Post-translational modifications

PC235

Katalin F. Medzihradszky
folkl@cgl.ucsf.edu

Post-translational Modification of Proteins
Expanding Nature's Inventory (2006)
by C.T. Walsh

ISBN 0-9747077-3-2

Just for introduction

Genome

↓

Transcription – mRNA

↓

Translation – proteins + co-translational modifications

↓

Post-translational modifications

Post-translational modifications I

- Enzymatic processing
- N-, O-, C-linked glycosylation – Asn, Ser/Thr/Hyl/Hyp, Trp
- Phosphorylation – Tyr, Ser, Thr, His, Asp
- Acylation
  - acetylation of the N-terminus
  - fatty acid anchors on Cys
- Cross-linkage – Lys, Trp, Tyr, Met
- Oxydation – Cys, Met, Trp, Tyr, His
- Methylation – N-terminus, Arg, Lys
- Ubiquitination - Lys

Post-translational modifications II

- Cannot be predicted – though consensus sequences have been reported for some of them
- Organism-dependent
- Can be tissue- or location-specific
- Stable or dynamic – high and low level
- Alters biological activity, *and* physical properties
- May alter the immune response

Sample preparation

- Will the modification survive?
- Can I get it down to a mass spectrometry friendly size?
- Isolation/enrichment?
- Losses – too hydrophilic/hydrophobic

Mass spectrometry

- Will the modification survive the ionization? the MS/MS activation?

- How unambiguous is the assignment?

→ modification assignment?

Ac vs. Me3; phosphate vs. sulfate etc.

→ site assignment?

Finding the N-termini

- 80-90% of the eukaryotic proteins is acetylated

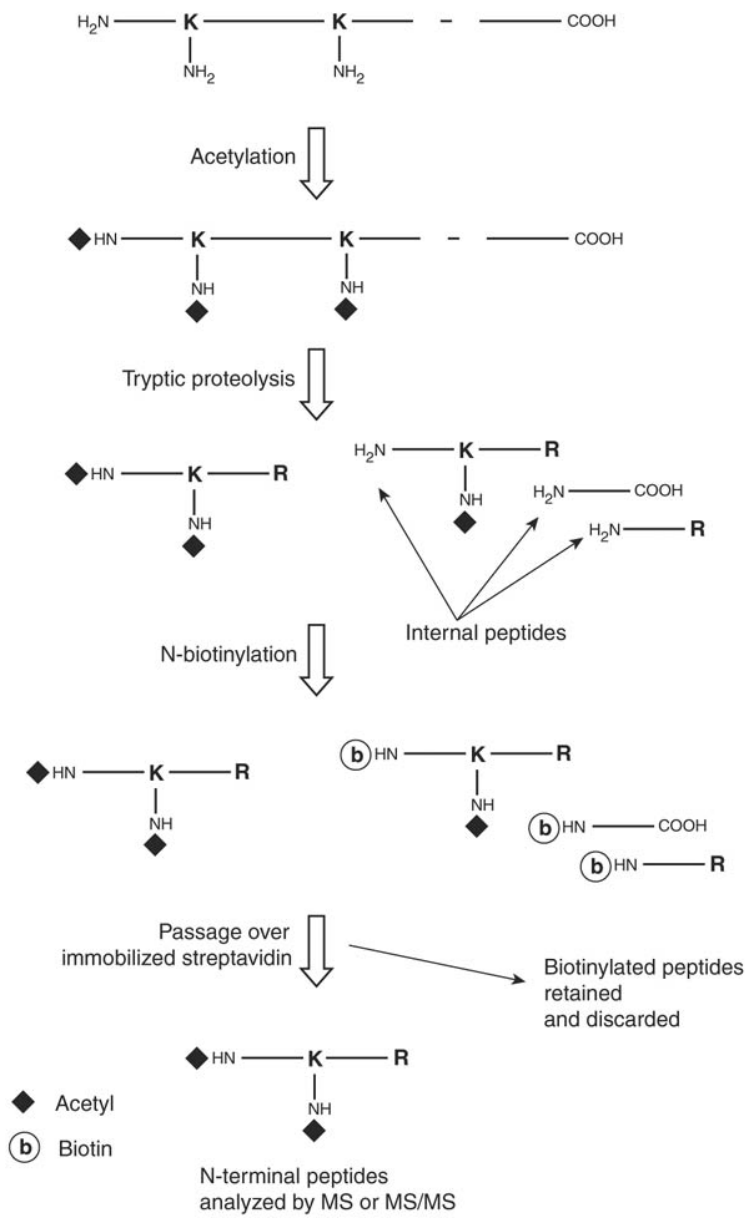
If 2nd aa: *G, A, S, C,* T, P, V

- Met-1 is clipped off; *Ac-* is added to *G, A, S,* T

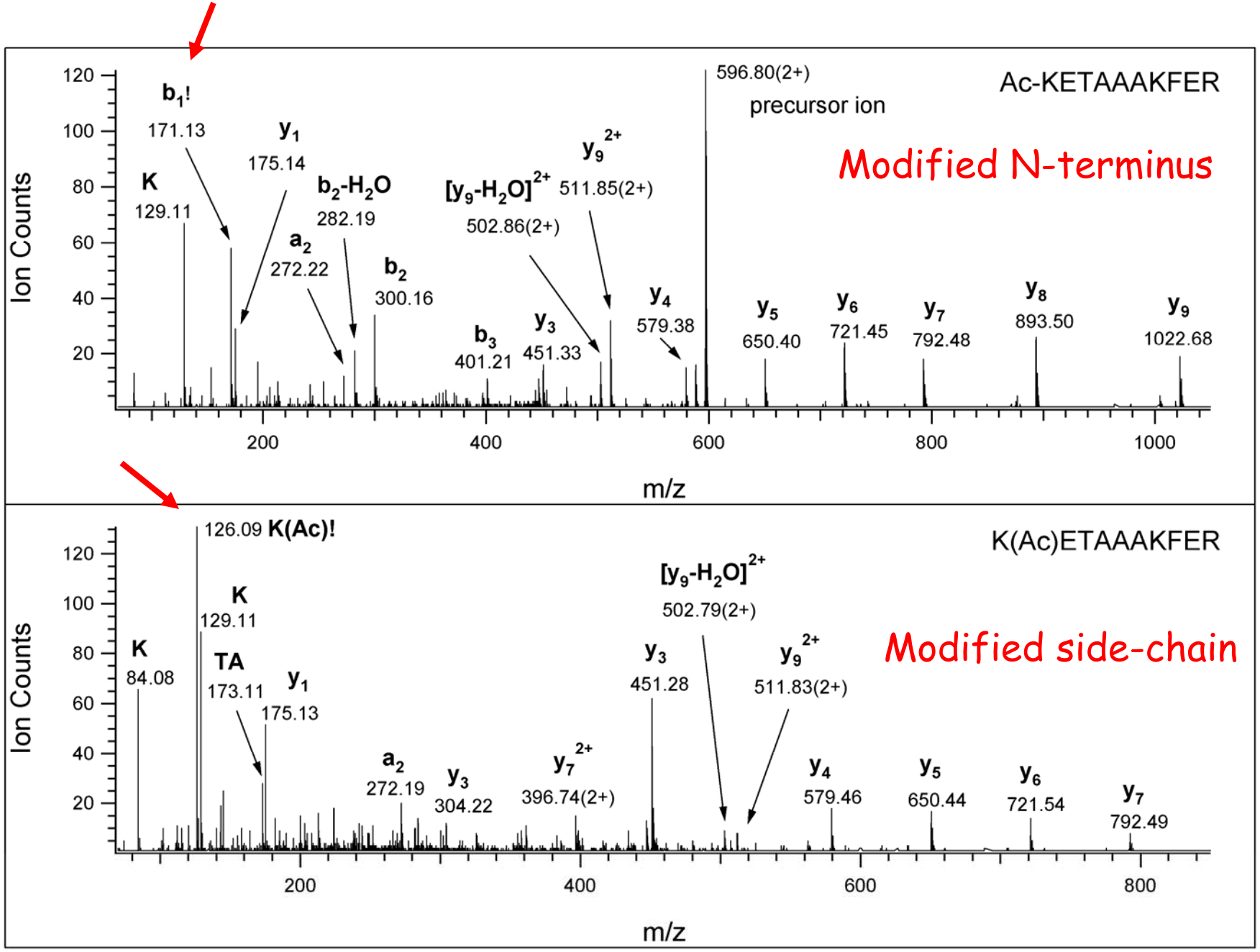
If 2nd aa: E, D, Q, M, I, L, W, F –

persisting Met gets acetylated

Finding the N-termini



- McDonald L, et al., Positional proteomics: selective recovery and analysis of N-terminal proteolytic peptides.
  Nat Methods. 2005 Dec;2(12):955-7. Epub 2005 Nov 18.



I don't think any search engine will tell the difference !

Methylation (mono, di, tri - +14, +28, +42 Da)

