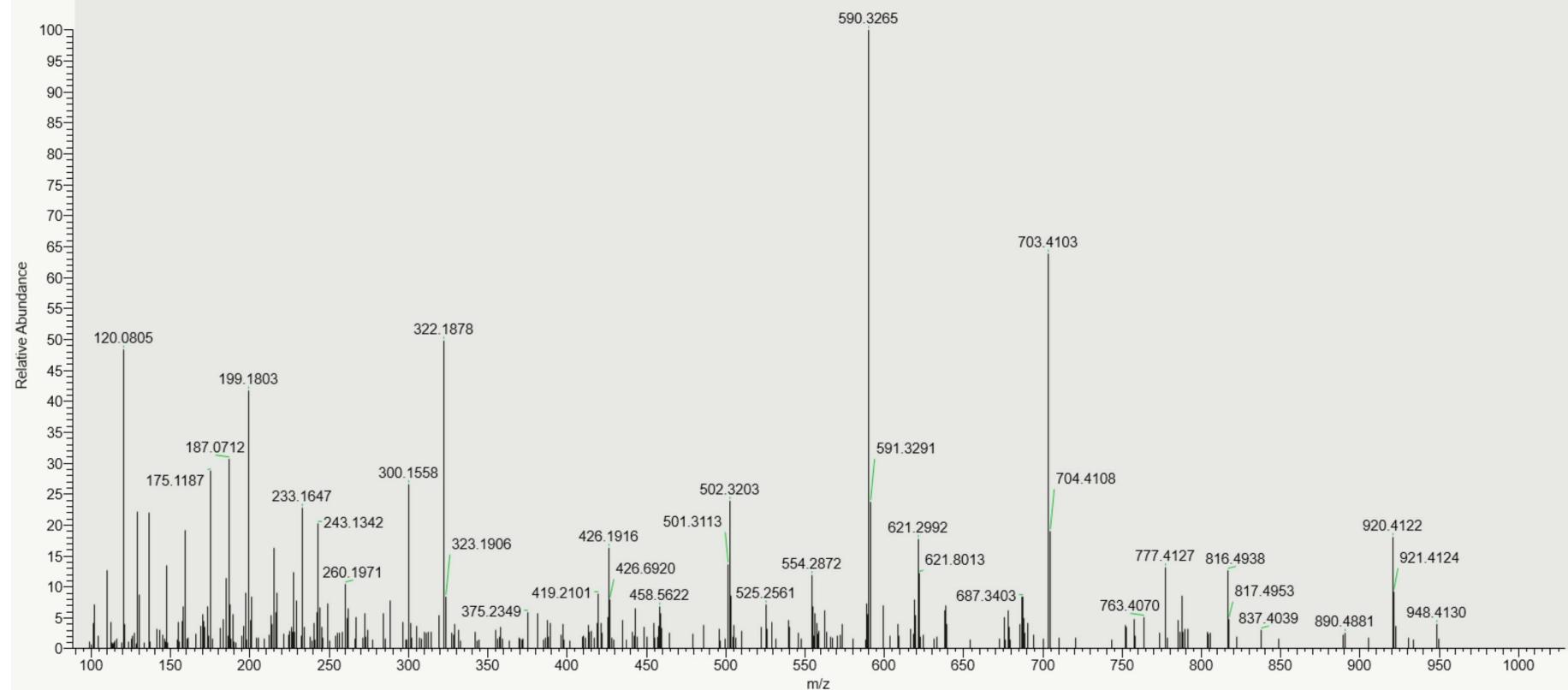
4. Fragment the ion

5. Acquire MS/MS scan

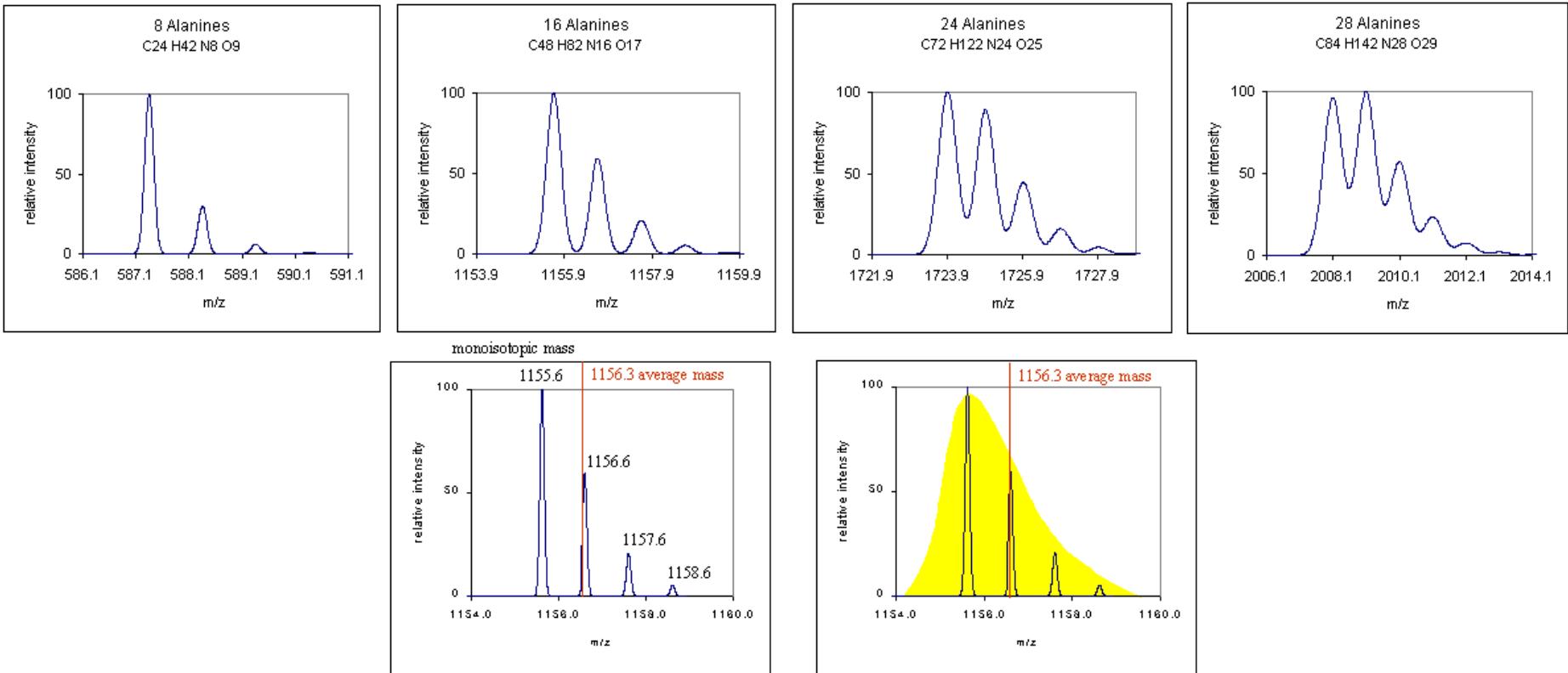


MS/MS spectrum

PRM_TZ2 #43405 RT: 32.27 AV: 1 NL: 7.59E+005
T: FTMS + c NSI Full ms2 501.7769@hcd30.00 [98.0000-1014.0000]

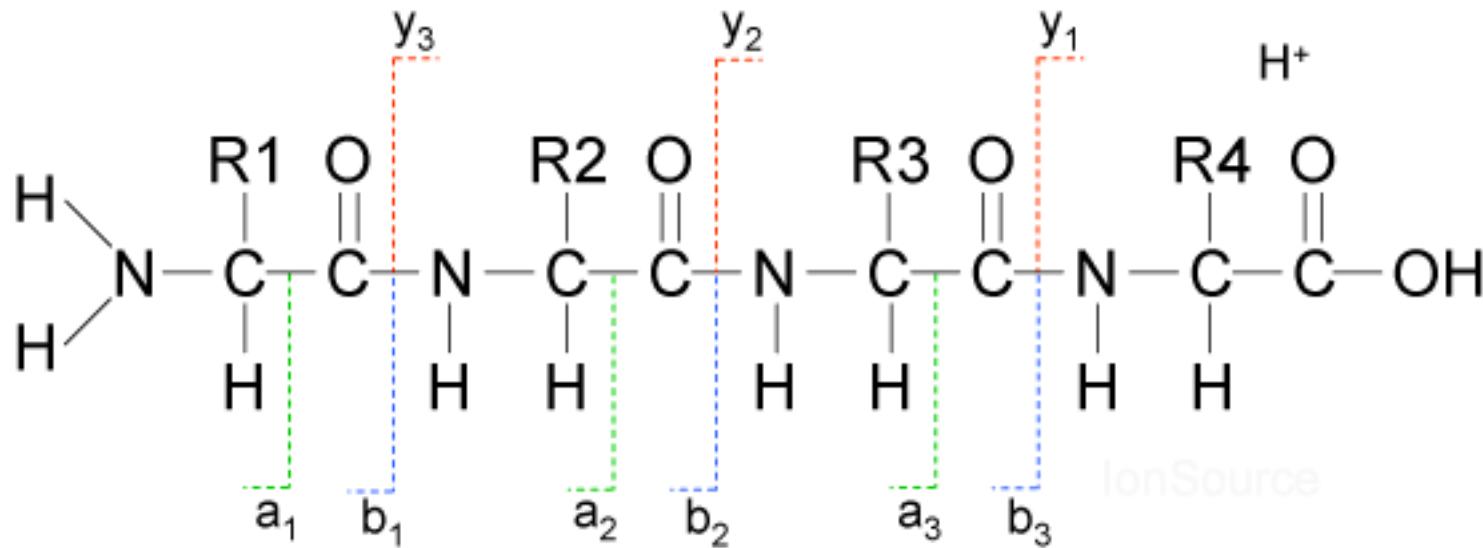


The isotope issue



- 1/100 C atoms is C13
- The more atoms a peptide contains, the more probably it is that one to several C atoms are C13

a, b and y ions after CID



- The most common peptide fragments observed in low energy collisions are **a**, **b** and **y** ions.
- The **b** ions appear to extend from the amino terminus (N-terminus), and **y** ions appear to extend from the carboxyl terminus (C-terminus).
- **a** ions occur at a lower frequency and abundance in relation to **b** ions.

adjusted from <http://www.ionsource.com/>

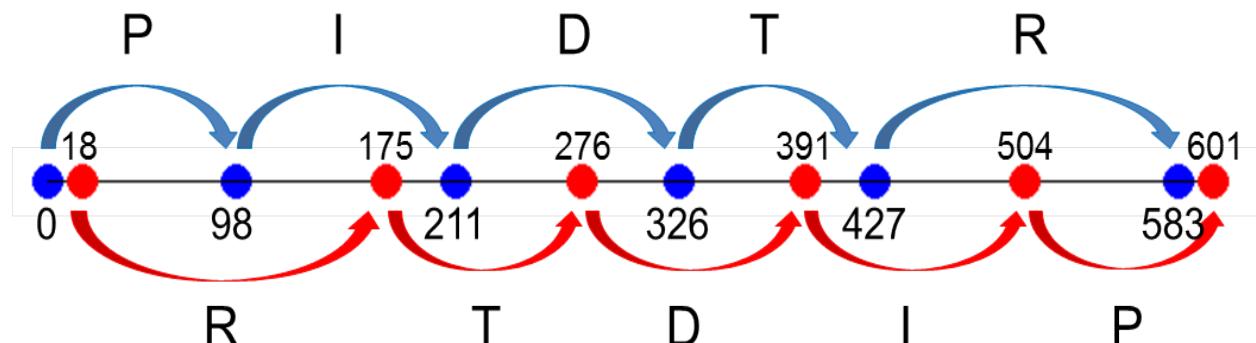
P – I – D – T – R

$m/z = 601.31$

Masses of b- and y- ions:

