Identification of Protein Modifications by Mass Spectrometry

PC219
Mike Baldwin

mbaldwin@cgl.ucsf.edu

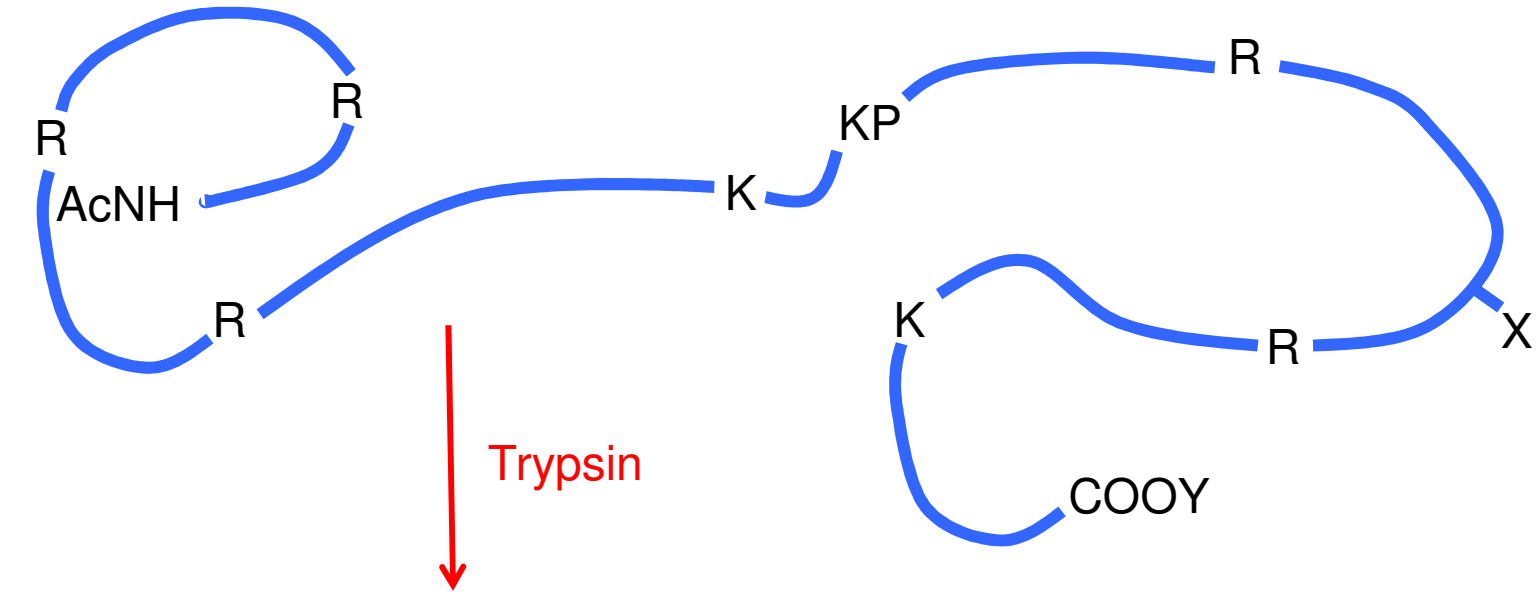
Why Mass Spectrometry?

Mass spectrometry is ideally suited to the identification of peptide and protein modifications:

- It is highly sensitive and is frequently effective down to the femtomole scale, sometimes to attomoles.
- Many modifications introduce a characteristic, recognizable mass shift.
- MS detects molecular mass shifts corresponding to modifications, at both the peptide and protein level.
- Tandem MS reveals peptide/protein sequence and identifies the locations of modifications.
- Accurate mass measurements can sometimes distinguish between modifications with the same nominal mass.

Modifications of most interest are either cotranslational or posttranslational modifications or mutations. However, these need to be distinguished from chemical modifications introduced during protein isolation and purification.

Identifying Modifications



1 AcNH — R — Predicted mass + 42

2 R — Predicted mass OK

3 R — Predicted mass OK

4 K — Predicted mass OK

5 (K) — Probably no peptide

6 KP — R — Missed cleavage

X

7 R — "Wrong" mass

8 K — Predicted mass OK

9 COOY — "Wrong" mass

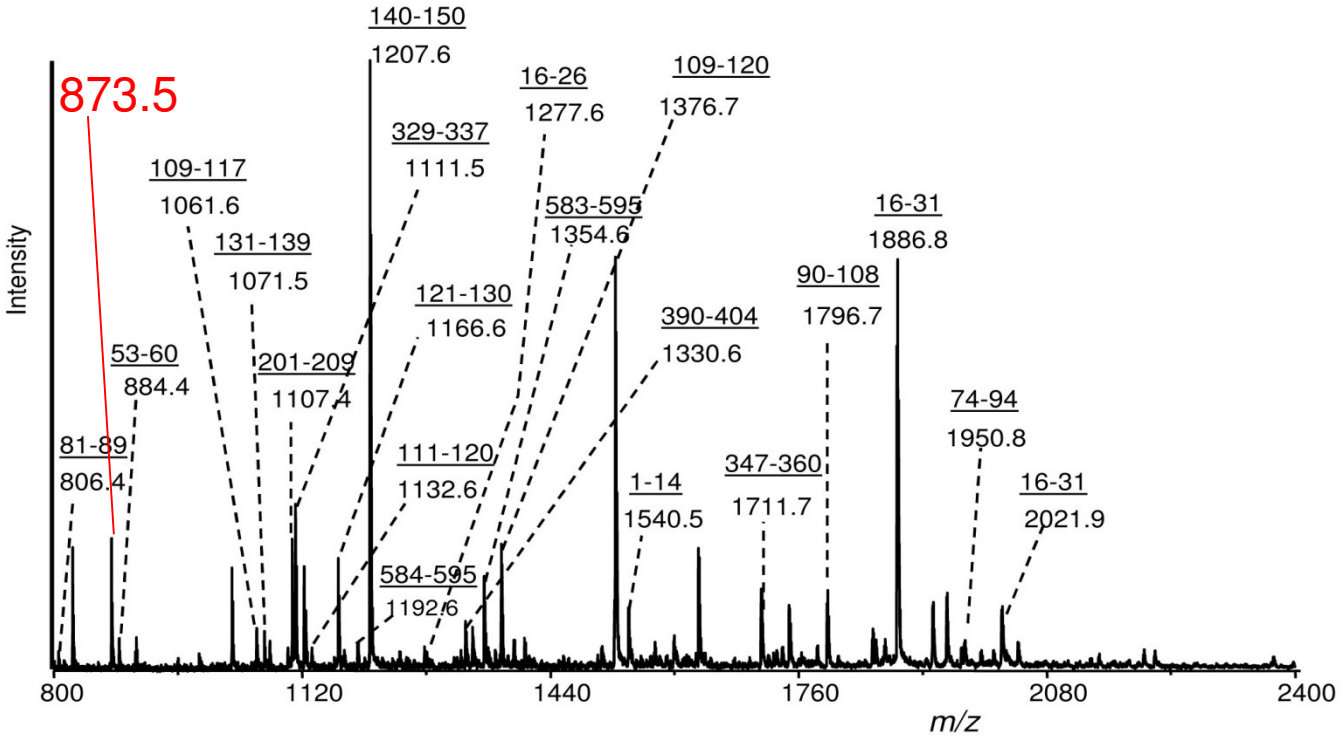
Identifying Modifications

Previous conclusion: Maybe 4 out of 9 predicted peptides will be identified as having the “correct” mass.

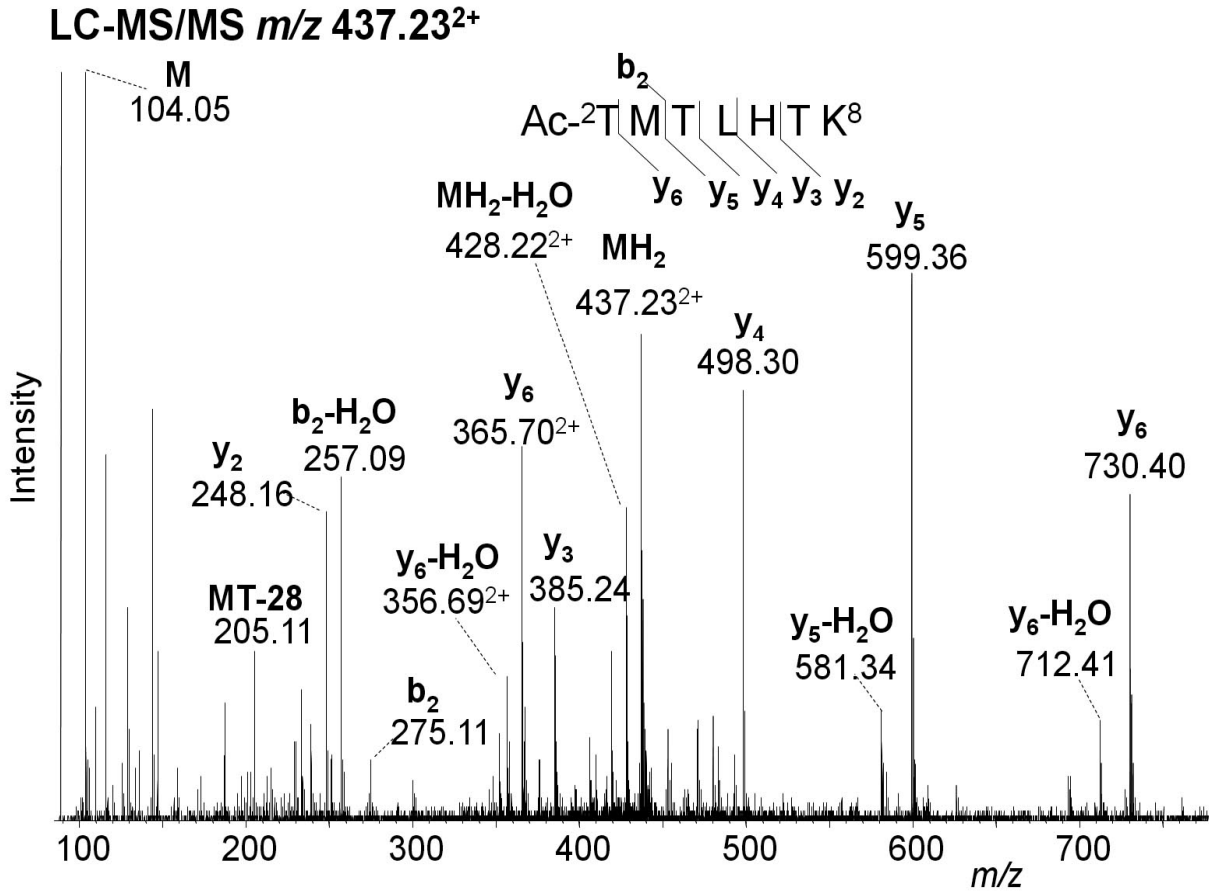
- Either PMF or MSMS will identify the “correct” peptides and will thereby identify the protein they came from.
- A theoretical digest of the identified protein will predict other peptides and allow for them to be searched for with anticipated modifications.
- The following example shows how the search for a modification can be narrowed down.

MALDI spectrum

of an unseparated mixture from a tryptic digest of a protein isolated from a human breast cancer cell.



- A database search identified the protein as the estrogen receptor (ER).
- A theoretical digest of the identified protein predicted other peptides and allowed for them to be searched for with various anticipated modifications.
- No peak was seen for the predicted N-terminal peptide MTMTLHTK, calc. m/z 962.48, but a peak was found for this peptide, less Met, plus acetyl: AcTMTLHTK, calc. m/z 873.45.
- This suggested but did not prove the modification, so the peptide was analyzed by MS/MS.



- This is an LC-MS/MS spectrum from a QTOF (QStar).
- The precursor ion is a doubly charged peptide of m/z 437.23, which matches the expected mass of the modified peptide AcTMTLHTK^{2+}.
- The strong y-ion series confirms the peptide sequence and identity. Note that the y-ions are predominantly singly charged.

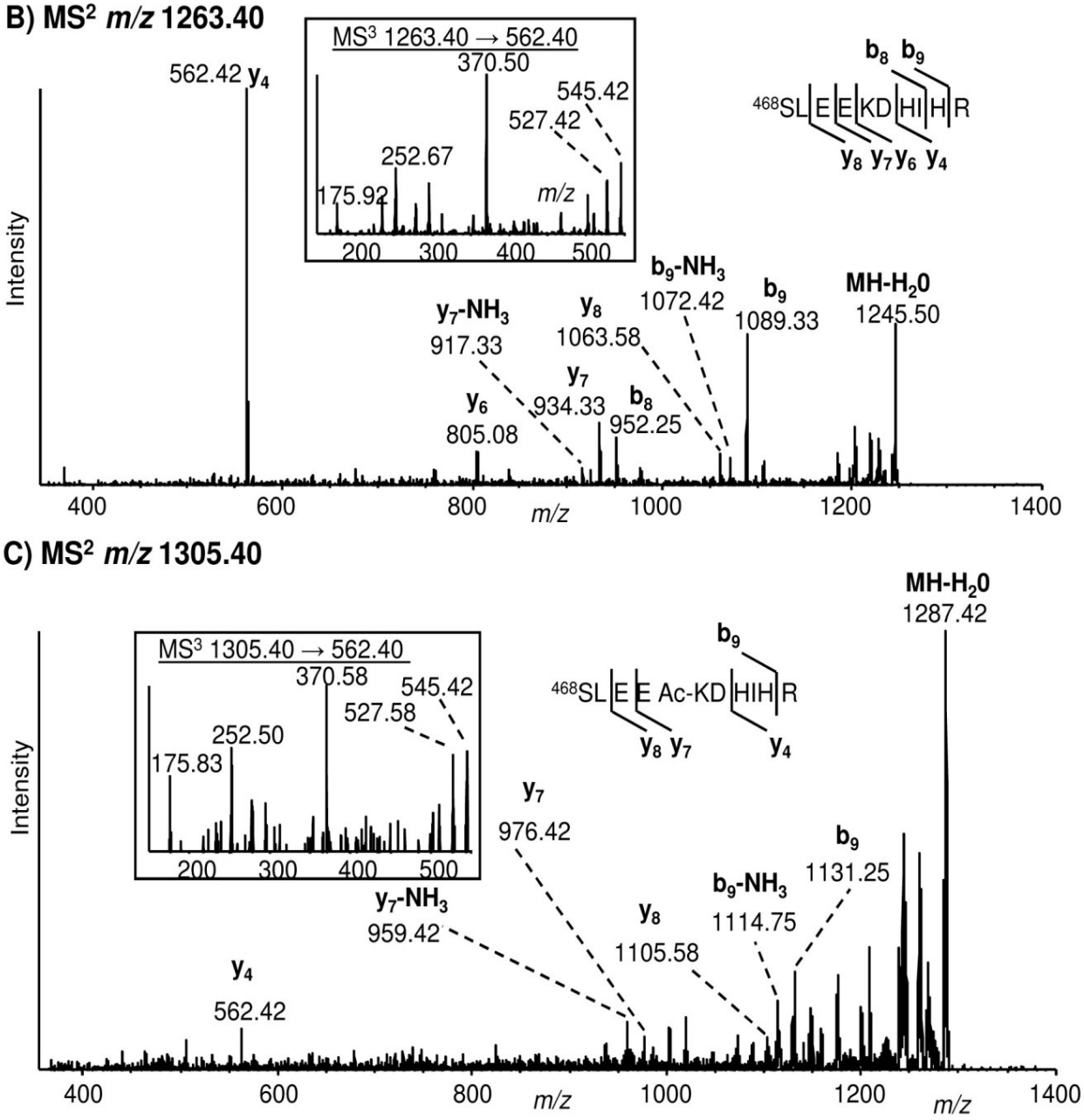
What did we learn?

- In theory the acetyl group could be on the lysine side-chain (residue 8), but this would give very different ion series.
- This confirmed the elimination of the N-terminal methionine and the acetylation of residue 2.
- Further evidence: In a subsequent experiment using Glu-C rather than trypsin, another peptide Ac-^{2}TMTLHTKASGMALLHQIQGNE was also identified. Database searching gave a Mascot score of 72.

Note: In separate experiments side-chain acetylation was identified at Lys-472. MS/MS/MS (MS^3) played an important role in confirming this.

Side-chain Acetylation

- Two peptides were seen with a mass difference of 42 Da, suggesting acetylation.
- MS/MS identified the lighter species as SLEEKDHIHR, with strong sequence ions.
- The heavier species gave some of the same ions but others such as y7 and y8 were shifted by 42 Da.
- This proved the modification was not N-terminal.
- MS^3 proved the m/z 562.4 ion y4 had the same structure in both cases.
- Therefore the modification was not C-terminal and could only be on EKD, i.e. Lys-472.



The Origins of Modifications

Genome

↓

Transcription – mRNA

↓

Translation – proteins + co-translational modifications

↓

Post-translational modifications

↓

Chemical changes during purification/isolation/digestion etc.