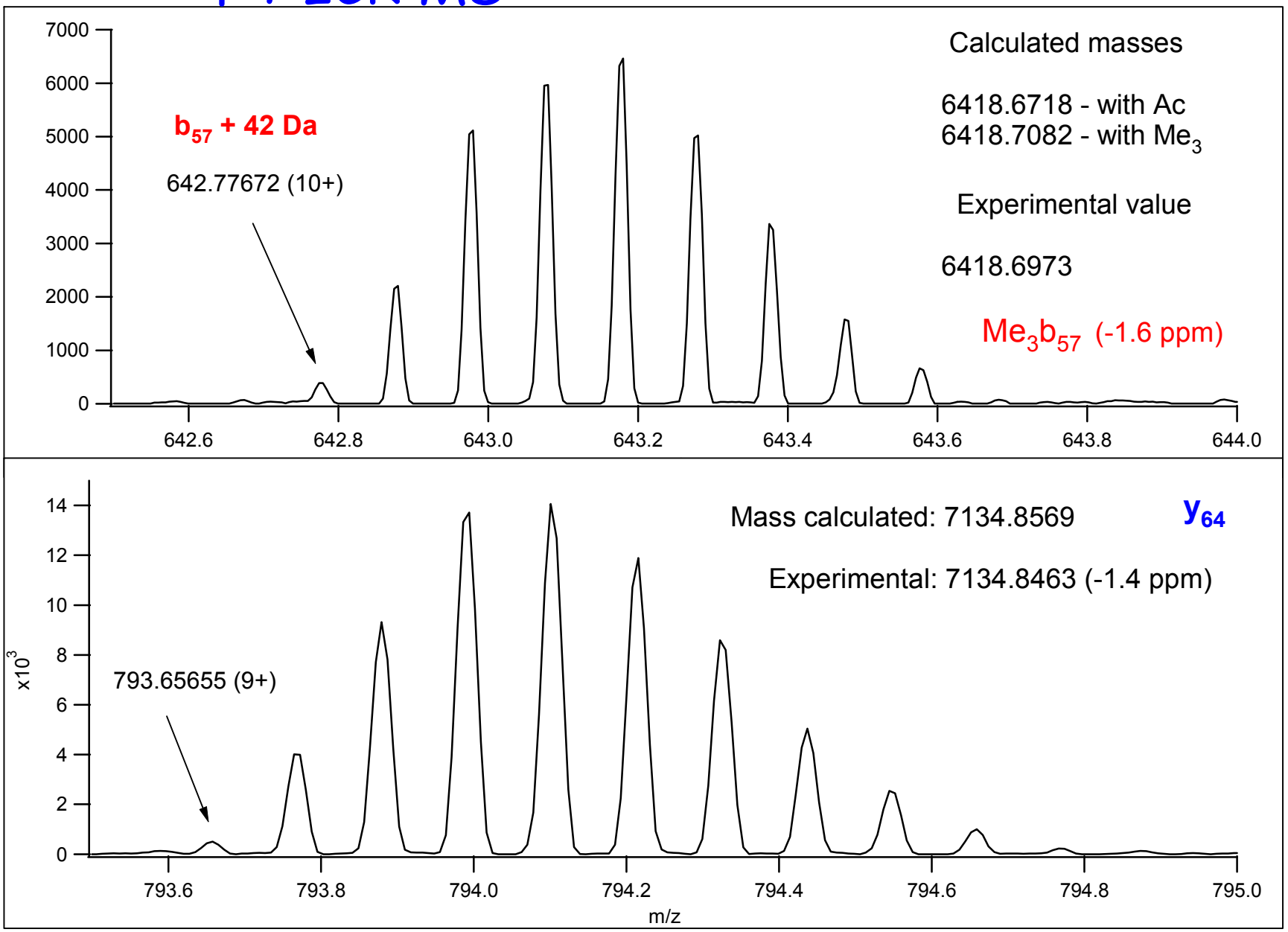
- N-terminus, Lys, Arg, His
- Trimethyl – acetyl = 36 mmu

Accurate mass measurement helps;

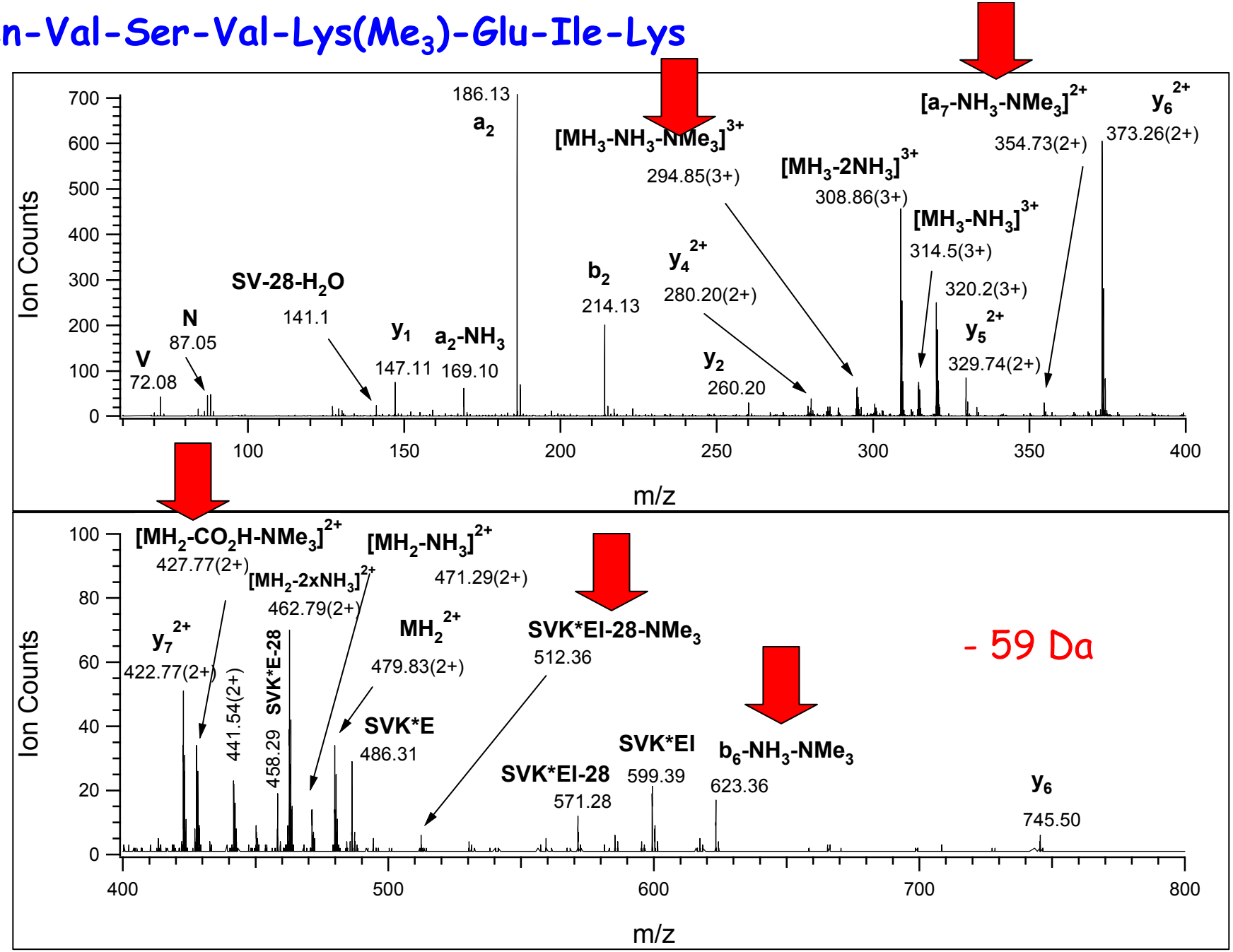
Fragmentation is different too

- Glu (Asp) may form Me-ester – upon CBB staining (MeOH + acid) + 14 Da

FT ICR MS


7000
6000
5000
4000
3000
2000
1000
0
642.6
642.8
643.0
643.2
643.4
643.6
643.8
644.0
b57 + 42 Da
642.77672 (10+)
Calculated masses
6418.6718 - with Ac
6418.7082 - with Me3
Experimental value
6418.6973
Me3b57 (-1.6 ppm)
14
12
10
8
6
4
2
0
x10^3
793.65655 (9+)
Mass calculated: 7134.8569
y64
Experimental: 7134.8463 (-1.4 ppm)
793.6
793.8
794.0
794.2
794.4
794.6
794.8
795.0
m/z