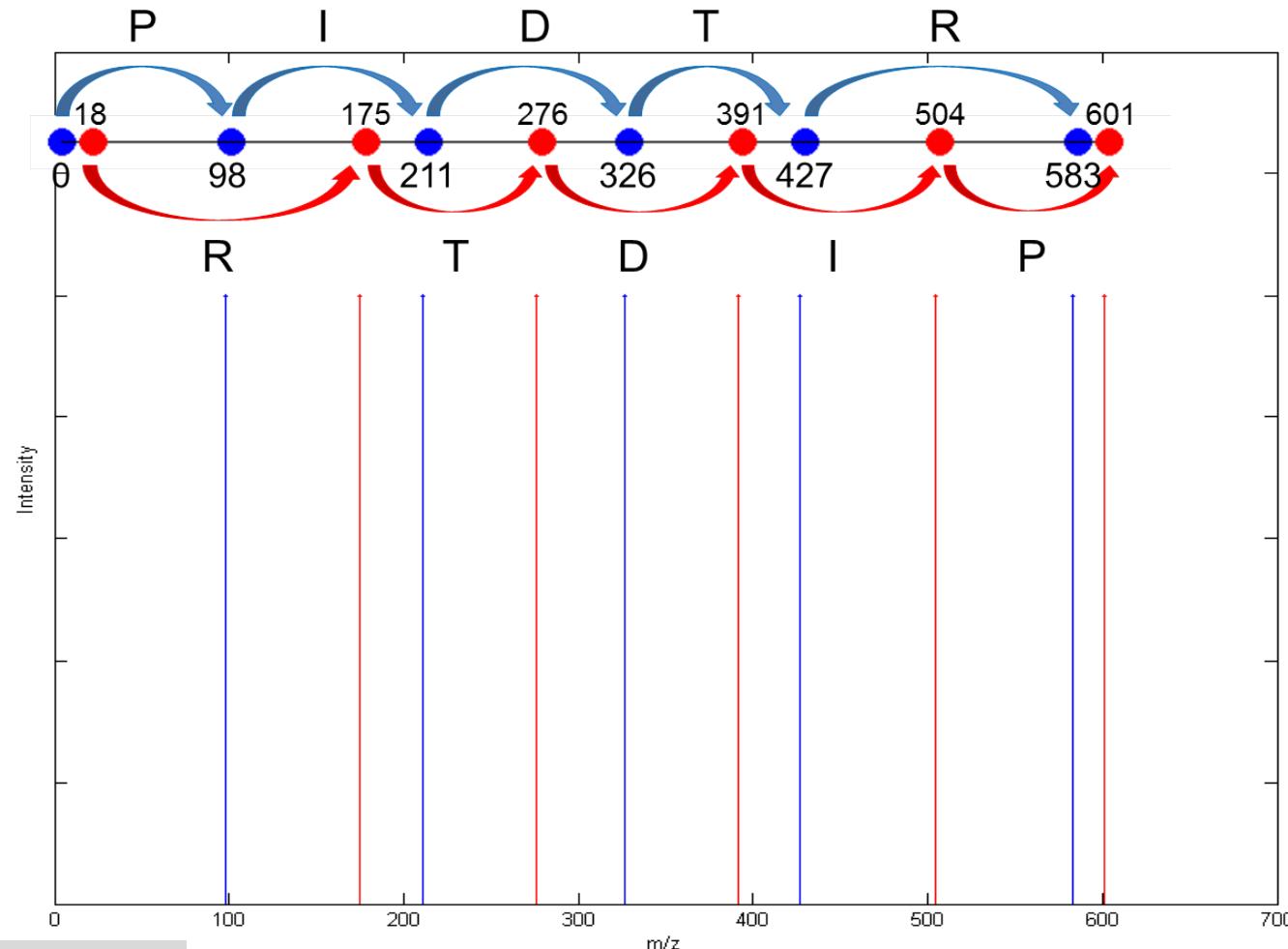
	b-ions			y-ions
		PIDTR	601.31	y5
b1	98.05	P ----- IDTR	504.26	y4
b2	211.14	PI ----- DTR	391.18	y3
b3	326.16	PID ----- TR	276.15	y2
b4	427.21	PIDT ----- R	175.10	y1
b5	583.31	PIDTR		

Fragment masses aligned along a spectrum graph:



Hypothetical fragment spectrum

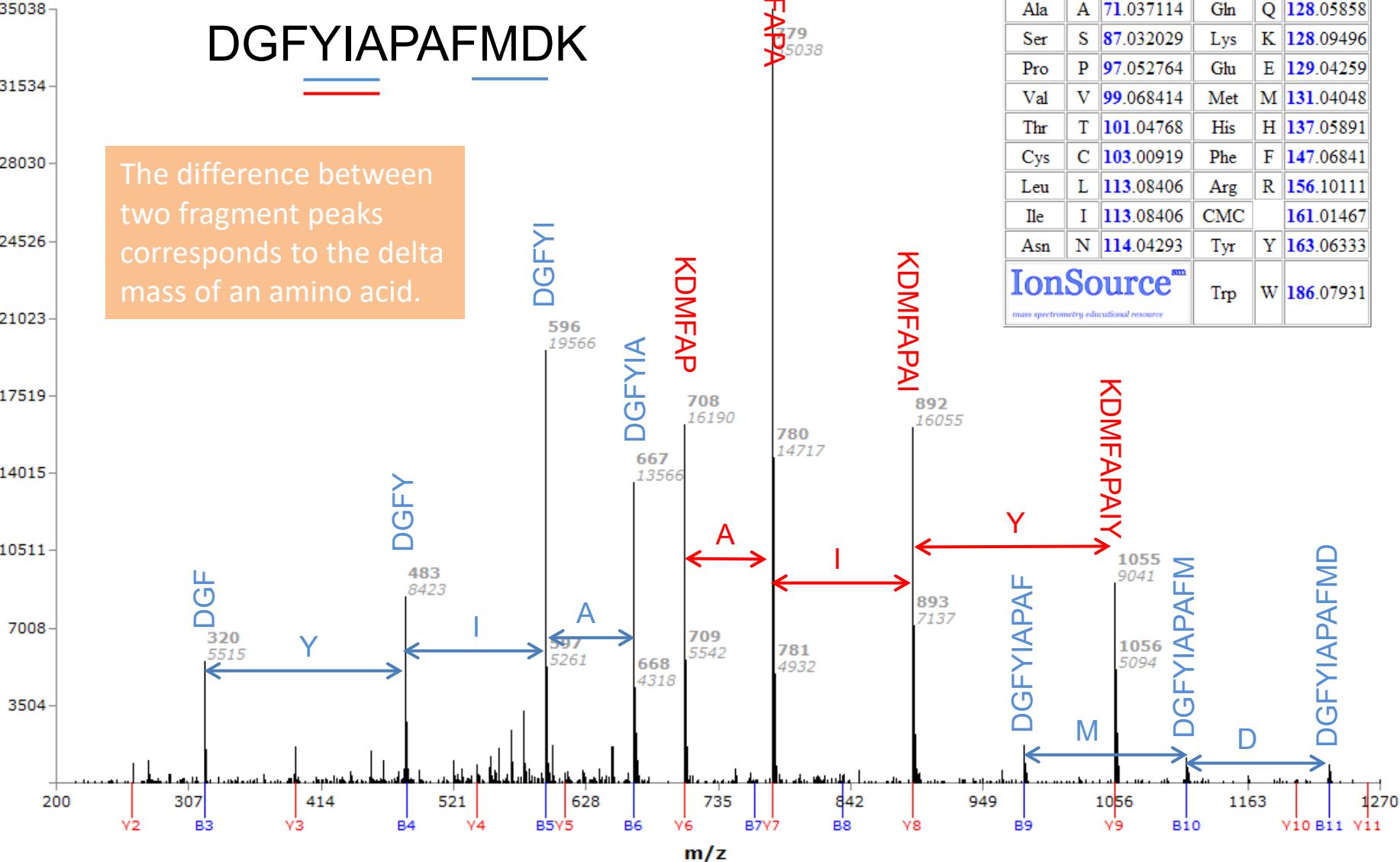
P – I – D – T – R
 $m/z = 601.31$



Analysing a spectrum for the fragments

DGFYIAPAFMDK

The difference between two fragment peaks corresponds to the delta mass of an amino acid.



AA Codes	Mono.	AA Codes	Mono.
Gly G	57.021464	Asp D	115.02694
Ala A	71.037114	Gln Q	128.05858
Ser S	87.032029	Lys K	128.09496
Pro P	97.052764	Glu E	129.04259
Val V	99.068414	Met M	131.04048
Thr T	101.04768	His H	137.05891
Cys C	103.00919	Phe F	147.06841
Leu L	113.08406	Arg R	156.10111
Ile I	113.08406	CMC	161.01467
Asn N	114.04293	Tyr Y	163.06333
IonSource sm		Trp W	186.07931
mass spectrometry educational resource			

In *de novo* peptide sequencing, all information about the peptide sequence resides in the MS/MS spectrum itself. It is therefore database-independent and can be done manually or with algorithms.

Manual *de novo* peptide sequencing:

- M = Mass of the peptide

- **Precursor ion** $M_{\text{Precursor}} \equiv \frac{(M + 2H)^{2+}}{2}$

- **Parent ion** $M_{\text{Parent}} \equiv (M + H)^{+}$

$$M_{\text{Parent}} = M_{\text{Precursor}} * 2 - 1 = \frac{(M + 2H)^{2+}}{2} * 2 - 1$$

- Monoisotopic mass of the parent ion

$$M_{\text{Parent mono}} = M_{\text{Parent average}} - \frac{M_{\text{Parent average}}}{1463}$$

- b- and y- ions:

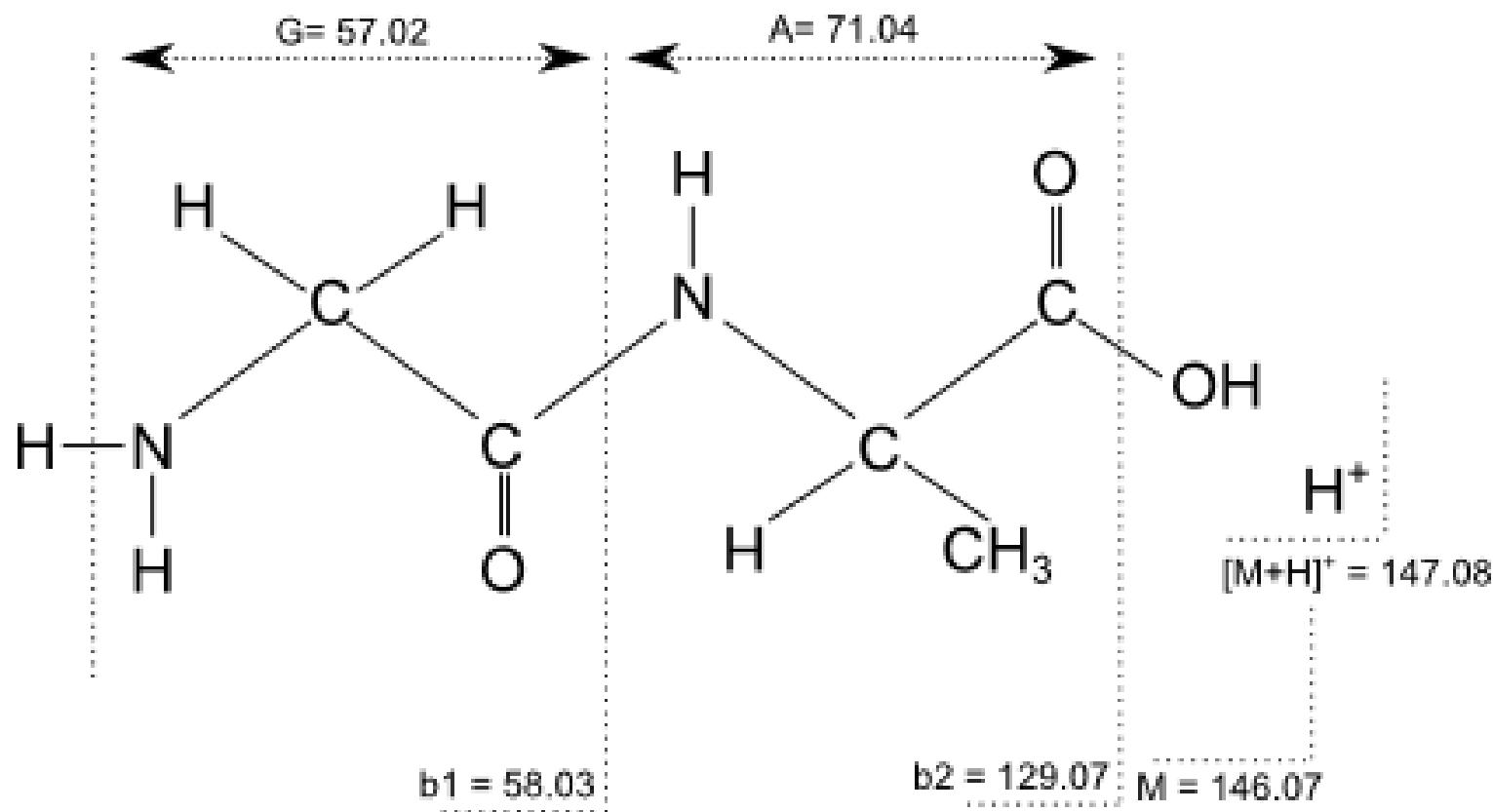
Zubarev and Bondarenko, 1991

$$y = (M + H)^{+} - b + 1 ; b = (M + H)^{+} - y + 1$$

- Find the b-ion without the C-terminal K or R: $(M + H)^{+} - 18$ ('lost' oxygen) – [KR]

