Proteomics Informatics – Protein characterization I: post-translational modifications (Week 10)

Post-translational modification

- Biologically important post-translational modification (phosphorylation, acetylation, glycosylation, etc.)

- Introduced on purpose during sample preparation (alkylation, iTRAQ, TMT etc.)

- Side-products of sample preparation (oxidation, deamidation, carbamylation, formylation etc.)

Post-translational modification

Table 1. Some common and important post-translational modifications

PTM type	ΔMass[a] (Da)	Stability[b]	Function and notes
Phosphorylation			Reversible, activation/inactivation of enzyme activity, modulation of molecular interactions, signaling
pTyr	+80	+++	
pSer, pThr	+80	+/++	
Acetylation	+42	+++	Protein stability, protection of N terminus. Regulation of protein–DNA interactions (histones)
Methylation	+14	+++	Regulation of gene expression
Acylation, fatty acid modification			Cellular localization and targeting signals, membrane tethering, mediator of protein–protein interactions
Farnesyl	+204	+++	
Myristoyl	+210	+++	
Palmitoyl	+238	+/++	
etc.			
Glycosylation			
N-linked	>800	+/++	Excreted proteins, cell–cell recognition/signaling
O-linked	203, >800	+/++	*O*-GlcNAc, reversible, regulatory functions
GPI anchor	>1,000	++	Glycosylphosphatidylinositol (GPI) anchor. Membrane tethering of enzymes and receptors, mainly to outer leaflet of plasma membrane
Hydroxyproline	+16	+++	Protein stability and protein–ligand interactions
Sulfation (sTyr)	+80	+	Modulator of protein–protein and receptor–ligand interactions
Disulfide bond formation	–2	++	Intra- and intermolecular crosslink, protein stability
Deamidation	+1	+++	Possible regulator of protein–ligand and protein–protein interactions, also a common chemical artifact
Pyroglutamic acid	–17	+++	Protein stability, blocked N terminus
Ubiquitination	>1,000	+/++	Destruction signal. After tryptic digestion, ubiquitination site is modified with the Gly-Gly dipeptide
Nitration of tyrosine	+45	+/++	Oxidative damage during inflammation

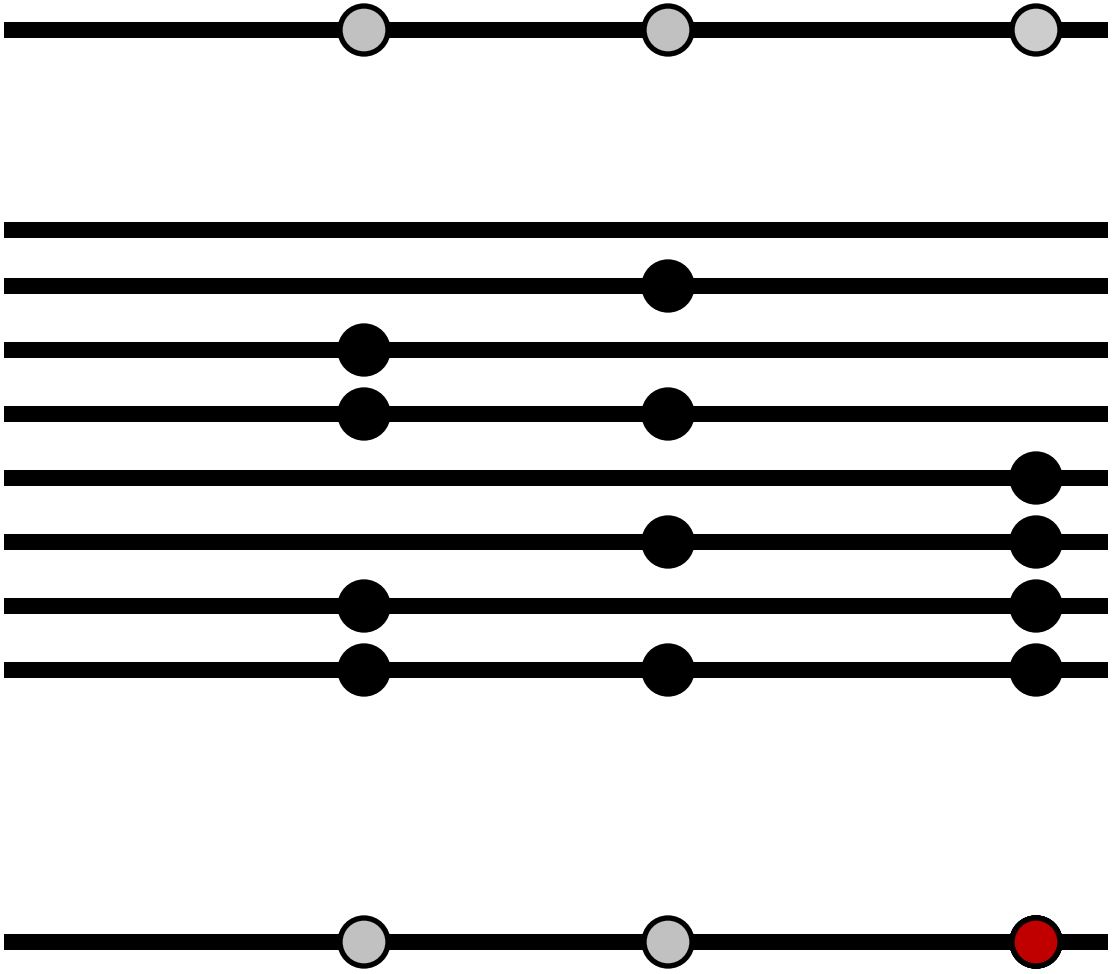
[a]A more comprehensive list of PTM Δmass values can be found at: http://www.abrf.org/index.cfm/dm.home
[b]Stability: + labile in tandem mass spectrometry, ++ moderately stable; +++ stable.

Mann and Jensen, Nature Biotech. 21, 255 (2003)

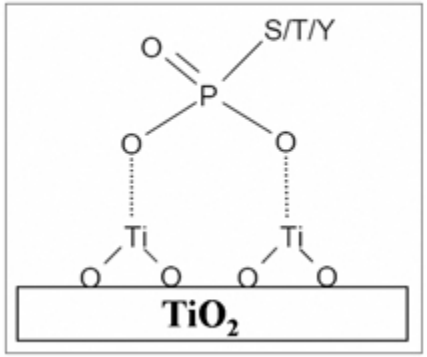
Phosphorylation examples

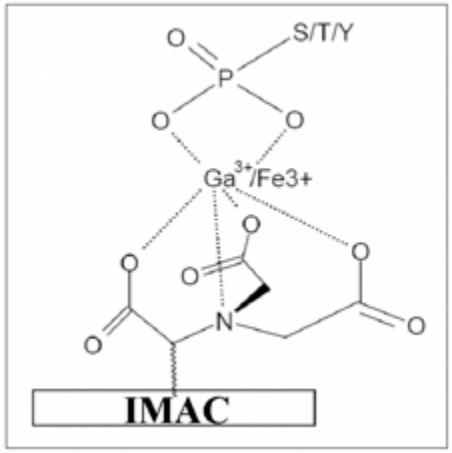
Unmodified				pS18				pT5			
b			y	b			y	b			y"
---	1	F	---	---	1	F	---	---	1	F	---
261.1556	2	I	2163.024	261.1556	2	I	2243.024	261.1556	2	I	2243.024
421.1862	3	C	2049.94	421.1862	3	C	2129.94	421.1862	3	C	2129.94
520.2546	4	V	1889.909	520.2546	4	V	1969.909	520.2546	4	V	1969.909
621.3022	5	T	1790.841	621.3022	5	T	1870.841	701.3022	5	T	1870.841
718.3549	6	P	1689.793	718.3549	6	P	1769.793	798.3549	6	P	1689.793
819.4025	7	T	1592.741	819.4025	7	T	1672.741	899.4025	7	T	1592.741
920.4502	8	T	1491.693	920.4502	8	T	1571.693	1000.45	8	T	1491.693
1080.481	9	C	1390.645	1080.481	9	C	1470.645	1160.481	9	C	1390.645
1167.513	10	S	1230.615	1167.513	10	S	1310.615	1247.513	10	S	1230.615
1281.556	11	N	1143.583	1281.556	11	N	1223.583	1361.556	11	N	1143.583
1382.603	12	T	1029.54	1382.603	12	T	1109.54	1462.603	12	T	1029.54
1495.687	13	I	928.4923	1495.687	13	I	1008.492	1575.687	13	I	928.4923
1610.714	14	D	815.4083	1610.714	14	D	895.4083	1690.714	14	D	815.4083
1723.798	15	L	700.3814	1723.798	15	L	780.3814	1803.798	15	L	700.3814
1820.851	16	P	587.2974	1820.851	16	P	667.2974	1900.851	16	P	587.2974
1951.891	17	M	490.2447	1951.891	17	M	570.2446	2031.891	17	M	490.2447
2038.923	18	S	359.2042	2118.923	18	S	439.2042	2118.923	18	S	359.2042
2135.976	19	P	272.1722	2215.976	19	P	272.1722	2215.976	19	P	272.1722
---	20	R	175.1195	---	20	R	175.1195	---	20	R	175.1195

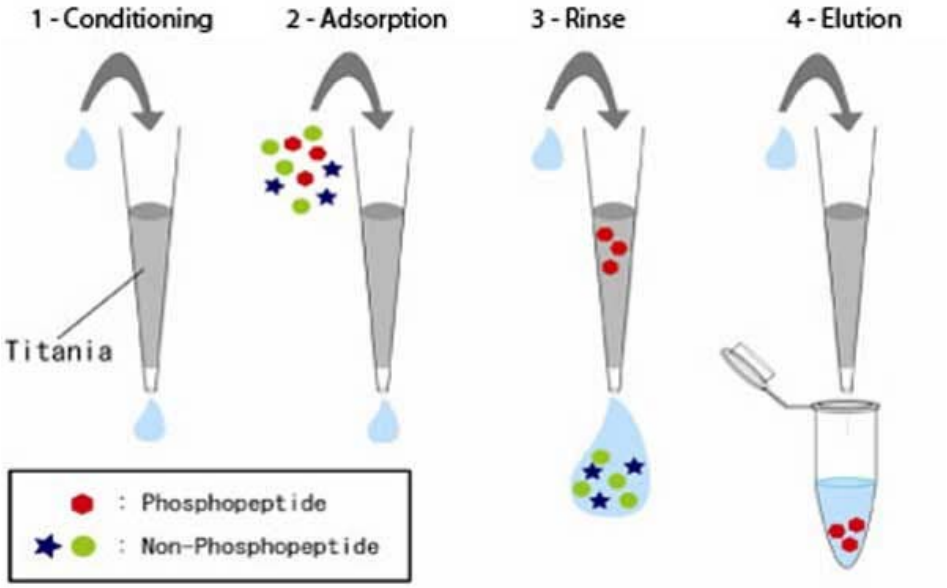
Potential modifications



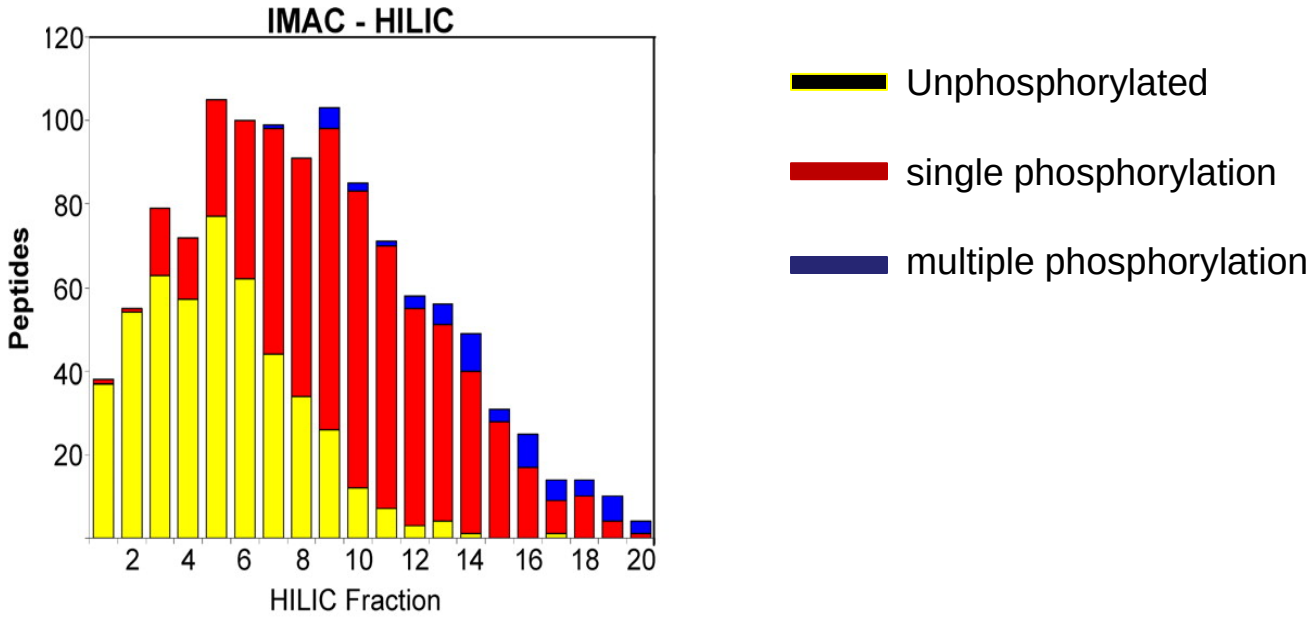
Enrichment Strategies for the Detection of Phosphorylated Peptides