Asn-Val-Ser-Val-Lys(Me_3)-Glu-Ile-Lys



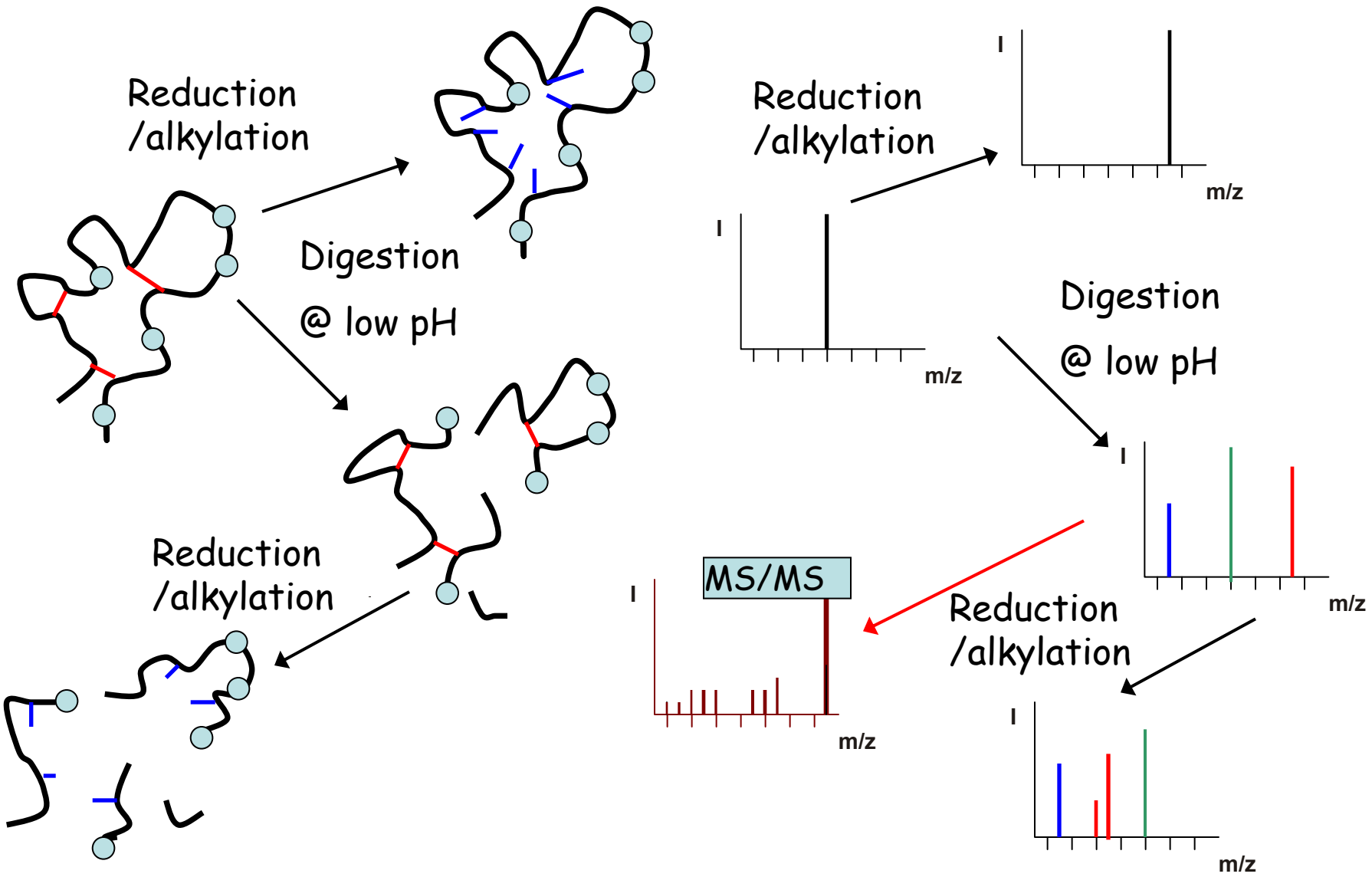
Disulfide-bridges

- in membrane and secreted proteins

important 3D structure feature

- prone to shuffling @ basic pH

Assigning disulfide-bridges



Synthetic Ac-TIMP-1(Ser^{175})126-184

ECTVFPCLSI PCKLQSGTHC LWTDQLLQGS
EKGFQSRHLA CLPREPGLCS WQSLRSQIA

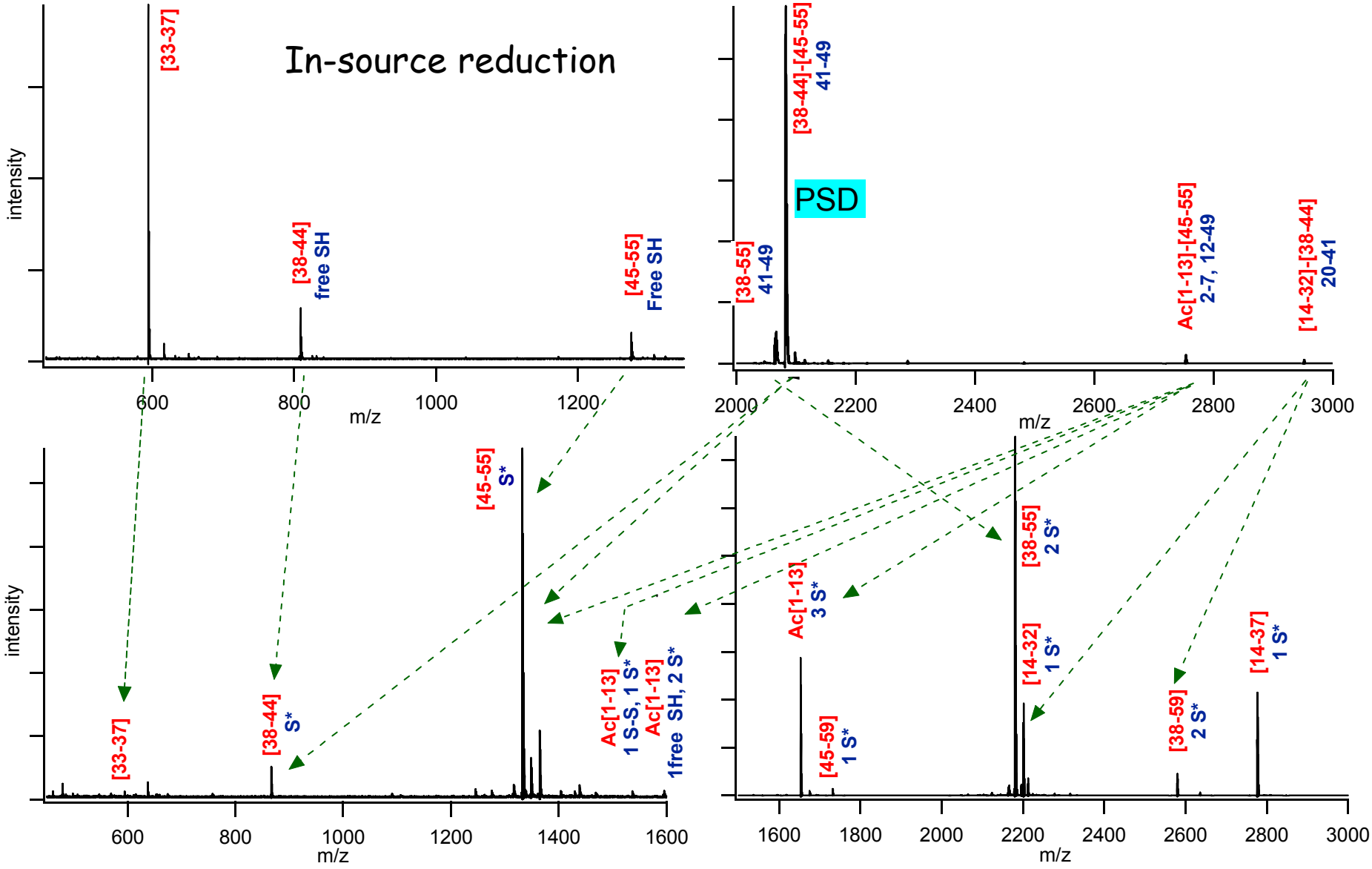
Where are the disulfide bridges?

* digestion with trypsin @ pH 6;

* with pepsin in acid

Bodi, N. et al., *J. Pept. Sci.* **9**, 430-441 (2003).

MALDI-TOF analysis of the tryptic digest



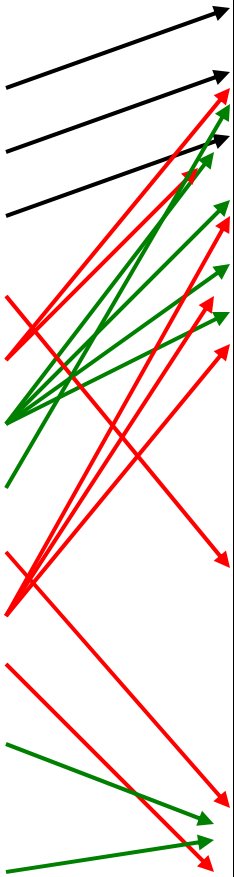
Results from the MALDI-TOF MS

Tryptic digest

594.30	[33-37]	
809.5	[38-44]	szabad SH
1275.7	[45-55]	szabad SH
2064.0	[38-55]	41-49
2082.1	[38-44]-[45-55]	41-49
2752.4	Ac[1-13]-[45-55]	2-49, 7-12
2950.6	[14-32]-[38-44]	20-41
3602.6	Ac[1-32]	2-7, 12-20
3620.6	Ac[1-13]-[14-32]	2-7, 12-20
4177.3	Ac[1-37]	2-7, 12-20
5665.1	Ac[1-32]-[37-55]+H_2O	2-49, 7-12, 20-41
5683.2	Ac[1-32]-[37-55]+2 H_2O	2-49, 7-12, 20-41

After reduction/alkylation

594.30	[33-37]	
866.5	[38-44]	1 CM Cys
1332.7	[45-55]	1 CM Cys
1536.8	Ac[1-13]	1 S-S, 1 CM Cys
1595.8	Ac[1-13]	1 SH, 2 CM Cys
1652.7	Ac[1-13]	3 CM Cys
1674.8	[45-59]	SH
1731.9	[45-59]	1 CM Cys
2144.2	[14-32]	SH
2180.1	[38-55]	2 CM Cys
2201.1	[14-32]	1 CM Cys
2579.3	[38-59]	2 CM Cys
2776.3	[14-37]	1 CM Cys
3834.7	Ac[1-32]	4 CM Cys
4409.7	Ac[1-37]	4 CM Cys



Disulfide bridges in TIMP-1 C-terminal domain: [38-44]-S-S-[45-55], PSD of MH^+ at m/z 2082.1