AA Codes	Mono.	AA Codes	Mono.		
Gly	G	57.021464	Asp	D	115.02694
Ala	A	71.037114	Gln	Q	128.05858
Ser	S	87.032029	Lys	K	128.09496
Pro	P	97.052764	Glu	E	129.04259
Val	V	99.068414	Met	M	131.04048
Thr	T	101.04768	His	H	137.05891
Cys	C	103.00919	Phe	F	147.06841
Leu	L	113.08406	Arg	R	156.10111
Ile	I	113.08406	CMC		161.01467
Asn	N	114.04293	Tyr	Y	163.06333
IonSource™		mass spectrometry educational resource			
	Trp		W	186.07931	

'Lost' oxygen



- Upon peptide bond formation, H_2O gets released and this mass is therefore not included in the Δ mass for the individual amino acids
- The largest b-ion is mass of the parent ion – 18, because H_2O gets released and the positive charge resides at the C-terminus

20070130_11_ppi1_B1 #1321 RT: 38.25 AV: 1 NL: 1.57E6
 T: + c d Full ms2 946.76@35.00 [250.00-1905.00]

$$M_{\text{Precursor}} = \frac{(M + 2H)^2}{2} = 946.76$$

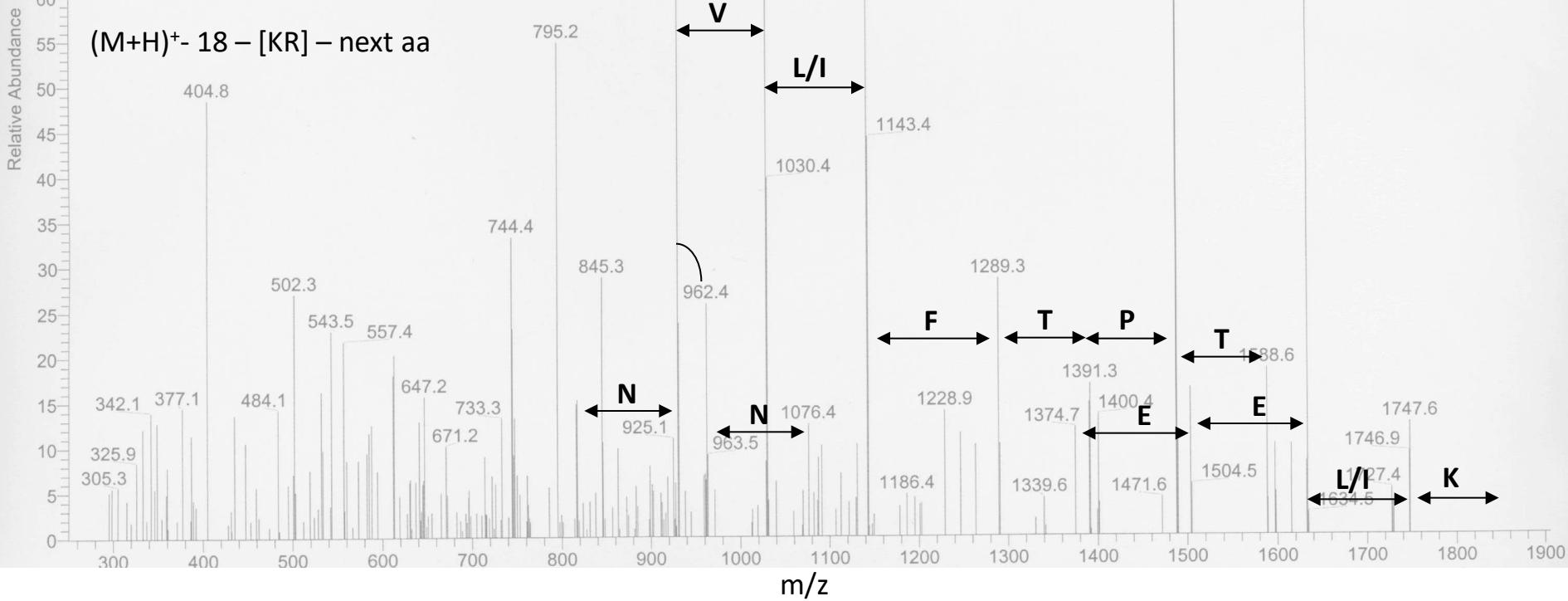
$$M_{\text{Parent average}} = (M+H)^+ = 1892.52$$

$$M_{\text{Parent mono}} = (M+H)^+ \sim 1892$$

$$y = (M+H)^+ - b + 1 ; b = (M+H)^+ - y + 1 ;$$

$(M+H)^+ - 18 - [\text{KR}]$

$(M+H)^+ - 18 - [\text{KR}] - \text{next aa}$



Neutral losses and peptide modifications

- Neutral losses
 - H₂O (-18 Da)
 - NH₂ (-17 Da)
 - CO (-28 Da)
 - H₃PO₄ (-98 Da)

- Modifications

Oxidation: M +16 Da

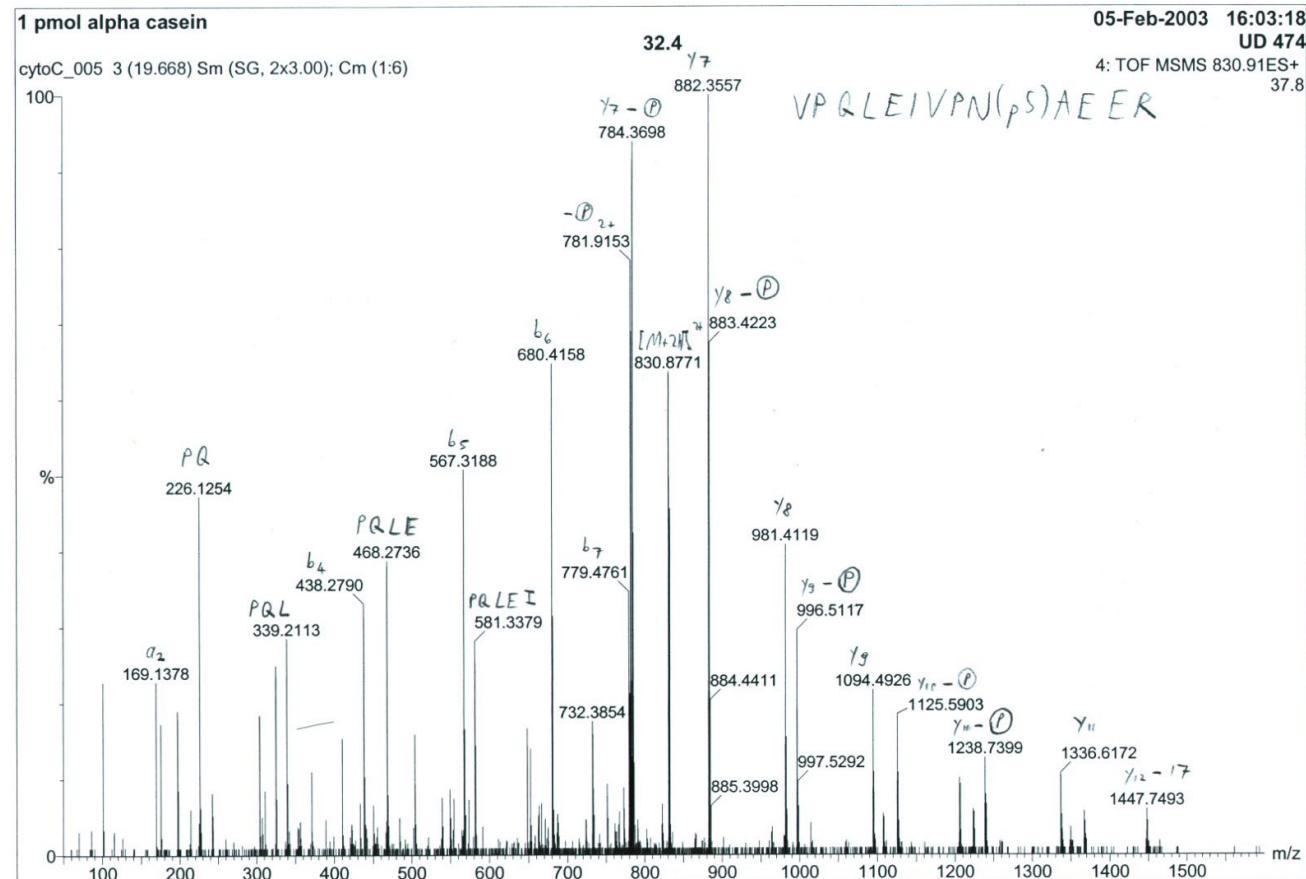
Deamidation: N → Q, -1 Da

Methylation: R +14 Da

Phosphorylation: S, T, Y +80 Da

Acetylation: K, R, N-terminus +42 Da

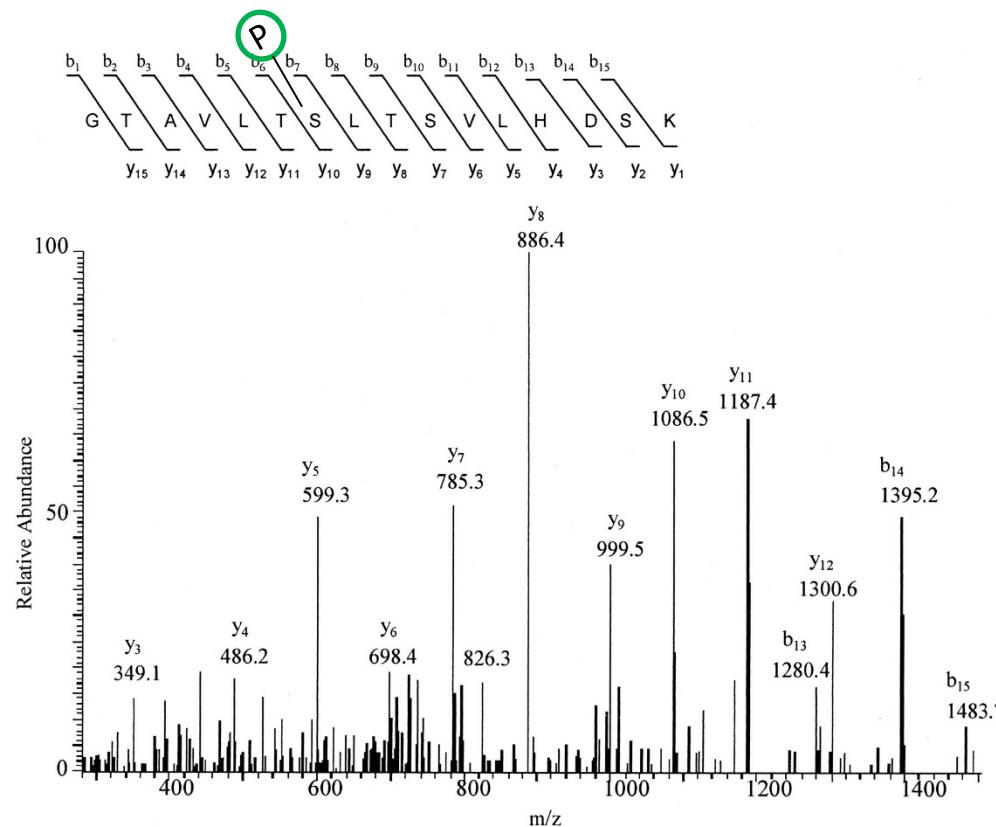
Carboxyamidomethylation: C +57 Da



De novo peptide sequencing – Exercise

Given the following peptide with the indicated fragment ions,

- 1) Which of the ions **b₂**, **b₇**, **b₁₁**, **b₁₅**, **y₁**, **y₆**, **y₁₀**, and **y₁₄** will change their m/z when serine 7 was phosphorylated?