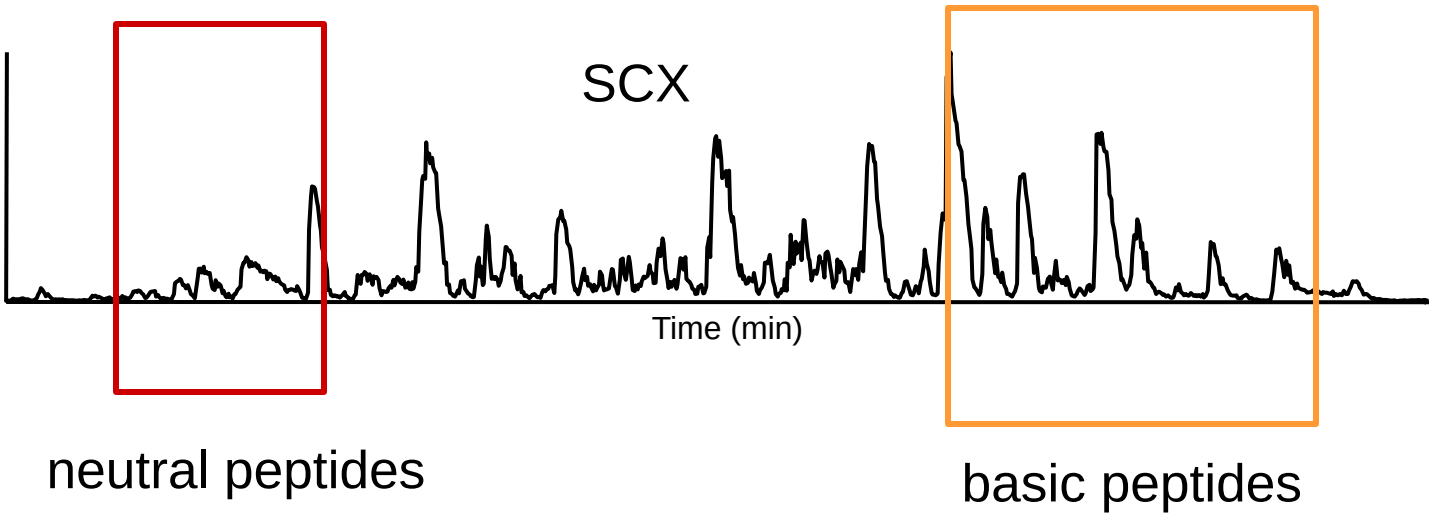


Enrichment Strategies for the Detection of Phosphorylated Peptides



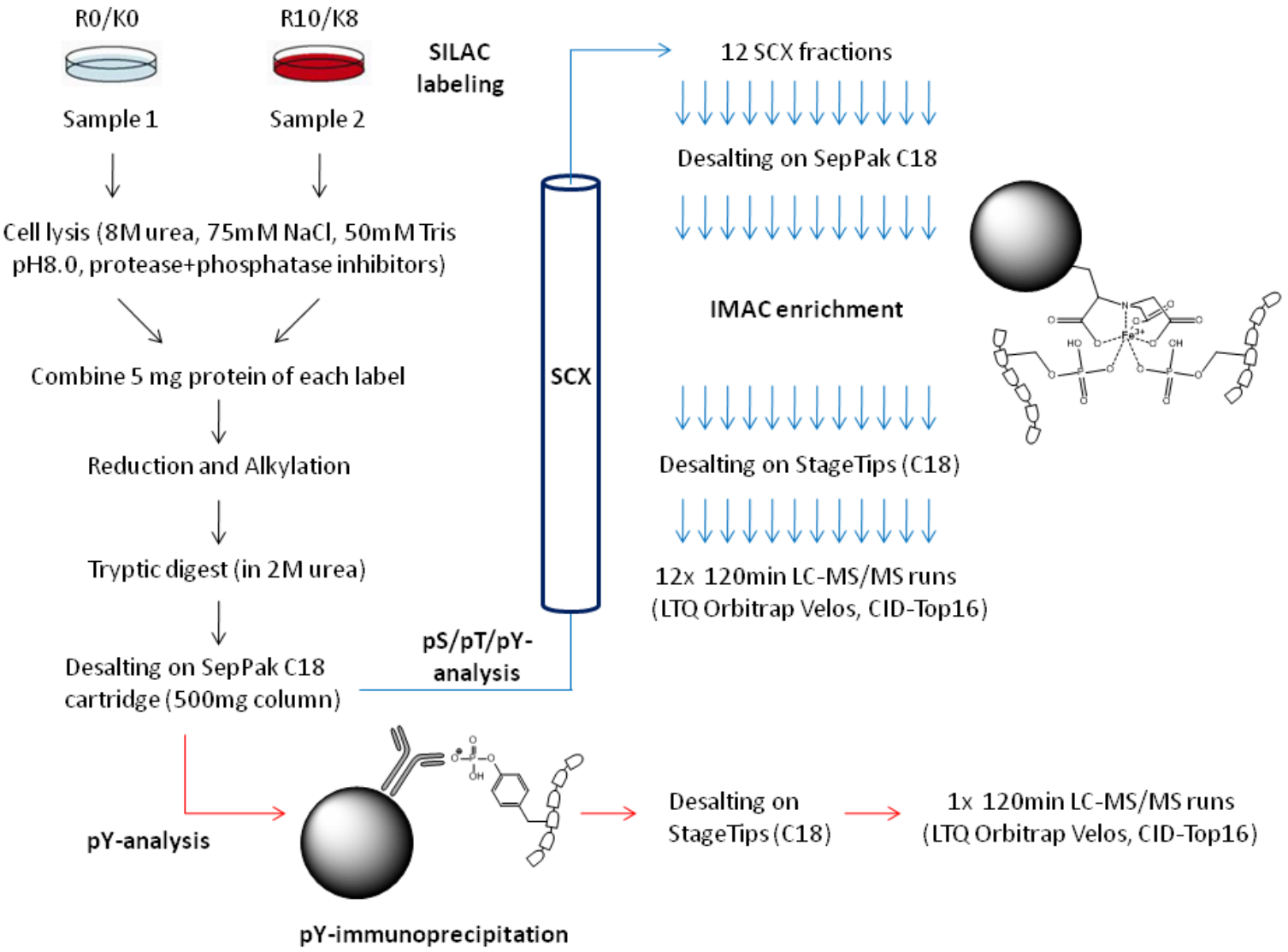
- Hydrophilic Interaction Chromatography (HILIC)
- Phosphopeptides elute later than their unphosphorylated counterparts
- Stationary phase is hydrophilic
- Mobile phase is hydrophobic

Enrichment Strategies for the Detection of Phosphorylated Peptides

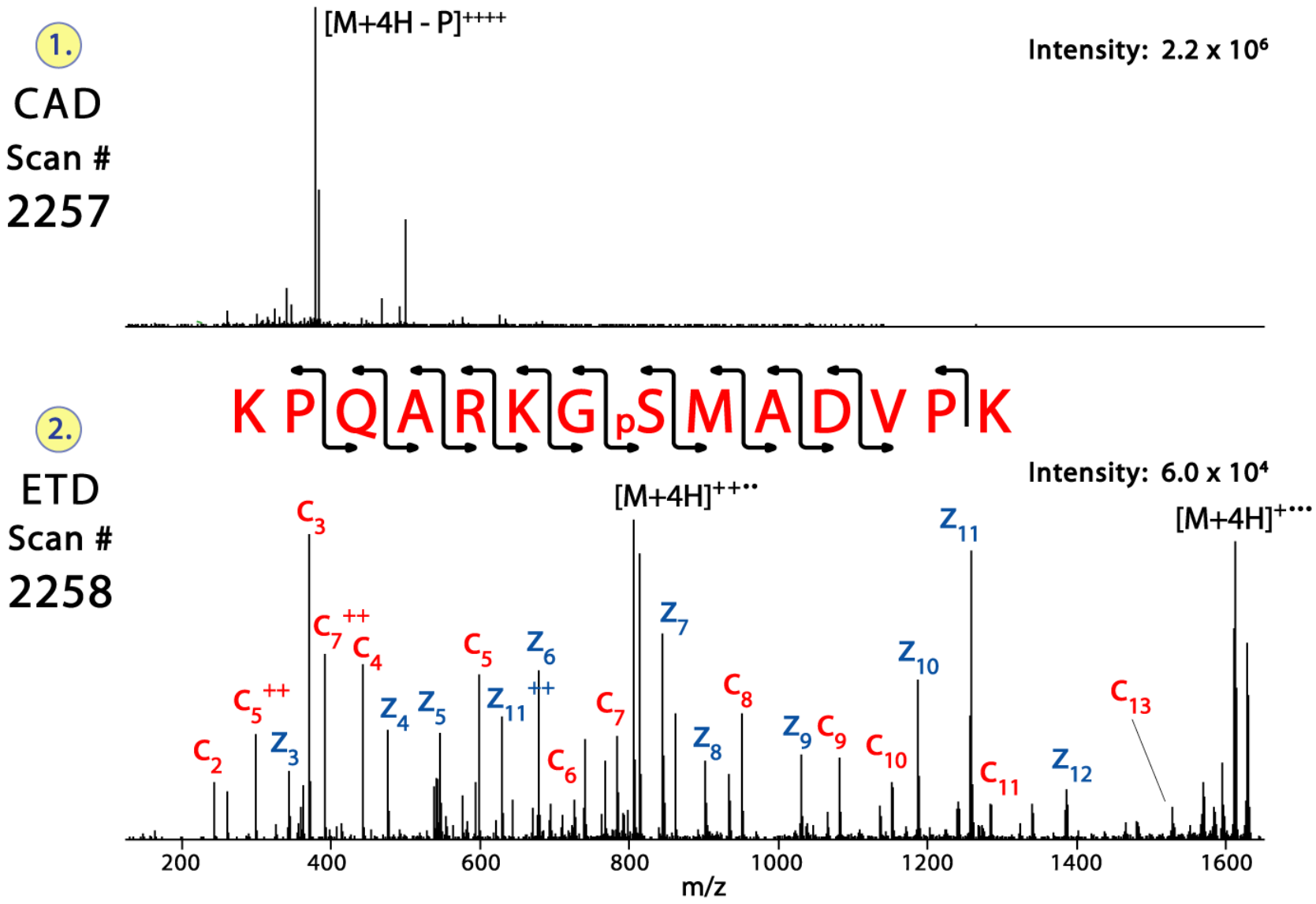


- Strong Cation Exchange Chromatography
- Stationary phase is negatively charged
- Mobile phase is a buffer that is increasing the pH (if peptide becomes neutral it elutes)
- Neutral peptides elute earlier: XXpSxxxxxR/K
- Positive peptides elute late: XXXXHXXXXR/K