MALDI-PSD/CID yields characteristic triplets

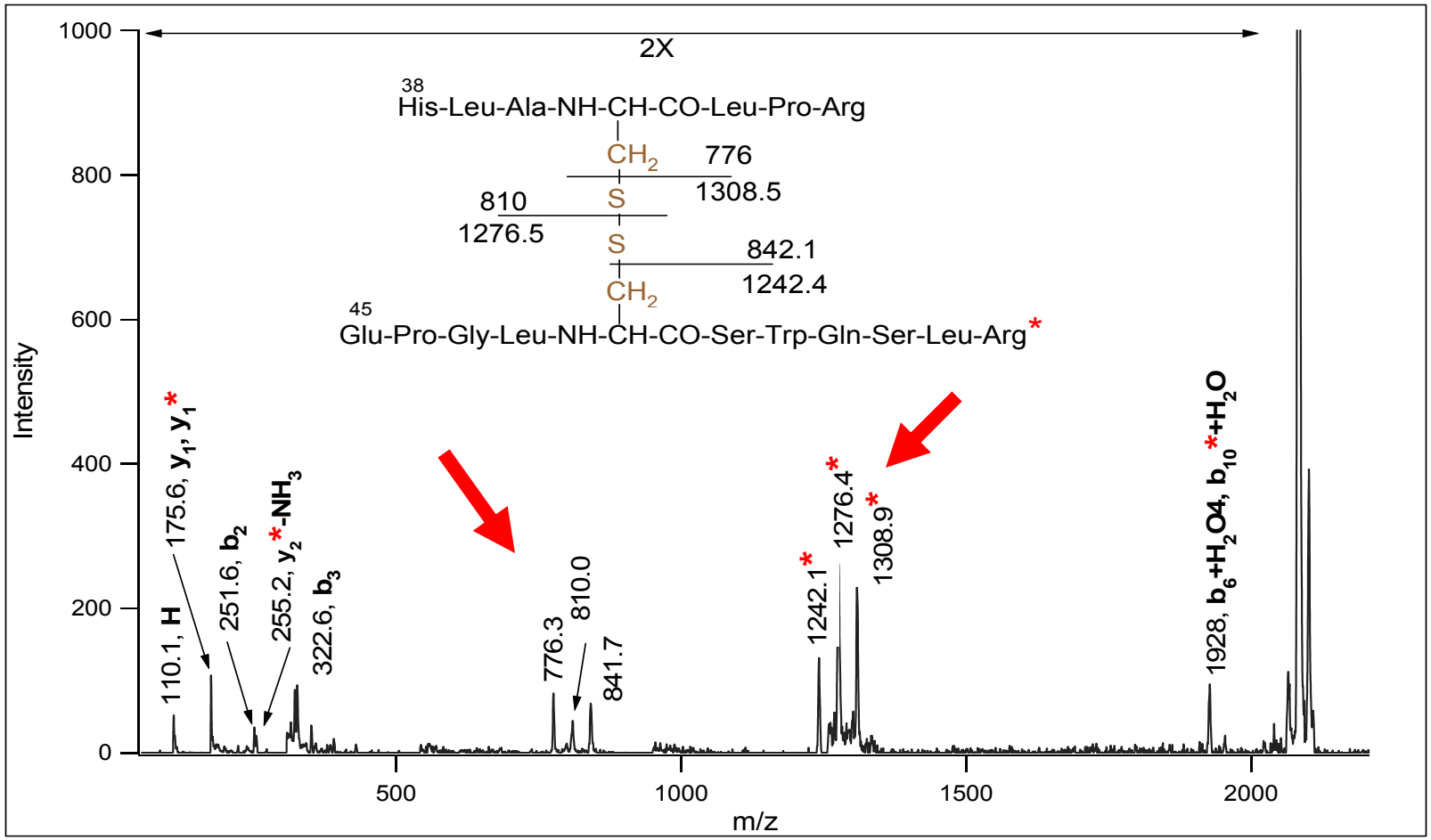
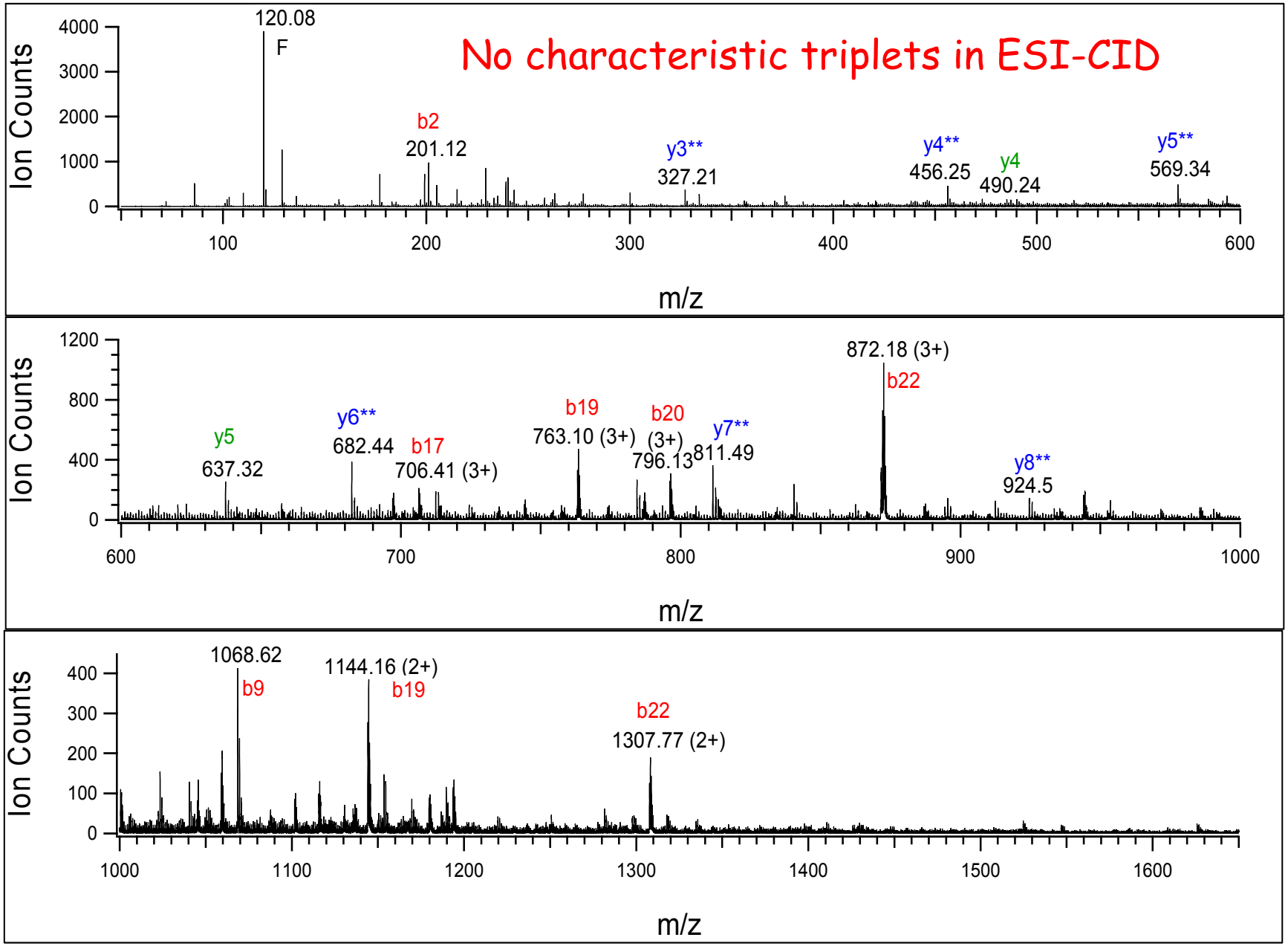
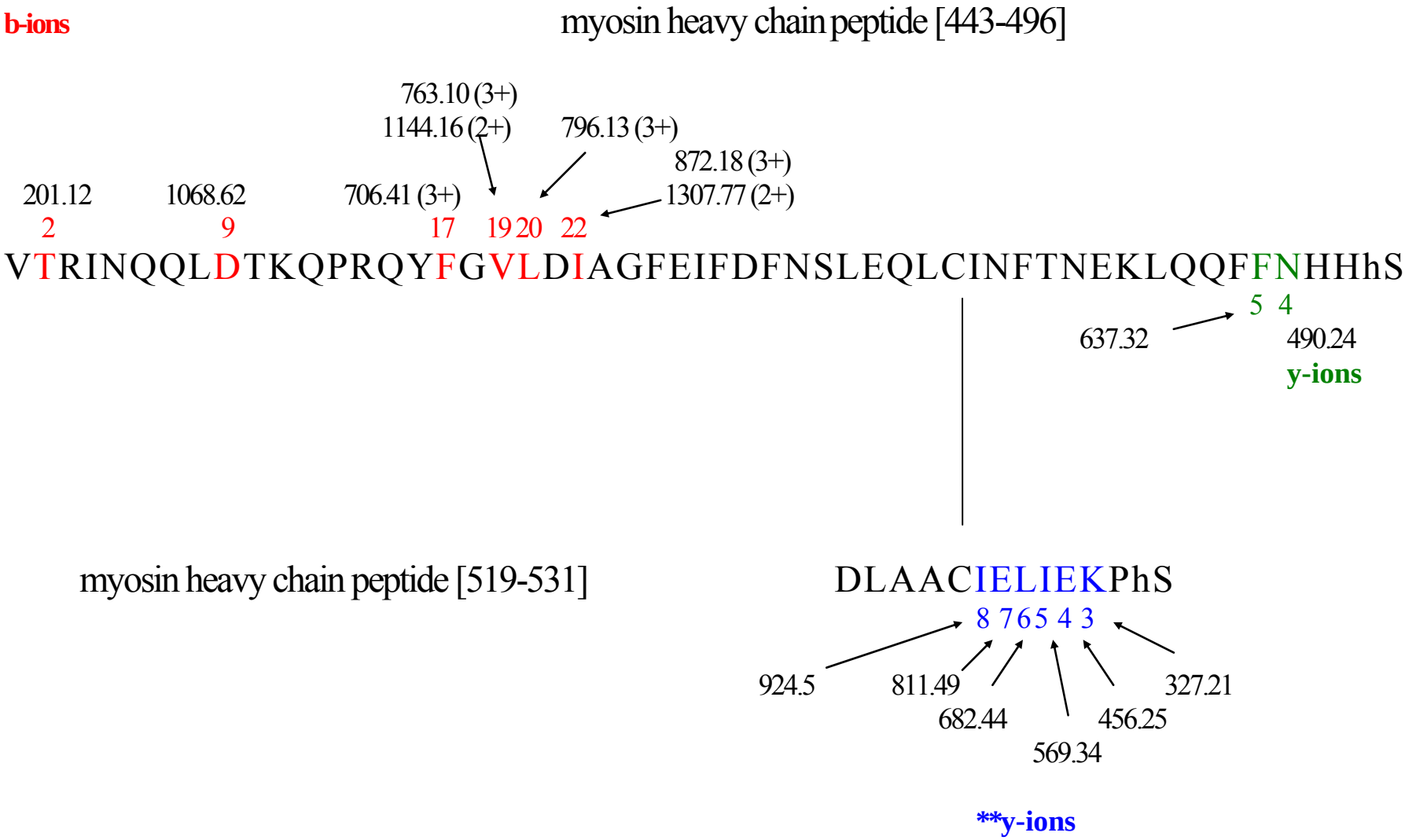


Figure 9. Low energy CID of m/z 871.79 (most abundant ion in 9+ cluster), MW (monoisotopic) 7835.03 Da. This molecule was identified as disulfide linked peptides [443-496] and [519-531] of myosin heavy chain. Sequence and fragmentation are shown on the next page.

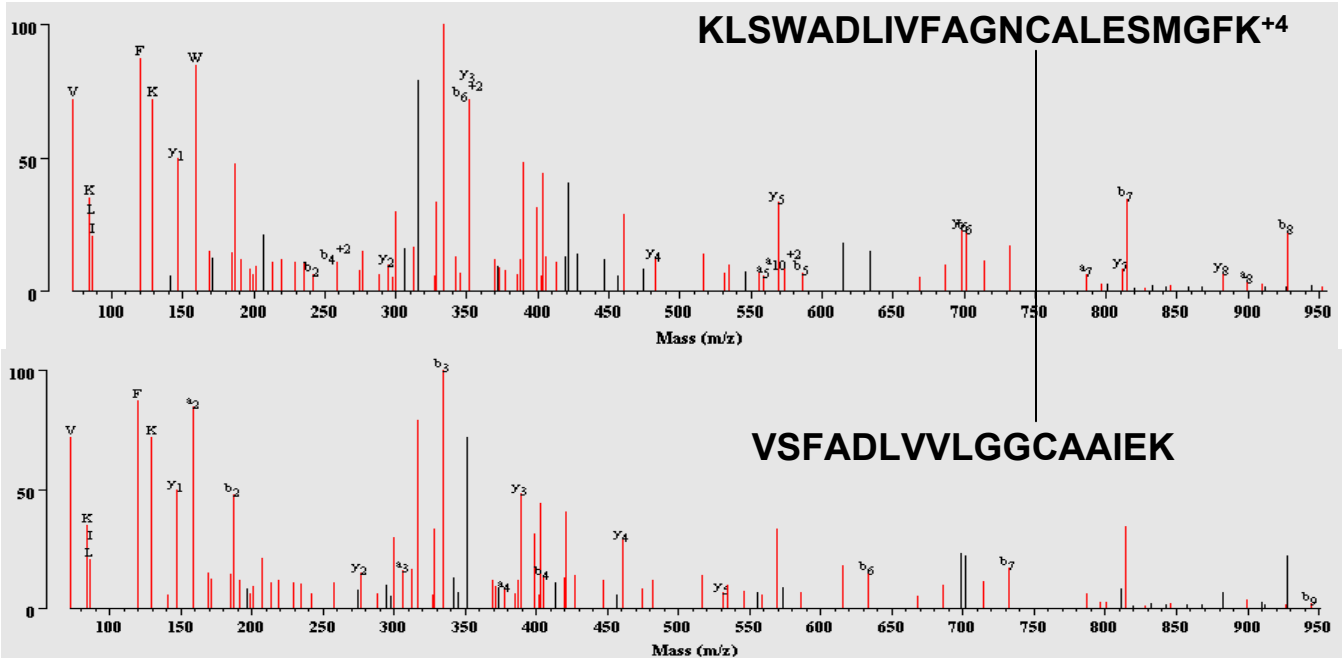


Fragment ions used for the identification of disulfide linked myosin peptides.
(see low energy CID spectrum in Figure 9.)