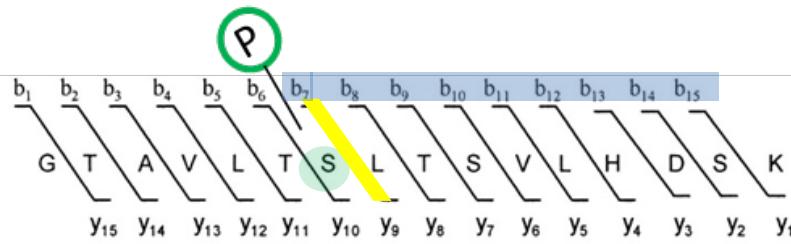


Nisar et al., *Drug Metabolism and Predisposition*, 2004
<https://doi.org/10.1124/dmd.32.4.382>

De novo peptide sequencing – Exercise

Given the following peptide with the indicated fragment ions,

- 1) Which of the ions **b2**, **b7**, **b11**, **b15**, **y1**, **y6**, **y10**, and **y14** will change their m/z when serine 7 was phosphorylated?



- All the b-ions (extending from the N-terminus) that contain the serine 7 will change their m/z; starting from b₇, where the amide bond between serine 7 and leucine 8 breaks:

b₆ GTAVLT – S_{LTSVLHDSK} y₁₀

b₇ GTAVLTS_LTSVLHDSK y₉

b₈ GTAVLTS_L – TSVLHDSK y₈

b₉ GTAVLTS_LT – SVLHDSK y₇, b₁₀, b₁₁, b₁₂, b₁₃, b₁₄, b₁₅

- All the y-ions (extending from the C-terminus) that contain the serine 7 will change their m/z; starting from y₁₀, where the amide bond between threonine 6 and serine 7 breaks:

b₇ GTAVLTS – L_{TSVLHDSK} y₉

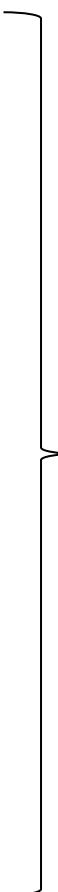
b₆ GTAVLT – S_{LTSVLHDSK} y₁₀

b₅ GTAVL – T_{SLSVLHDSK} y₁₁

b₄ GTAV – L_{TSLTSVLHDSK} y₁₂

b₃ GTA – V_{LTSLSVLHDSK} y₁₃, y₁₄, y₁₅

Different database-dependent peptide identification search algorithms:

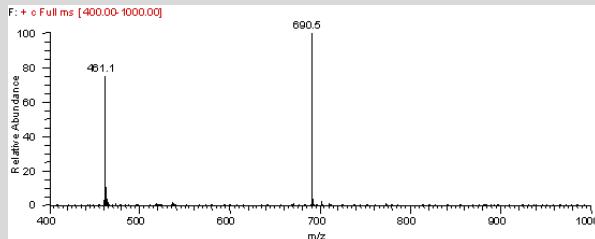
- Sequest
 - Mascot
 - PepSplice
 - OMSSA
 - X!Tandem
 - Phenyx
 - ProteinPilot
 - SpectrumMill
 - ProbID
 - PepFrag
 - InSpect
 - ...
- 
- Peptide sequences with one or more scores with which to evaluate the likelihood that the resulting sequence is correct.
 - Even though each implementation is different, they operate under the same general principle.

Database-dependent peptide identification

Experimental

Protein sample

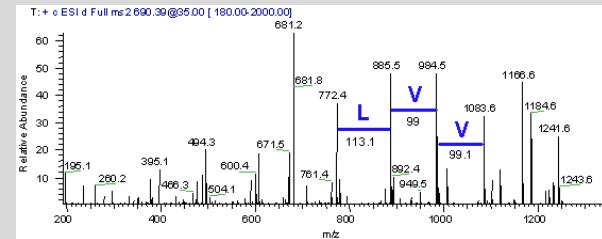
tryptic digest



Full Scan [MS]:

Mass measurement of full peptides

fragmentation



MS/MS-Scan:

Measurement of peptide fragments

1) Peptides from a protein database are matched to the measured mass

2) Theoretical spectrum is cross-correlated to the measured one

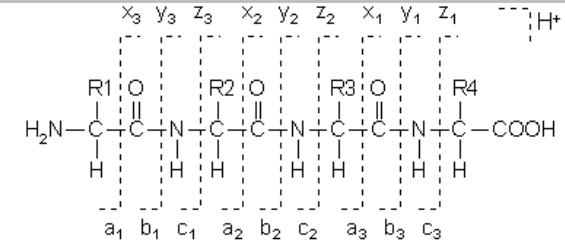
Protein database

in silico digest

Peptide database

in silico dissociation

Theoretical tryptic peptides
from a protein database



Theoretical spectra
assuming peptide bond breakage

In silico

Peptide spectrum assignment

- The goal is to identify the best sequence match to the spectrum
- The details of this implementation differ among the algorithms. In addition, the methods used to assign scores are very different.
- Several approaches have been developed to model matches to the sequences: descriptive models, interpretative models, stochastic models and statistical and probability models (Sadygov, Cociorva and Yates, 2004)

Descriptive algorithms are based on a mechanistic prediction of how peptides fragment in a tandem mass spectrometer, which is then quantified to determine the quality of the match between the prediction and the experimental spectrum. Mathematical methods such as correlation analysis have been used to assess match quality.

- Sequest is an example of a program using a descriptive model:
 - S_p , sums the peak intensity of fragment ions matching the predicted sequence ions and accounts for the continuity of an ion series and the length of a peptide
 - Xcorr, is a cross-correlation score of the experimental and theoretical spectra
 - ΔC_n gives the normalised difference of Xcorr values between the best sequence and lower-scoring matches and is useful to determine the uniqueness of the match

Sadygov, Cociorva and Yates, 2004

Empirical statistical model to estimate the accuracy of peptide identifications

HTML-SUMMARY v.8 (rev 0), Copyright 1996 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

file:///Snapserver/proteomesoft/users/brian/sergei_digest_A_full_01.html

PSWebMail Gmail - Inbox Slashdot Google News PSWebPage JDK 1.4 API

HTML-SUMMARY v.8 (rev 0), Copyright 1996

Molecular Biotechnology, Univ. of Washington, J.Eng/J.Yates
Compiled for use by the Aebersold lab @ Univ. of Washington
03/19/02, 07:20 AM, /data/search/akeller/databases/new_hum_plus_proteinmix.db, AVG/AVG

#	File	MH+	XCorr	dCn	Sp	RSp	Ions	Ref	Sequence
325	./sergei_digest_A_full_01.1001.1003.3	2511.7 (-0.6)	7.5863	0.450	4546.1	✓ 1	42/ 84	sp P00921 CAH2_BOVIN	R.MVMNGHSFNVVEYDDSQDKAVLK.D
462	./sergei_digest_A_full_01.1239.1241.3	2584.7 (+2.0)	6.1808	0.382	3126.2	✓ 1	39/ 84	sp P00921 CAH2_BOVIN	R.LVQFHFMWGSSBBQGSEHTVDR.K
1070	./sergei_digest_A_full_01.2335.2337.3	2254.5 (+0.5)	6.0682	0.491	2166.9	✓ 1	37/ 84	sp P00921 CAH2_BOVIN	K.YGDFGTAAQQPDGLAVVGVLK.V
1105	./sergei_digest_A_full_01.2405.2407.3	2254.5 (+0.8)	6.0041	0.511	1873.6	✓ 1	35/ 84	sp P00921 CAH2_BOVIN	K.YGDFGTAAQQPDGLAVVGVLK.V
510	./sergei_digest_A_full_01.1317.1325.3	2584.7 (+1.5)	5.9521	0.403	2488.3	✓ 1	38/ 84	sp P00921 CAH2_BOVIN	R.LVQFHFMWGSSBBQGSEHTVDR.K
1219	./sergei_digest_A_full_01.2617.2619.3	2187.6 (-0.1)	5.7343	0.502	2282.1	✓ 1	33/ 72	sp P02666 CASE_BOVIN	R.DMP1QAFLLYQEPEVLPGPVR.G
894	./sergei_digest_A_full_01.2013.2015.3	2314.7 (+0.0)	5.5636	0.418	1260.3	✓ 1	33/ 76	sp P02754 LACB_BOVIN	R.VYVEELKPTPEGDLEILLQK.W
812	./sergei_digest_A_full_01.1873.1875.3	2314.7 (+0.5)	5.5466	0.428	1407.6	✓ 1	35/ 76	sp P02754 LACB_BOVIN	R.VYVEELKPTPEGDLEILLQK.W
1142	./sergei_digest_A_full_01.2471.2473.2	2254.5 (-0.8)	5.5372	0.526	771.0	✓ 1	24/ 42	sp P00921 CAH2_BOVIN	K.YGDFGTAAQQPDGLAVVGVLK.V
856	./sergei_digest_A_full_01.1943.1945.3	2314.7 (+0.9)	5.4579	0.426	1581.3	✓ 1	34/ 76	sp P02754 LACB_BOVIN	R.VYVEELKPTPEGDLEILLQK.W
1289	./sergei_digest_A_full_01.2765.2771.2	2709.1 (+0.4)	5.3678	0.495	1654.1	✓ 1	24/ 50	sp P02754 LACB_BOVIN	K.VAGTWYSLAMAADISLILLAQSAPLR.V
1220	./sergei_digest_A_full_01.2621.2623.2	2187.6 (-0.5)	5.3391	0.461	1646.5	✓ 1	22/ 36	sp P02666 CASE_BOVIN	R.DMP1QAFLLYQEPEVLPGPVR.G
1153	./sergei_digest_A_full_01.2491.2493.3	2219.5 (+2.1)	5.3167	0.276	1640.4	✓ 1	31/ 72	sp P00921 CAH2_BOVIN	R.TLNFnAEAGEPELLMLANUR.P
1102	./sergei_digest_A_full_01.2401.2403.2	2254.5 (-0.6)	5.1675	0.495	1009.0	✓ 1	24/ 42	sp P00921 CAH2_BOVIN	K.YGDFGTAAQQPDGLAVVGVLK.V
1067	./sergei_digest_A_full_01.2329.2333.2	2254.5 (-0.9)	5.1492	0.546	688.2	✓ 1	23/ 42	sp P00921 CAH2_BOVIN	K.YGDFGTAAQQPDGLAVVGVLK.V
125	./sergei_digest_A_full_01.0681.0681.2	2100.2 (+0.8)	4.9146	0.481	1779.2	✓ 1	22/ 34	sp P00921 CAH2_BOVIN	R.MVMNGHSFNVVEYDDSQDK.A
1020	./sergei_digest_A_full_01.2237.2239.2	1568.7 (-0.1)	4.8921	0.413	1764.0	✓ 1	19/ 24	sp P02769 ALBU_BOVIN	K.DAFLGSFLYEYSR.R
981	./sergei_digest_A_full_01.2163.2165.3	2147.3 (-0.7)	4.8738	0.452	1515.4	✓ 1	26/ 38	sp P02666 CASE_BOVIN	E.LNVPGEIVESLSSSEESITR.I
533	./sergei_digest_A_full_01.1361.1367.3	2906.0 (+0.7)	4.8712	0.301	655.0	✓ 1	29/ 92	sp P02666 CASE_BOVIN	K.FQSEEQQQTEDELQDKIHFAQTQ.S
815	./sergei_digest_A_full_01.1879.1883.2	2314.7 (-0.6)	4.7827	0.419	438.7	✗ 1	19/ 38	sp P02754 LACB_BOVIN	R.VYVEELKPTPEGDLEILLQK.W
760	./sergei_digest_A_full_01.1771.1773.2	2034.2 (-0.1)	4.7587	0.465	1047.9	✓ 1	26/ 36	sp P02666 CASE_BOVIN	L.NVPGEIVESLSSSEESITR.I
1048	./sergei_digest_A_full_01.2289.2291.2	2804.0 (+0.5)	4.7564	0.433	1497.0	✓ 1	26/ 48	sp P02666 CASE_BOVIN	A.RELEELNVPGEIVESLSSSEESITR.I
157	./sergei_digest_A_full_01.0735.0737.3	2100.2 (-0.5)	4.7471	0.300	1496.2	✓ 1	32/ 68	sp P00921 CAH2_BOVIN	R.MVMNGHSFNVVEYDDSQDK.A
91	./sergei_digest_A_full_01.0607.0609.2	1983.0 (-0.4)	4.6602	0.358	3224.6	✗ 1	24/ 30	sp P02666 CASE_BOVIN	K.FQSEEQQQTEDELQDK.I
625	./sergei_digest_A_full_01.1523.1525.3	2046.3 (-0.7)	4.6539	0.307	1901.4	✓ 1	29/ 60	sp P02769 ALBU_BOVIN	R.RHPFYFAPELLYYANK.Y
117	./sergei_digest_A_full_01.0663.0665.3	2100.2 (+0.0)	4.5901	0.286	1627.3	✓ 1	36/ 68	sp P00921 CAH2_BOVIN	R.MVMNGHSFNVVEYDDSQDK.A
712	./sergei_digest_A_full_01.1675.1681.2	1832.0 (-0.3)	4.5772	0.393	1148.8	✗ 1	21/ 34	sp Q29443 TRFE_BOVIN	K.GEADAMSLDGGLYLYIAGK.C
651	./sergei_digest_A_full_01.1569.1571.3	3106.3 (+0.4)	4.5737	0.362	537.5	✗ 1	26/100	sp P02666 CASE_BOVIN	K.FQSEEQQQTEDELQDKIHFAQTQL.V
942	./sergei_digest_A_full_01.2085.2089.3	2314.7 (+1.4)	4.5707	0.299	1348.2	✗ 1	32/ 76	sp P02754 LACB_BOVIN	R.VYVEELKPTPEGDLEILLQK.W
1045	./sergei_digest_A_full_01.2277.2279.3	2804.0 (-0.5)	4.5012	0.352	1236.7	✓ 1	31/ 96	sp P02666 CASE_BOVIN	A.RELEELNVPGEIVESLSSSEESITR.I
861	./sergei_digest_A_full_01.1951.1953.2	2314.7 (-0.4)	4.4874	0.348	482.4	✗ 1	20/ 38	sp P02754 LACB_BOVIN	R.VYVEELKPTPEGDLEILLQK.W
626	./sergei_digest_A_full_01.1529.1531.2	1480.7 (-0.5)	4.4192	0.315	1731.6	✗ 1	20/ 24	sp P02769 ALBU_BOVIN	K.LGEYGFQNALIVR.Y
176	./sergei_digest_A_full_01.0767.0767.2	1983.0 (+0.7)	4.4109	0.414	1966.2	✗ 1	22/ 30	sp P02666 CASE_BOVIN	K.FQSEEQQQTEDELQDK.I
328	./sergei_digest_A_full_01.1009.1011.3	1440.7 (+0.8)	4.4102	0.302	2269.3	✗ 1	28/ 44	sp P02769 ALBU_BOVIN	R.RHPEYAVSVLLR.L