Some Typical Posttranslational Modifications

- Enzymatic processing. May be cotranslational e.g. loss of N-terminal Met (-131 Da), usually with acetylation (-131 + 42 = -89 Da)
- N-, O-, C-linked glycosylation (various)
- Phosphorylation of Tyr, Ser, Thr, His, Asp (+80 Da)
- Acylation , e.g. acetylation of the N-terminus (+42 Da), or fatty acid anchors on Cys (various)
- Cross-linking of Lys, Trp, Tyr, Met
- Disulfide formation between Cys residues (-2, -4, etc.)
- Other oxidation of Cys, Met, Trp, Tyr, His (e.g. +16, 32, etc.)
- Methylation of N-terminus, or Arg, Lys (e.g. +14, 28, 42)
- Ubiquitination of Lys (Tryptic peptide mass increases by 114 (GG) or 383 (LRGG))

Predicting Post-translational modifications and their behavior

- Generally the presence of PTM's is not easily predicted, although some consensus sequences have been derived, e.g. various algorithms predict phosphorylation with modest success. Also, eukaryotic proteins are mostly acetylated at the N-terminus. For 2nd aa G, A, S, C, T, P or V, Met-1 is clipped off. For 2nd aa E, D, Q, M, I, L, W or F, there is no clipping and Met-1 gets acetylated.
- Usually organism-dependent and often tissue- or location-specific.
- Can be stable or dynamic and may exhibit high or sometimes very low occupancy.
- May alter biological activity, *and* physical properties.
- Differences in physical properties may facilitate the separation of modified and unmodified forms, which can be very important for low occupancy modifications.
- May be homogeneous or highly heterogeneous, displaying a wide range of masses, e.g. N-linked sugars.

Some questions to ask before starting analysis

- Will the modification survive the purification / isolation / digestion procedures?
- Will the modification survive the ionization and MS/MS activation?
- Can the site of modification be isolated in a peptide of a size suitable for mass spectrometric analysis?
- Does it have chemical or physical properties that will facilitate the isolation and enrichment?
- Will its physical properties cause peptide losses, e.g. will the peptide be too hydrophilic or hydrophobic for LC separation?
- Will the use of a chemical modification increase the success of separation and analysis, and at what stage should it be modified?

Phosphorylation is an important regulatory event

- Turns proteins on and off.
- Induces or prevents other post-translational modifications in the same protein.
- Involved in signaling pathways and phosphorylation cascades.
- BUT:
- Dynamic process are dependent on kinase vs. phosphatase. Ideally *both* should be blocked during isolation.
- The stoichiometry / occupancy is often very low (<5%) so we need high sensitivity or we need enrichment protocols.
- Low ionization efficiency: phosphopeptide ions are often suppressed relative to unmodified peptide.
- It can be labile and may be lost in the separation, purification or mass analysis.

Phosphorylation Analysis - Advantages

- Phospho group confers physical / chemical properties that may facilitate separation and enrichment of phosphoproteins or phosphopeptides.
- Phosphorylation causes a slight shift in LC retention times so phosphopeptides usually elute before the non-phospho analog.
- It only occurs on a subset of amino acids: Ser, Thr, Tyr, (His, Asp). A peptide without one of these residues cannot be phosphorylated.
- It is homogeneous, i.e. a single phosphorylation always adds 80 Da.
- It often can survive MS/MS activation and analysis.
- Chemical elimination of phosphoric acid may occur to give a new species 98 Da less than the phosphopeptide, but this gives an identifiably different MSMS spectrum compared with the unmodified peptide.
- Special methods have been developed for MS^n, i.e. MS^3 and MS^4.

Is the Estrogen Receptor a Phosphoprotein?

- The estrogen receptor is a 66-kDa nuclear transcription factor that mediates transcriptional regulation of genes involved in cell proliferation and differentiation.
- It plays a pivotal role in the development and progression of breast cancer.
- Based on biochemical analysis including autoradiography, Edman sequencing, etc., it was reported that phosphorylation of serine, threonine, and tyrosine residues plays a critical role in modulating the activity of ER.
- But all previous biochemical identification of phosphorylated ER residues was limited to protein artificially overexpressed in transfected cell lines.
- There was strong evidence for phosphorylation but it had never been observed by mass spectrometry, not even in over-expressed protein.
- The analysis of endogenous protein would be much more challenging.

ER: Analytical Protocol

1. Human breast cancer cells were treated with estradiol (or HGF) as their interactions with ER are biologically significant and have been reported to increase phosphorylation at some sites.

2. ER was immunoprecipitated with an agarose-bound anti-ER antibody in the presence of phosphatases to inhibit removal of phospho groups.

3. The protein was separated by 1-D PAGE, cut out and digested in-gel with various enzymes to enhance the sequence coverage.

4. MALDI-MS/MS and LC-ES-IMS/MS gave 94% sequence coverage.

5. Peptides containing previously reported phosphorylation sites, e.g. Ser-118, Ser-167, were probed by MS^n for phosphorylation.

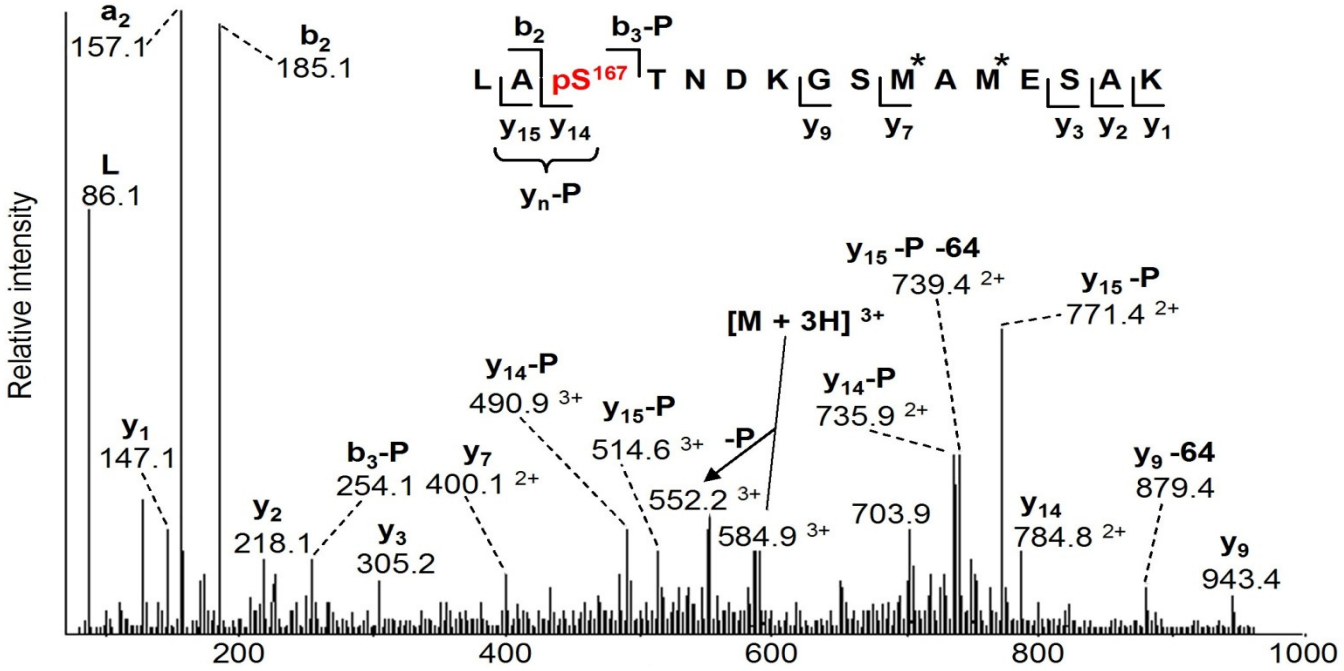
6. All other peptides were examined for +80 Da adducts.

7. For all serine-containing peptides, tandem MS experiments were carried out at 80 Da higher than the molecular ion, even when no precursor ion was visible.

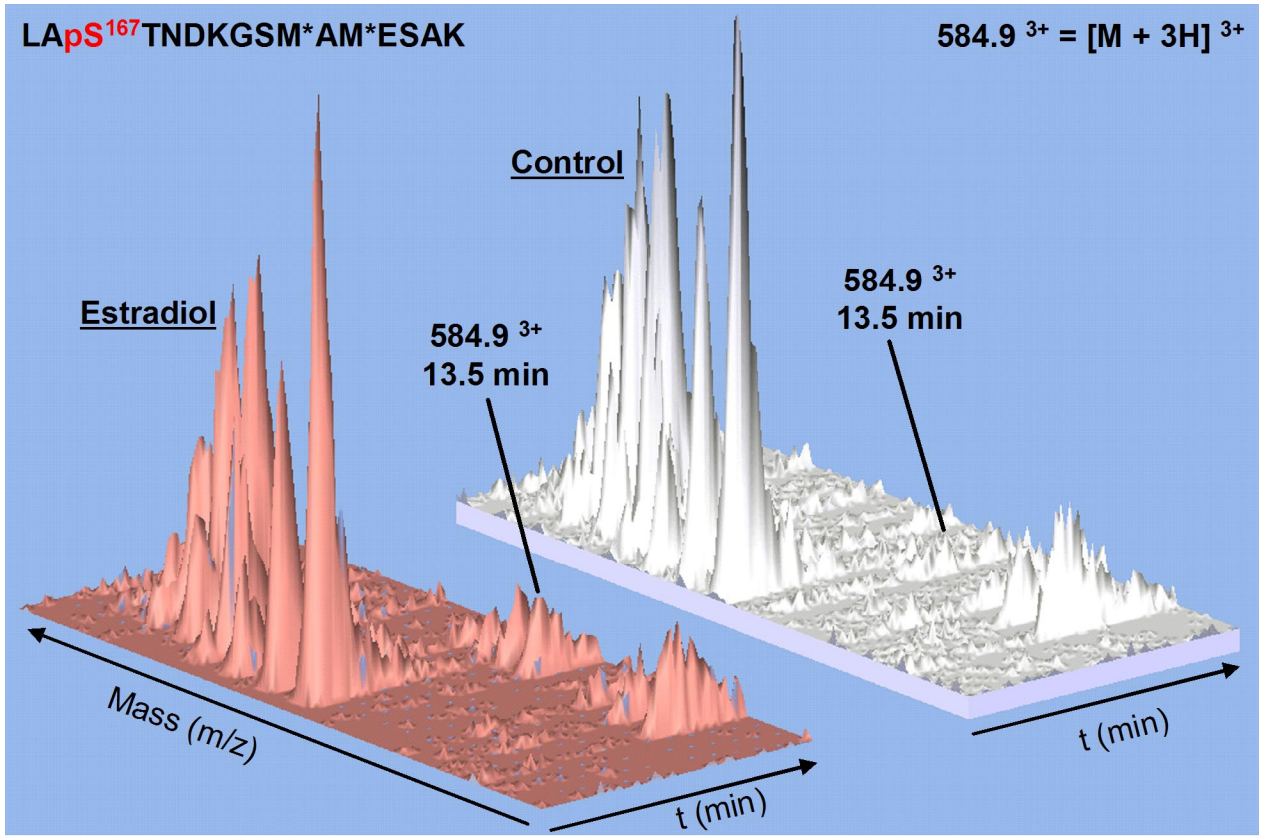
Locating Phos. in a peptide with 3 Serines

- A peptide containing Ser-167 was 112 Da higher in mass than predicted, based on residue masses alone.
- ESI-MSMS of the triply charged ion at m/z 584.9 identified phosphorylation (+80) and 2 oxidized methionines (+16 Da each). Note: Loss of 64 Da indicates Met-ox.
- The y9 ion confirmed that Ser-173 and Ser-178 could not be phosphorylated.
- The b3 ion lost phosphoric acid, giving a new ion 18 Da lower than b3 from the unmodified peptide, as did y14 and y15.

This experiment confirmed earlier reports that Ser-167 is phosphorylated.



Phosphorylation was stimulated by Estradiol



- In a semi-quantitative experiment, ER isolated from cells treated with estradiol was compared to ER from untreated control.
- Estradiol treatment gave increased ion current for the triply charged phosphorylated peptide *m/z* 584.9, eluting at 13.5 min.

Analysis of Protein Phosphorylation by Hypothesis-Driven Multiple-Stage Mass Spectrometry

E. J. Chang et al, *Anal. Chem*. 2004, 76, 4472-4483

A quote from the abstract:

- In this strategy, we postulate that any or all of the potential sites of phosphorylation in a given protein may be phosphorylated.
- Using this assumption, we calculate the *m/z* values of all the corresponding singly charged phosphopeptide ions that could, in theory, be produced by the enzyme employed for proteolysis.
- We test ions at these *m/z values for the presence of phosphoserine or phosphothreonine* residues using tandem mass spectrometry (MS2) in a vacuum MALDI ion trap mass spectrometer
- The neutral loss of the elements of H_3PO_4 (98 Da) provides a sensitive assay for the presence of phosphopeptides.
- Subsequent MS3 analysis of the $(M+H-98)^+$ peaks allows us to confirm or reject the hypotheses that the putative phosphopeptides are present in the sample.

Neutral loss scanning for Phos. In ER

- Earlier studies suggested Ser-154 might be a phosphorylation site.
- The unphosphorylated peptide EAGPPAFYRPNS^{154}DNRR (*m/z* 1847.0) was seen by MALDI but neither ESI-MS/MS nor vMALDI-MS showed ions for the phosphopeptide at +80 Da.
- "Hypothesis-driven" neutral loss scans (-98 Da) were used to interrogate all peptides with unmodified serine residues, to determine whether the "unmodified" molecular ion had a corresponding phospho analog which would lose phosphate, i.e. loss of 98 Da from an "invisible" species 80 Da higher.

Example: Peptide EAGPPAFYRPNSDNRR containing Ser-154 was identified by MSMS with its molecular peak at m/z 1847.0.

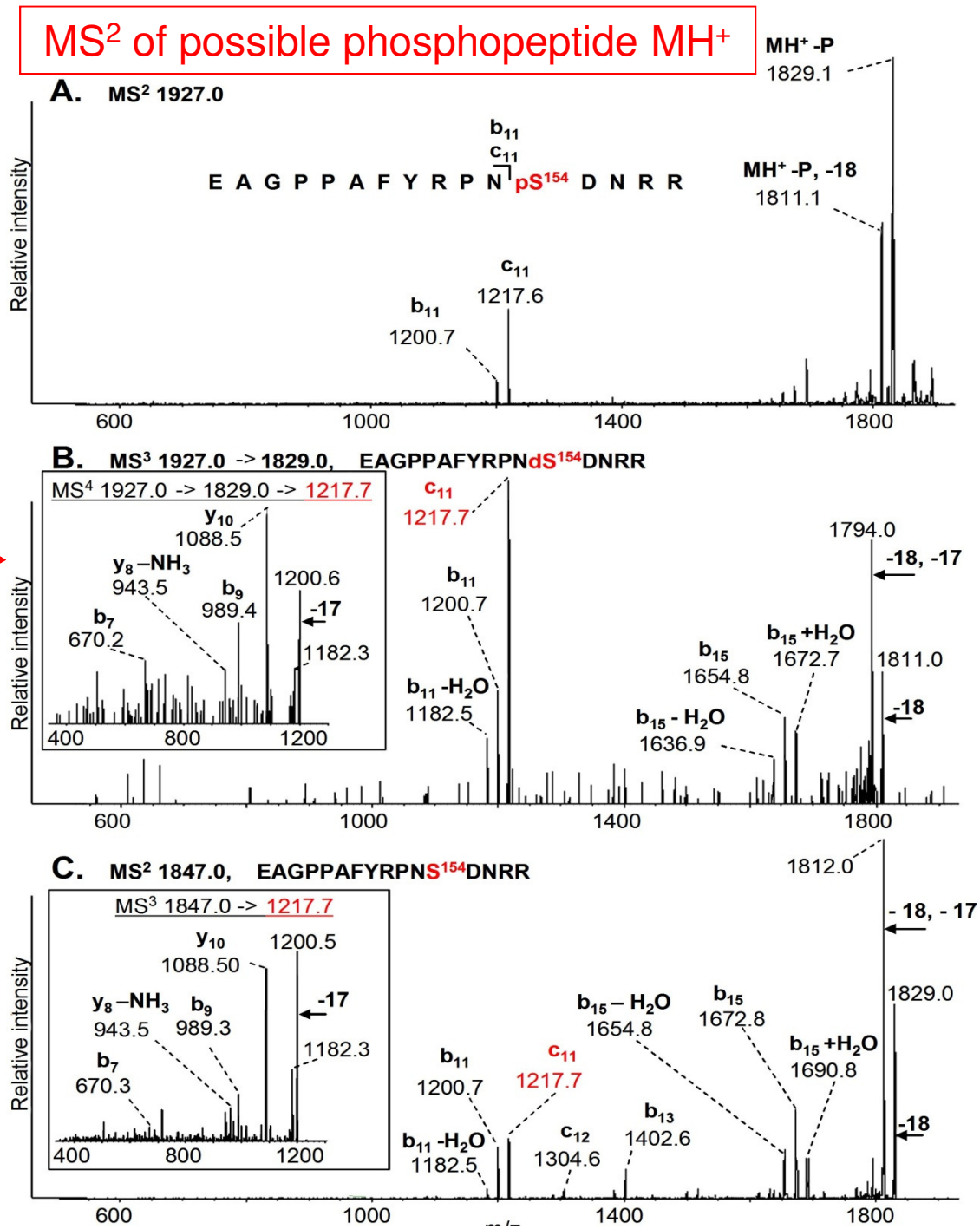
- MSMS was carried out by selecting precursor mass 1927.0, even though no peak could be seen at this mass.
- A fragment ion was observed at m/z 1829.0, i.e. loss of 98 Da.
- MS^3 confirmed Ser-154 as a phosphorylation site.