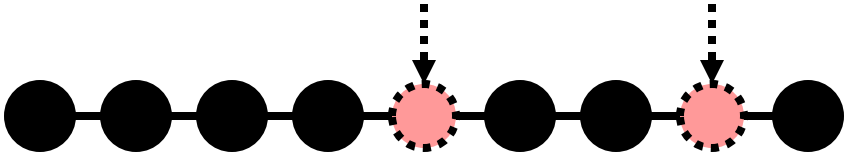
Several Strategies are often combined

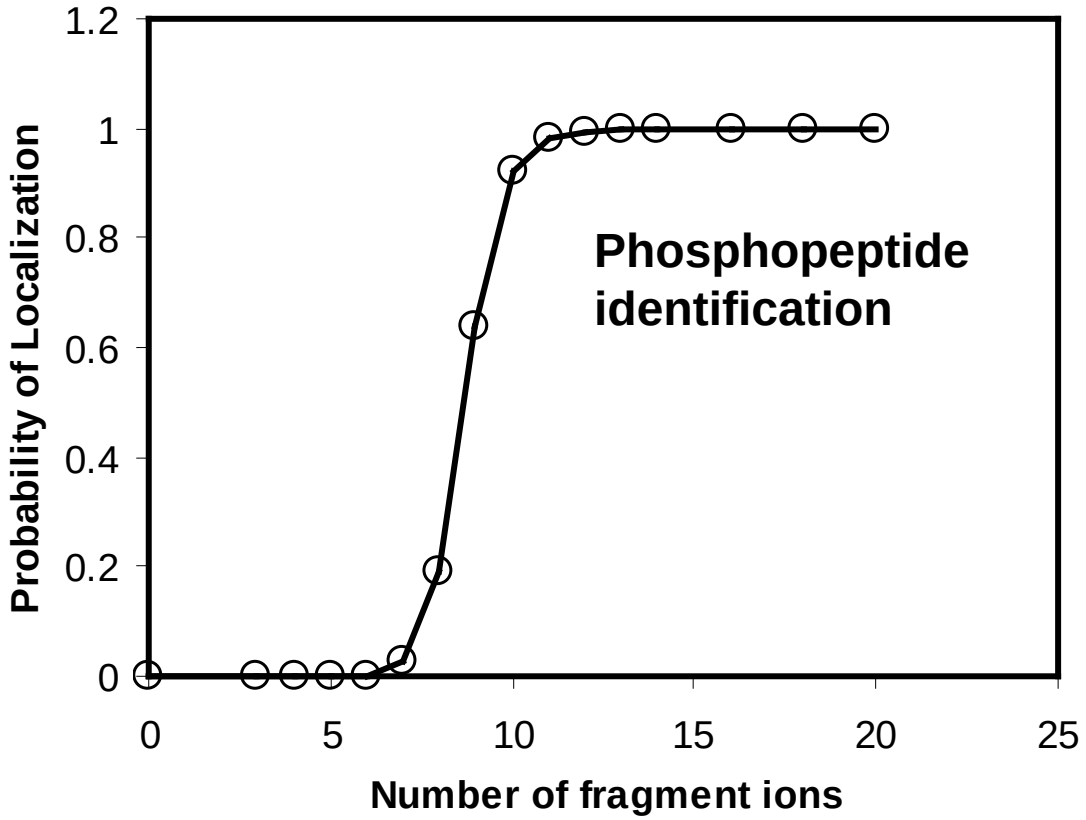


Loss of the phosphate group



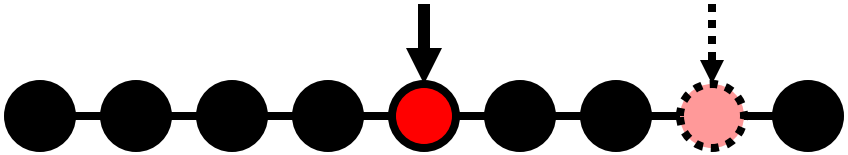
Localization of modifications

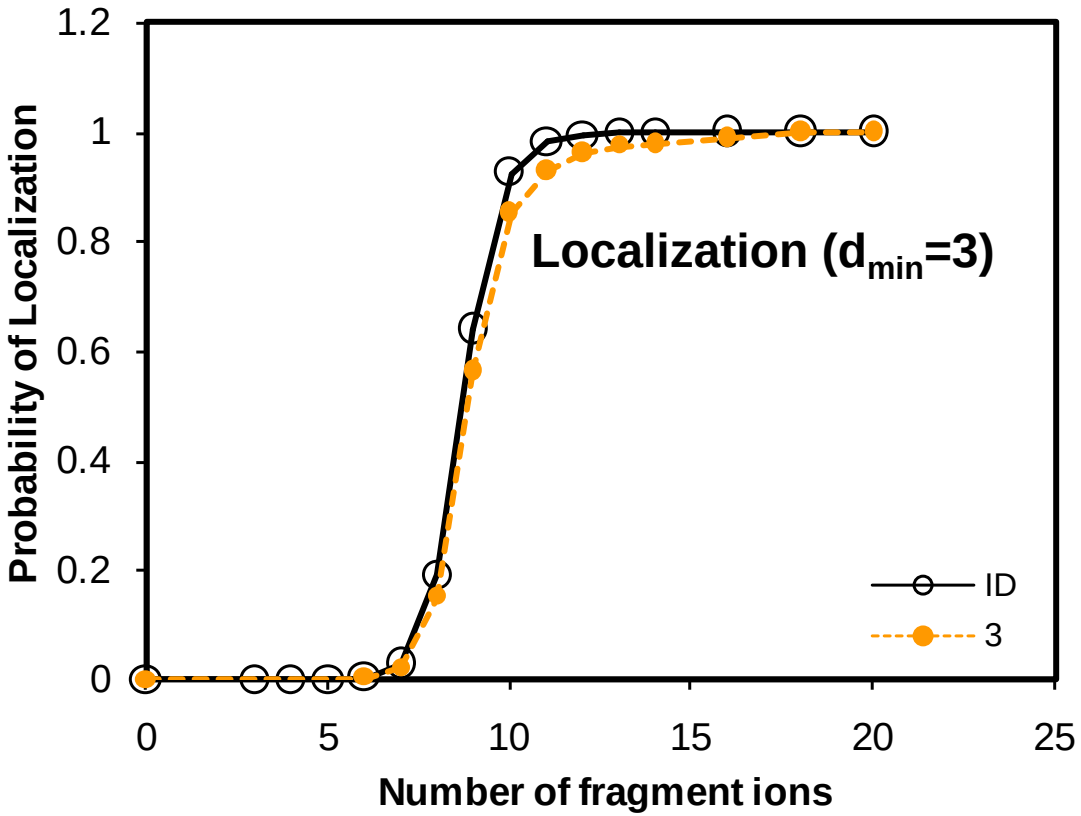




$m_{precursor}$ = 2000 Da
$\Delta m_{precursor}$ = 1 Da
$\Delta m_{fragment}$ = 0.5 Da
Phosphorylation

Localization of modifications

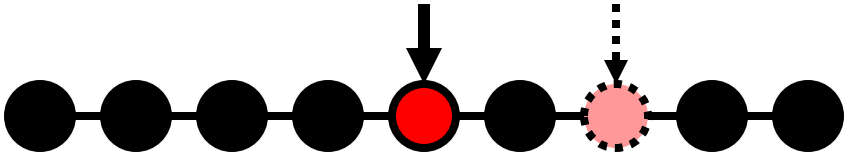


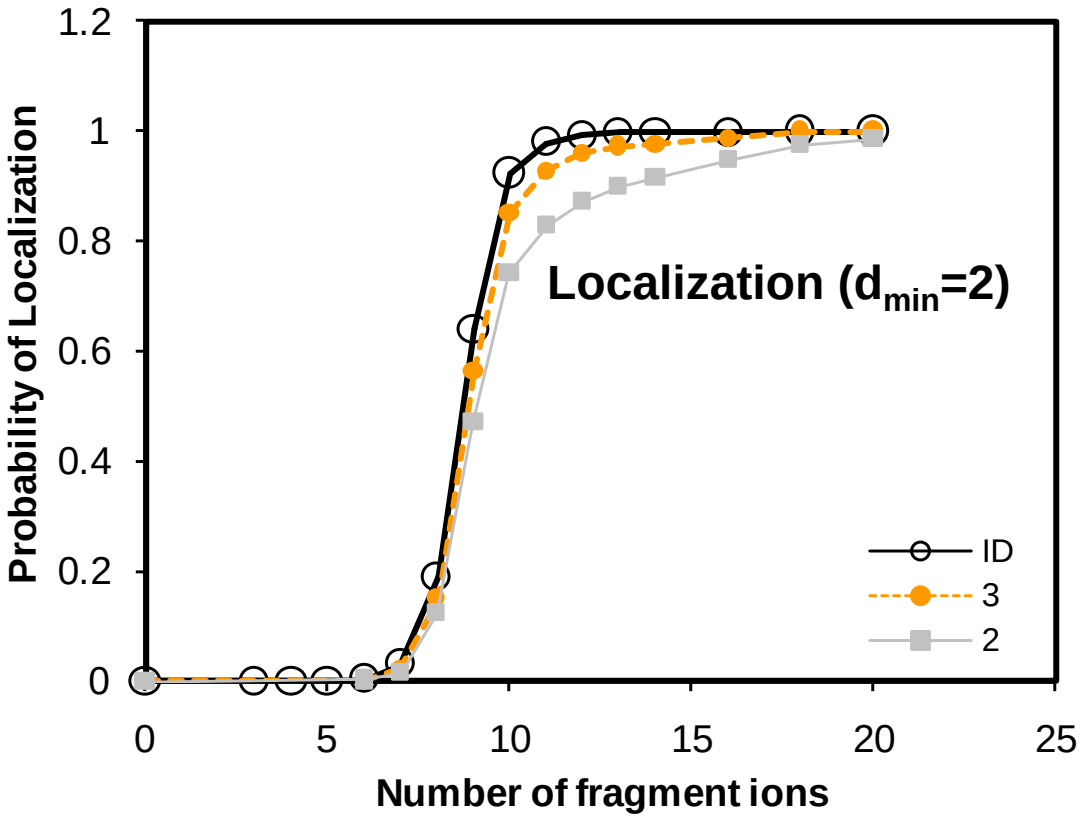


d_{min}>=3 for 47% of human tryptic peptides

$m_{precursor}$ = 2000 Da
$\Delta m_{precursor}$ = 1 Da
$\Delta m_{fragment}$ = 0.5 Da
Phosphorylation

Localization of modifications

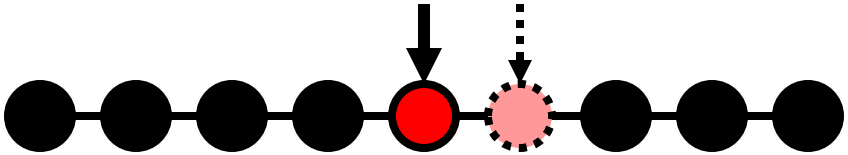


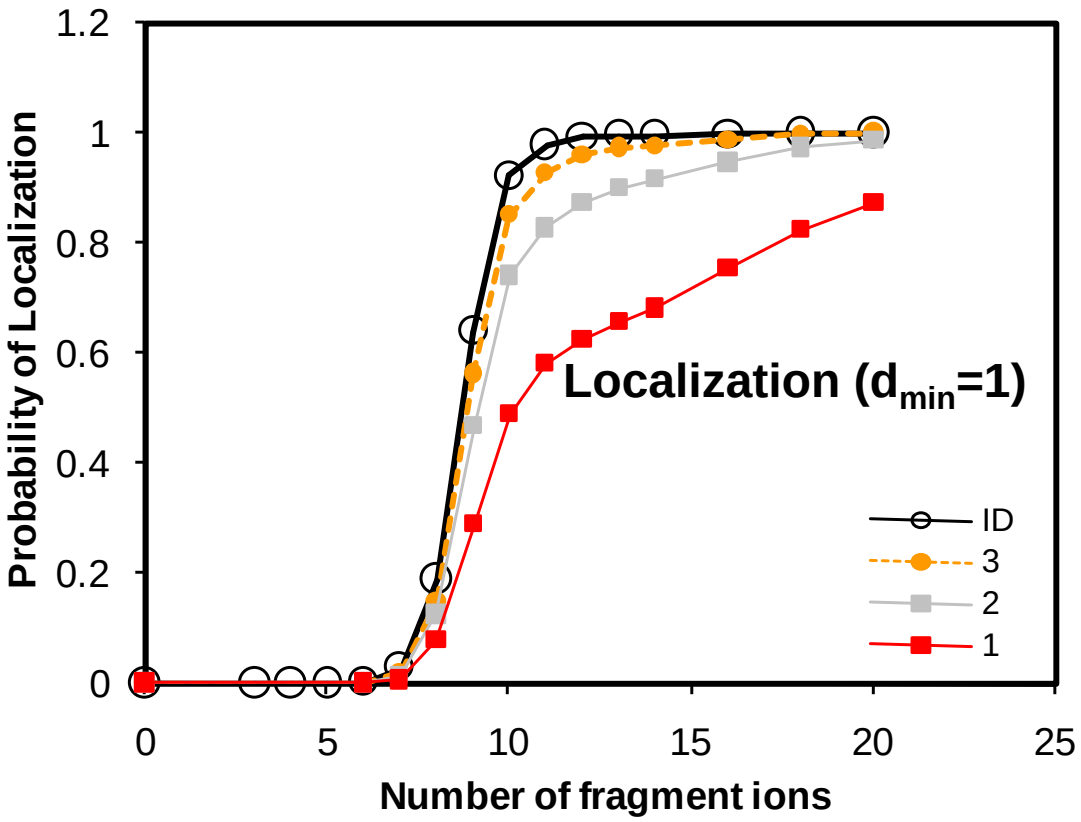


d_{min}=2 for 33% of human tryptic peptides

$m_{precursor}$ = 2000 Da
$\Delta m_{precursor}$ = 1 Da
$\Delta m_{fragment}$ = 0.5 Da
Phosphorylation

Localization of modifications

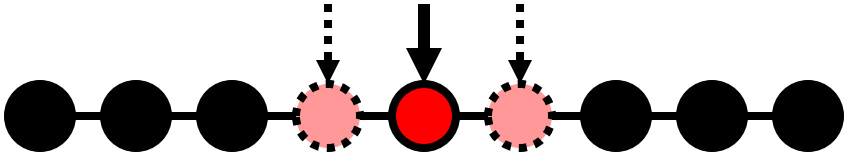


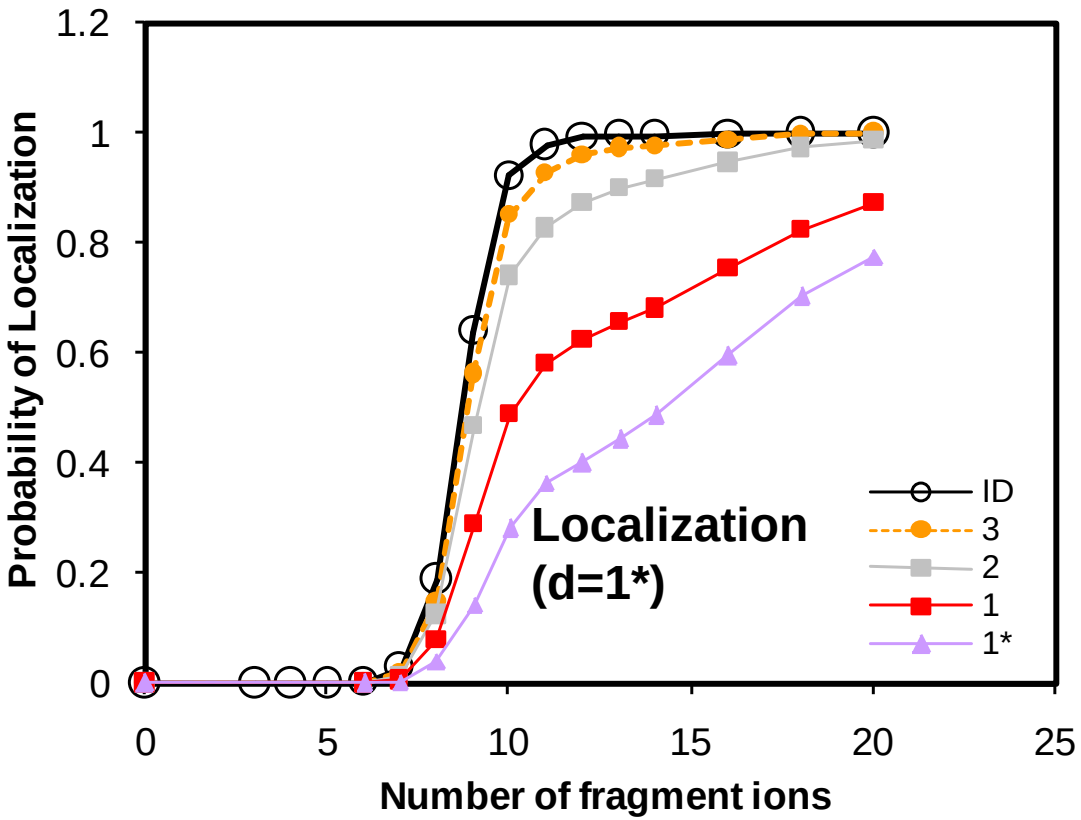


d_{min}=1 for 20% of human tryptic peptides

$m_{precursor}$ = 2000 Da
$\Delta m_{precursor}$ = 1 Da
$\Delta m_{fragment}$ = 0.5 Da
Phosphorylation