Empirical statistical model to estimate the accuracy of peptide identifications

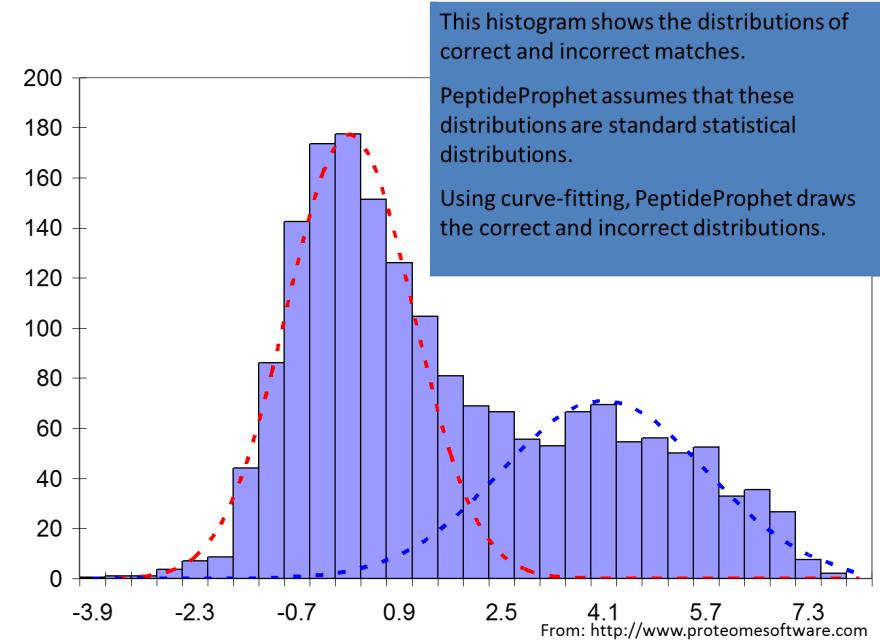
Task:

Derive a list of identified peptides from database search results carried out with a large number of MS/MS spectra.

- This entails distinguishing correct peptide assignments from false identifications.
- For high-throughput analysis and consistent data analysis a statistical model is needed to assess the validity of peptide identifications made by MS/MS database searches.

- PeptideProphet computes for each peptide assignment to a spectrum a probability of being correct.

- A discriminant function analysis is used to combine together any number of database search scores into a single discriminant score that best separates training data into correct and incorrect identifications. The discriminant score F is a weighted combination of the database search scores.



- Bayes' theorem gives the probability that a particular peptide assignment with a specific discriminant score is correct:

$$p(+|F) = p(F|+)p(+) / (p(F|+)p(+) + p(F|-)p(-))$$

where $p(+|F)$ = probability that the peptide assignment with discriminant score F is correct

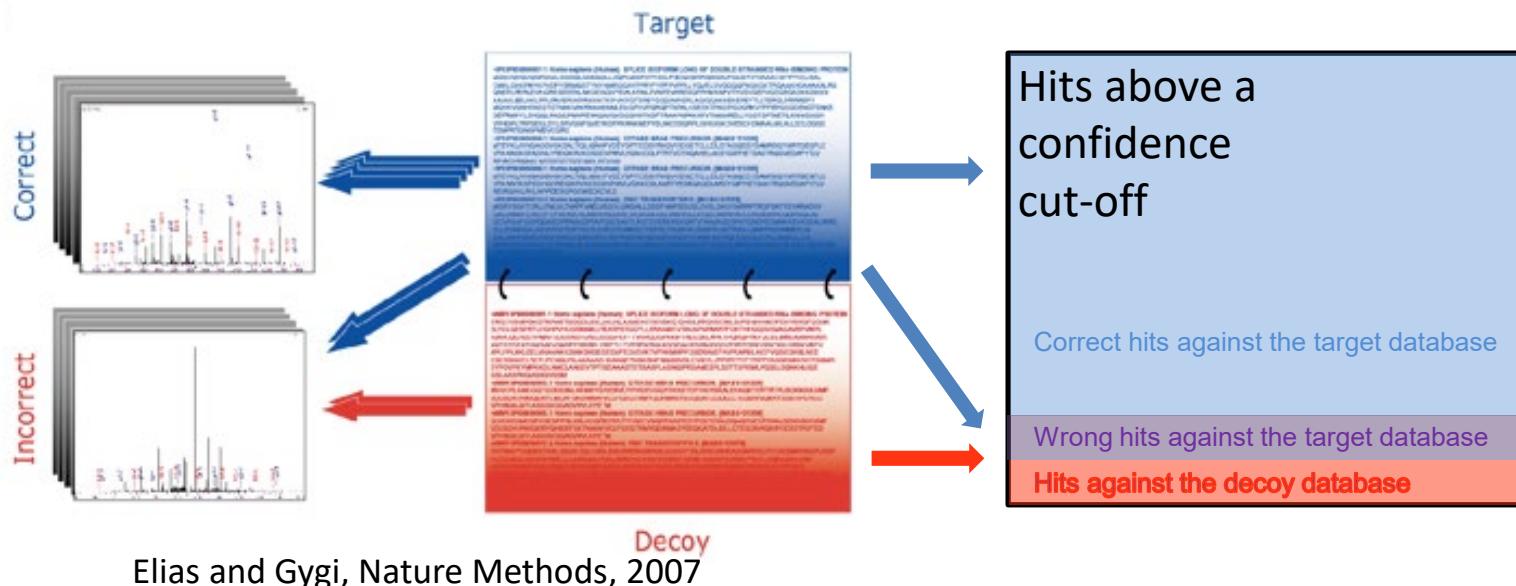
Keller et al., 2002

False positives are a concern and can occur because:

- There is usually a mess of peaks observed in a fragment spectrum, which is a reflection of the population of fragment ions produced in the collision cell of a mass spectrometer (not only b and y ions are produced)
- Spectra can be single peptide ions, chemical noise, non-peptide molecules for mixtures of co-eluting isobaric peptides
- Peptides are often present at a wide range of concentrations in a sample, and peptides present at the limit of detection can produce poor quality spectra
- Chemistry of peptide fragmentation is not completely understood
- There are amino acid sequences that do not produce a unique fragmentation pattern but share enough of the same fragment ions to be indistinguishable from one another

Target-decoy search strategy

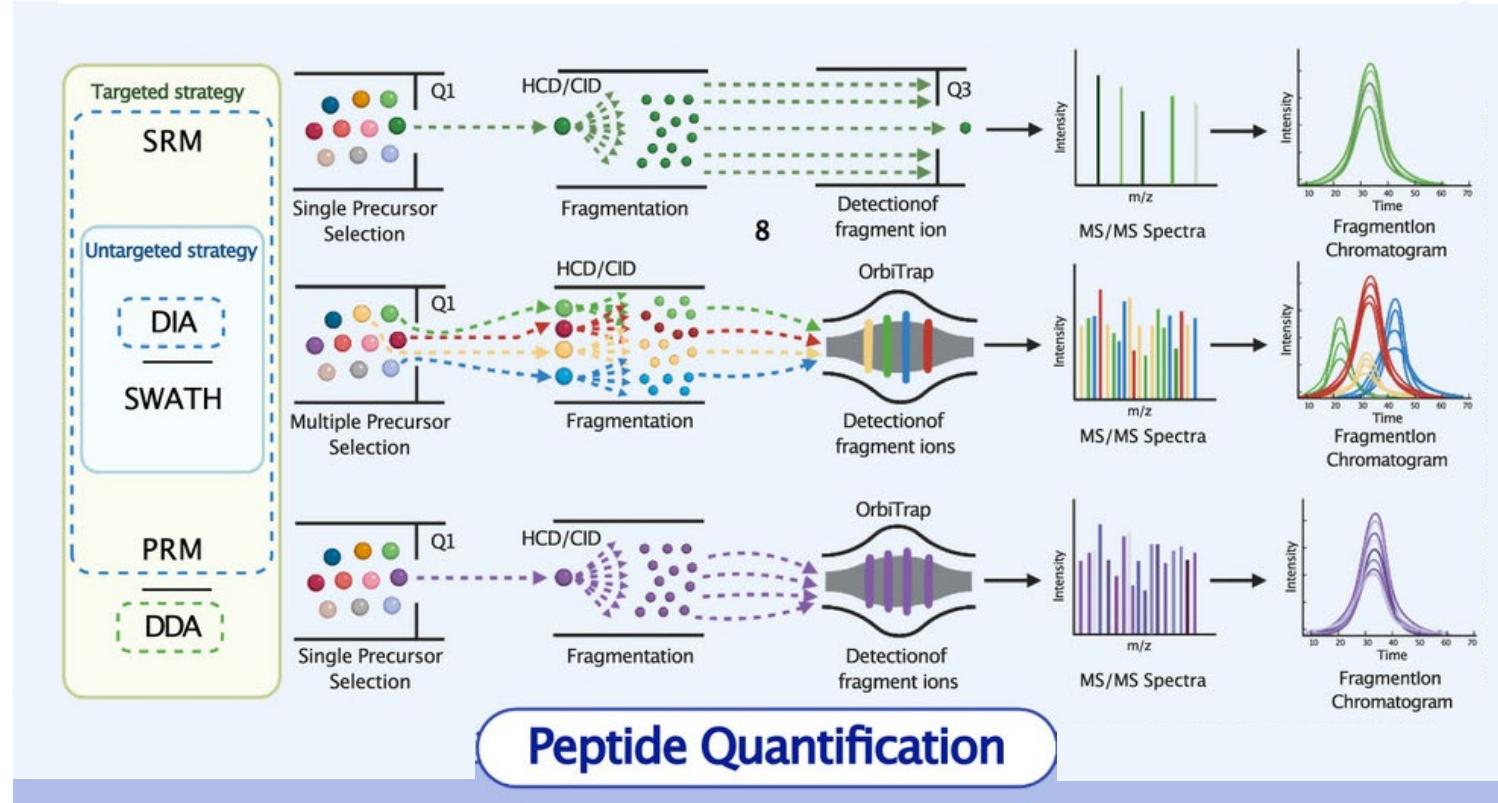
1. All the spectra are searched against a database that consists of the target database concatenated to a decoy database (either randomized or reversed target database)



1. The hits against the decoy database are clearly wrong as these sequences don't exist
2. It can then be assumed that the number of noticeable wrong hits against the decoy database equals the number of non-visible wrong hits against the target database

Question: In a PeptideProphet Search applying a 95% probability cut-off, 1500 spectra were matched to peptides; of these, 30 were matched against peptides from the decoy database. How many incorrect peptide spectrum matches against the target database do you expect?