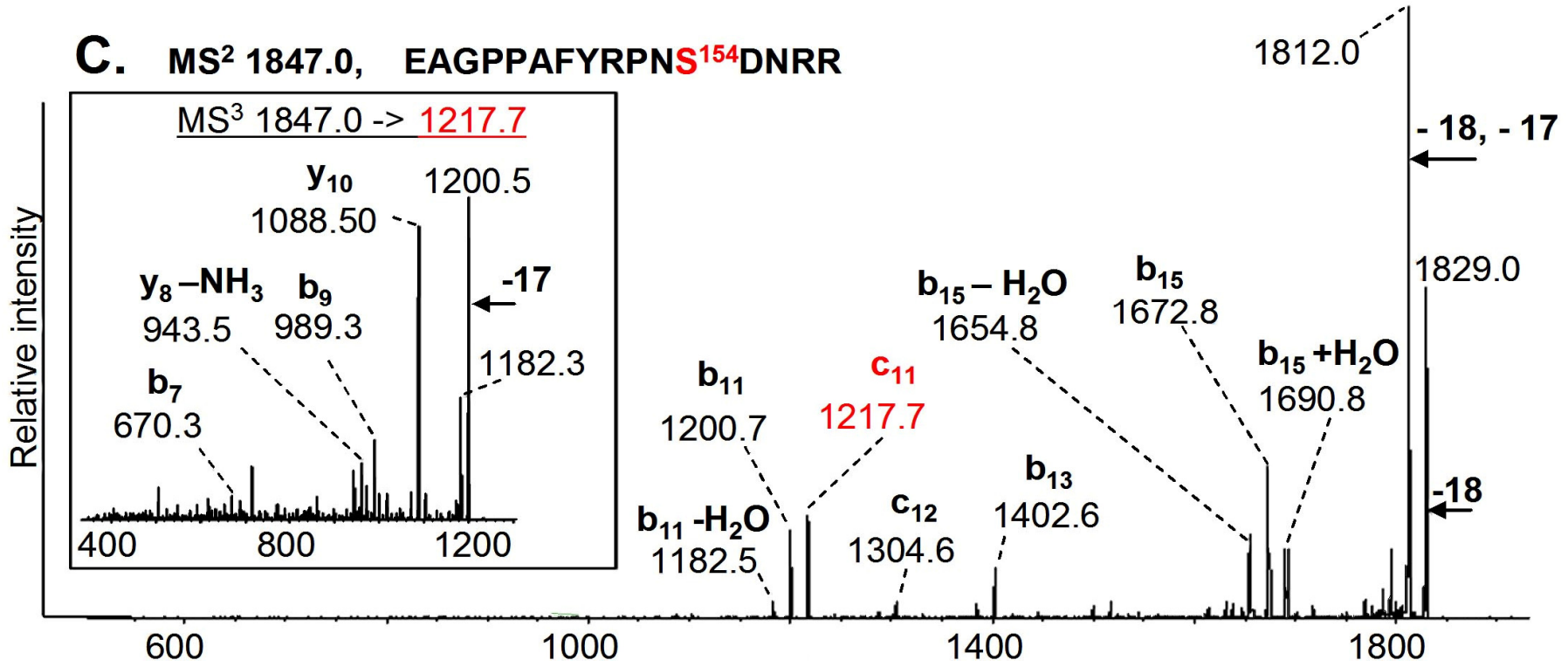
Multiple modes of tandem mass spectrometry using a vMALDI-LTQ, proved the existence of phosphorylation at Ser-154

MS^2 of possible phosphopeptide MH^+

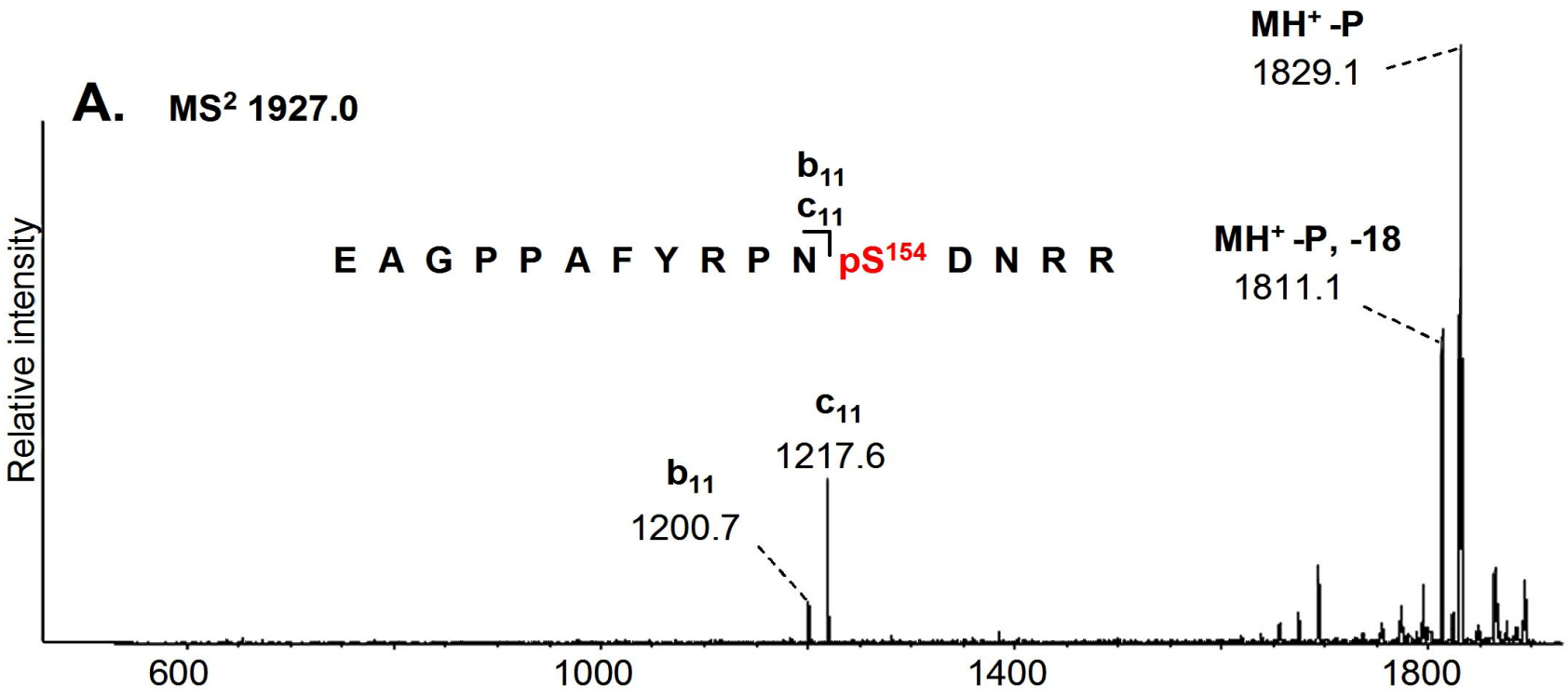
MS^3 after phosphoric acid loss ($MH^+ - H_3PO_4$)

MS^2 of unmodified peptide MH^+

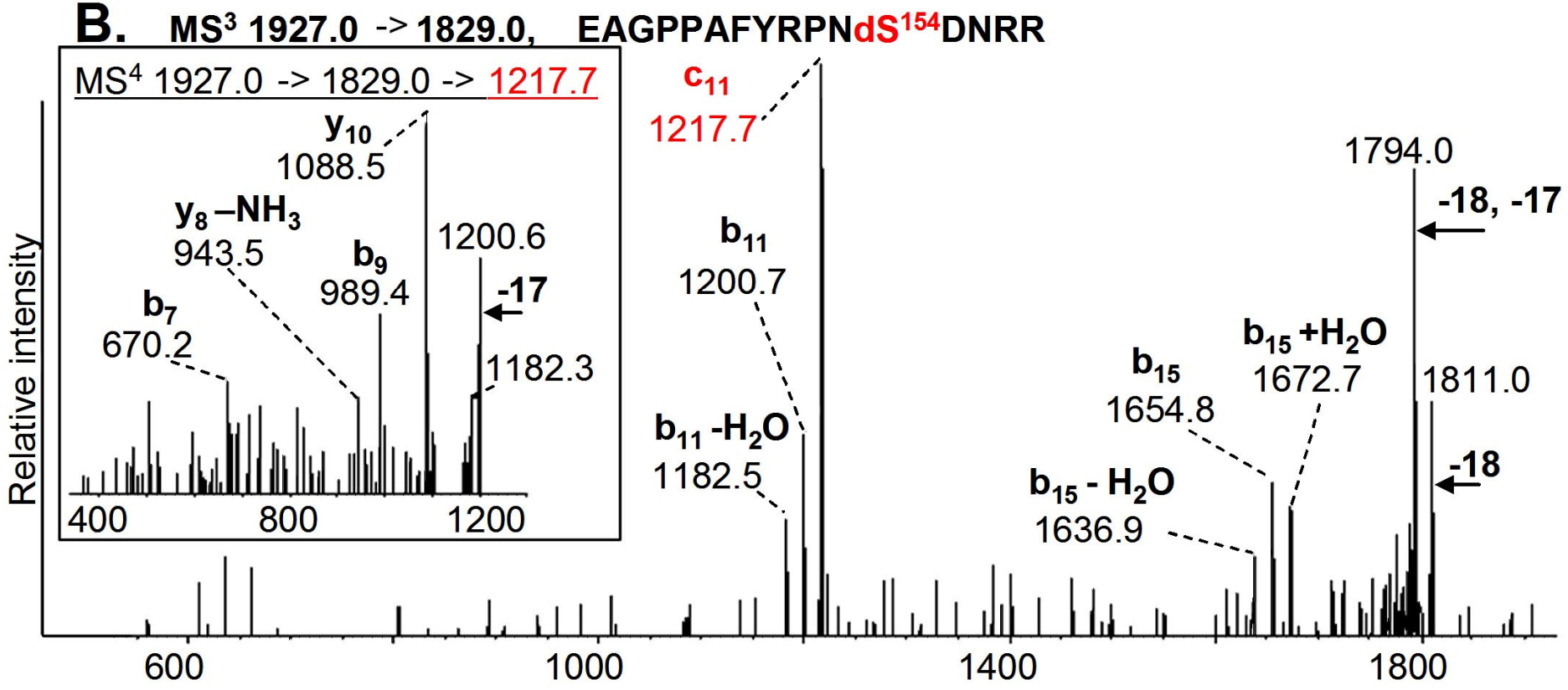




- MS/MS of the unmodified peptide at *m/z* 1847.0. The spectrum was stronger than that of the phosphopeptide as the stoichiometry was low for the modification.
- The c11 ion was fragmented (MS^3; see inset), giving a spectrum virtually identical to that for the M-Phos c11 from the modified peptide (see later slide).
- Note: MALDI gave only singly charged ions for each of the peptides whereas ESI of the same peptides gave multiply charged ions +3 and +4.

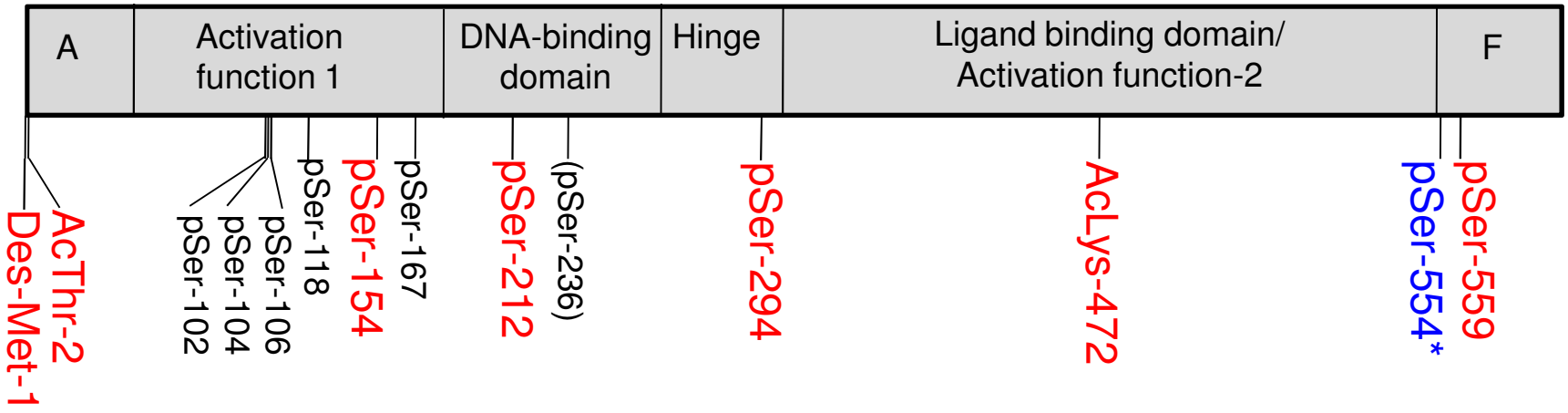


- MS/MS of the precursor ion of the phosphopeptide at *m/z* 1927.0.
- -P represents loss of phosphoric acid, (-98 Da), followed by further loss of water (-18 Da).
- The strong c-ion is relatively unusual but was found to be useful for diagnostic purposes. It proved that the modification was not in the 11-amino acid N-terminal region.



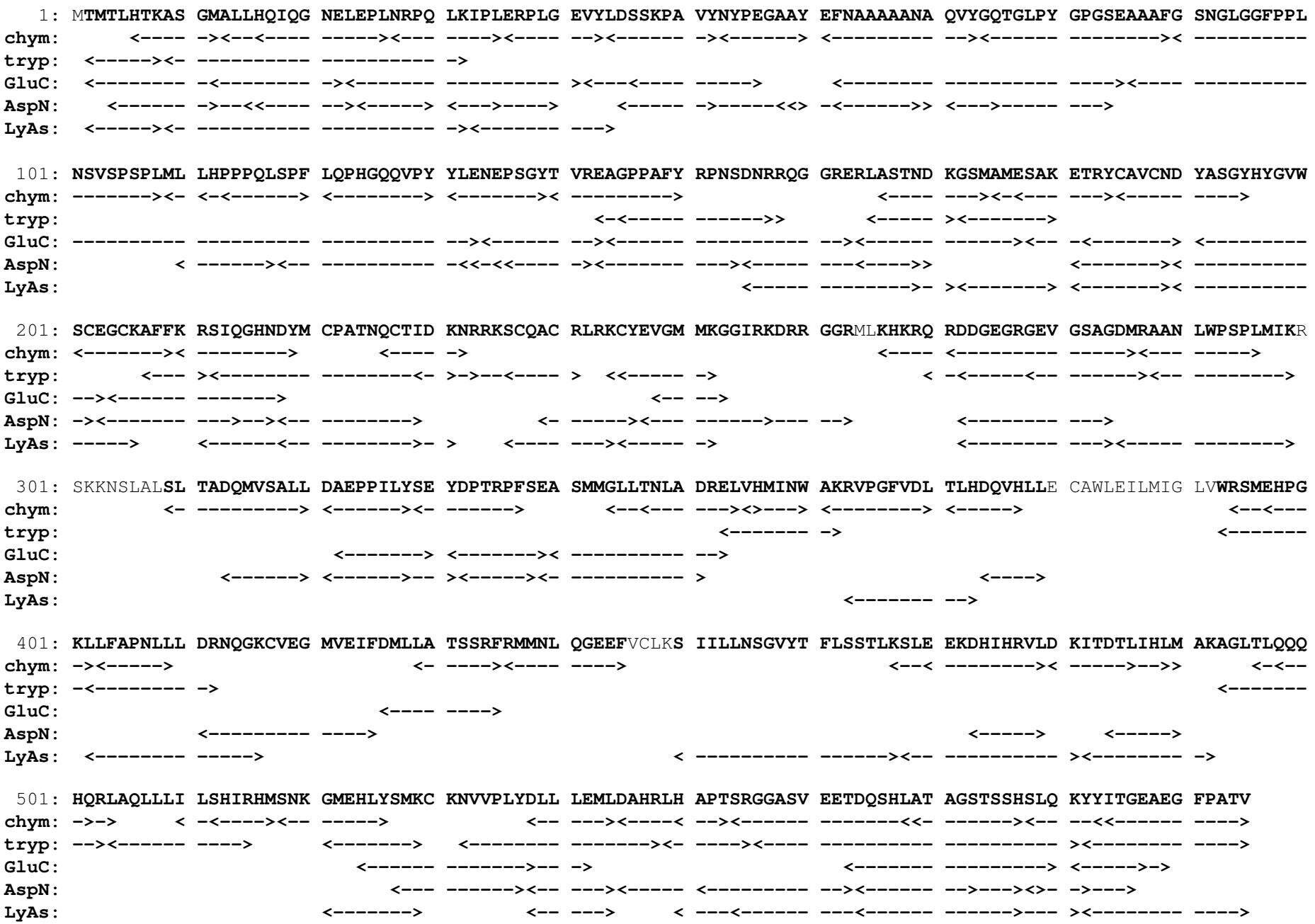
- MS^3 spectrum of *m/z* 1829.0 [MH-98]$^+$. The cluster of ions associated with b15 are all 18 Da less than for the unmodified peptide as Ser-154 was converted to dehydro-alanine by the loss of phosphoric acid.
- The c11 fragment ion (1217.7 *m/z*) was fragmented for additional sequence information (MS^4; see inset). It gave a spectrum virtually identical to the unmodified peptide.

Summary of ER-PTM's found in this study



Red: Novel finding in endogenous ER
Blue: Novel finding in recombinant ER
Black: Confirmation of previous report

94% sequence coverage did not include all Ser residues



Are all these findings unambiguously correct?

Not Really!