Localization of modifications

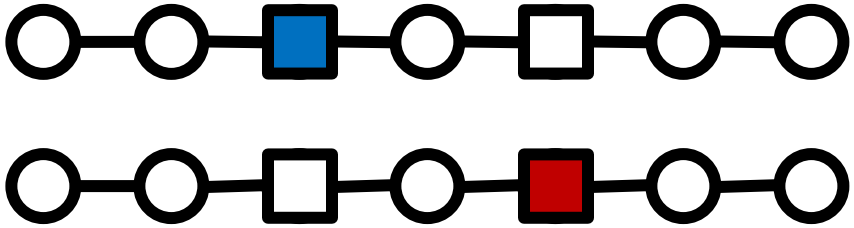




$m_{precursor}$ = 2000 Da
$\Delta m_{precursor}$ = 1 Da
$\Delta m_{fragment}$ = 0.5 Da
Phosphorylation

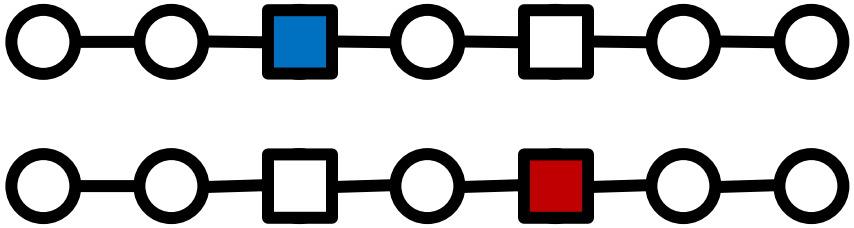
Localization of modifications

Peptide with two possible modification sites

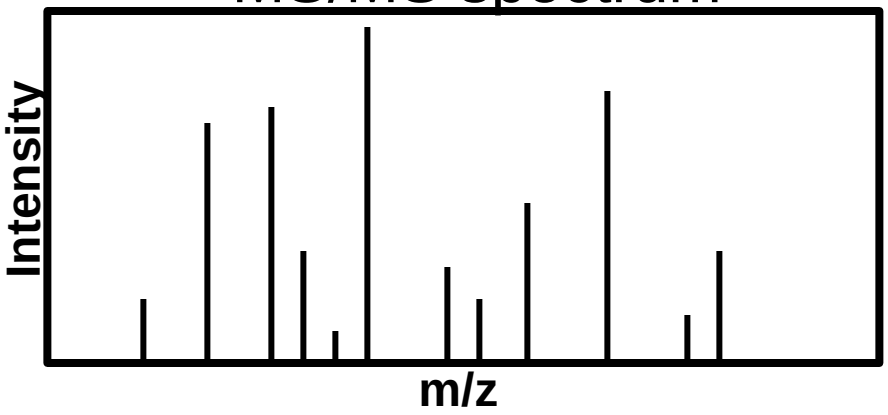


Localization of modifications

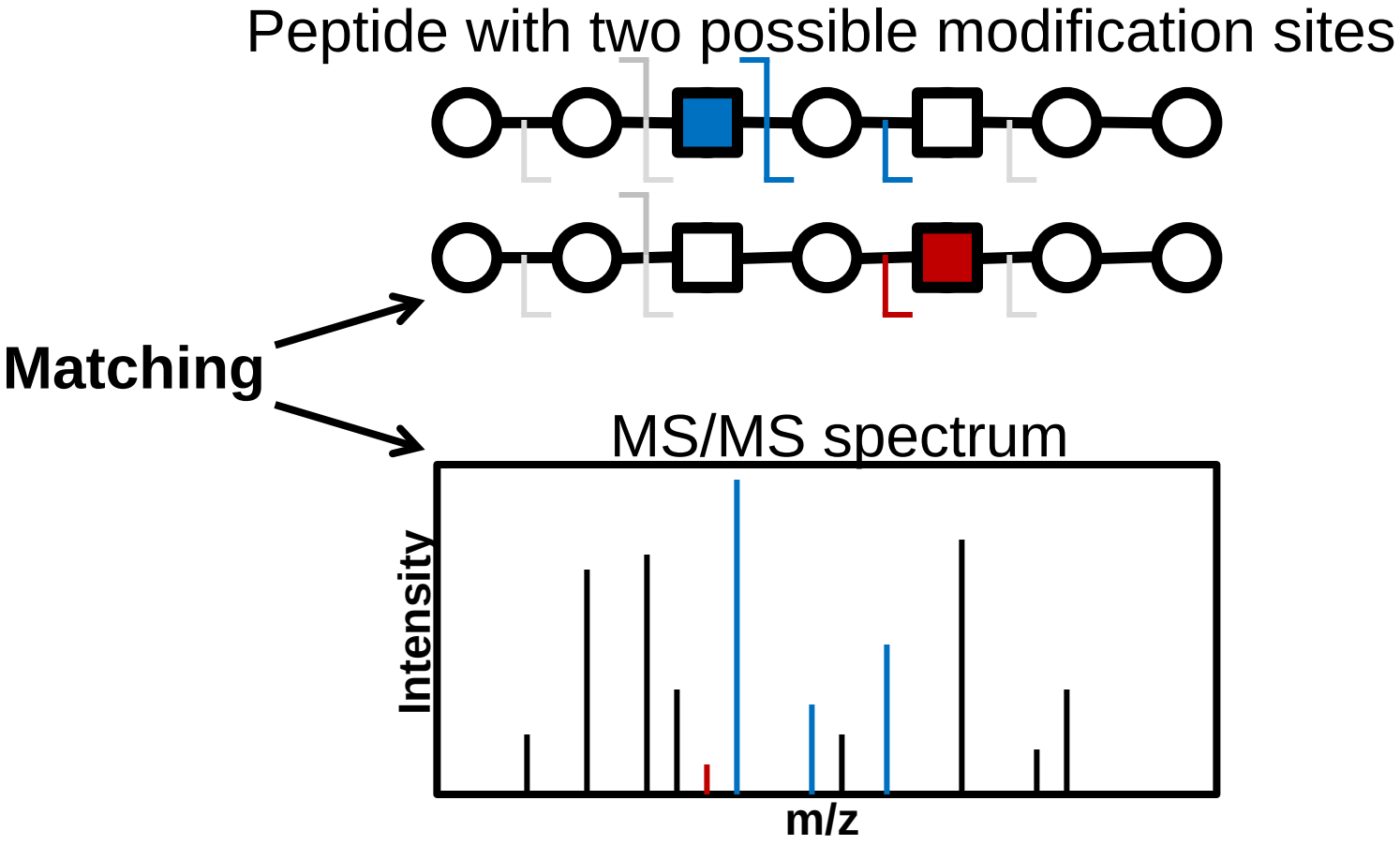
Peptide with two possible modification sites



MS/MS spectrum

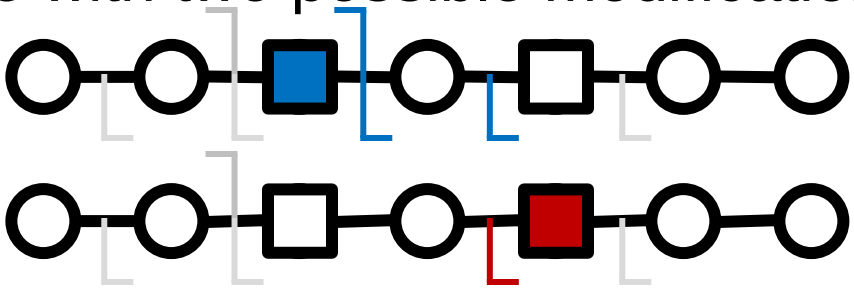


Localization of modifications



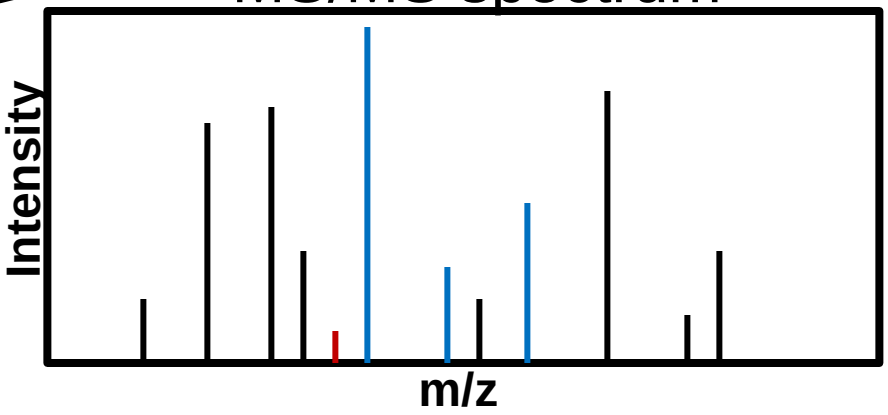
Localization of modifications

Peptide with two possible modification sites



Matching

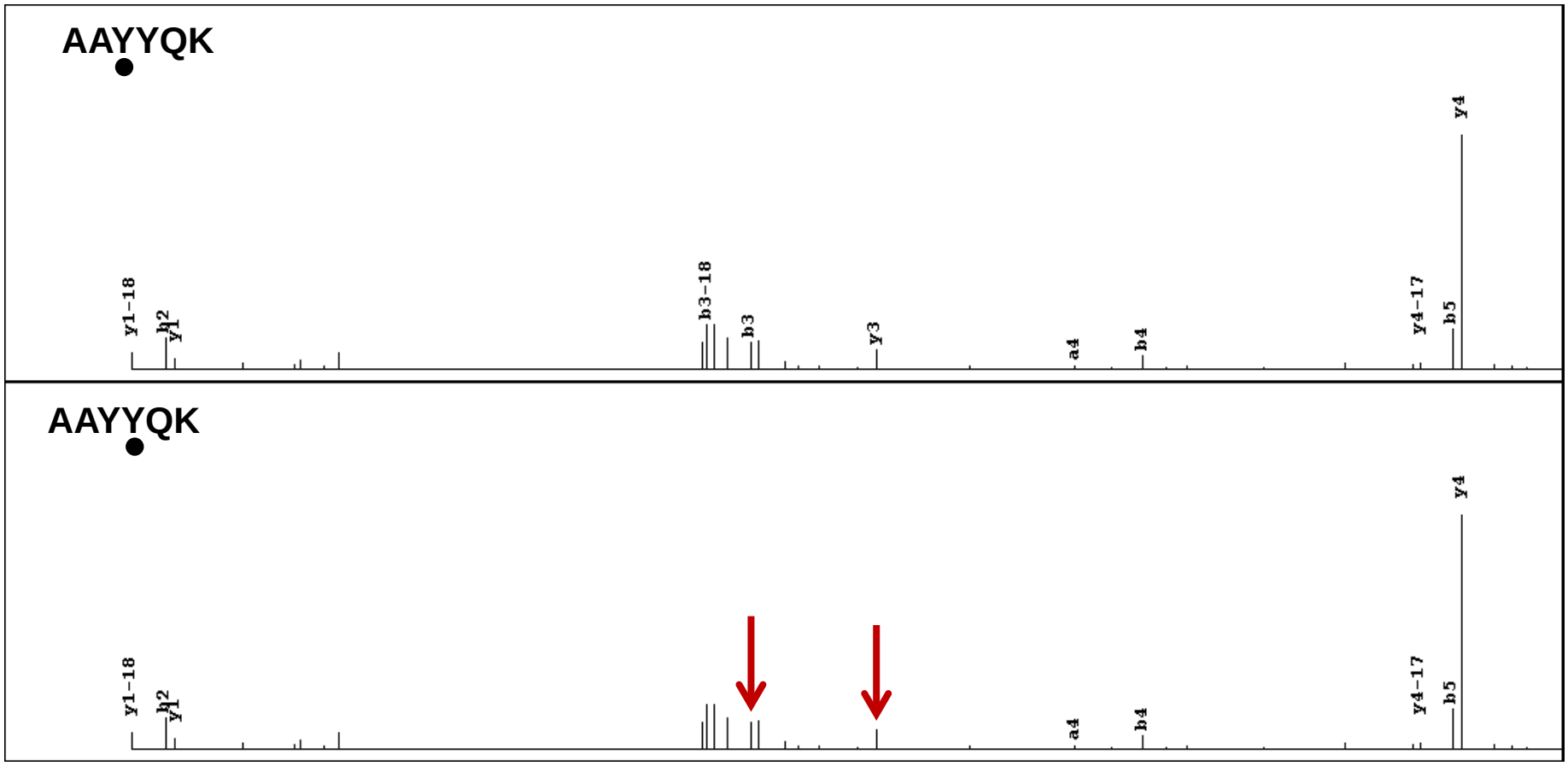
MS/MS spectrum



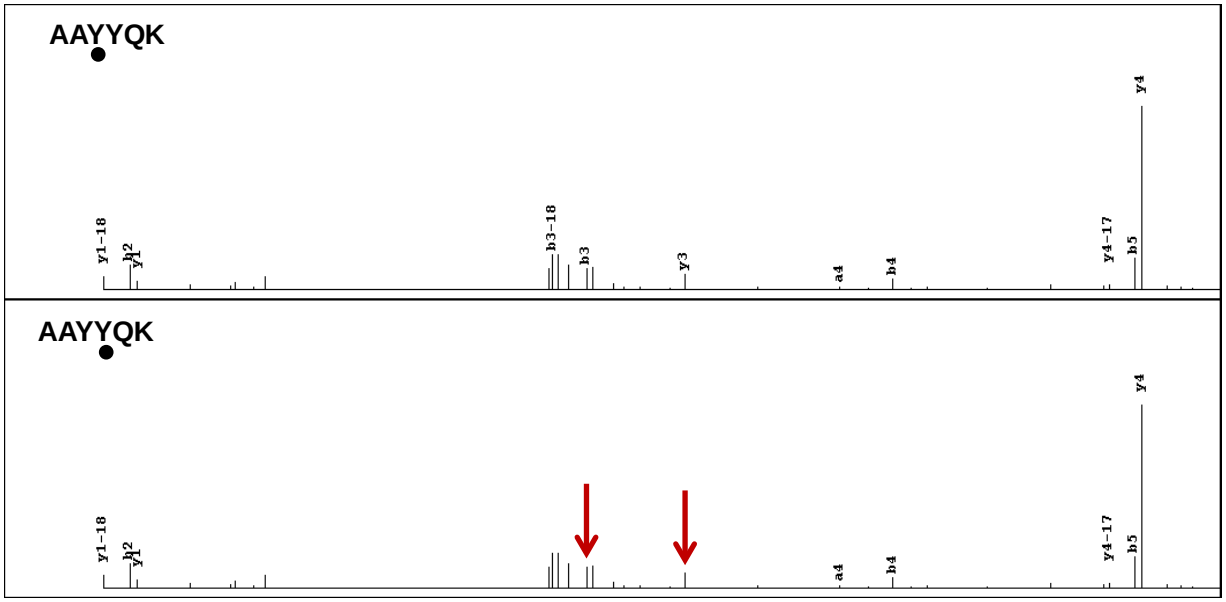
Which assignment does the data support?