Hypothesis-driven, targeted bottom-up proteomics approaches



Radzikowska et al., *Omics technologies in allergy and asthma research: An EAACI position paper*, Allergy, 2022

S/MRM: Selected/Multiple Reaction Monitoring; the proteins are pre-selected and provide information on the characteristic peptide precursor and fragment ion signals (transitions)

DIA/SWATH: Data Independent Acquisition/Sequential Windowed Acquisition of All Theoretical Mass Spectra

PRM: Parallel Reaction Monitoring; similar to S/MRM, but all resulting fragment ion signals from a precursor ion are monitored

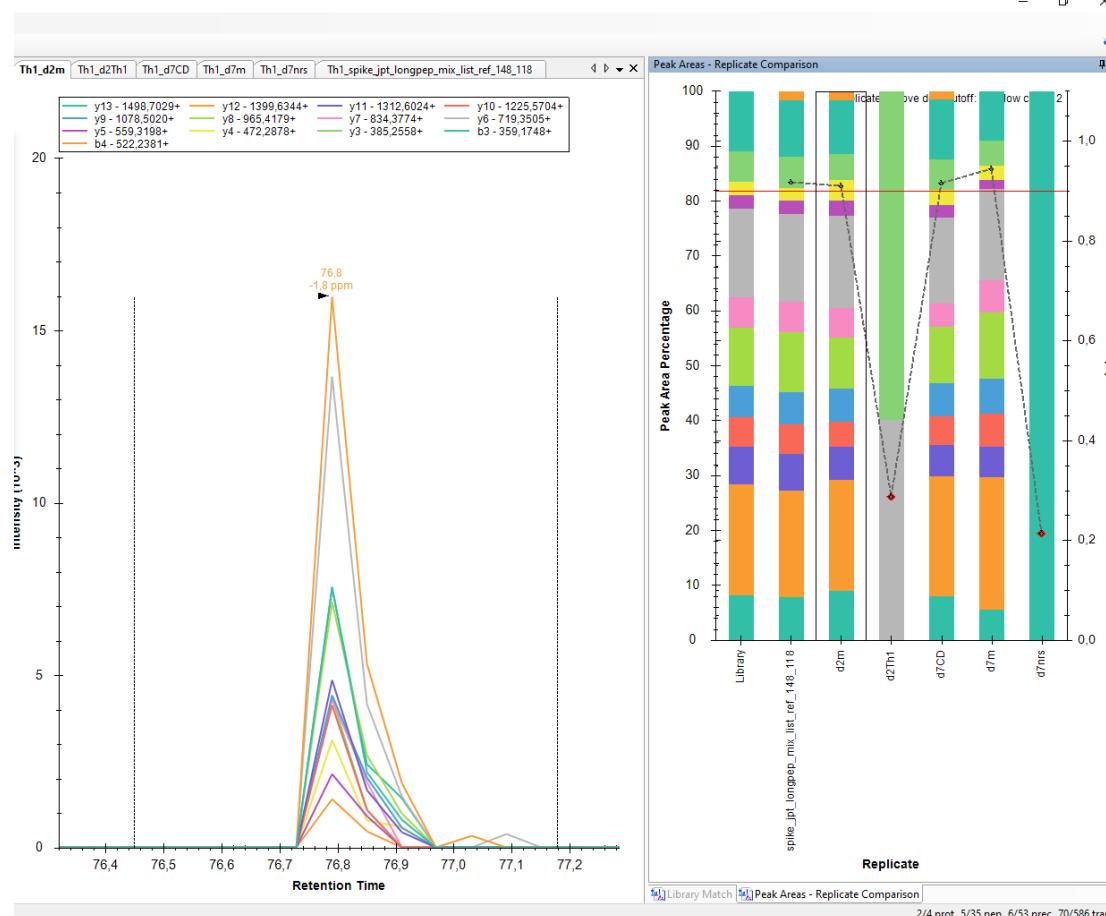
Peptide identification and quantification with Parallel Reaction Monitoring (PRM)



→ With PRM, a set of predefined proteins/peptides can be detected and quantified

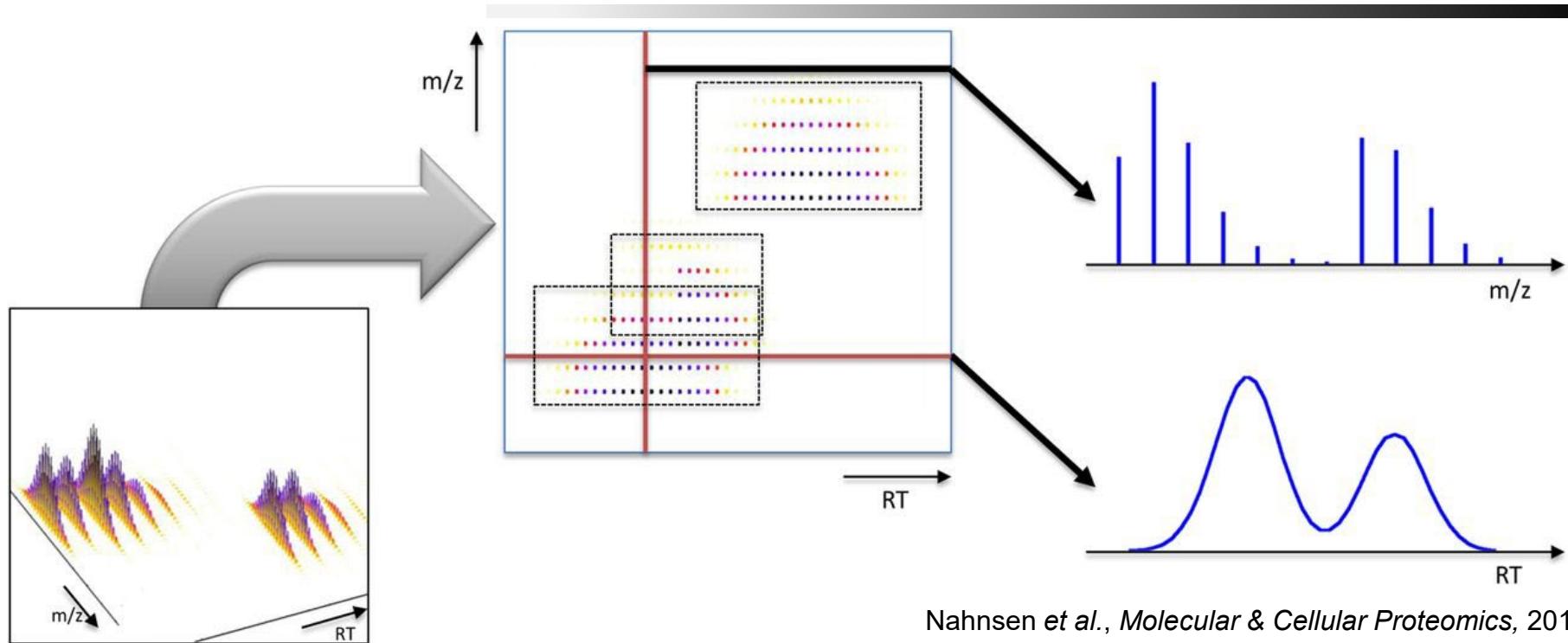
Targeted search for novel peptides encoded by lncRNAs

SQAF



Using Parallel Reaction Monitoring (PRM), we can also specifically search for peptides that are encoded in long noncoding RNAs (lncRNAs). These were identified using translatomics technologies and are not yet known to even exist.

Label-free quantification (data dependent analysis DDA)

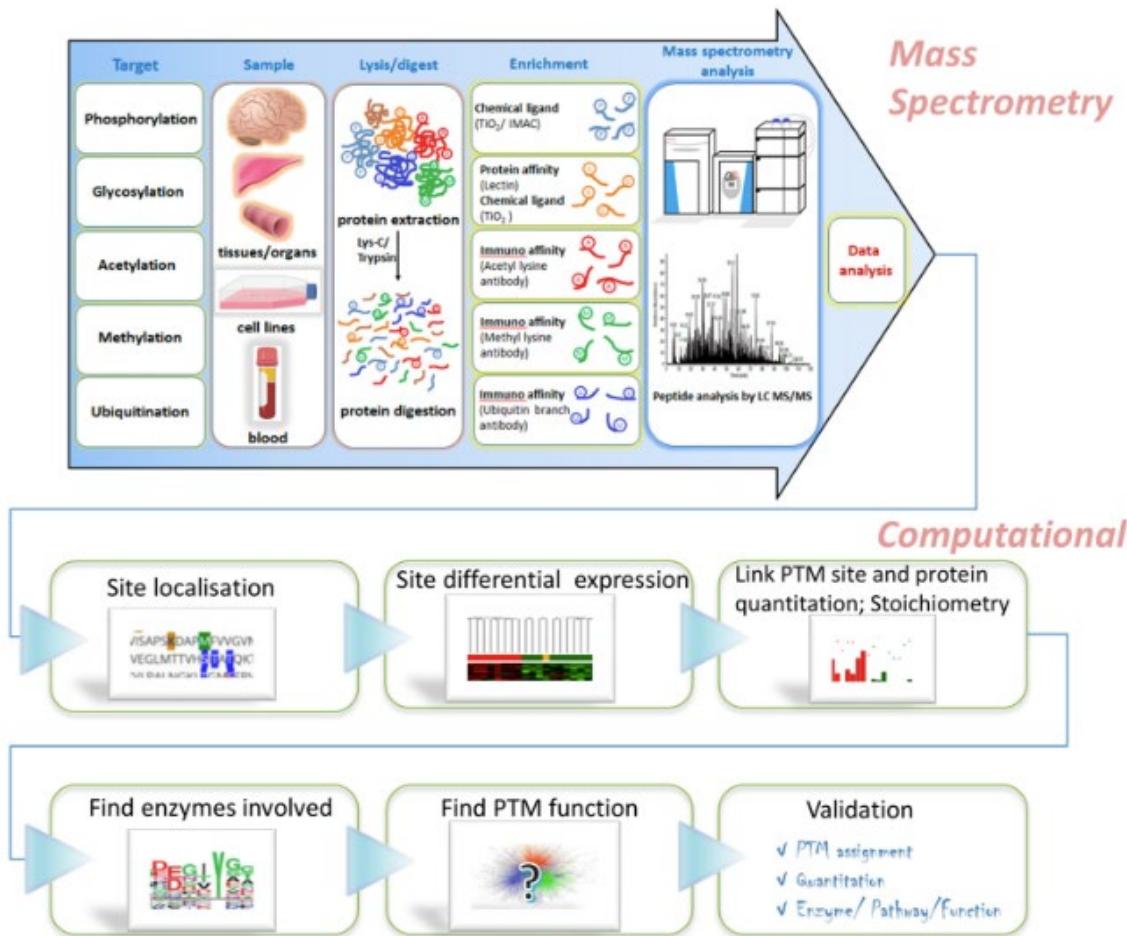


Nahnsen et al., *Molecular & Cellular Proteomics*, 2013

LC/MS data consist of individual MS spectra accumulated over (retention) time. Stacked side by side, these spectra form two-dimensional maps.