Disulfide Bridges

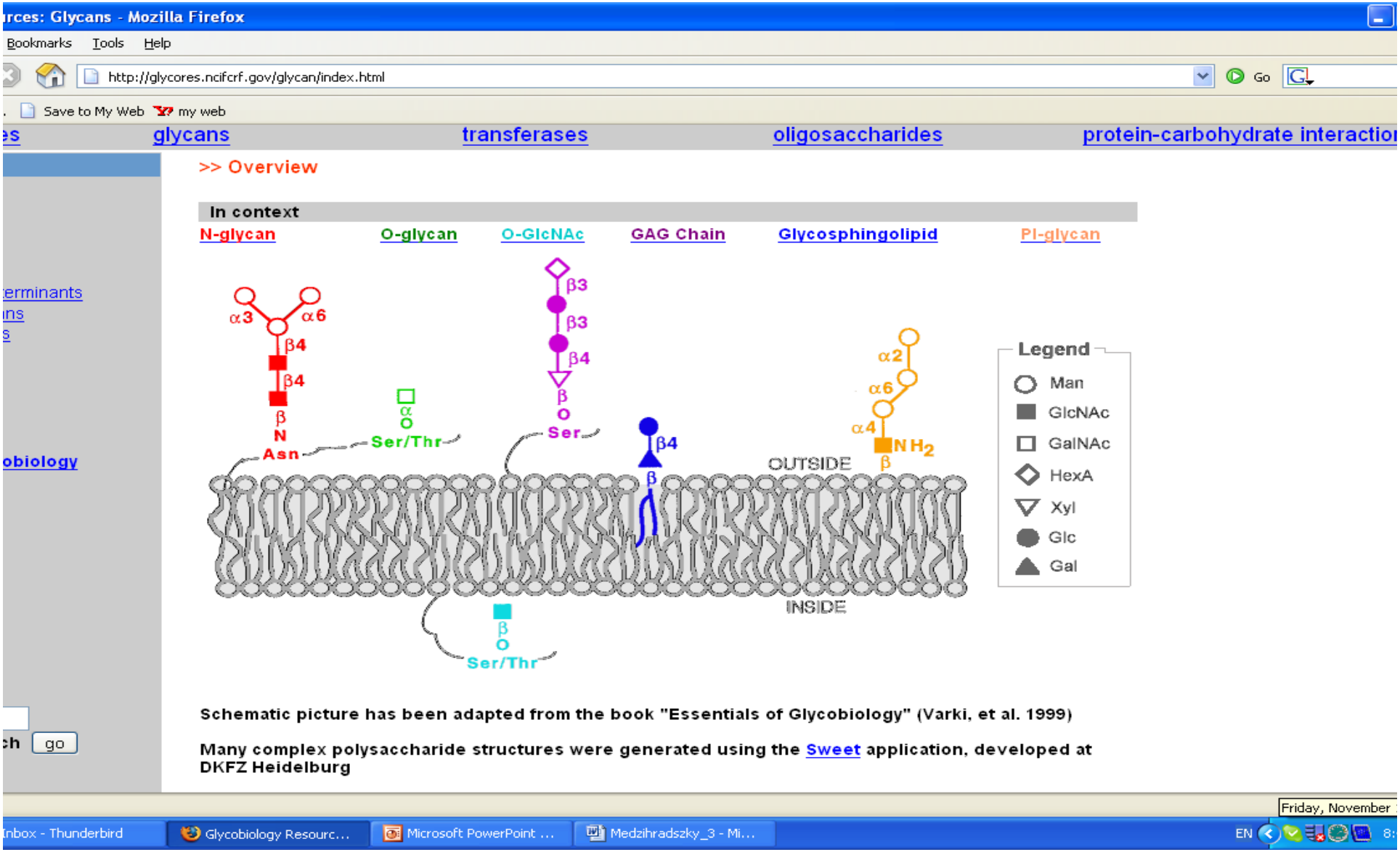
The ProteinProspector mass modification search can be used in conjunction with MS-Bridge to find peptides with disulfide bridges. For this example, mass shifts between 0-2000 Da were considered. (1023.3276^{+4}).



Glycosylation

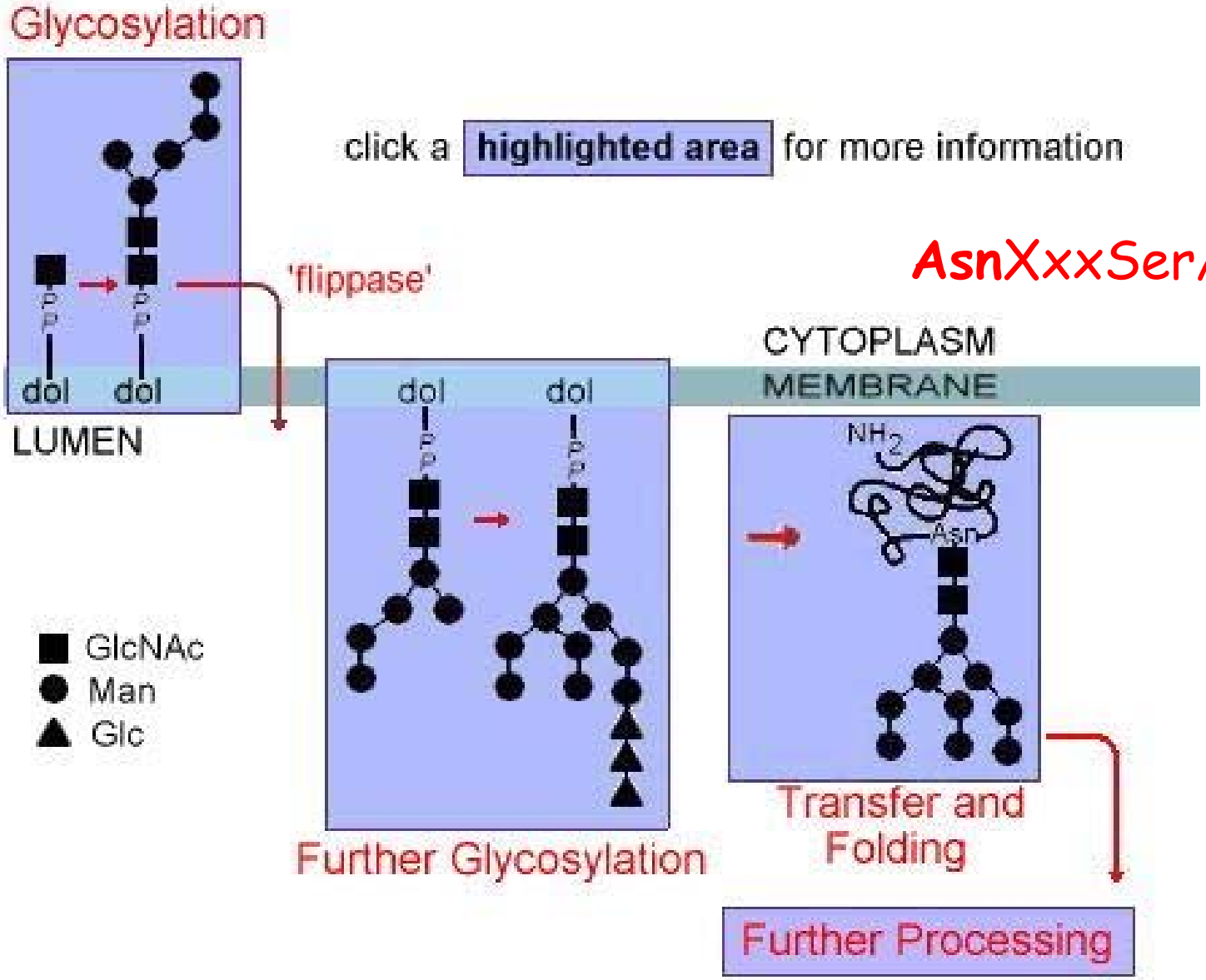
http://glycores.ncifcrf.gov/

Reference: Essentials of Glycobiology by Varki et al.