1, 1 or 2, or 1 and 2?

Visualization of evidence for localization



Visualization of evidence for localization



Rank	AAYYQK	Total (matched)	Difference (matched)
1	AAYYQK	8	-
2	AAYYQK	6	2 0

Rank	AAYYQK	Total (intensity)	Difference (intensity)
1	AAYYQK	0.809	-
2	AAYYQK	0.735	0.074 0

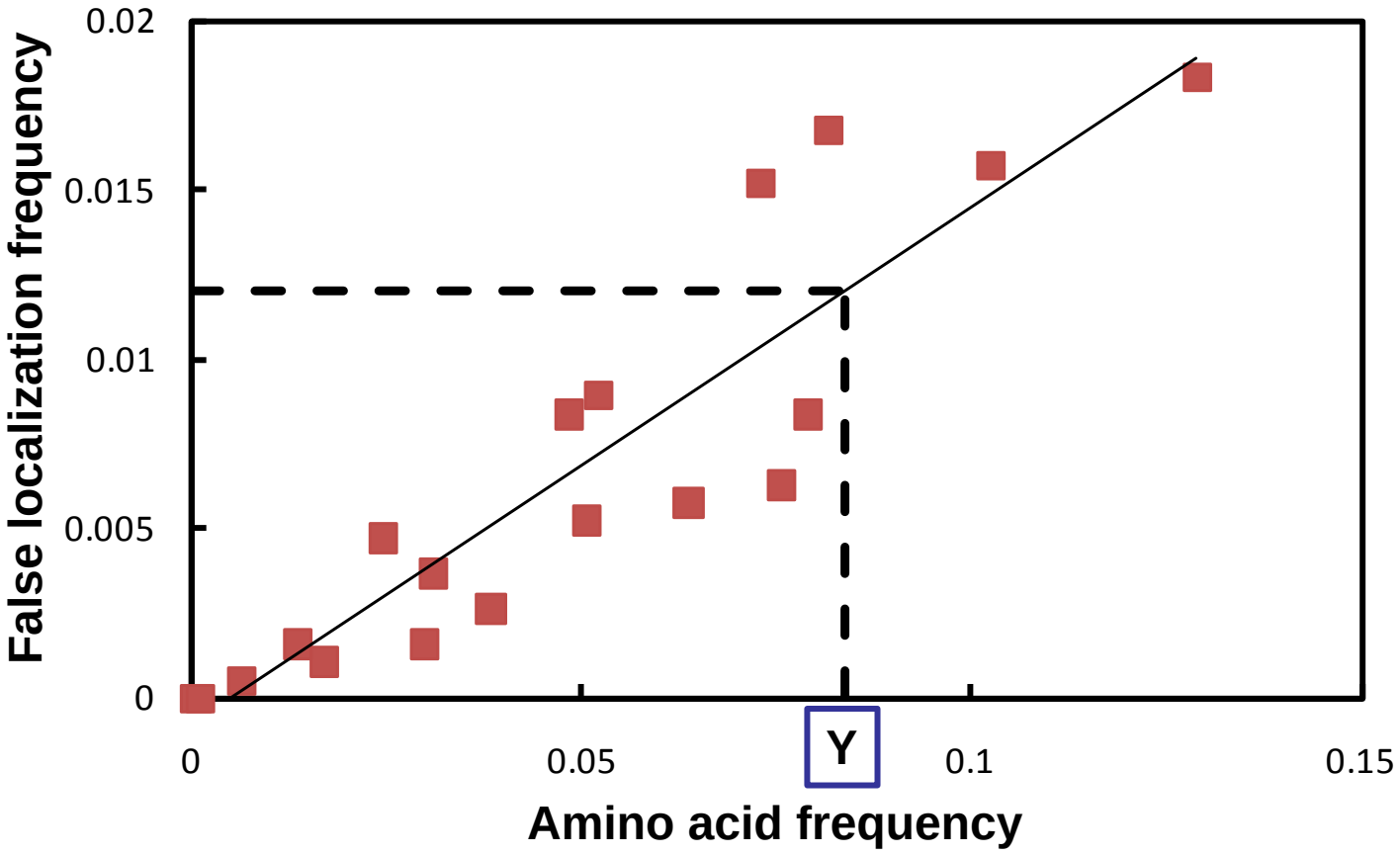
Visualization of evidence for localization

Rank	AAVPSGASTGIYEALELR	Total (matched)	Difference (matched)
1	AAVPSGASTGIYEALELR	15	-
2	AAVPSGASTGIYEALELR	14	3 2
3	AAVPSGASTGIYEALELR	13	4 2

Rank	AAVPSGASTGIYEALELR	Total (intensity)	Difference (intensity)
1	AAVPSGASTGIYEALELR	0.315	-
2	AAVPSGASTGIYEALELR	0.29	0.025 0
3	AAVPSGASTGIYEALELR	0.187	0.167 0.039

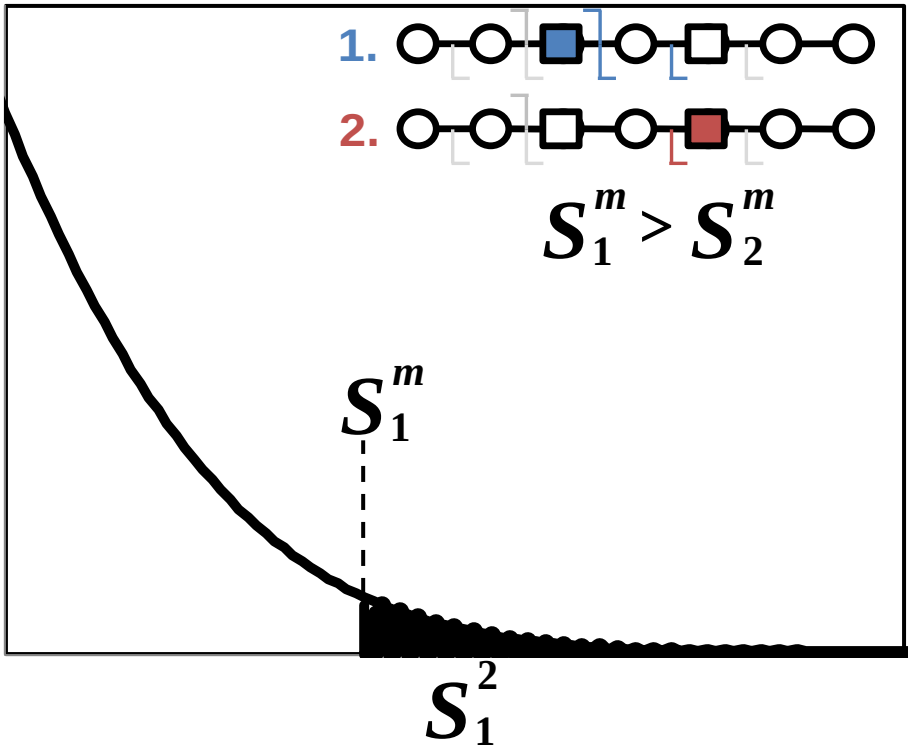
Estimation of global false localization rate using decoy sites

By counting how many times the phosphorylation is localized to amino acids that can not be phosphorylated we can estimate the false localization rate as a function of amino acid frequency.



How much can we trust a single localization assignment?

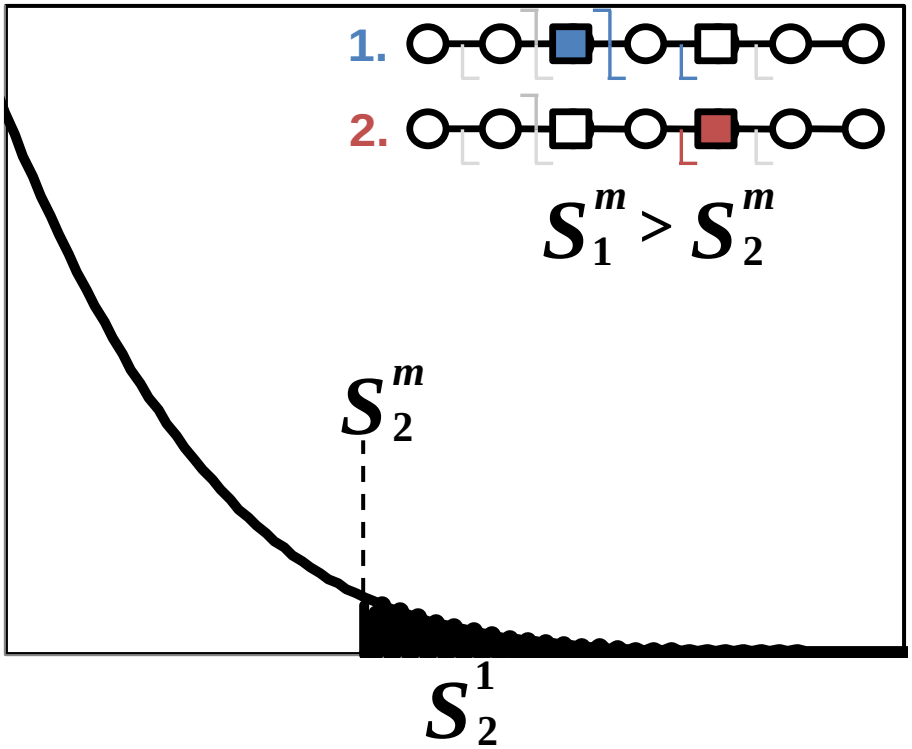
If we can generate the distribution of scores for assignment 1 when 2 is the correct assignment, it is possible to estimate the probability of obtaining a certain score by chance for a given peptide sequence and MS/MS spectrum assignment.



$$p_1^2 = \frac{\int_0^{S_1^m} F^2(S_1^2)\,dS_1^2}{\int_0^{\infty} F^2(S_1^2)\,dS_1^2}$$

Is it a mixture or not?

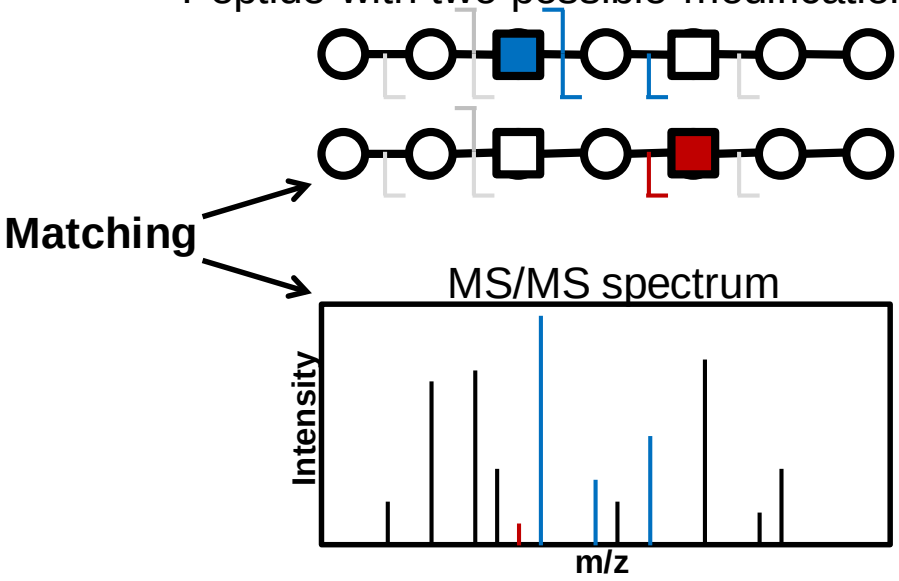
If we can generate the distribution of scores for assignment 2 when 1 is the correct assignment, it is possible to estimate the probability of obtaining a certain score by chance for a given peptide sequence and MS/MS spectrum assignment.



$$p_2^1 = \frac{\int_0^{S_2^m} F^1(S_2^1)\,dS_2^1}{\int_0^{\infty} F^1(S_2^1)\,dS_2^1}$$

Localization of modifications

Peptide with two possible modification sites



Which assignment does the data support?
1, 1 or 2, or 1 and 2?