- Sulfate adds the same nominal mass as phosphate (80)
- Trimethyl adds the same nominal mass as acetyl (42)

But

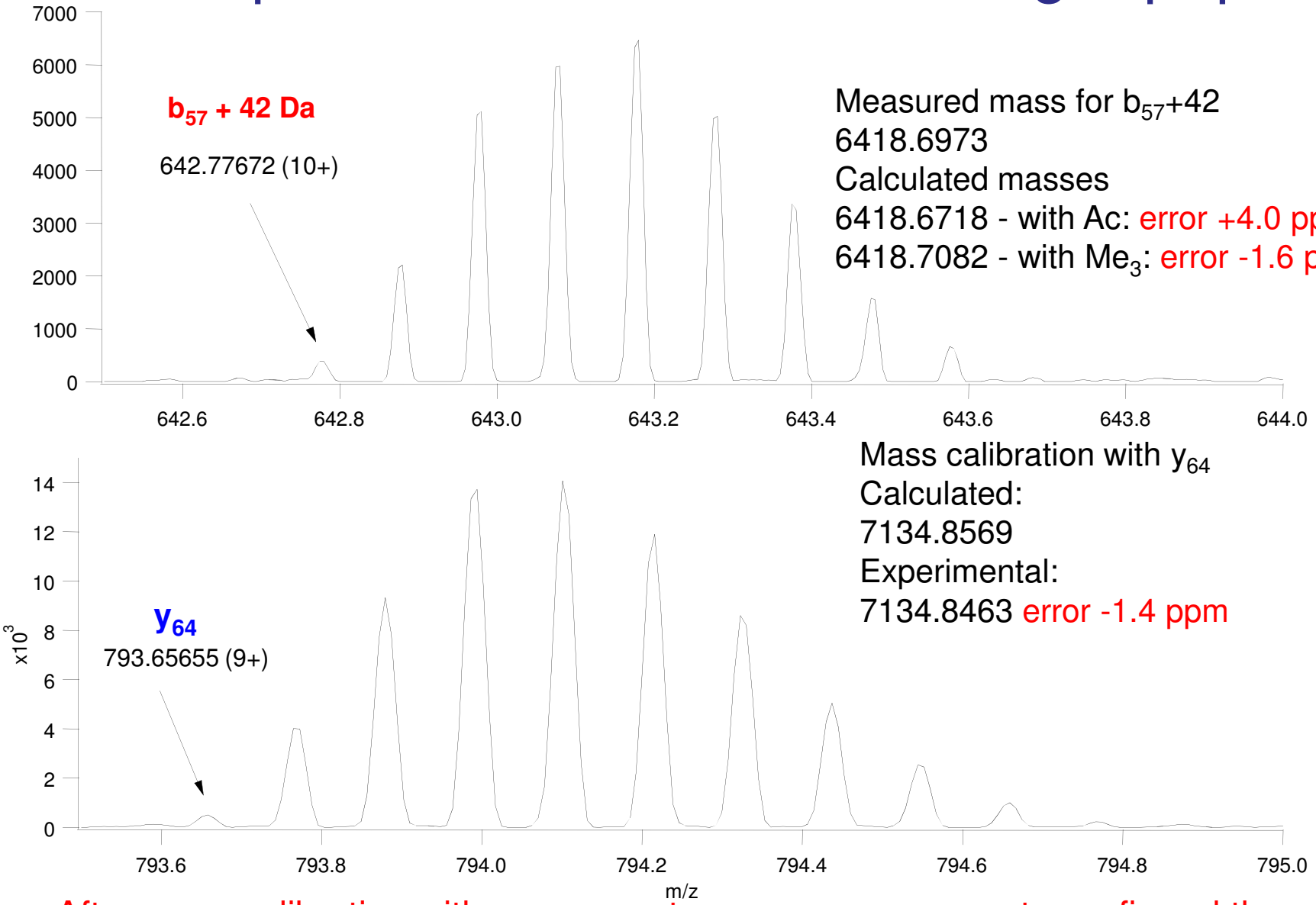
- The biology may give strong reasons for believing in one PTM rather than its alternative of the same nominal mass.
- Other MS characteristics may differentiate PTM's.
- Accurate mass measurements may differentiate PTM's.

Methylation can be mono (+14), di (28) or tri (42)

- Methylation can occur at the protein N-terminus or at the side chains of Lys, Arg or His.
- Trimethyl / acetyl mass difference = 0.036 Da, which can be differentiated by accurate mass measurement.
- For the modified peptide SLEEKDHIHR of nominal mass 1305, the mass difference is 28 ppm. Measuring masses to within 10 ppm or better is well within the capabilities of many modern instruments.

Note: For a fixed mass difference such as between Acetyl and trimethyl, the ppm difference is inversely proportional to the mass of the peptide, thus larger peptides have a smaller ppm mass difference.

An example from FTICR-MS of a larger peptide



After mass calibration with y_{64}, accurate mass measurements confirmed the modification was trimethylation, not acetylation.

+80 Da: phosphate or sulfate?

Phosphorylation

- Ser and Thr lose H_3PO_4 (98 Da) in CID
- Tyr does not lose phosphate in CID

Sulfation

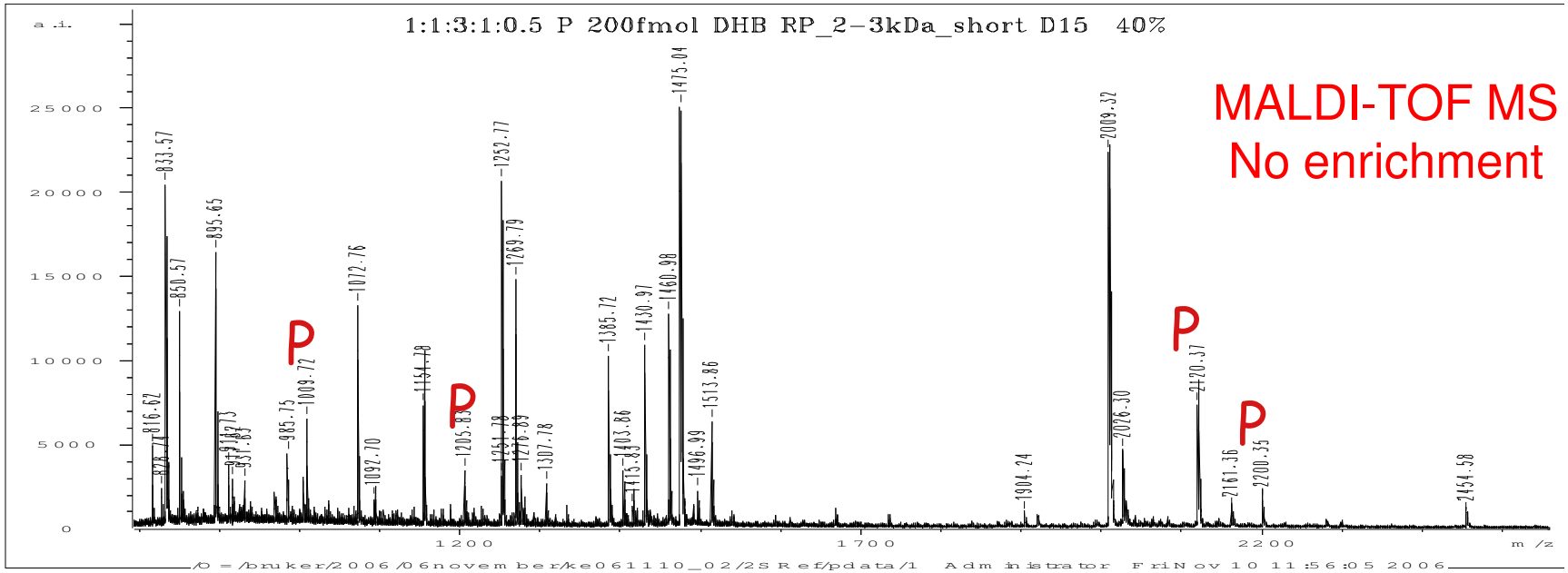
- Tyr only
- Significant SO_3 (80 Da) loss even in MS

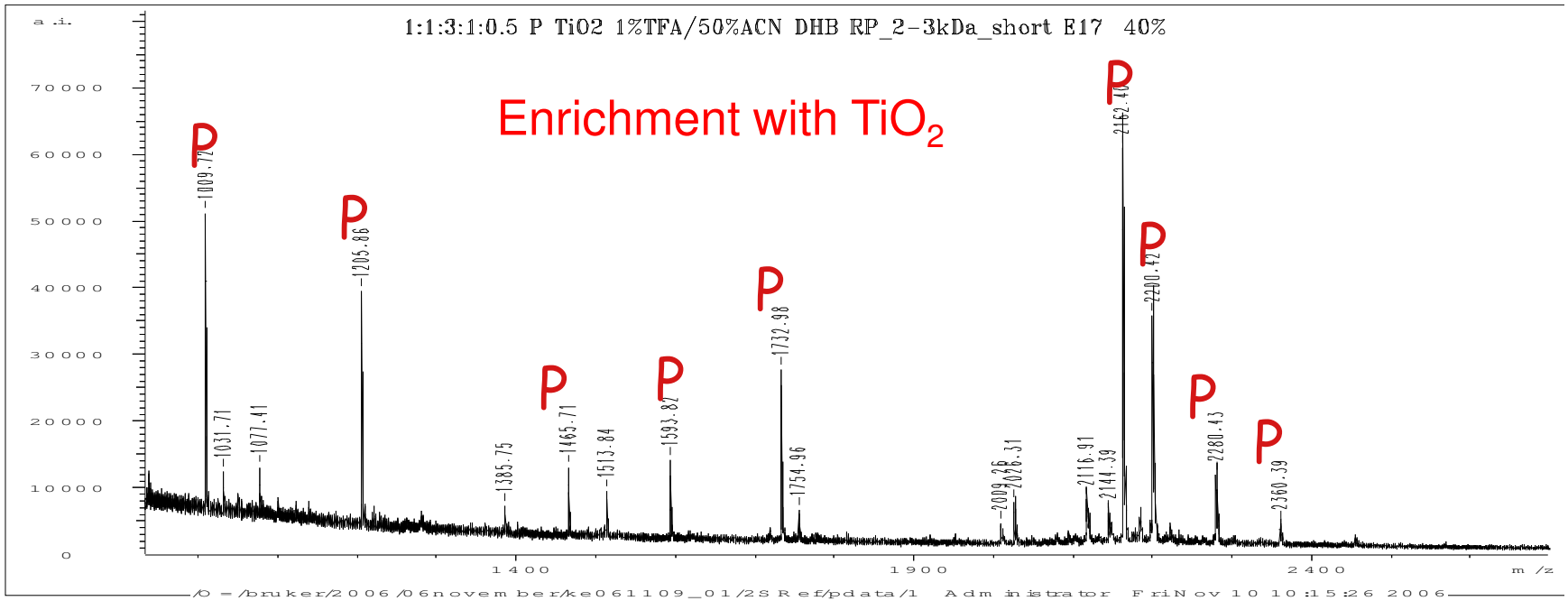
Sulfopeptides may be misidentified as phosphopeptides!

- The mass difference is only 0.009 Da, i.e. 10 ppm at mass 900
- They show “identical” behavior by ESIMS, chromatography and under basic conditions.
- But they have different CID fragmentation

Phosphopeptide / Phosphoprotein Enrichment

- Ion exchange on SCX
- IMAC : Fe(3+), Ga(3+)…
  - binding at low pH
  - methyl-esterification prior to IMAC
- TiO_2, ZrO_2
- Immunoprecipitation (only for pTyr)


1:1:3:1:0.5 P 200fmol DHB RP_2-3kDa_short D15 40%
MALDI-TOF MS
No enrichment
P
P
P
P
833.57
895.65
850.57
816.62
1009.72
1072.76
1092.70
1205.83
1252.77
1269.79
1307.78
1385.72
1430.97
1460.98
1475.04
1513.86
1904.24
2009.32
2026.30
2120.37
2161.36
2200.35
2454.58
1200
1700
2200
m/z

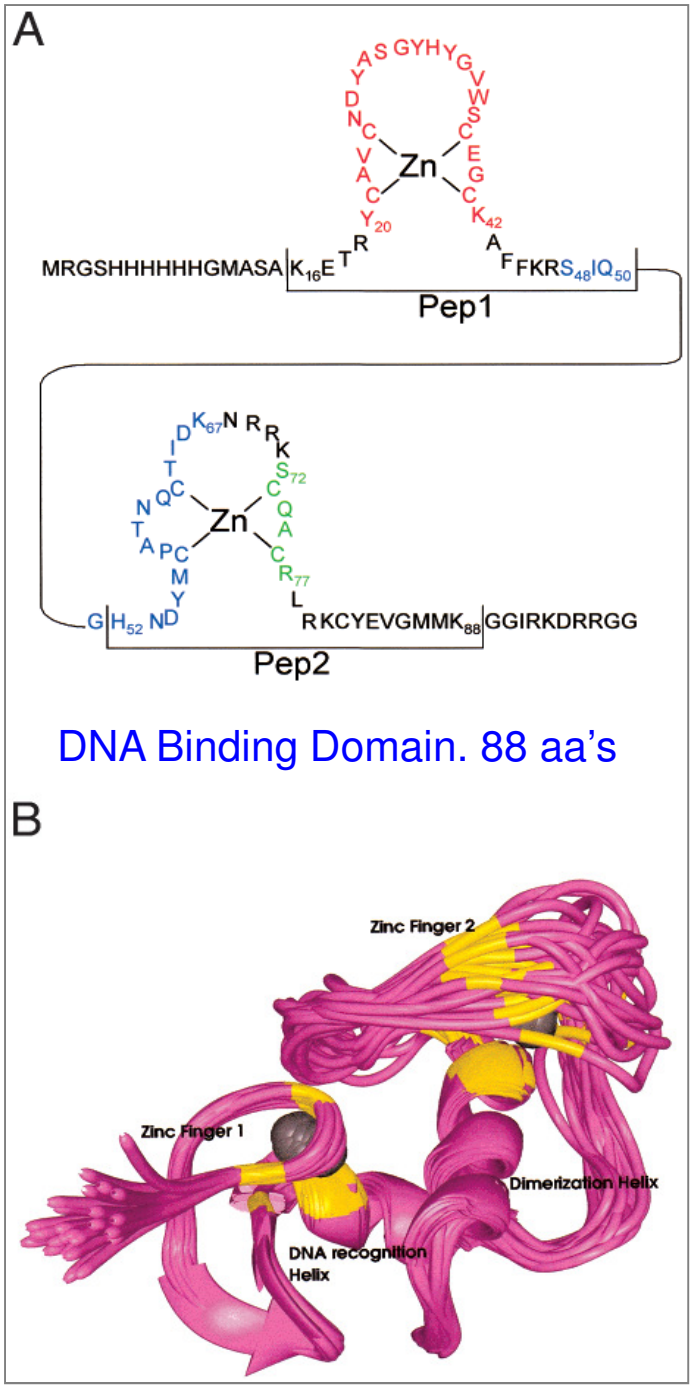

1:1:3:1:0.5 P TiO2 1%TFA/50%ACN DHB RP_2-3kDa_short E17 40%
Enrichment with TiO2
P
P
P
P
P
P
P
1009.77
1031.71
1077.41
1205.86
1385.75
1465.71
1513.84
1593.82
1732.98
1754.96
2009.26
2026.31
2116.91
2144.39
2162.40
2220.47
2280.43
2360.39
1400
1900
2400
m/z

Confirming phosphorylation and locating sites

- Western blotting: pTyr is large enough for sequence independent recognition; generally works well. pSer, pThr are not reliable.
- Dyes have questionable reliability.
- Phosphatase treatment + isoelectric focusing show a pI shift.
- *In vitro/in vivo* assays with radioactive phosphate.
- Edman microsequencing with radioactive P.
- Mutation studies.
- Mass spectrometry
  - MS spectrum should show 80 Da shift
  - MSMS fragmentation:
    - pSer, pThr give H_3PO_4 loss: –98 Da
    - pTyr invariably retains the modification but it gives a characteristic immonium ion at m/z 216.

Determining disulfide bridges

- ER is a transcription factor that binds to DNA.
- The DNA-binding domain contains two zinc fingers, each having 4 cysteine residues. These are responsible for ER dimerization.
- In oxidant stressed protein these cysteines may form disulfide bonds, losing the ability to bind zinc and preventing dimerization.
- This would impair the ability of ER to bind to DNA and prevent transcription.



Are cysteine residues oxidized or reduced?

1. Expose ER to various oxidizing conditions over a range of concentrations to create some disulfide bonds.
2. Derivatize remaining thiol groups on free cysteine residues with iodoacetic acid.
3. Reduce disulfide bonds then derivatize newly generated thiol groups with an isotopic analog [$^{13}C_2$]-bromoacetic acid.