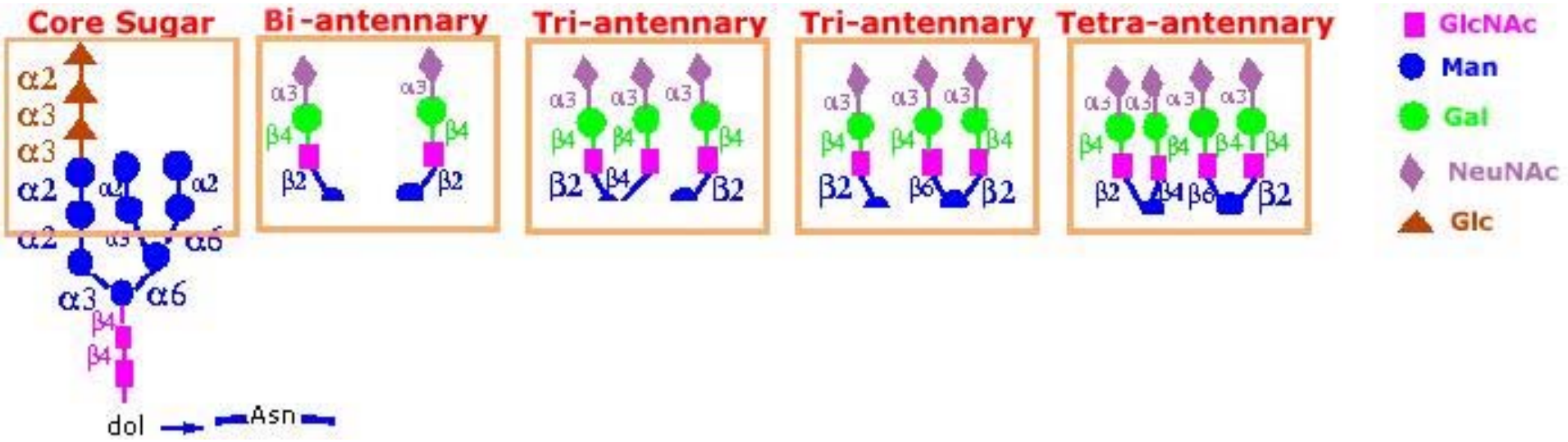


N-linked

AsnXxxSer/Thr/Cys



Further processing



N-linked glycosylation

- consensus sequence
- $GlcNAc_2Man_3$ – core

oligomannose structure – just Man units

complex sugars– GlcNAc-Gal–SA antennae

hybrid structures

core fucosylation

sulfate, phosphate modifications

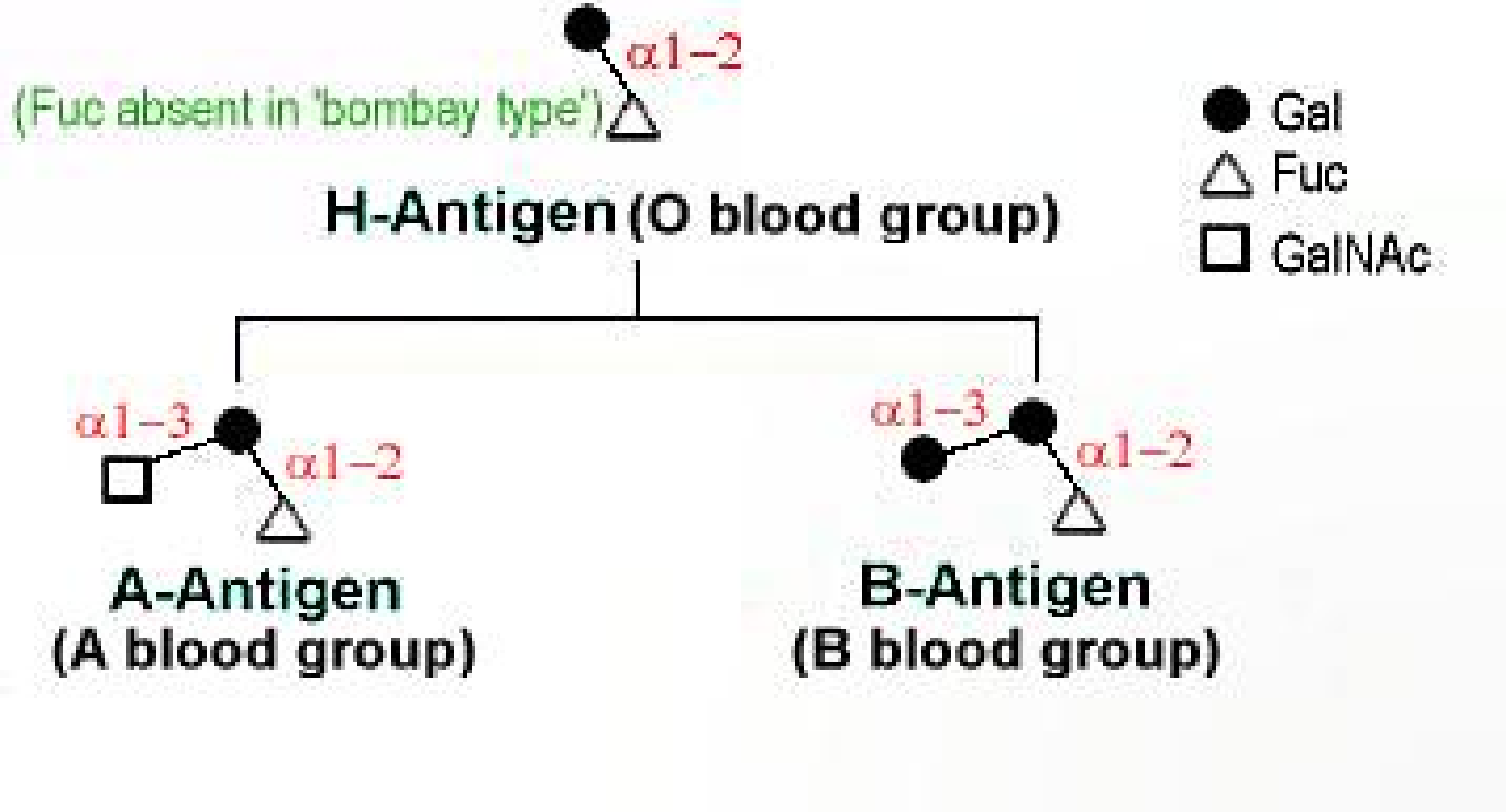
- PNGase F removes all N-linked structures; Asn $\rightarrow$ Asp

N-linked glycosylation

- Incredible heterogeneity: a site may be only partially occupied and may display numerous different carbohydrates
- species-, tissue-, cell-type-specific modification, physiological changes, diseases may alter the sugars

certain structures are immunogenic

Gal α1-3 capping, Fuc α1-3 on inner GlcNAc; blood group determinants



N-linked glycosylation

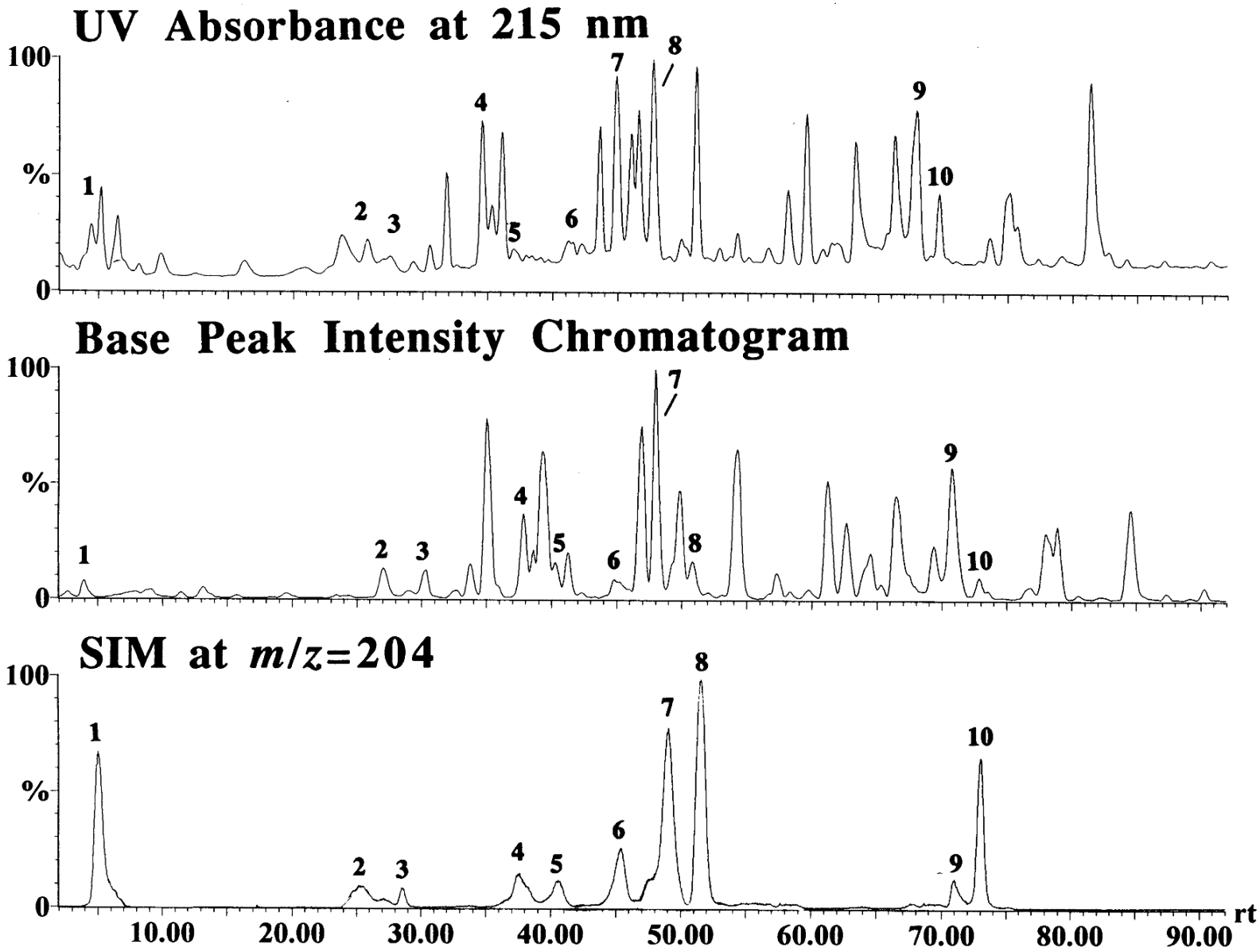
- Identification from diagnostic fragments:

* HexNAc m/z 204
* HexHexNAc m/z 366

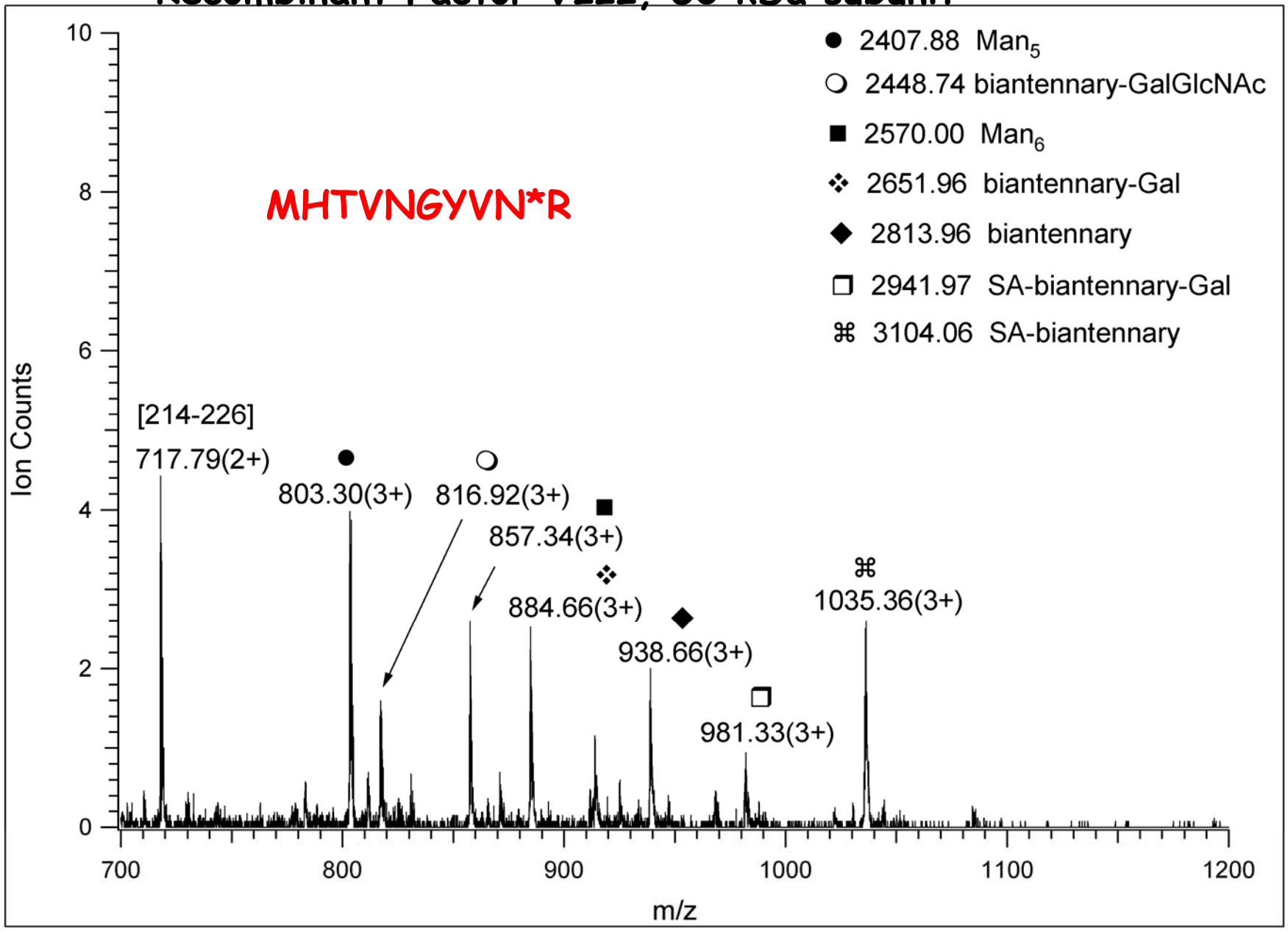
Precursor scan, or „ping-pong" acquisition