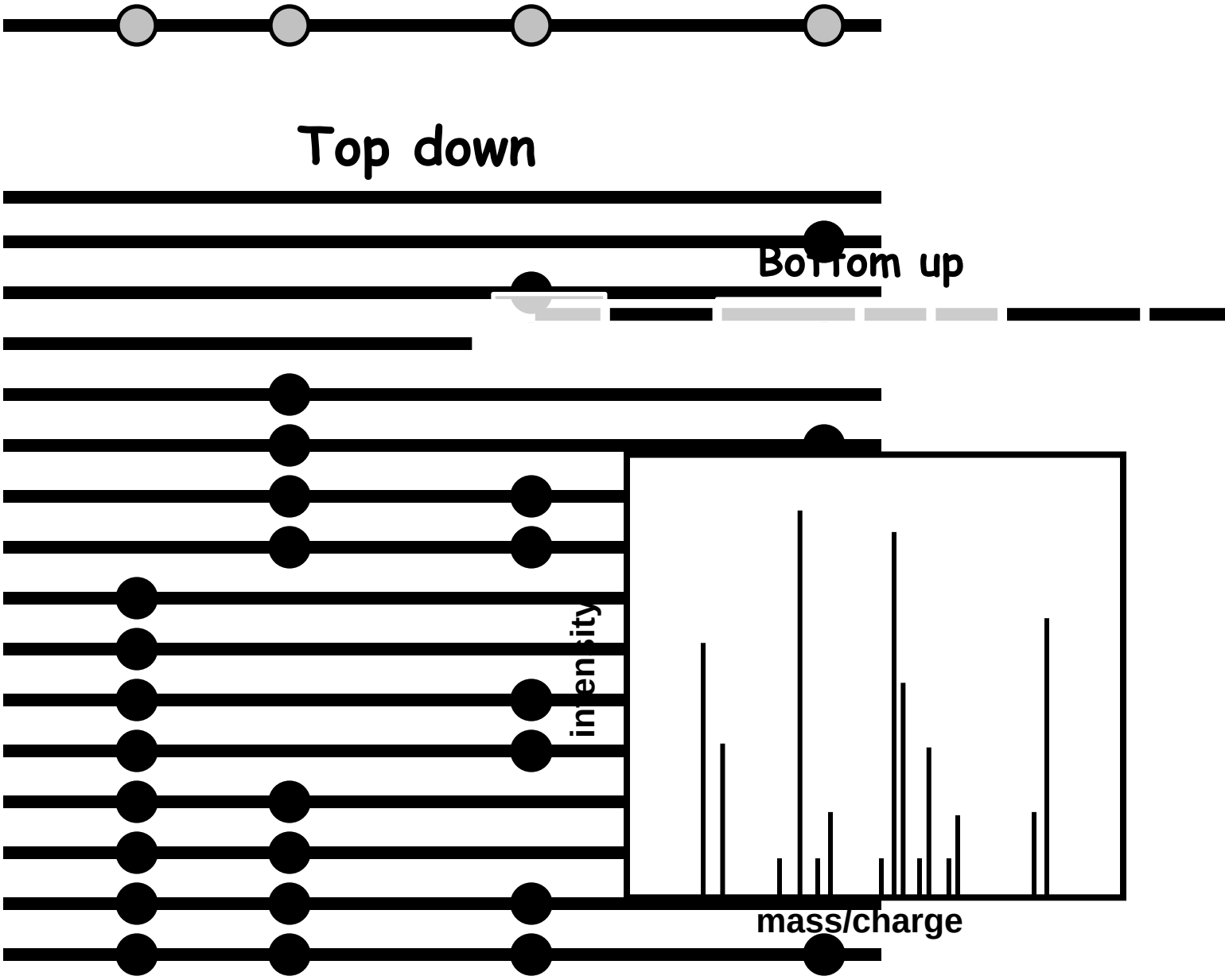
$$p_1^2 \le p_{th} \text{ and } p_2^1 \le p_{th} \Rightarrow \textbf{1 and 2}$$

$$p_1^2 \le p_{th} \text{ and } p_2^1 > p_{th} \Rightarrow \textbf{1}$$

$$p_1^2 > p_{th} \text{ and } p_2^1 \le p_{th} \Rightarrow \emptyset \quad (S_1^m \ge S_2^m \Rightarrow p_1^2 \le p_2^1)$$

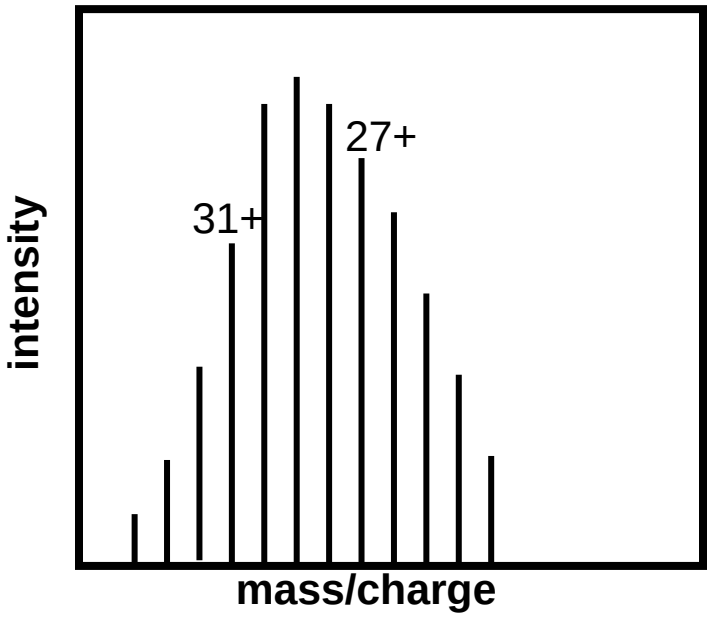
$$p_1^2 > p_{th} \text{ and } p_2^1 > p_{th} \Rightarrow \textbf{1 or 2}$$