- In spectral counting the basic assumption is that protein abundance is proportional to the number of spectra (after normalization)
- Quantification can also be based on the comparison of features, which can be defined as all mass-spectrometric signals (peaks) caused by the same peptide

Workflow for post-translational modifications (PTMs)

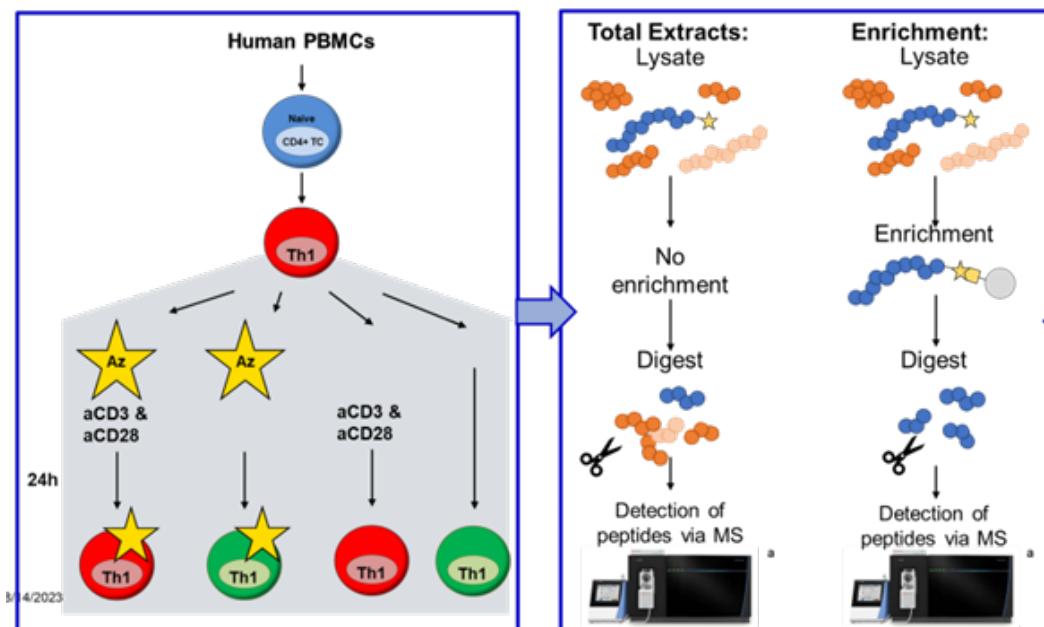


- If PTMs are analysed, the modified proteins or peptides need to be enriched first, as the modified peptides are strongly in the minority
- Searching a lot of variable modifications in your peptide spectrum matching will result in a huge search space and a lot of false positive identifications, so only search for enriched or abundant modifications
- To claim changes in protein/peptide modifications, the PTM data need to be linked to measurements of the total proteome without PTM enrichment

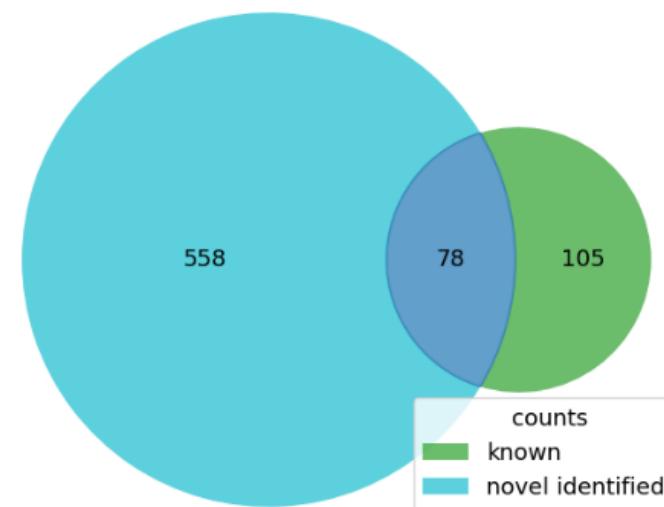
Figure 1 in Pascovici et al., Int. J. Mol. Sci. 2019,
doi.org/10.3390/ijms20010016

A novel view on protein prenylation

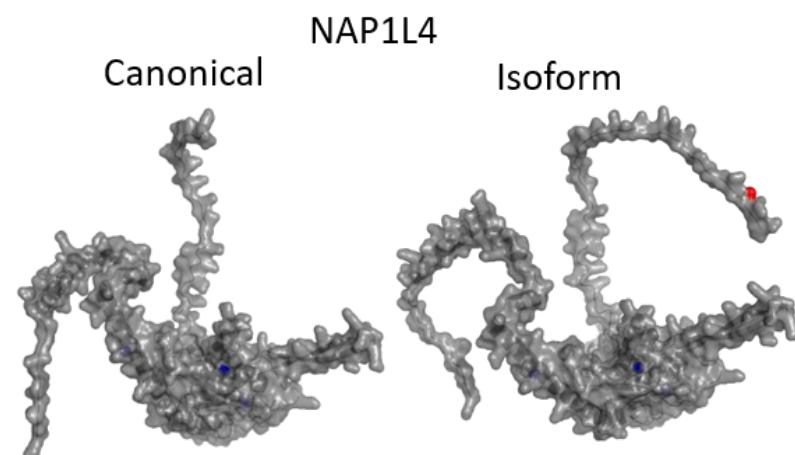
SIAF



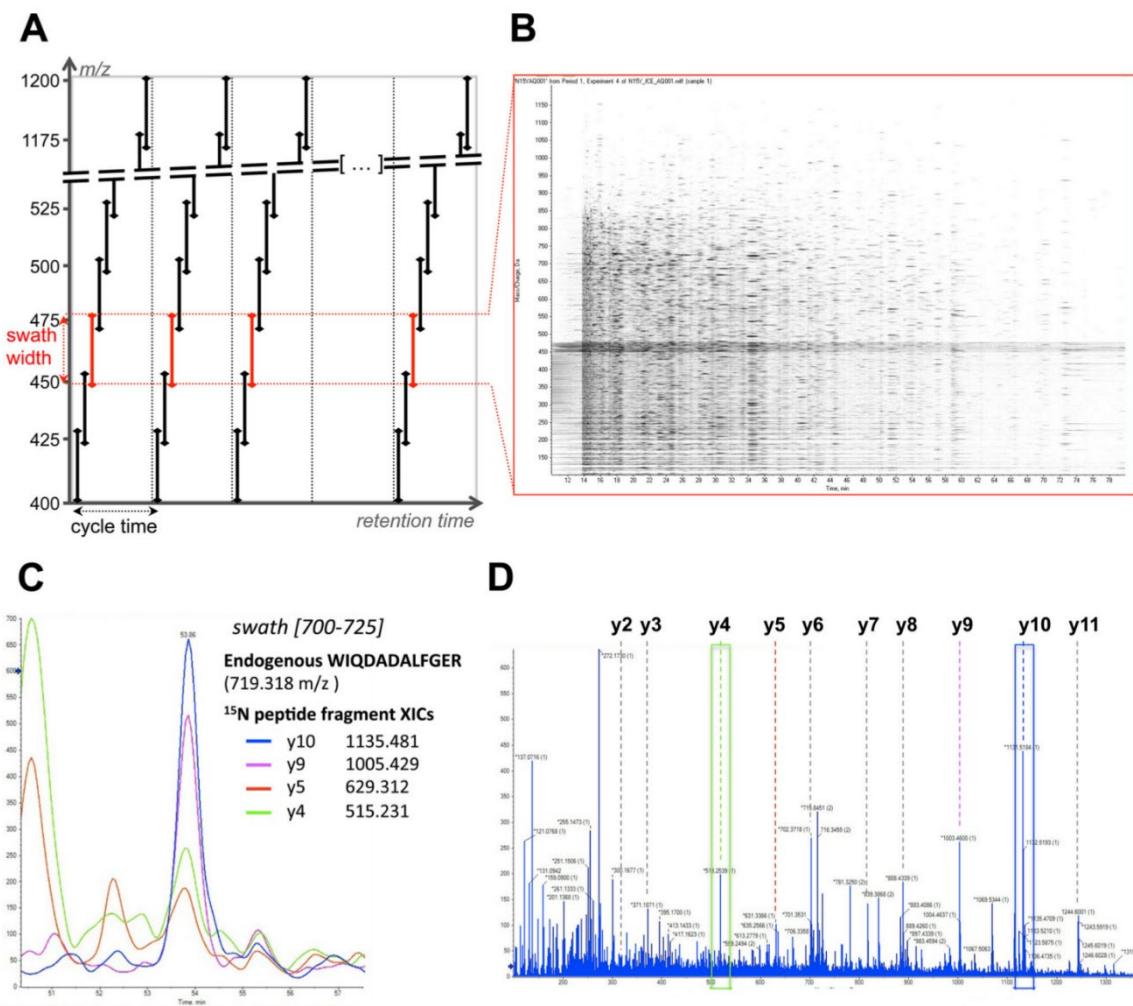
All known and identified prenylated proteins



- Prenylated proteins can be enriched by adding azide-labelled isoprenoid moieties to cells and then using click chemistry to covalently link the prenylated proteins to a matrix.
- Analysis of the identified prenylated proteins revealed novel prenylated proteins and novel sites for prenylation including internal prenylation.



SWATH-MS (data independent analysis DIA)

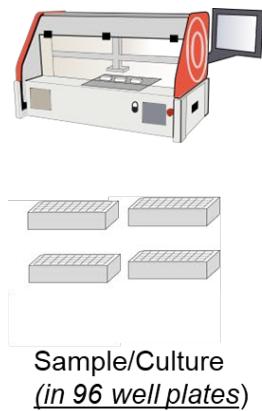


- A)** the mass spectrometer steps through a set of precursor acquisition windows
- B)** In each cycle it fragments all precursors from all the respective quadrupole isolation windows and in each isolation window it records a complete, high accuracy fragment ion spectrum of all precursors
- C)** The data are analyzed by reconstructing the lineage of precursor and fragment ions based on their chromatographic elution profile, and with software for automated targeted data analysis

@ <http://www.imsb.ethz.ch/research/aebersold/research/swath-ms.html>

How to measure thousands of proteome profiles

Christoph Messner
Center for Precision Proteomics



Throughput:

Up to 768 samples/day

> 100 samples/day
on one Instrument

> 2,000 samples/day

Sample preparation:
96-well based semi-automated
Protein extraction, digestion
and SPE clean-up

Data acquisition:
- "High" flow LC
- "scanning" SWATH
- 1-20 minute gradient

Data processing with DIA-NN:
- based on neural networks
- open source
- github.com/vdemichev/diann

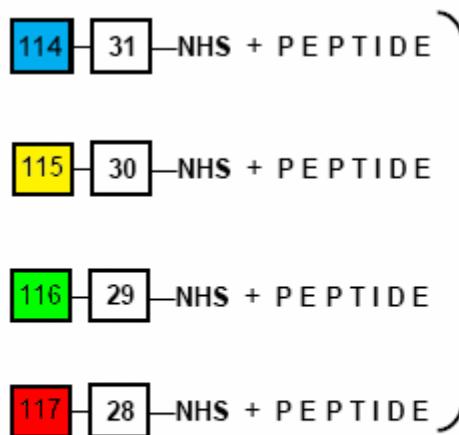
Messner et al, Nature Biotechnology, 2021