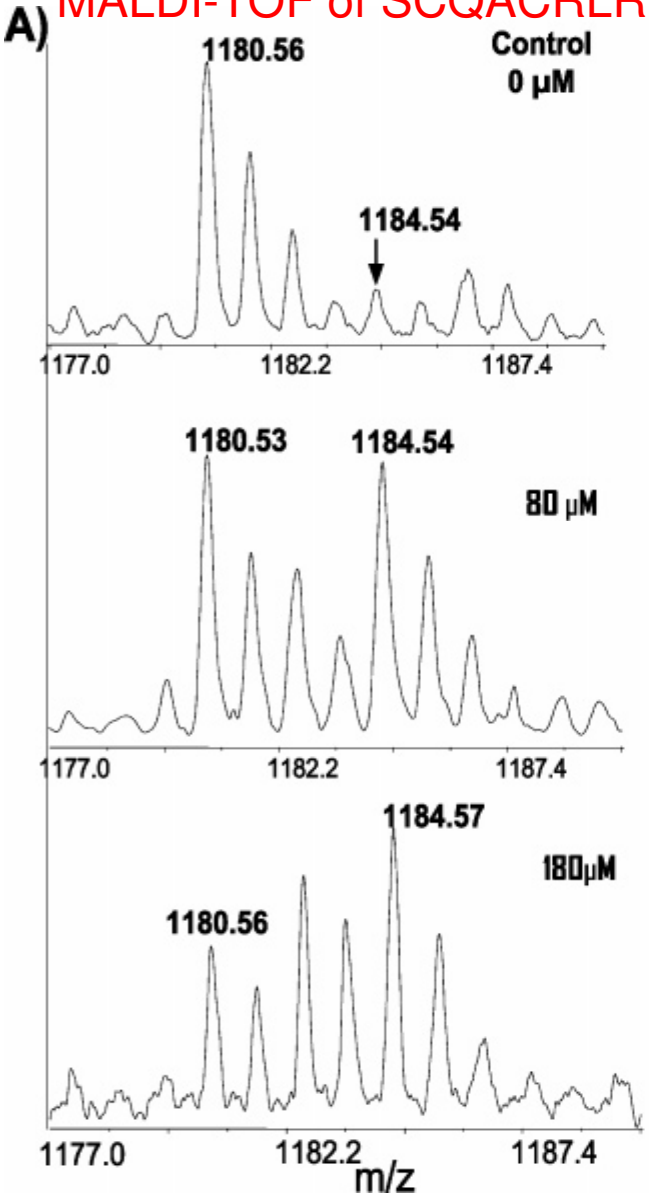
$I^{12}CH_2{}^{12}COOH$	$Br^{13}CH_2{}^{13}COOH$
Iodoacetic acid	[$^{13}C_2$]-Bromoacetic acid
+58 Da per SH	+60 Da per SH

4. Digest the protein with Lys C / Asp N to give peptides that contain pairs of Cys residues and analyze by MS and MSMS.
5. Peptides should display clusters of peaks separated by 0, 2 or 4 Da corresponding to initial states of 2 free thiols,1 free thiol or no free thiols.
6. Determine the proportion of free to disulfide-bound cysteines as a function of oxidant concentration.
7. Apply protocol to ER extracted from breast cancer tissue.

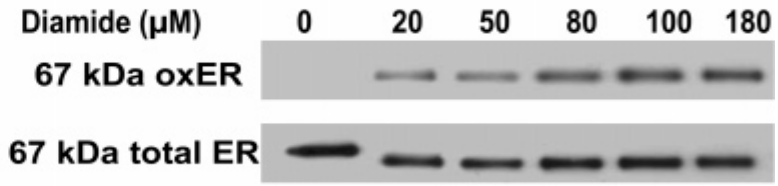
Atsriku et al, *Anal. Chem.*, 2007, 79, 3083-3090.

Titration of recombinant ER with diamide

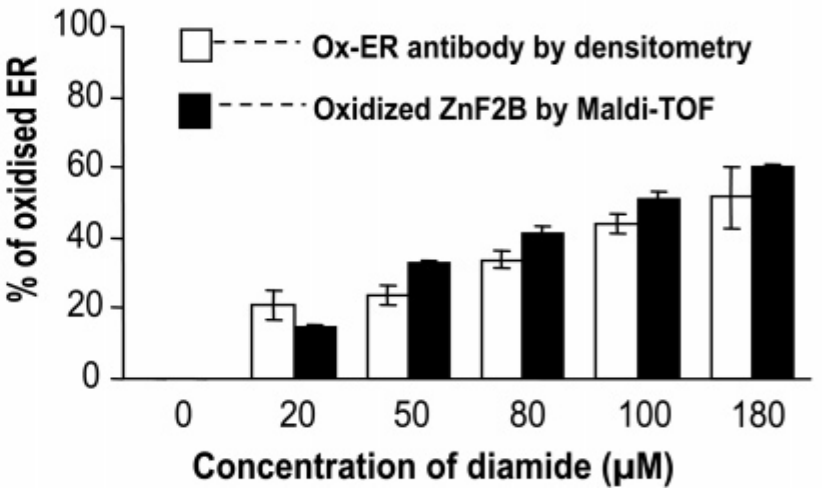
A) MALDI-TOF of SCQACRLRK



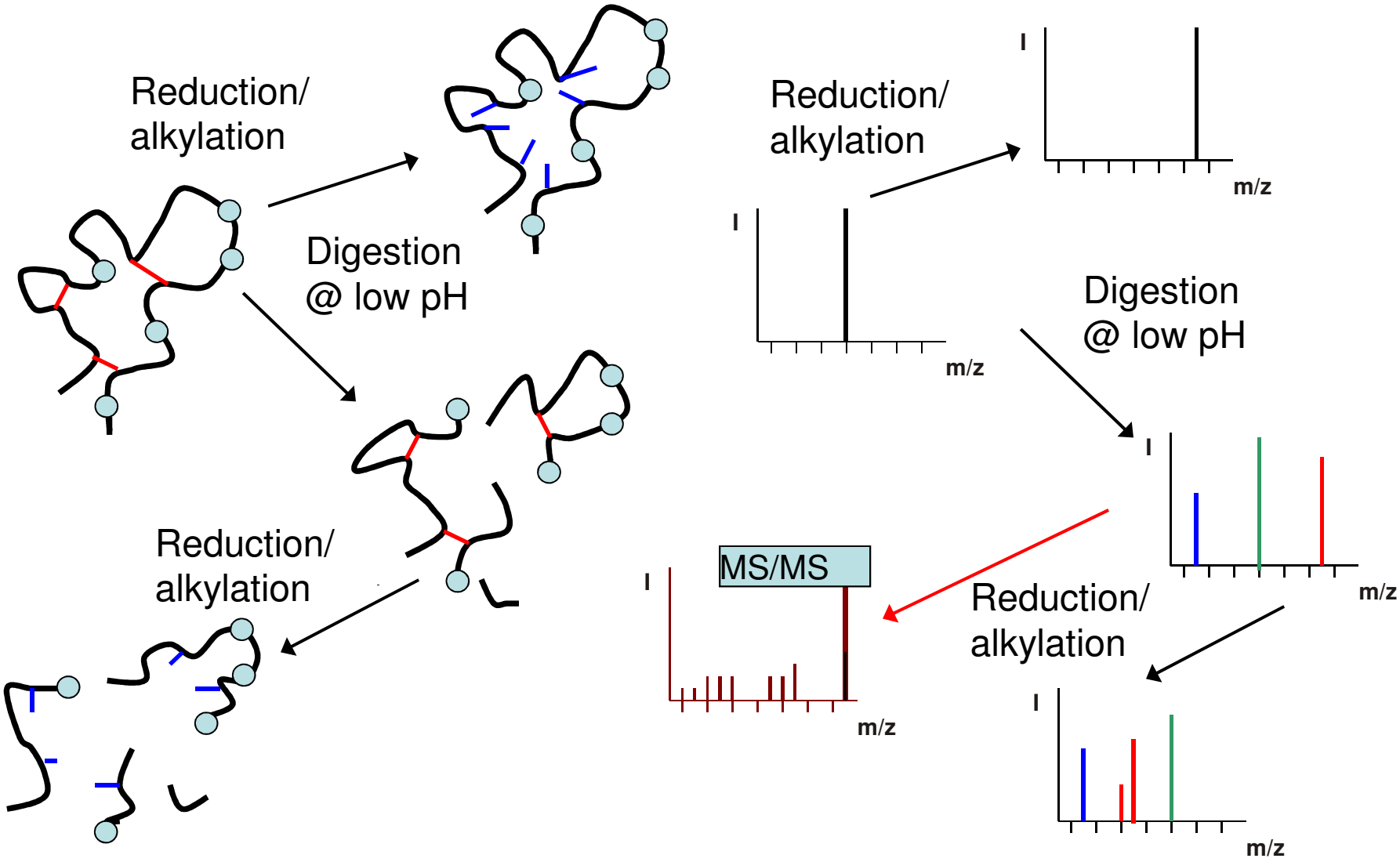
B) Western blot analysis



C) Comparison of MS and Western blot

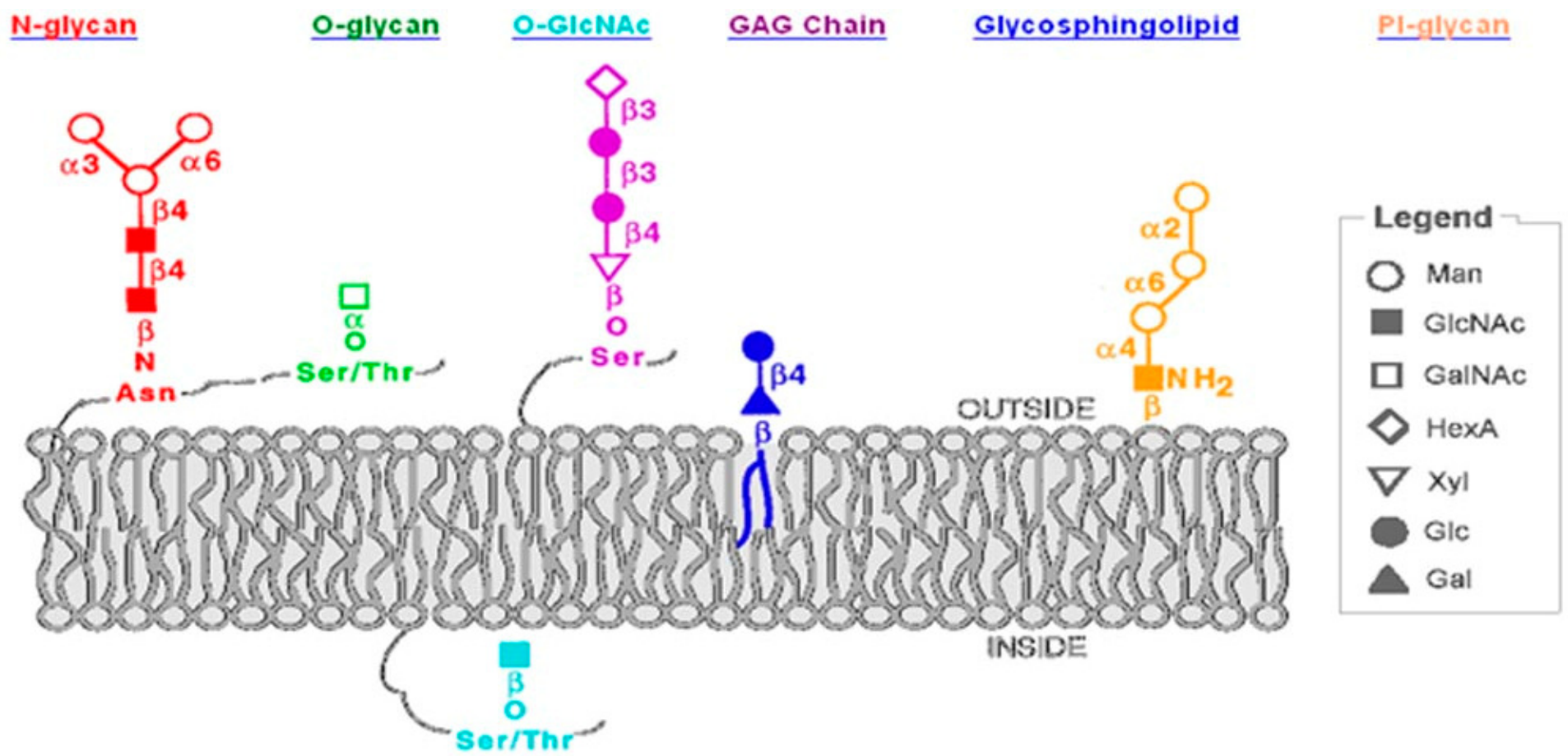


Assigning disulfide-bridges



Protein Prospector has a mass modification search which can be used in conjunction with MS-Bridge to find peptides linked together by disulfide bridges.

Protein Glycosylation

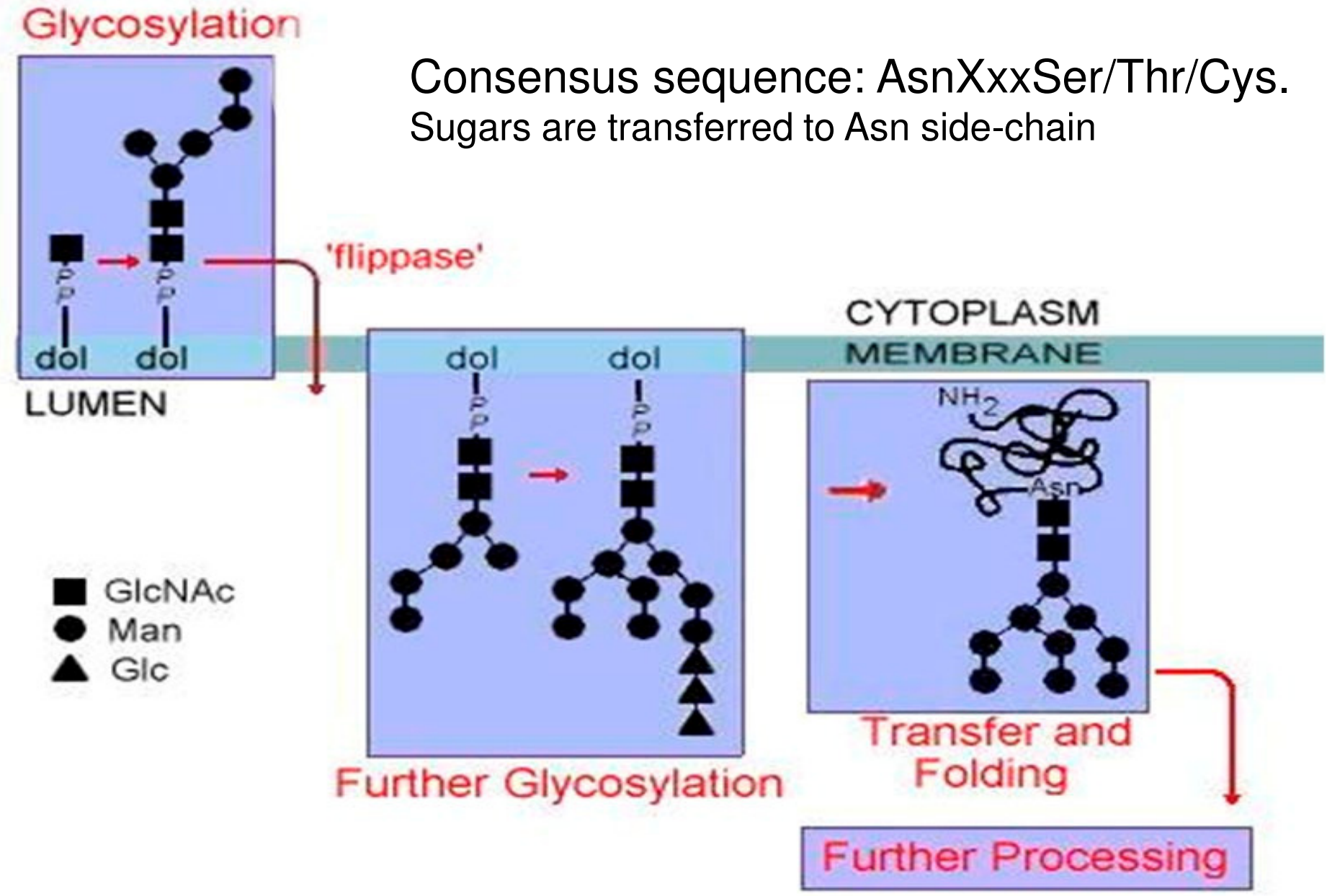


Reference: Essentials of Glycobiology Varki et al, 1999.

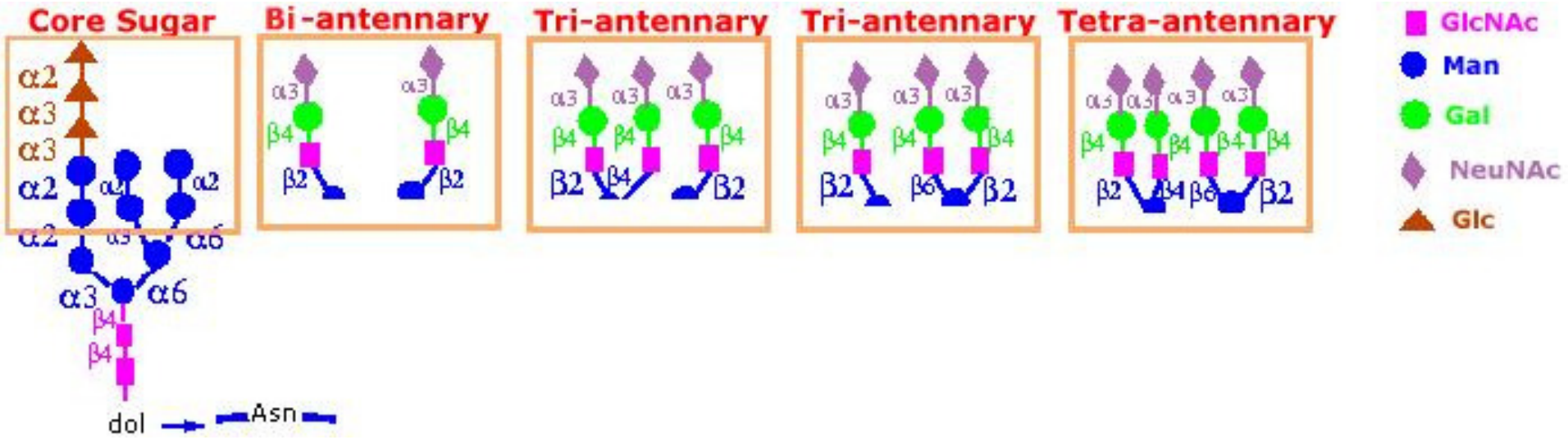
N-linked sugars

Consensus sequence: AsnXxxSer/Thr/Cys.