Top down / bottom up

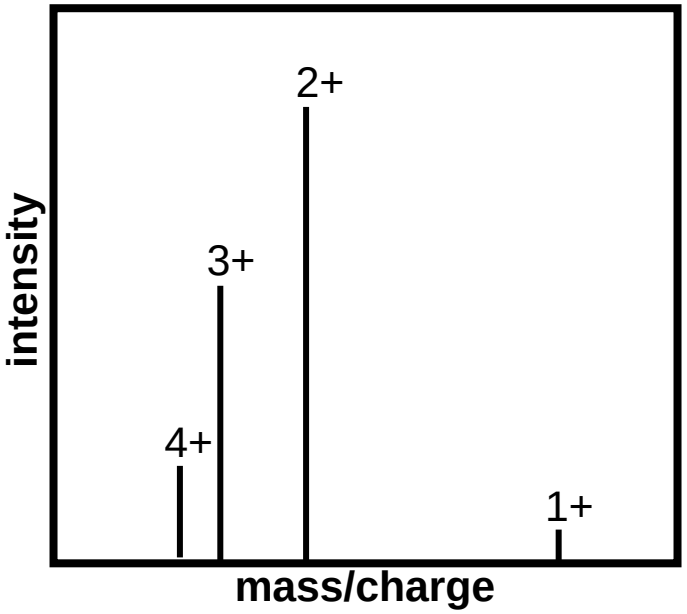


Charge distribution

Top down

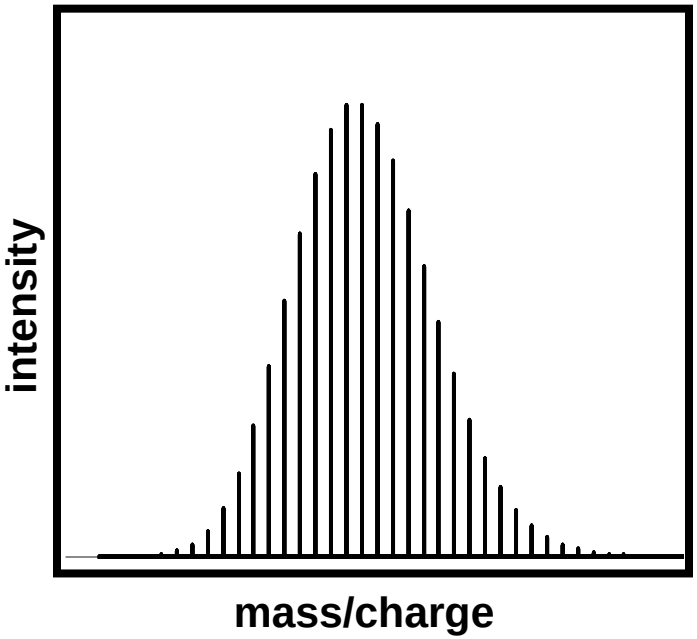


Bottom up

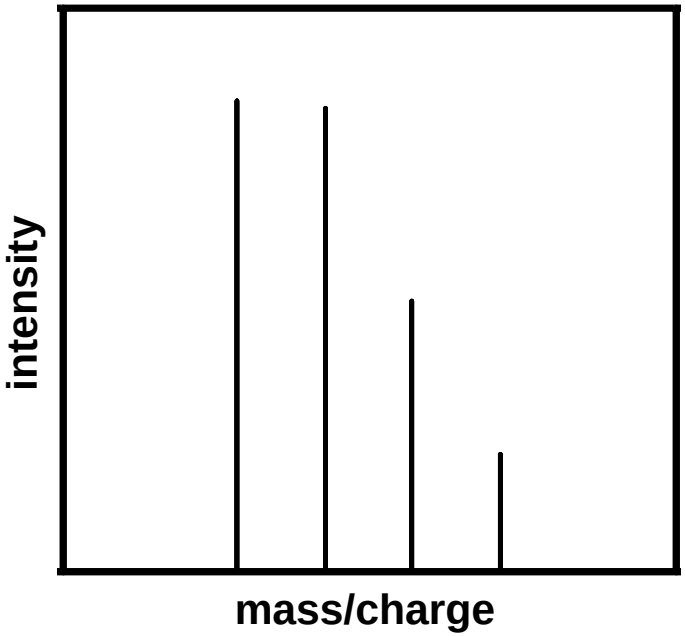


Isotope distribution

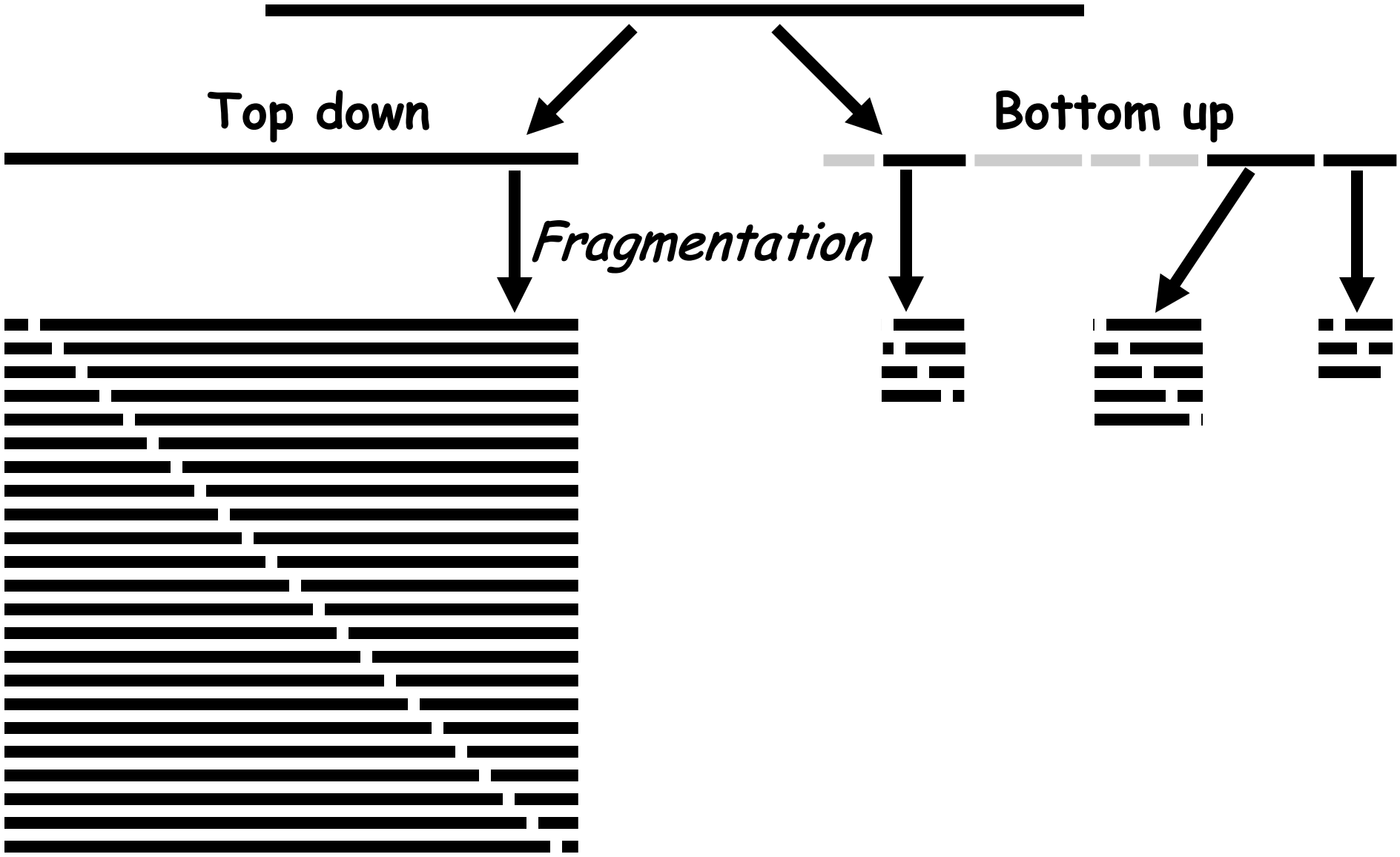
Top down



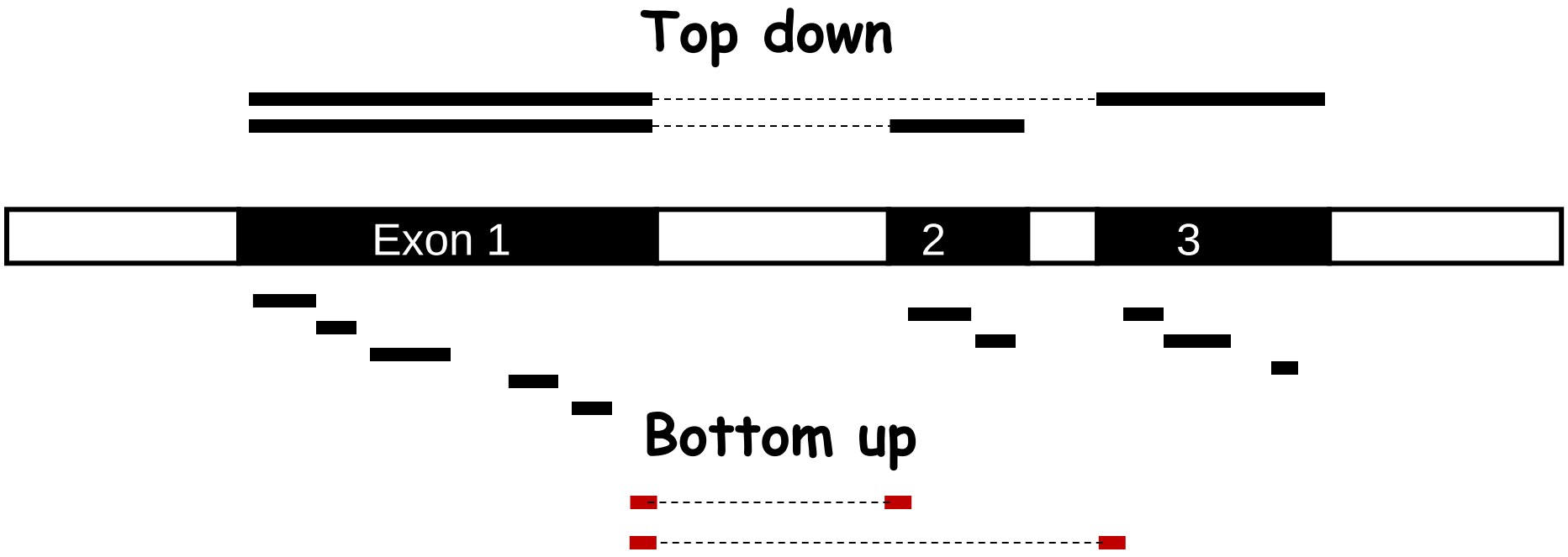
Bottom up



Fragmentation

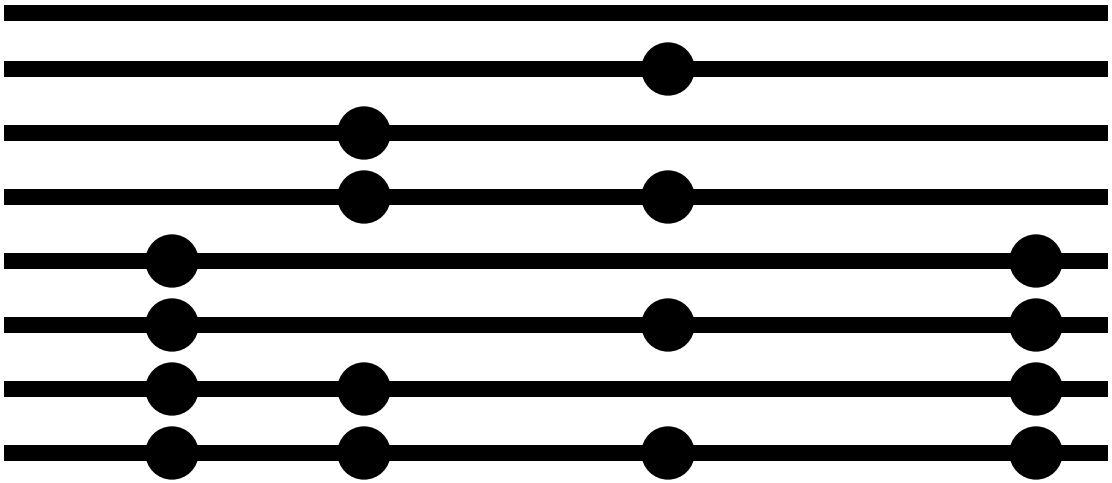


Alternative Splicing



Correlations between modifications

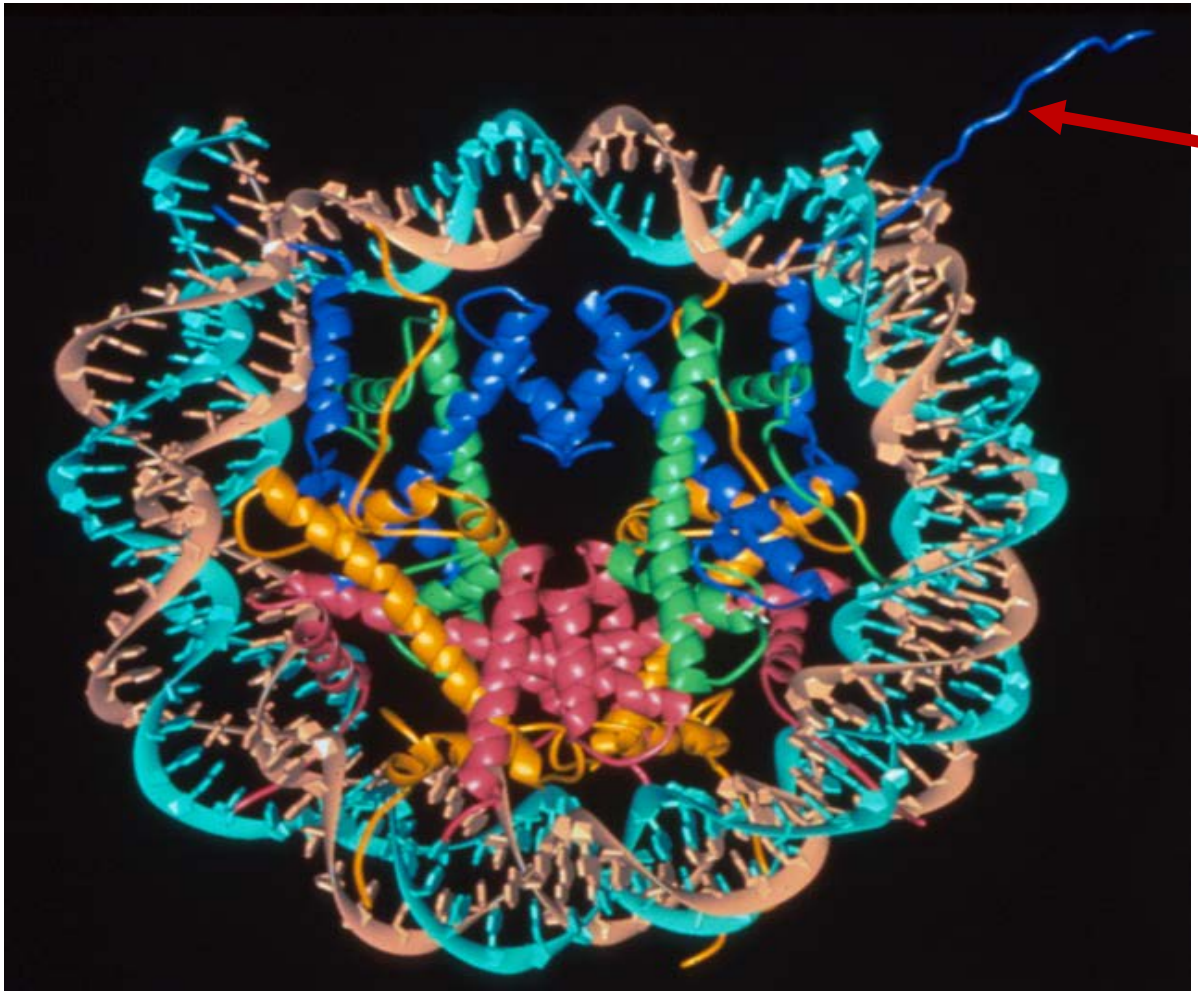
Top down



Bottom up

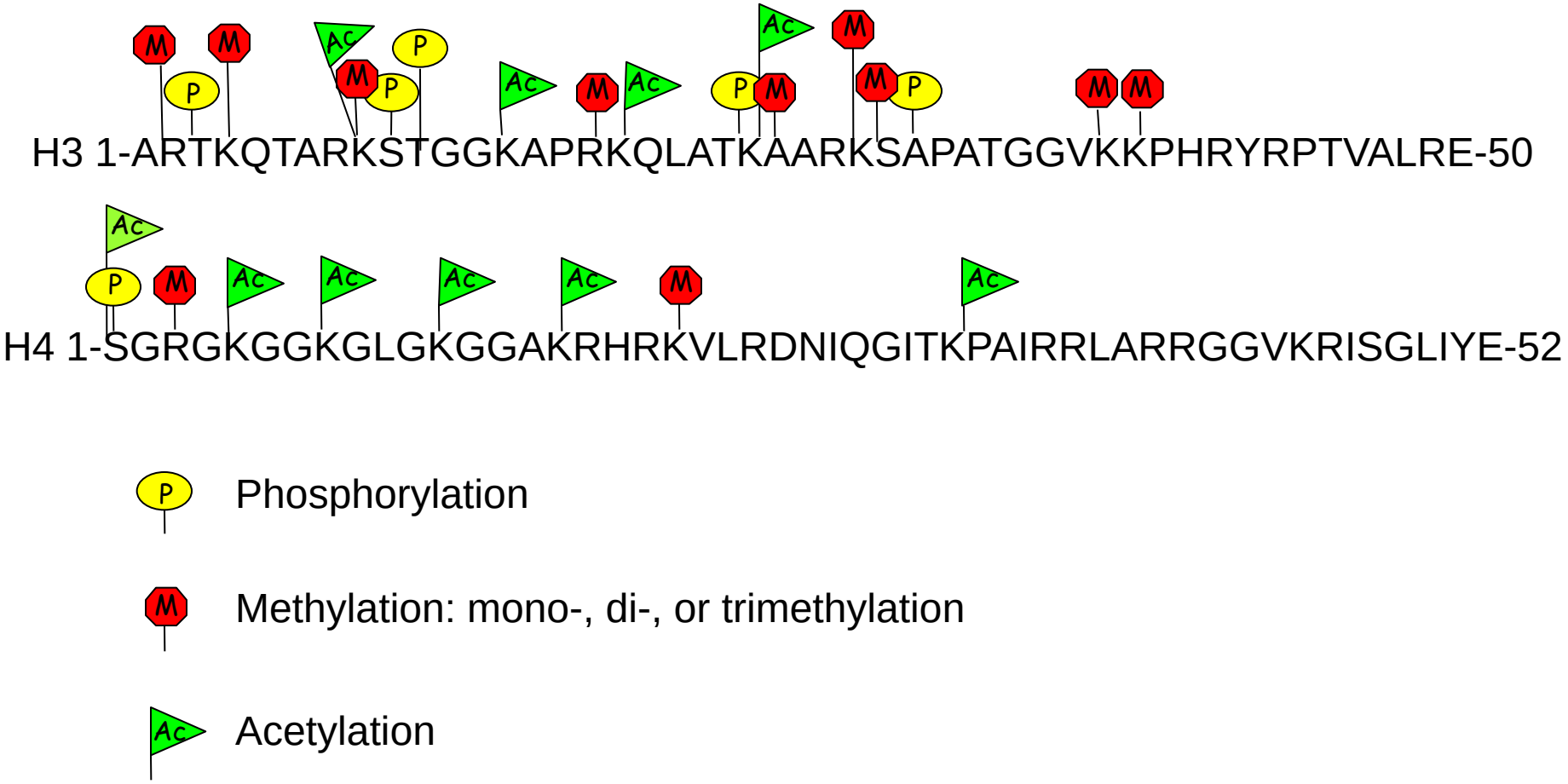
The Nucleosome Core Complex

H3
H4
H2A
H2B



Luger et al., Nature, 389, 251-260, 1997

The N-terminal Tails of Histone H3 and H4

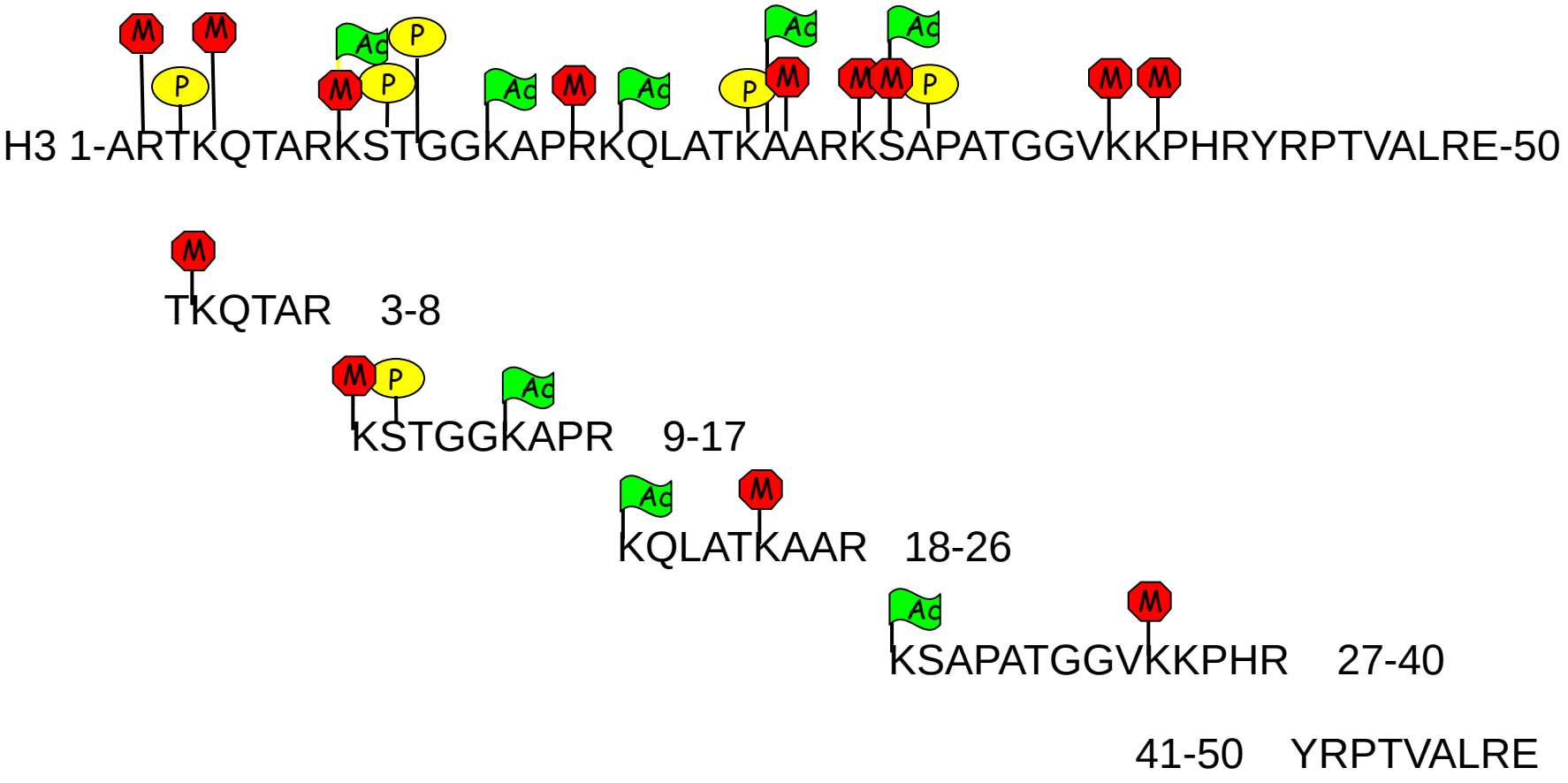


The Histone Code Hypothesis

Specific post translational modifications (PTMs) of the N-terminal tails of histones function as a scaffold for binding of protein factors leading to transcriptional activation or inactivation.