Messner et al, Cell Systems, 2020

Demichev, Messner et al, Nature Methods, 2020

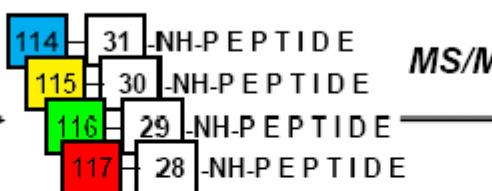
Quantification with labelling: iTRAQ

The iTRAQ reagent



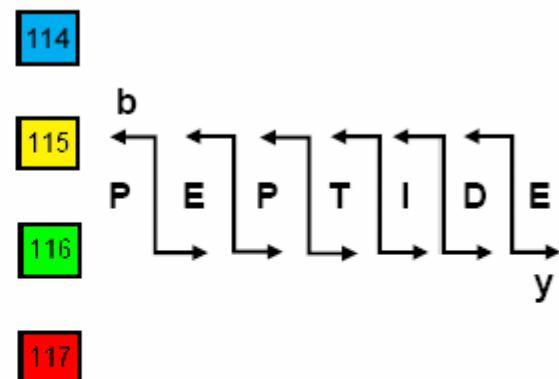
MS (first dimension)

MS



MS/MS

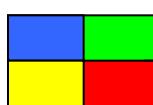
MS/MS (second dimension)



Reporter-Balance-Peptide INTACT
- 4 samples *identical m/z*

-Peptide fragments **EQUAL**
-Reporter ions **DIFFERENT**

Reporter



Linker



Intensity

Parent Ion

m/z

Intensity

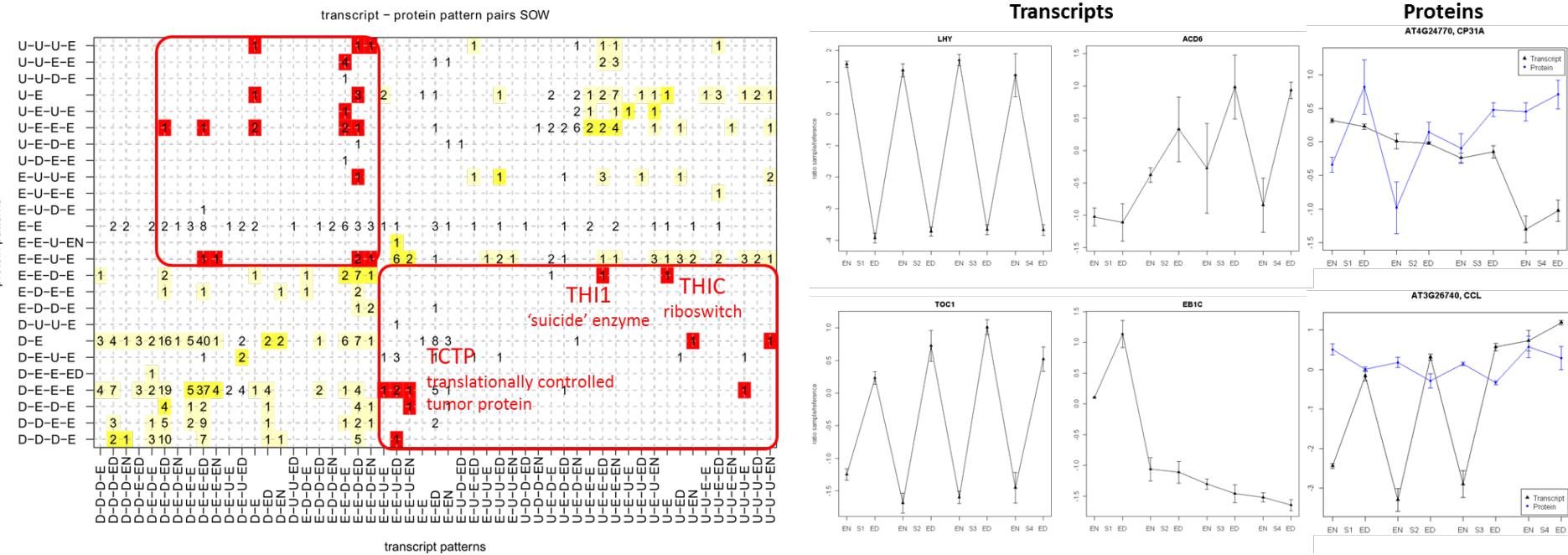
fragments (b- and y-ions)

Reporter
(114-117)

m/z

Protein quantification with iTRAQ und correlation of protein and transcript data

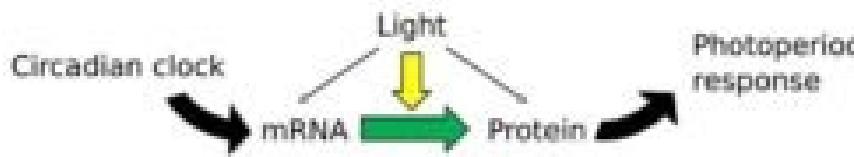
In a project on plant leaf growth, quantitative protein and transcript data were acquired from Arabidopsis leaves at four growth stages, each at the end of the day and end of the night. Correlation of the transcript and protein data revealed a strong correlation across the growth stages, but essentially no correlation for the daily fluctuations of the transcript data.



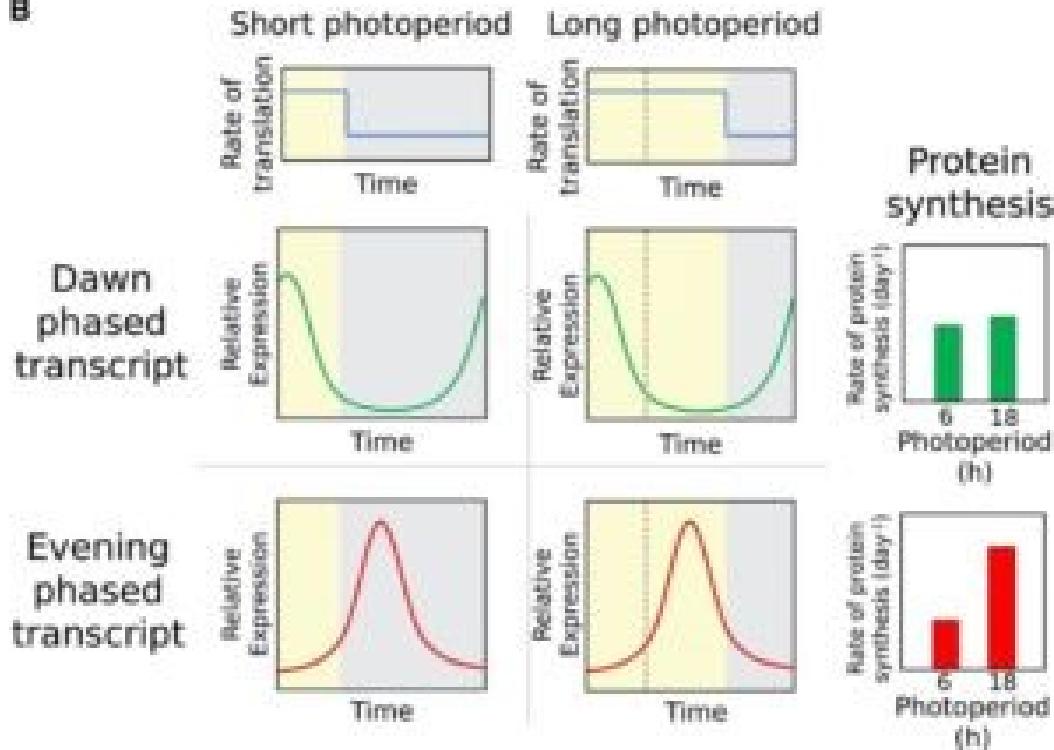
Baerenfaller et al., 2012

Translational coincidence mechanism

A



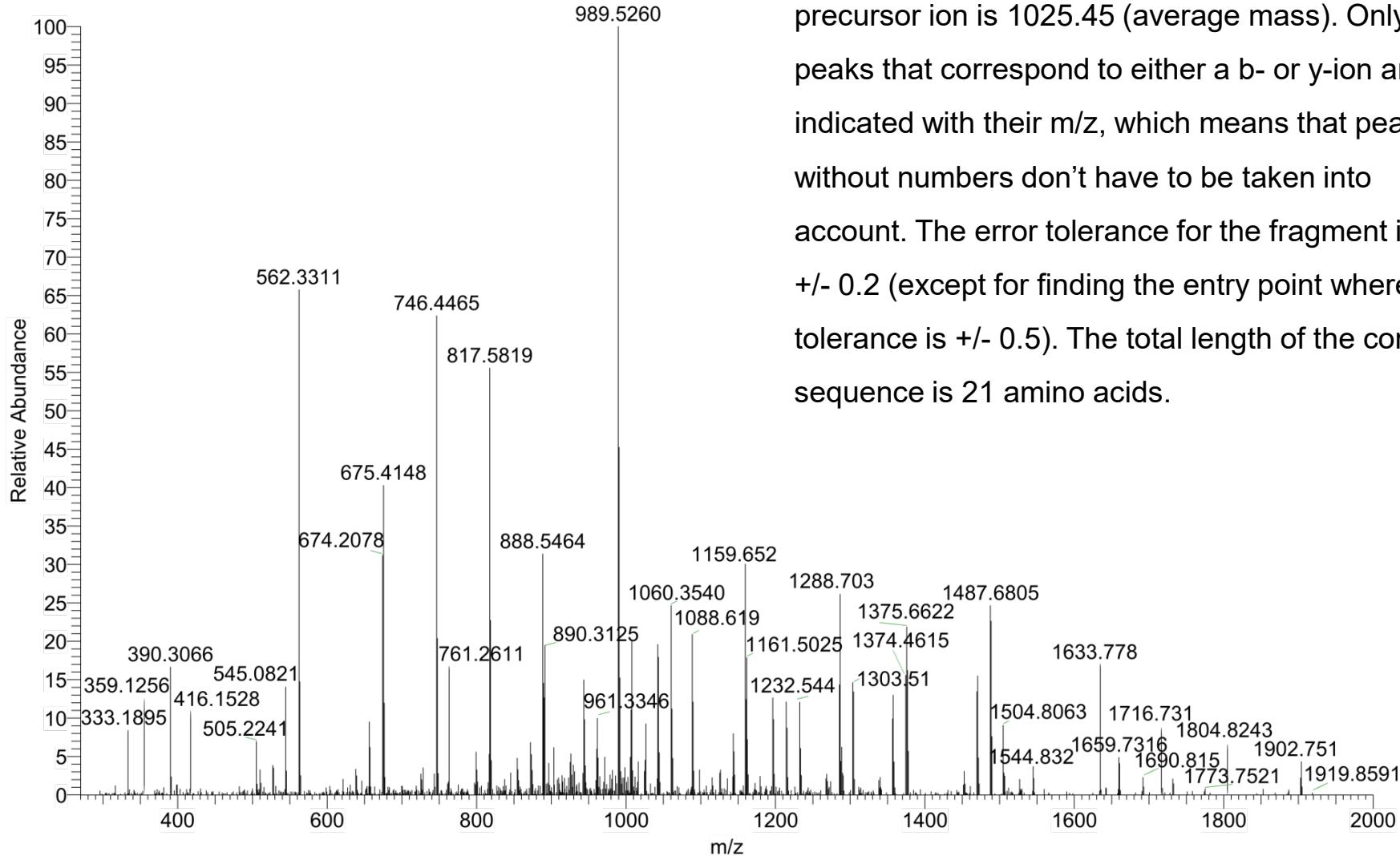
B



- Quantitative protein and transcript data of *Arabidopsis* in four different photoperiods have revealed that proteins with clock-regulated, evening-peaking RNAs tended to increase in abundance under longer daylengths, whereas proteins with morning-peaking RNAs did not.
- A “translational coincidence” model predicted the experimental results, because high, light-induced translation rates will coincide with high levels of an evening-expressed RNA only under long days, not short days.
- Many clock-controlled genes might gain seasonal control of protein levels via translational coincidence.

De novo peptide sequencing – Exercise

20110819_35_UBA_urea_5 #4691 RT: 43.76 AV: 1 NL: 5.97E2
T: ITMS + c NSI d Full ms2 1025.45@cid28.00 [270.00-2000.00]



Derive an amino acid sequence from the idealized spectrum. The measured m/z of the doubly charged precursor ion is 1025.45 (average mass). Only peaks that correspond to either a b- or y-ion are indicated with their m/z , which means that peaks without numbers don't have to be taken into account. The error tolerance for the fragment ions is ± 0.2 (except for finding the entry point where the tolerance is ± 0.5). The total length of the correct sequence is 21 amino acids.