Sugars are transferred to Asn side-chain



Further processing



Aspects of N-linked glycosylation

- Extreme heterogeneity is common: a site may be only partially occupied and may display numerous different carbohydrates.

- Species-, tissue-, cell-specific modifications, physiological changes and disease may alter the sugars.

- The basic core is $GlcNAc_2Man_3$ but there may be more complex sugars: GlcNAc-Gal–sialic acid antennae, hybrid structures, core fucosylation, sulfate, phosphate modifications.

- Treatment with PNGase F removes all N-linked structures and converts Asn $\rightarrow$ Asp. Works for proteins and peptides. Allows separate analysis of peptides and oligosaccharides.

- Because of the high incidence of isomeric structures, e.g. glucose, mannose and galactose, all of which are hexoses and each of which can be linked differently, the identity of the sugar units and their linkage positions CANNOT be determined by MS alone. NMR, exo- and endoglycosidases are needed for complete analysis.

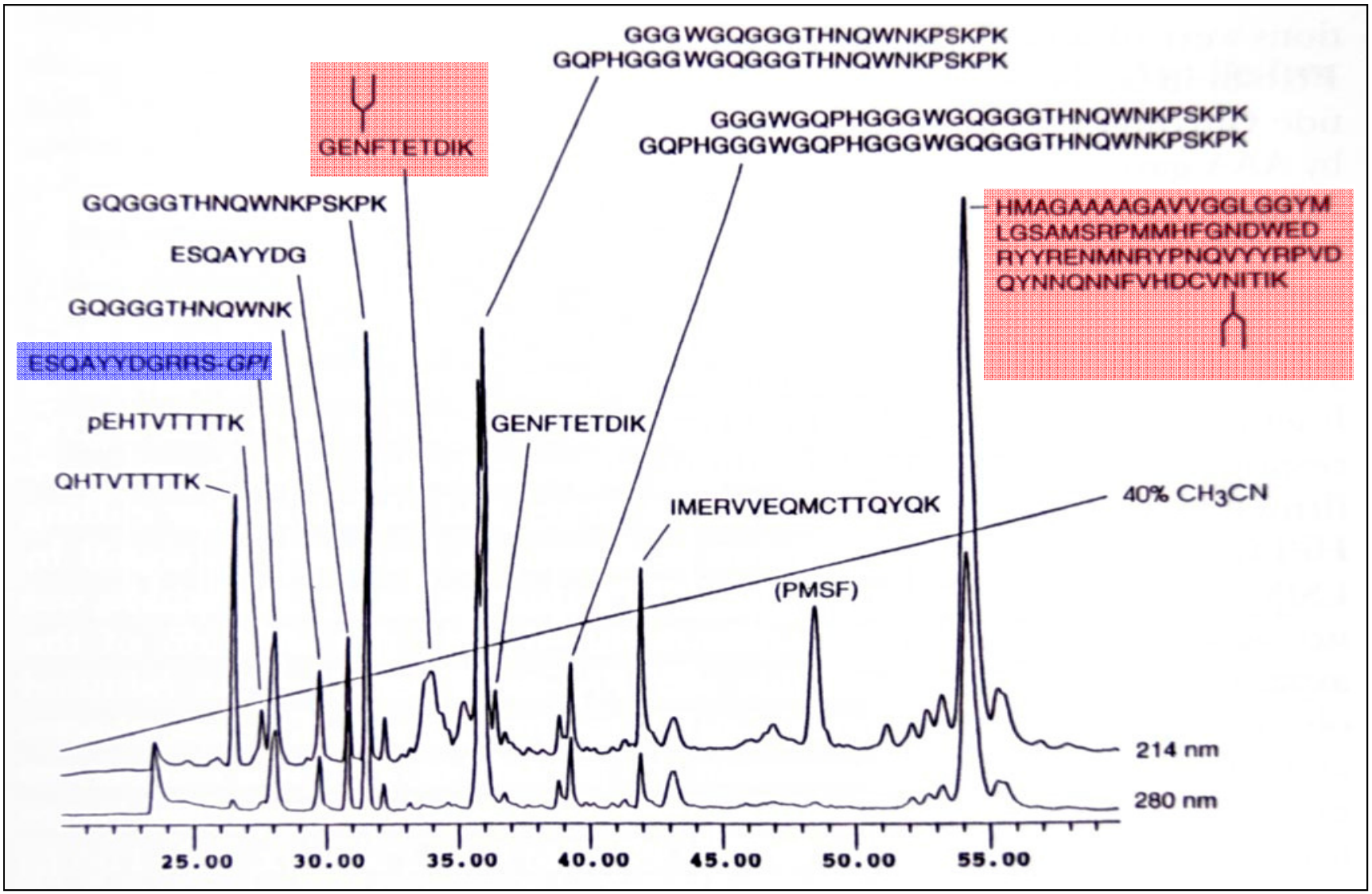
Example: the Prion Protein PrP

- PrP is a normal protein expressed in brain that can be refolded to form a pathogenic isoform. Responsible for transmissible, fatal, neurodegenerative diseases.
- Identified by cloning as having a 254 amino acid open reading frame.
- The pathogenic isoform is protease resistant: an N-terminally truncated form was isolated from brains of hamsters infected with scrapie.
- Digestion with Lys-C, HPLC separation and off-line analysis of collected fractions confirmed the majority of the amino acid sequence.
- Some peptides with modifications required further analysis.

Question: PrP exists in two forms, the insoluble pathogenic form (PrP^{Sc}) that forms plaques in the brain, and the normal soluble form (PrP^{C}). Could any modifications be identified that explained the different behavior?

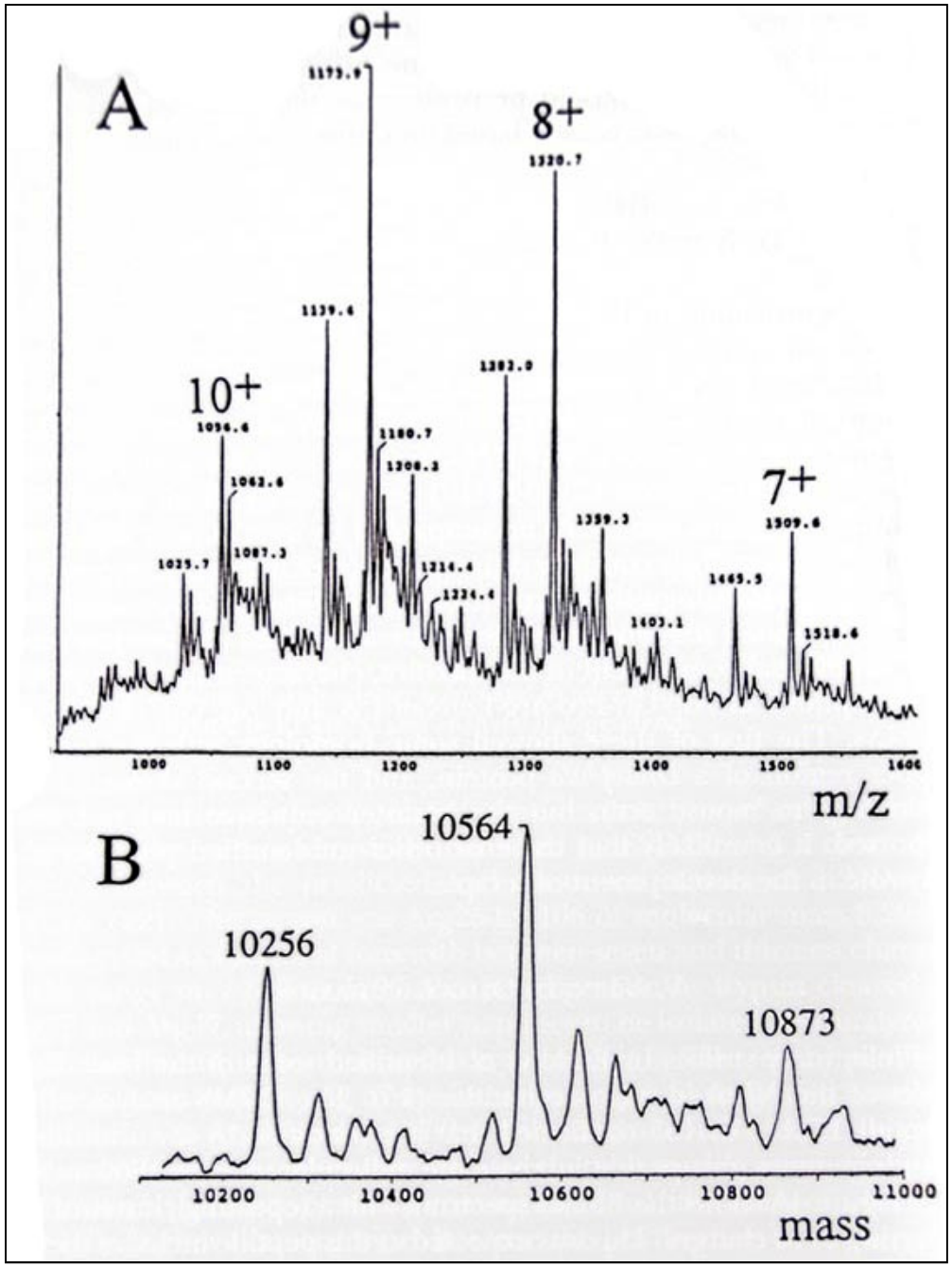
HPLC of a Lys-C Digest of PrP^{Sc}

Peptides were identified by off-line MS and MSMS. No mutations were found.

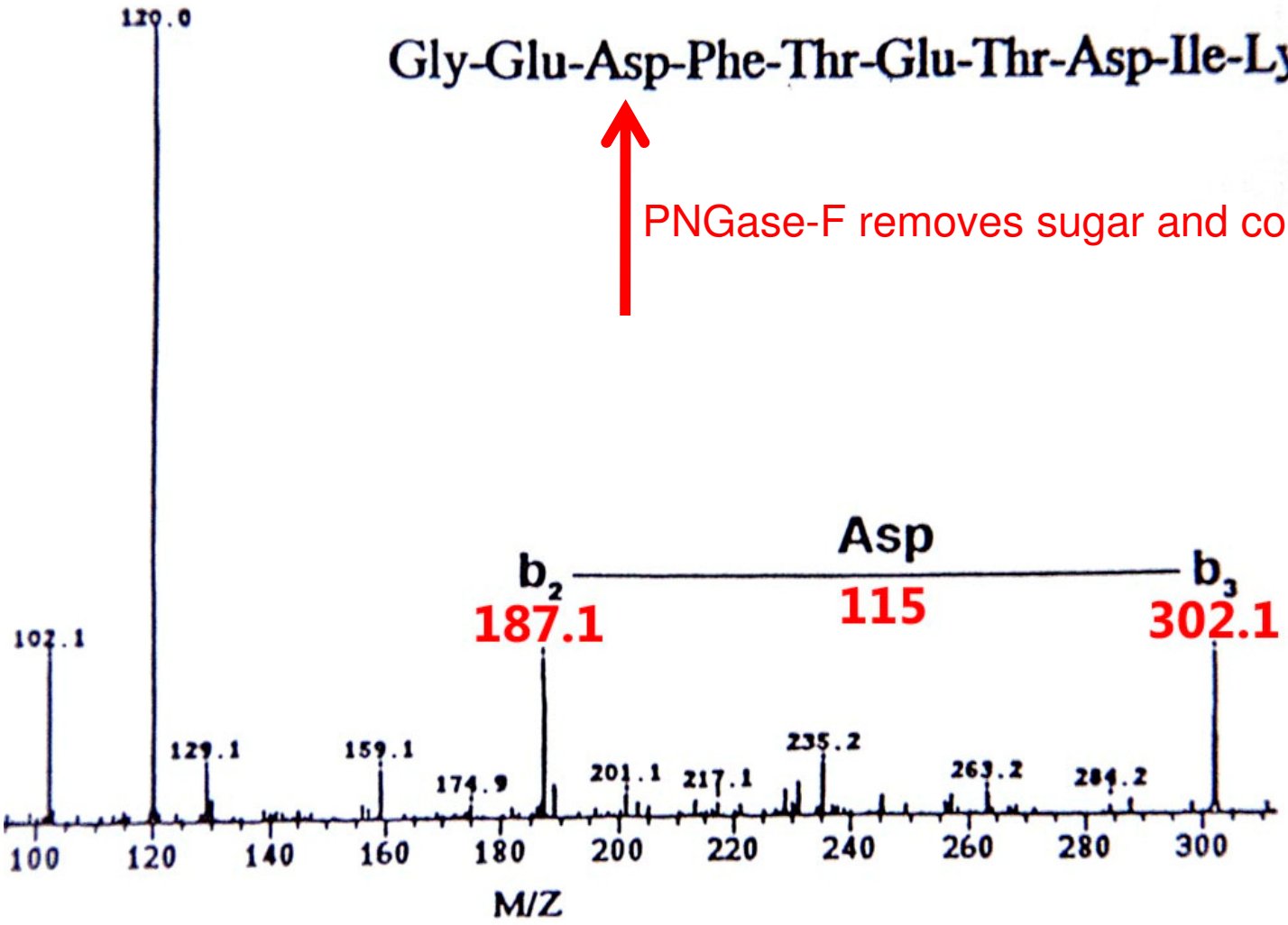


Glycopeptide 111-185 ESIMS

- The original spectrum shows repeating patterns of multiply charged ions +7 through +10.
- A “deconvoluted” spectrum shows ion masses converted to uncharged molecular masses.
- In a separate experiment the glycan was removed with PNGase F showing the peptide had an N-linked sugar. The peptide mass was measured as 8608.6, which was 1 Da high as Asn-181 was converted to Asp.
- The sugars were permethylated and analyzed independently



MSMS of peptide 195-204 confirmed Asn-197 to be the second glycosylation site



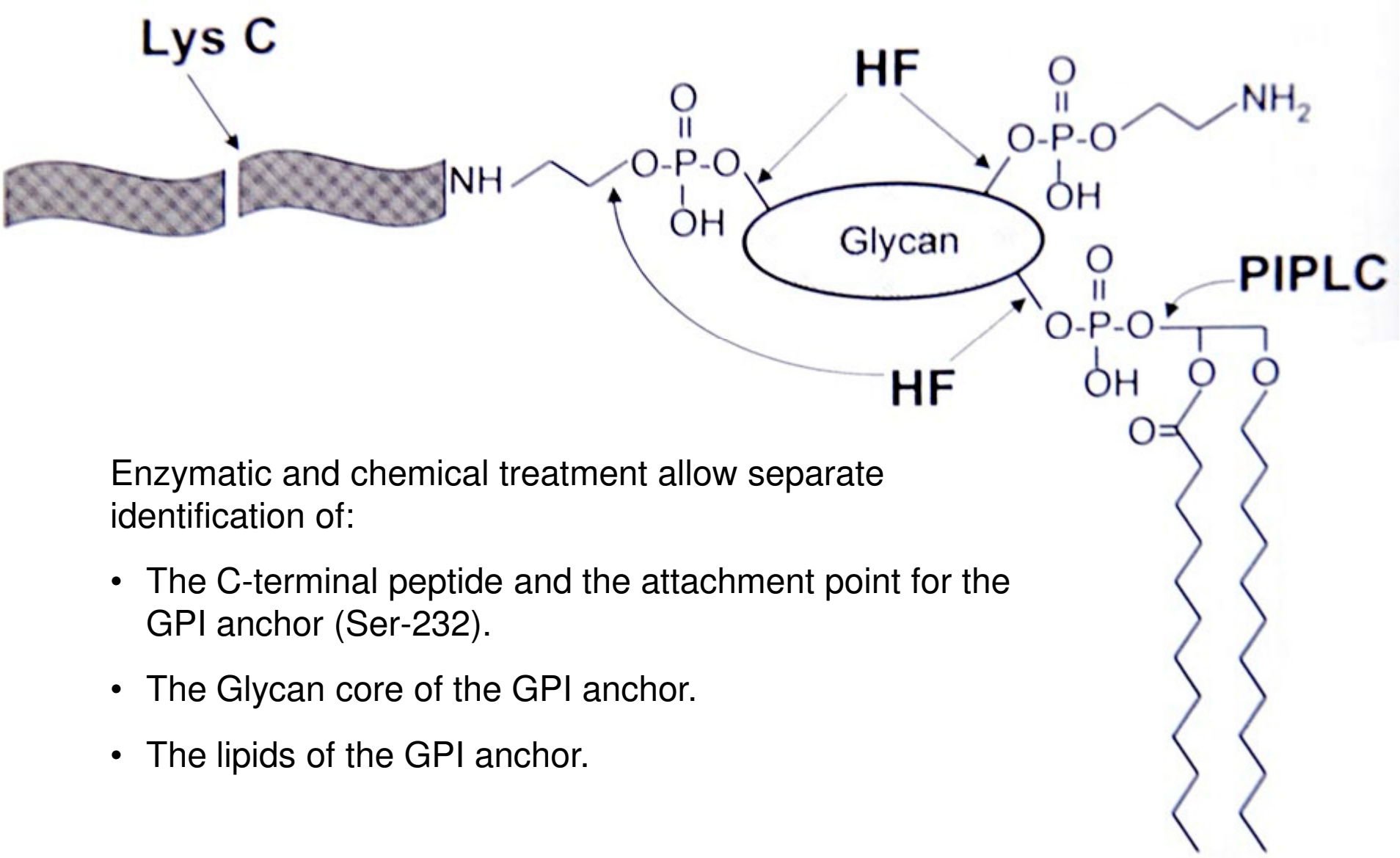
N-linked glycosylation

- The action of PNGase-F confirmed the presence of N-linked sugars
- The sites of glycosylation were identified by MSMS as 181 and 197
- The heterogeneous sugars were permethylated and analyzed independently for both sites by MSMS.