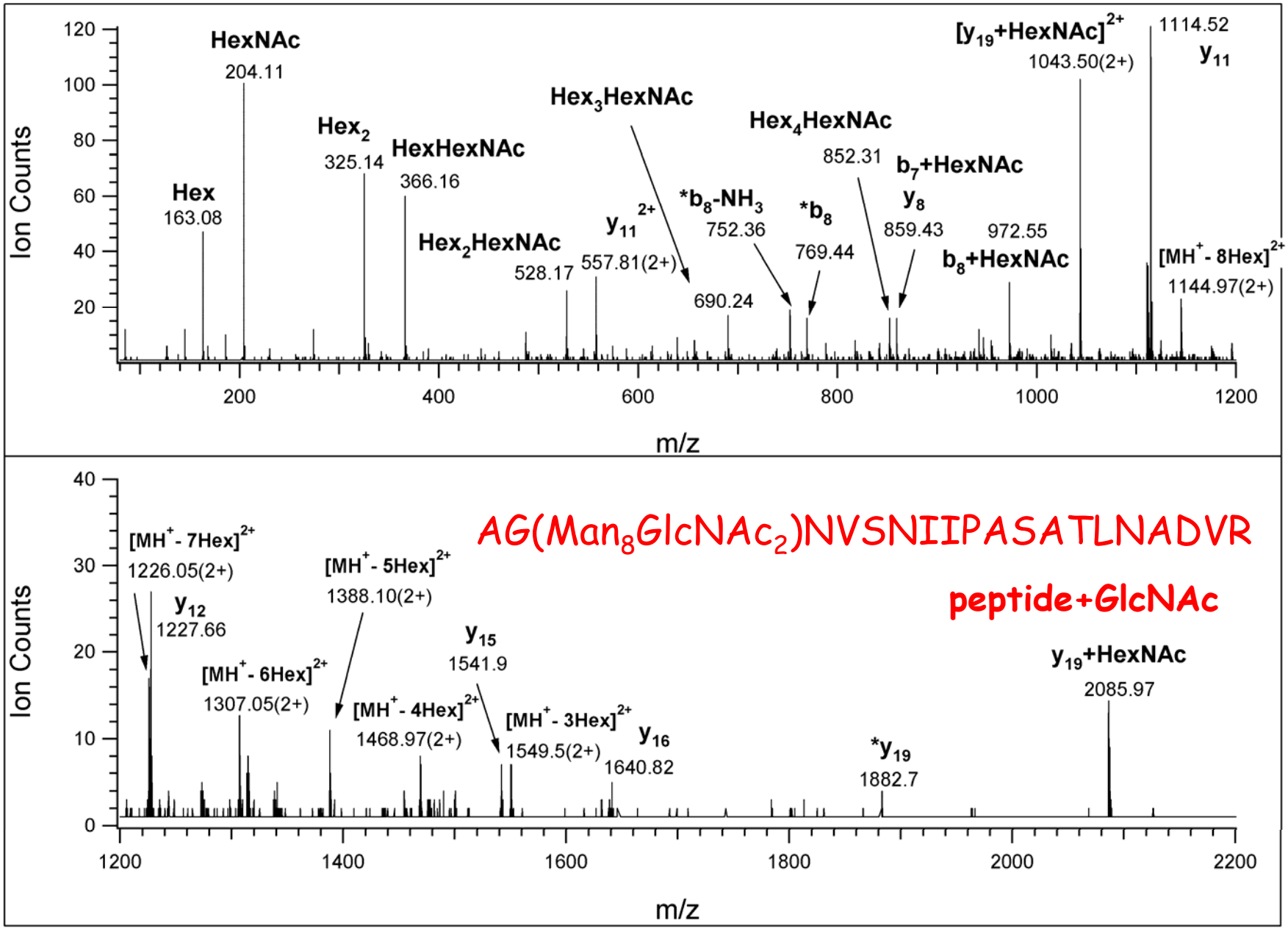
- Identification from oligosaccharide heterogeneity
- enrichment by HILIC or lectin-chromatography

human lecithin:cholesterol acyltranferase and apolipoprotein D, tryptic digest, LC/MS analysis



Recombinant Factor VIII, 50 kDa subunit

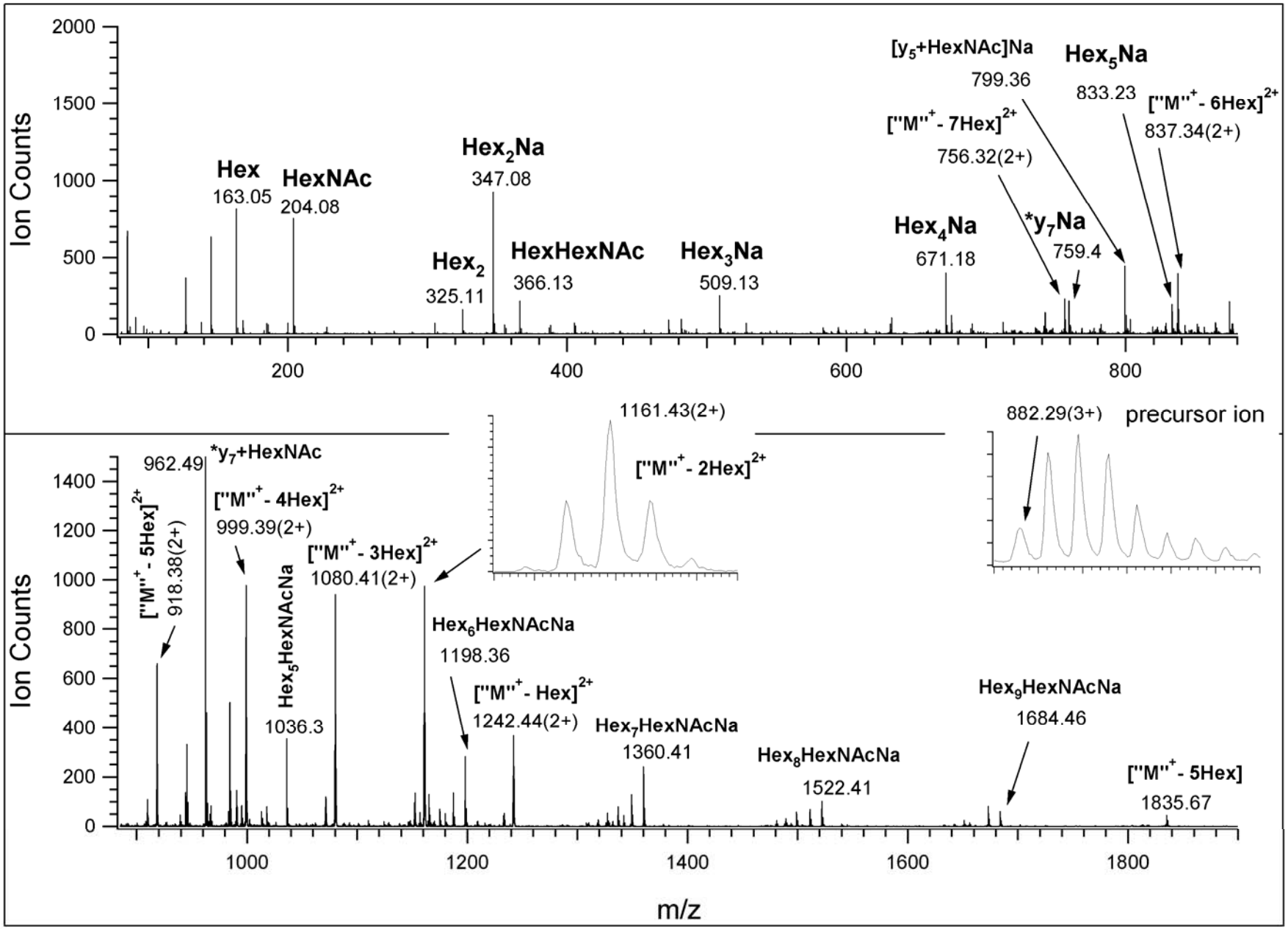



Ion Counts
m/z
HexNAc
204.11
Hex
163.08
Hex2
325.14
HexHexNAc
366.16
Hex2HexNAc
528.17
y11 2+
557.81(2+)
Hex3HexNAc
690.24
*b8-NH3
752.36
Hex4HexNAc
*b8
769.44
852.31
b7+HexNAc
y8
859.43
b8+HexNAc
972.55
[y19+HexNAc]2+
1043.50(2+)
y11
1114.52
[MH+ - 8Hex]2+
1144.97(2+)
AG(Man8GlcNAc2)NVSNIIPASATLNADVR
peptide+GlcNAc
[MH+ - 7Hex]2+
1226.05(2+)
y12
1227.66
[MH+ - 6Hex]2+
1307.05(2+)
[MH+ - 5Hex]2+
1388.10(2+)
[MH+ - 4Hex]2+
1468.97(2+)
y15
1541.9
[MH+ - 3Hex]2+
1549.5(2+)
y16
1640.82
*y19
1882.7
y19+HexNAc
2085.97