Jenuwein, T., Allis, C.D., Science, 293, 2001

Interdependence of Modifications is lost in Standard Mass Spectrometry Analysis



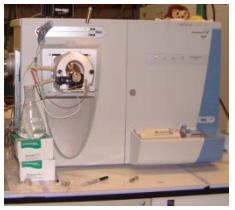
Histone Proteins are a Highly Complex Mixture of a Single Protein….

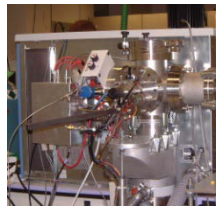
…… and many many more!

Protocol

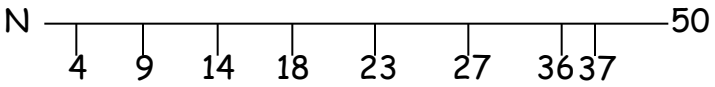
LTQ-FTMS

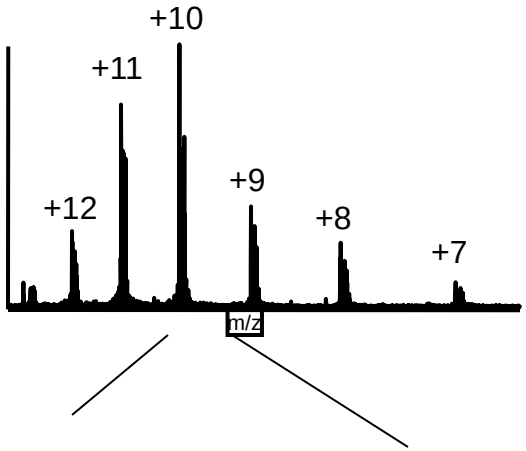
LTQ-ETD/PTR



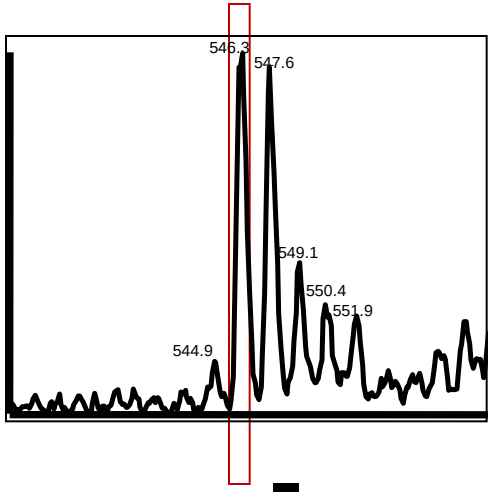


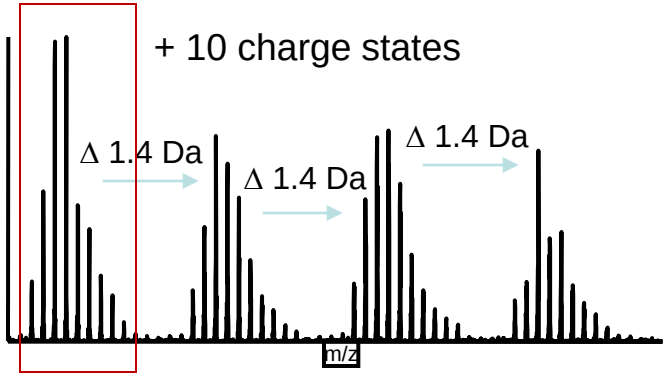
Glu-C generated
N-terminal H3 peptide (1-50)

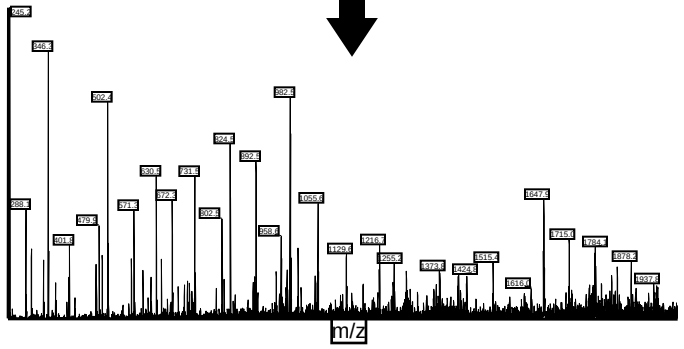




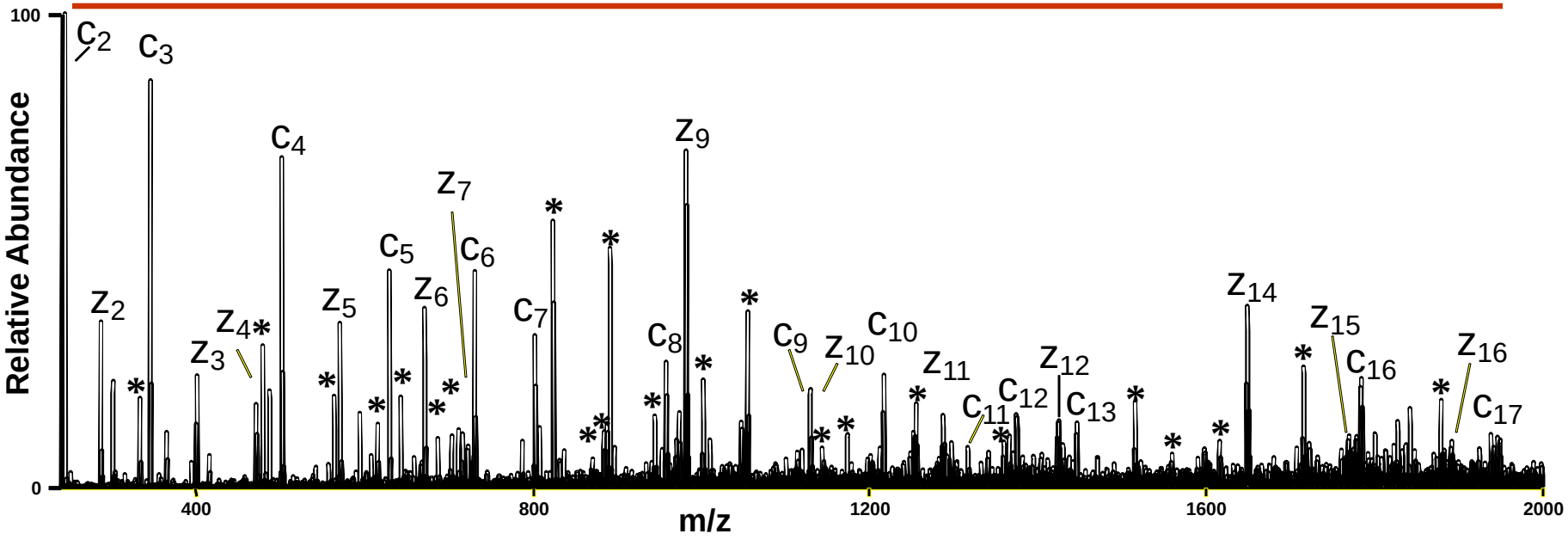
- Isolate m/z ± 0.5 Da
- 60 ms ETD
- ~ 3 min acquisition

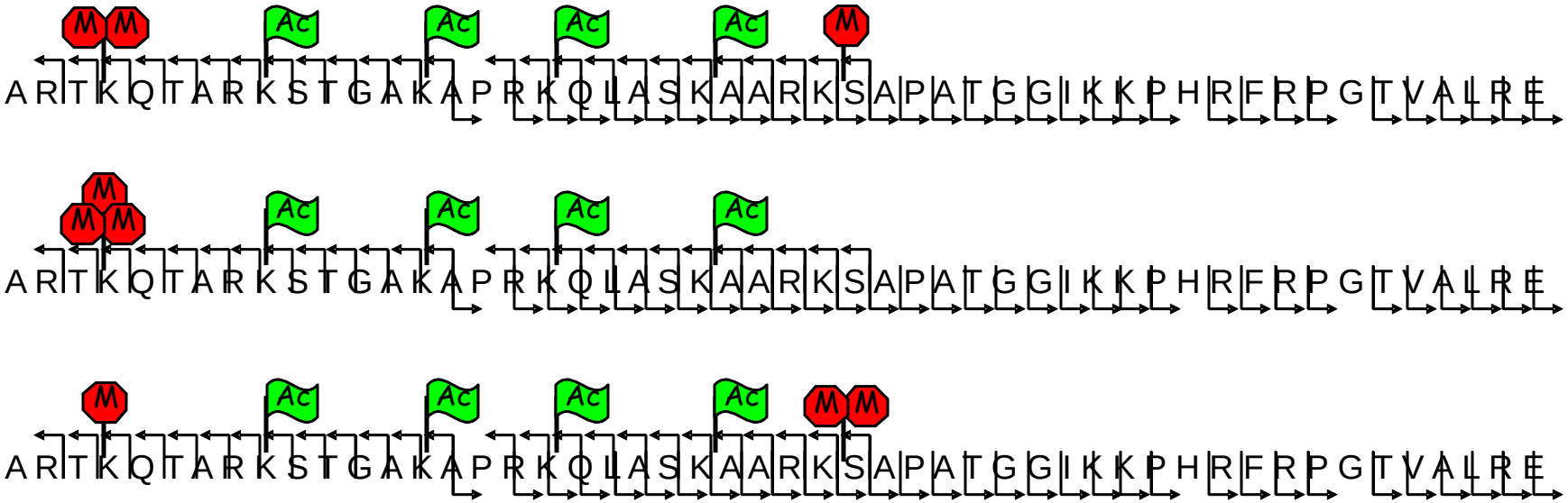




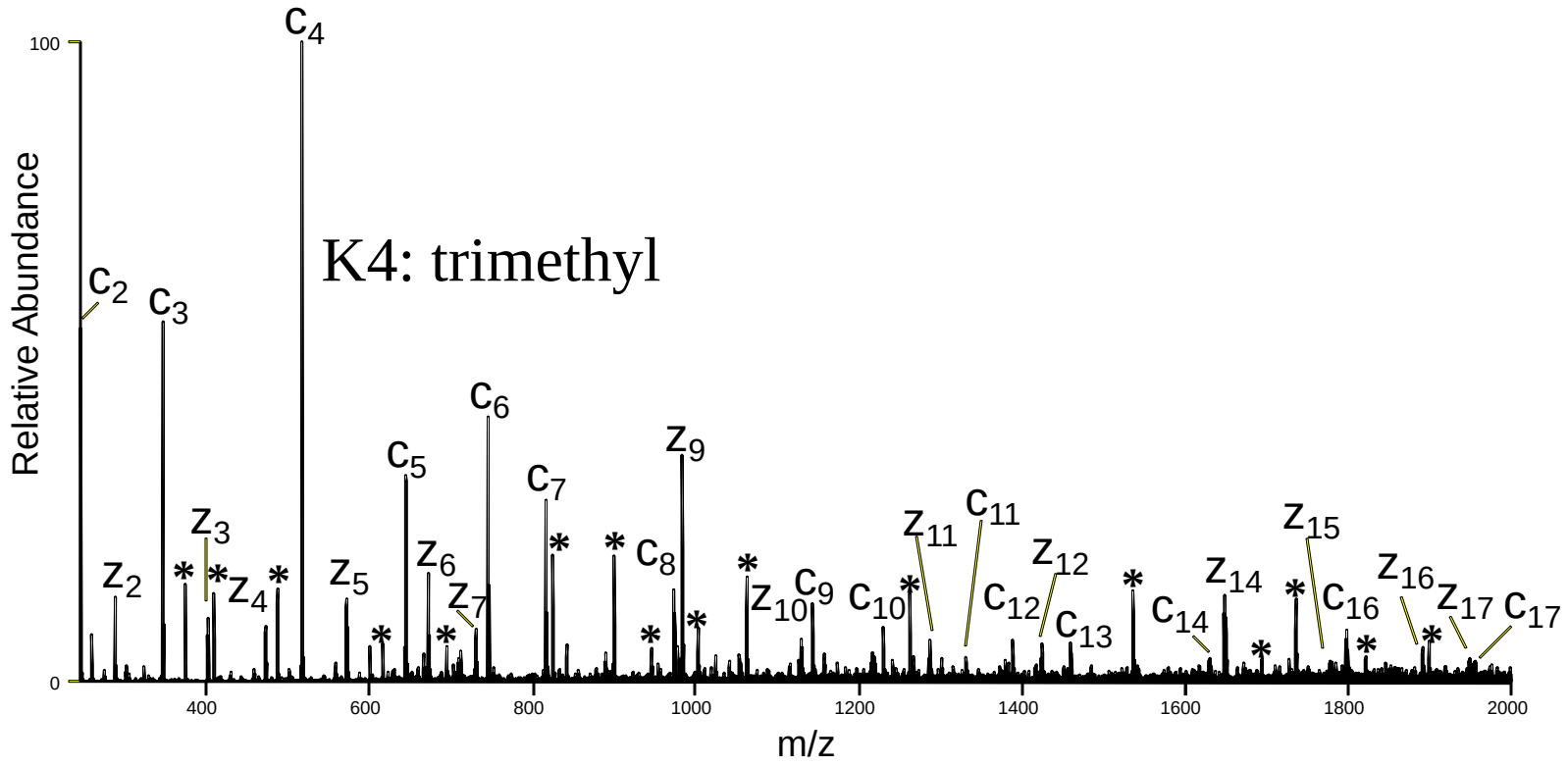


Group '4': 4 Acetyl Groups





Group '5': 5 Acetyl Groups



Proteomics Informatics – Protein characterization I: post-translational modifications (Week 10)