An additional modification

- The intact protein was lipid soluble and amino acid analysis identified ethanolamine, suggesting a glycosylphosphatidylinositol anchor (GPI)
- The action of PIPLC released the lipids and changed the solubility of the protein.

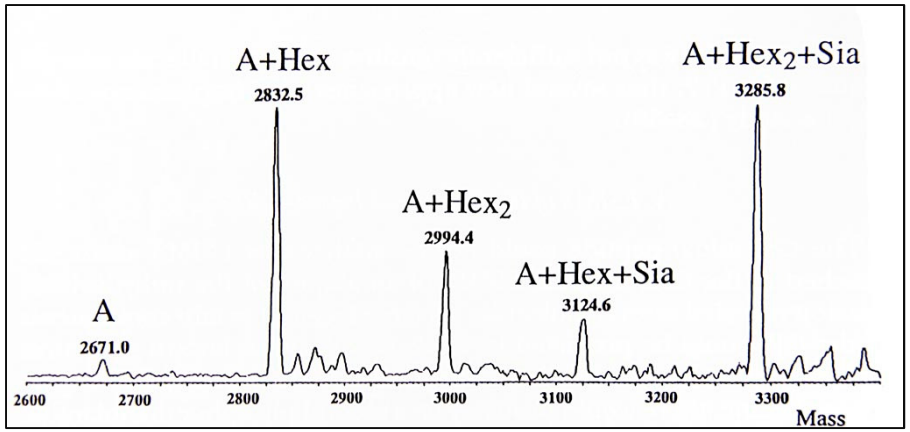
The C-terminus is truncated and modified with a glycosylphosphatidylinositol anchor (GPI)



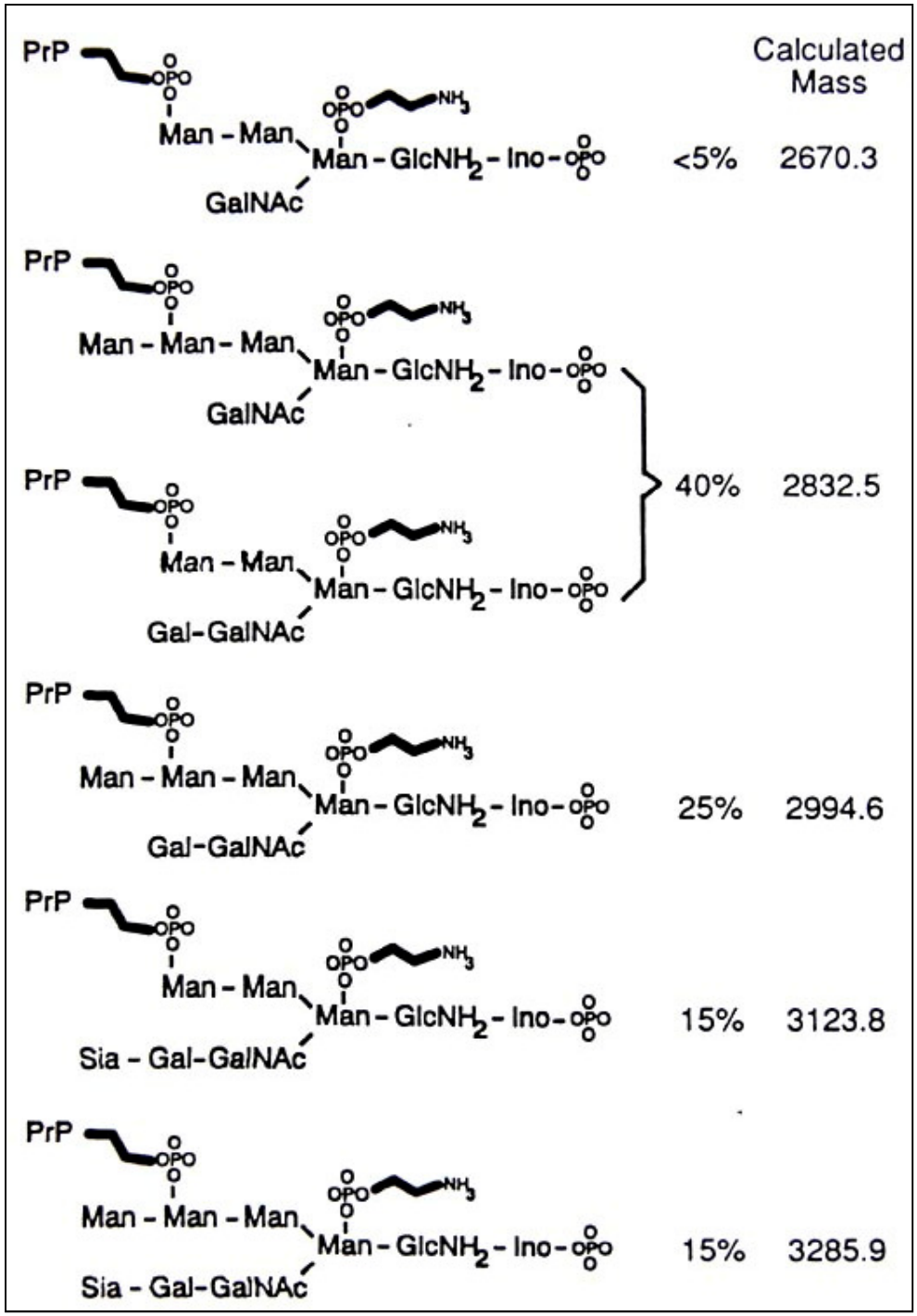
Enzymatic and chemical treatment allow separate identification of:

- The C-terminal peptide and the attachment point for the GPI anchor (Ser-232).
- The Glycan core of the GPI anchor.
- The lipids of the GPI anchor.

GPI structures



- ESIMS of the C-terminal peptide was obtained after digestion with Lys-C and PIPLC.
- The glycan core proved to be heterogeneous. MSMS analysis identified six related structures.
- Unlike peptides, oligosaccharides are largely made up of isomeric units so additional experiments are required to fully characterize them.
- This was the first GPI found to have sialic acid.



Modifications Identified

- The N-terminus was truncated in several places giving multiple start sites.
- PrP^{Sc} deposited in plaques was further processed, mostly at residue 90, but the pathogenicity was independent of this.
- N-linked sugars identified at Asn-181 and 197 were the same for PrP^{C} and PrP^{Sc}.
- The C-terminus was truncated and modified at Ser-232 with a glycosylphosphatidylinositol anchor (GPI) that was the same for PrP^{C} and PrP^{Sc}.

No differences in the primary structures of PrP^{C} and PrP^{Sc} were identified. It was concluded that prion diseases result from abnormal protein folding.

O-linked sugars

- No consensus sequence.
- No common core structure.
- No universal enzyme.
- β-elimination is effective (NaOH) but sugars must be reduced upon release.
- Detection is problematic – because of heterogeneity; variable site occupancy. Site assignment is even harder.

Other O-linked core structures

- Fucose: Harris, R.J. & Spelmann, M.W. (1993) *Glycobiology,* **3,** 219-224.
- Glucose: Nishimura, H et al., (1989) *J. Biol. Chem.* **264,** 20320-20325.
- Mannose – in yeast
- GlcNAc – single unit; INSIDE the cell

Characterization of protein populations

Two alternative approaches

“Bottom up”

- Digestion followed by structural analysis of peptides, from which conclusions are drawn about the proteins.

“Top Down”

- Direct analysis of the intact proteins yielding extensive structural information.

- Both approaches are effective. Top down is limited in the size of proteins that can be analyzed and requires high performance instruments (FTICR). Bottom up has no size limit but it may be impossible to see combinations of modifications on a specific protein sub-population.

Requirements for Top-Down Analysis

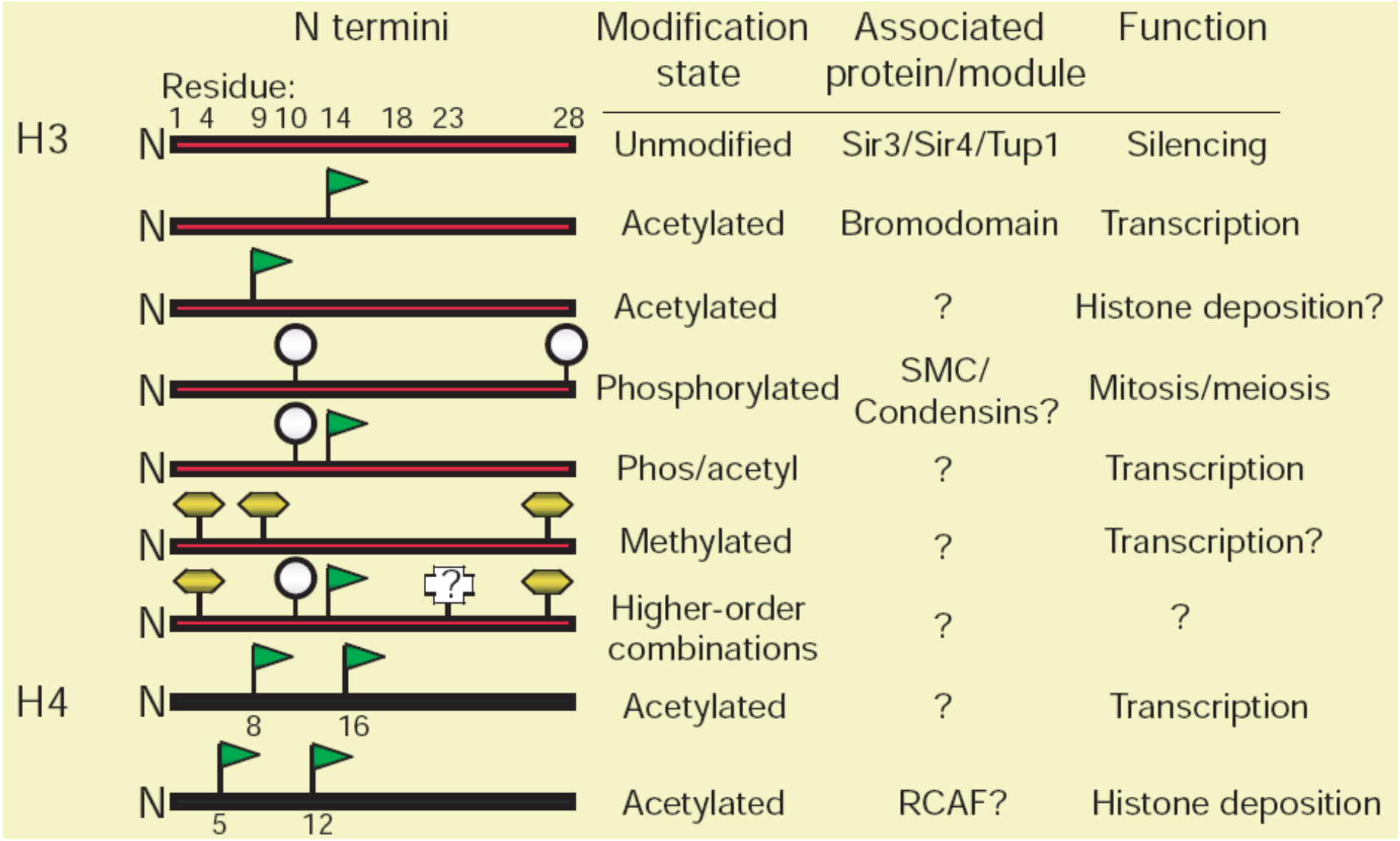
- Reasonable amount of protein, typically pmoles, and mol mass not too high.
- Protein suited to ESI giving relatively efficient fragmentation: ECD > CID.
- Good ion statistics and deconvolution software (to predict isotope pattern to determine monoisotopic mass).

Histones – Ideal samples?

- Small proteins <15 kDa.
- Extensive PTM's: phosphorylation, acetylation, methylation, ubiquitination
- Histone code: PTM state regulates gene transcription 'on' and 'off' state

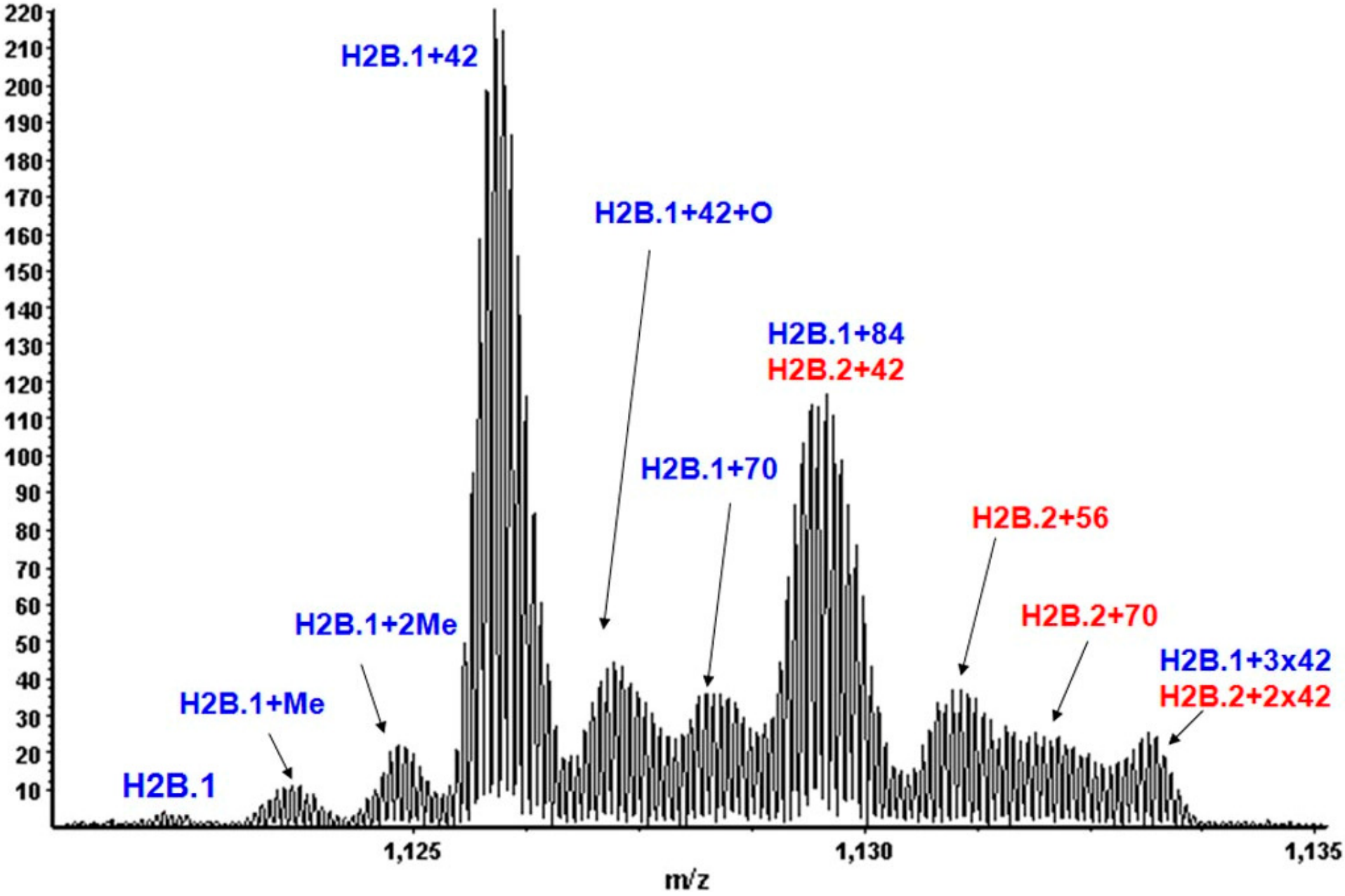
Tetrahymena histone H2B variants studied

`Histone Code' Hypothesis

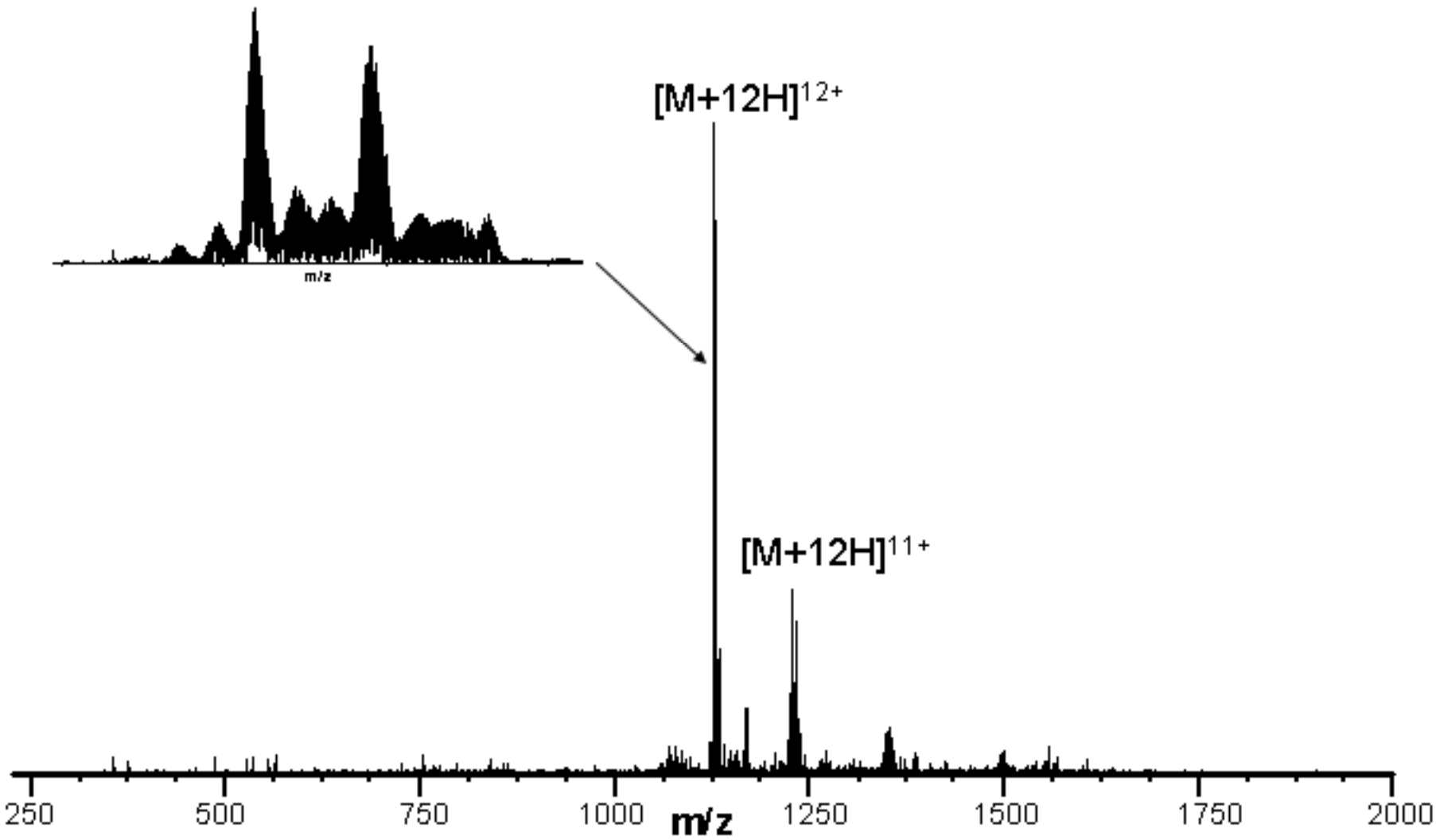


Strahl and Allis, Nature (2000) 403, 41-45.

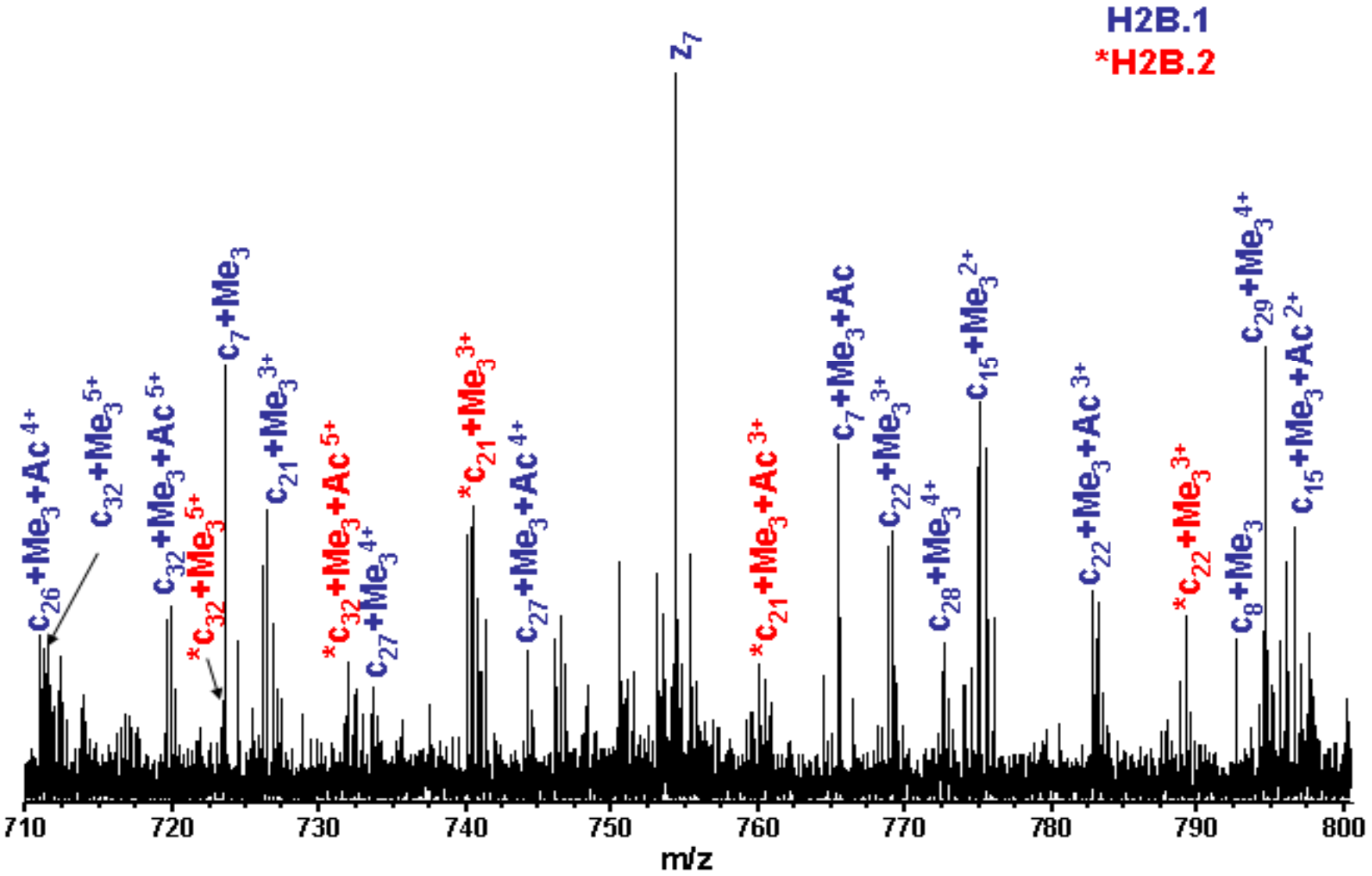
12+ Molecular ions *Tetrahymena* histone H2B



Electron-capture dissociation of H2B



Same ECD spectrum – “zoom-in”



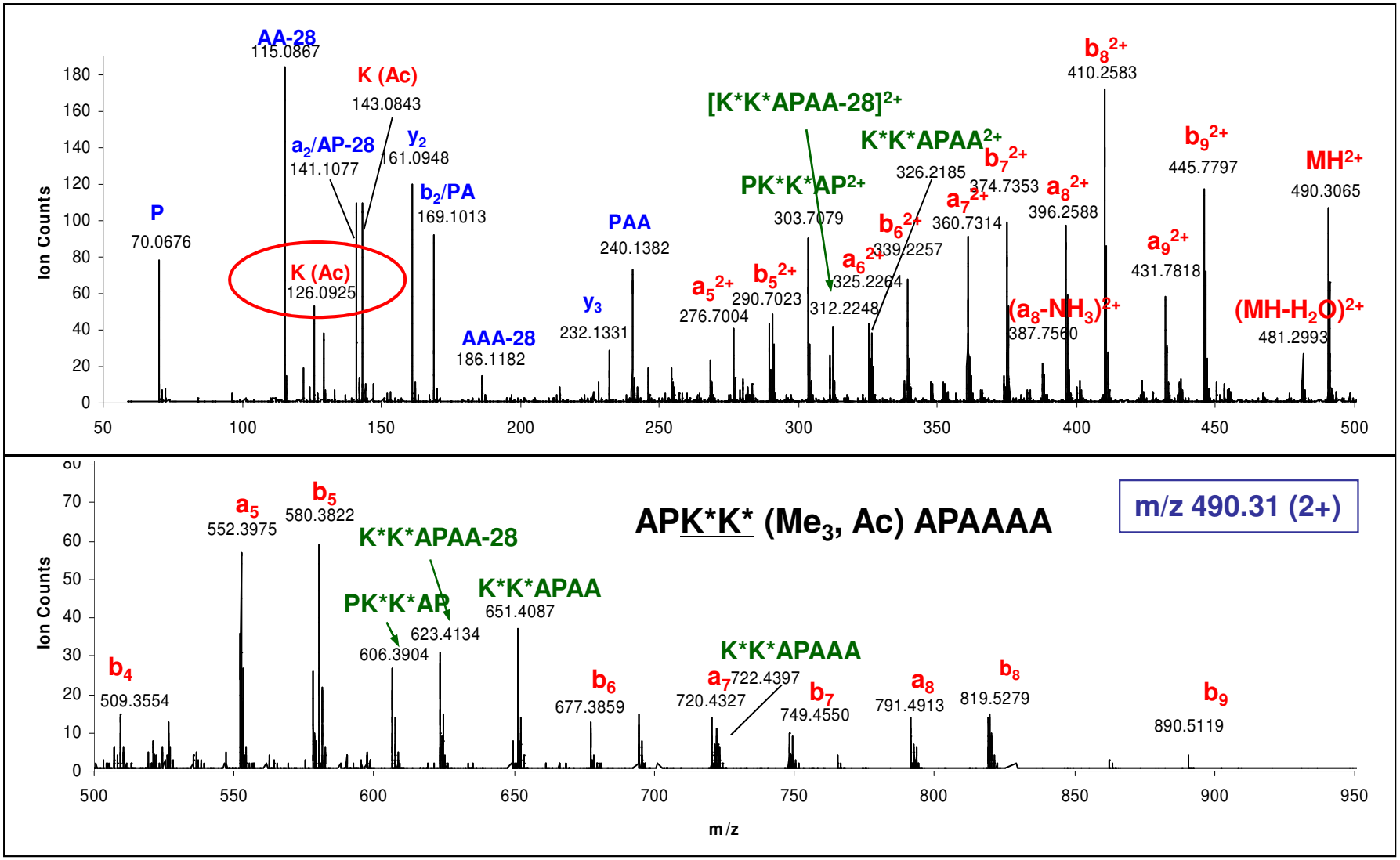
Mass accuracy distinguishes trimet. from acetyl

Calculated masses for Ac / Me_3 (+42) or 2Ac / Ac+Me_3 / 2Me_3 (+84)

Mass obs.		**Calculated Masses**	**Δ[ppm]**
356.2667	c_3+42	356.2298 /**356.2661**	+104/ **+1.6**
526.3735	c_4+2x42	526.3354/**526.3718**/526.4082	+72/ **+3.2**/ -66
723.4889	c_7+42	723.4518/**723.4882**	+51/ **+ 0.9**
754.3535	z_7	**754.3497**	**-5**
• • • • •			
1065.6380	c_{11}+42	1065.6066/**1065.642**	+29.5/ **-3.7**
1107.6556	c_{11}+2x42	1107.6171/**1107.6525**/1107.6889	+34.8/ **+2.7**/ -30.1
1167.6792	*c_{11}+2x42	1167.6372/**1167.6736**/1167.71	+36/ **+4.7**/ -26.4
1193.7422	c_{12}+42	1193.7006/**1193.737**	+34.8/ **+4.3**
1235.7426	c_{12}+2x42	1235.7111/**1235.7475**/1235.7839	+25.5/ **-3.9**/ -33.4
1253.7572	*c_{12}+42	1253.7217/**1253.7581**	+ 28.3/ **-0.7**
• • • • •			
9383.2622	c_{82}+42	**9383.2877**/9383.3241	**-2.7**/ -6.6
10563.767	z_{94}	**10563.7159**	+4.8
11548.262	z_{102}	**11548.324**	-5.3

- A single tri-methylation was either at the N-terminus or on Lys-3.
- Lys-4 in both H2B.1 and H2B.2 can also be modified and the modification is acetylation.
- No C-terminal modification was observed.

“Bottom up” CID of a +84 Da modified H2B.1 Peptide



- ✓ Characteristic K(Ac) immonium ion 126.09 indicates the presence of Lys-acetylation
- ✓ Internal fragments show the modifications are on the two lysines, instead of the N-terminus.

Fragment ion accurate masses confirm +84 Da as one trimethylation and one acetylation

	APxxAPAAAA			APwwAPAAAA			APwx (or xw) APAAAA		
$Mass_{obs}$ (Da)	Ions	$Mass_{cal}$ (Da)	Error (ppm)	Ions	$Mass_{cal}$ (Da)	Error (ppm)	Ions	$Mass_{cal}$ (Da)	Error (ppm)
70.0676	P	70.0657	27						
115.0867	AA-28	115.0871	-3						
126.0925	K(Ac)	126.0919	27						
141.1077	a_2	141.1028	35						
143.0843	AA	143.0821	15						
161.0948	y_2	161.0926	14						
169.1013	b_2 or PA	169.0977	21						
232.1331	y_3	232.1297	15						
240.1382	PAA	240.1348	14						
552.3975	a_5	552.3510	**84**	a_5	552.4237	**-47**	a_5	552.3873	**18**
580.3822	b_5	580.3459	**63**	b_5	580.4186	**-63**	b_5	580.3823	**0**
578.4262	PxxAP-28	578.3666	**103**	PwwAP-28	578.4394	**-23**	PwxAP-28	578.4030	**40**
580.3968	b_5	580.3459	**88**	b_5	580.4186	**-38**	b_5	580.3823	**25**
606.4080	PxxAP	606.3615	**77**	PwwAP	606.4343	**-43**	PwxAP	606.3979	**17**
649.4450	a_6	649.4037	**64**	a_6	649.4765	**-49**	a_6	649.4401	**8**
677.4436	b_6	677.3986	**66**	b_6	677.4714	**-41**	b_6	677.4350	**13**
720.4550	a_7	720.4408	**20**	a_7	720.5136	**-81**	a_7	720.4772	**-31**
748.4628	b_7	748.4357	**36**	b_7	748.5085	**-61**	b_7	748.4721	**-12**
791.5098	a_8	791.4779	**40**	a_8	791.5507	**-52**	a_8	791.5143	**-6**
819.5088	b_8	819.4729	**44**	b_8	819.5456	**-45**	b_8	819.5092	**0**
Error for Common Ions (AVG ± STD)			**20±9**						
Error for Modification Ions (AVG ± STD)			**60±23**			**49±16**			**15±10**

x—K(Ac); w– $K(Me_3)$

Summary of H2B PTM Characterization

Intact Protein Analysis			Proteolytic Digest Analysis	
Molecular Weight		ECD- FT-ICR MS	Trypsin	Asp-N
Most Abundant	H2B.1- Me_3	**(N*-or K3)-Me_3**		[1-45]- Me_3
	H2B.2- Me_3	**(N* or K3)-Me_3**		[1-45]- Me_3
	H2B.1- Me_3 Ac	(N* or K3)-Me_3 + K4-Ac	**N*-Me_3 + K4-Ac** and **(K3K4)- Me_3 Ac**	(K3K4)-Me_3 Ac
	H2B.2- Me_3 Ac	(N* or K3)-Me_3 + K4-Ac	**N*-Me_3 + K4-Ac** and **(K3K4)-Me_3 Ac**	(K3K4)-Me_3 Ac
Less Abundant	H2B.1 + 56 Da			**(K3K4)- Me Ac**
	H2B.1 + 70 Da		**N*-Me_2 + K4-Ac**	**(K3K4)- Me_2 Ac**
	H2B.2 + 56 Da			**(K3K4)- Me Ac**
	H2B.2 +70 Da			**(K3K4)- Me_2 Ac**
	H2B.1- Me			[1-45] Me
	H2B.1- Me_2			[1-45] Me_2
	H2B.2- Me			[1-45] Me
	H2B.2- Me_2			[1-45] Me_2
Least Abundant			**H2B.1/H2B.2-K41-Ac**	
			H2B.1 K111-Me_{2-3}	
			H2B.2 K111-Me_3/Ac	

N* indicate N-terminus of the protein

Using both methods gave more information

The “bottom-up” approach:

- More sensitive: ~300 fmoles injected.
- Reveals more modifications.

BUT unlike the “top-down” approach:

- Misses the major protein component!
- Does not show the relative distribution of differently modified populations.

Conclusions

Mass spectrometry provides a sensitive, unbiased approach to the analysis of a wide variety of peptide/protein modifications.

- Isolation and MS analysis need to be adjusted to suit the PTM of interest.
- For single protein/simple mixture PTM analysis, it is best to use no enrichment, to fragment as many components as possible, then try to find spectra of modified peptides.
- Enrichment methods can be effective but they require relatively large amounts of protein.
- Peptide-level analysis may not accurately reflect the composition of protein populations in bottom up experiments. Top down may be a useful adjunct.