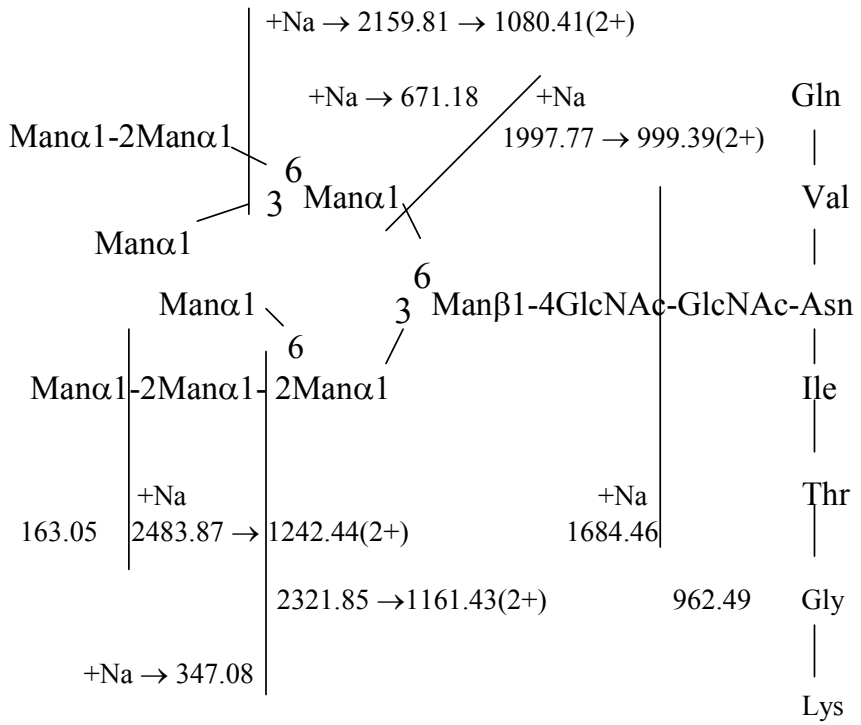
About the structures of N-linked glycopeptides

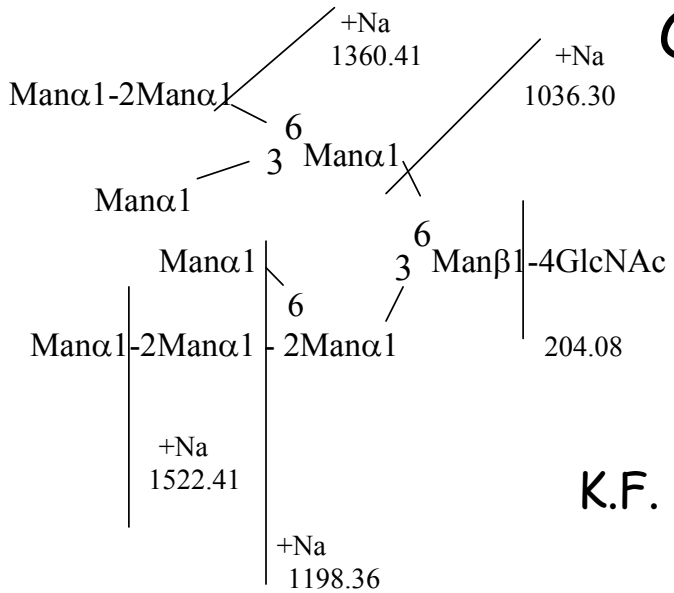
- from the measured mass, and the CID spectrum the modified peptide can be identified + the size and class of the sugar
- the identity of the sugar units and their linkage positions CANNOT be determined
- NMR, exo- and endoglycosidases are needed to complete the job

$[MHNa_2]^{3+}$ of QV($Man_{10}GlcNAc_2$)NIT and $[MH_2Na]^{3+}$ of QV($Man_9GlcNAc_2$)NITGK



Scheme 1. Fragments observed in the low energy CID spectrum of $[MH_2Na]^{3+}$ of glycopeptide QV($Man_9GlcNAc_2$)NITGK (Figure 6)





One component from the previous slide

K.F. Medzihradszky *Meth. Enzymol.* **405**, 116-138 (2005).

O-linked sugars

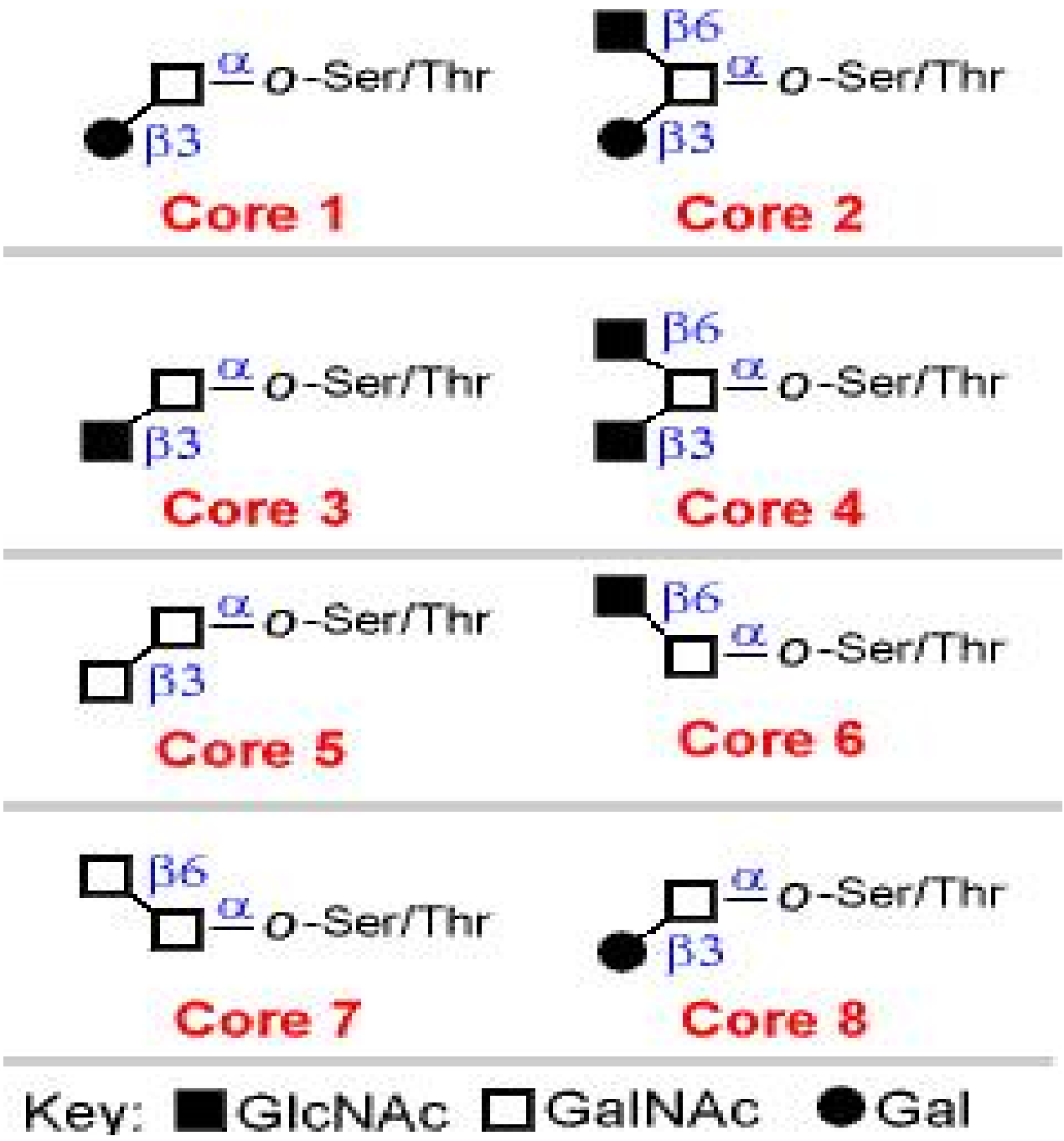
- No consensus sequence
- No common core structure
- No universal enzyme

β-elimination works (NaOH)

sugars have to be reduced upon release

Detection is problematic – because of heterogeneity; variable site occupancy

Site assignment is even harder



O-linked sugars

Other O-linked core structures

- Fuc

Harris, R.J. & Spelmann, M.W. (1993)
Glycobiology, **3,** 219-224.

- Glc

Nishimura, H et al., (1989)
J. Biol. Chem. **264,** 20320-20325.

- Man – in yeast

- GlcNAc – single unit; INSIDE the cell

J. Am. Soc. Mass Spectrom. **7**, *1996, 319-328.*

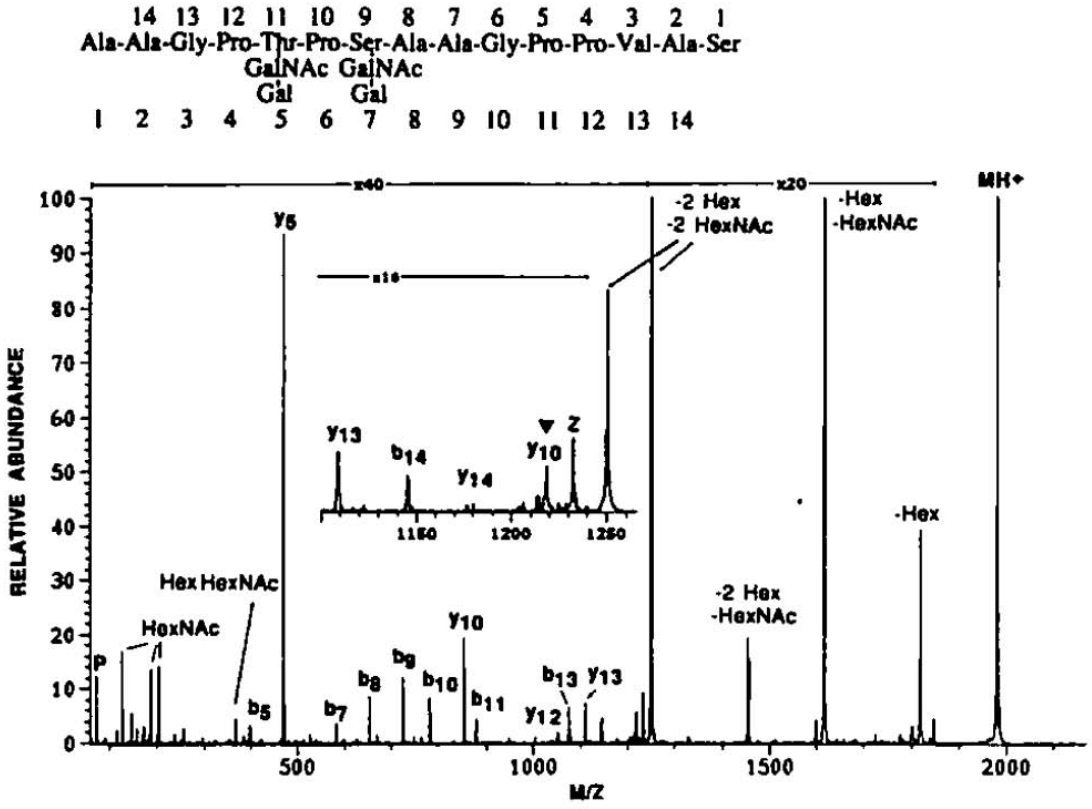


Figure 2. High energy CID spectrum of an asialo fetuin glycopeptide isolated from a tryptic digest. $MH^+ = 1980.8$. Fragment ions are labeled as in Figure 1. Fragment y_{10} (indicated in the inset) shows the presence of the oligosaccharide on the Ser(7) residue.

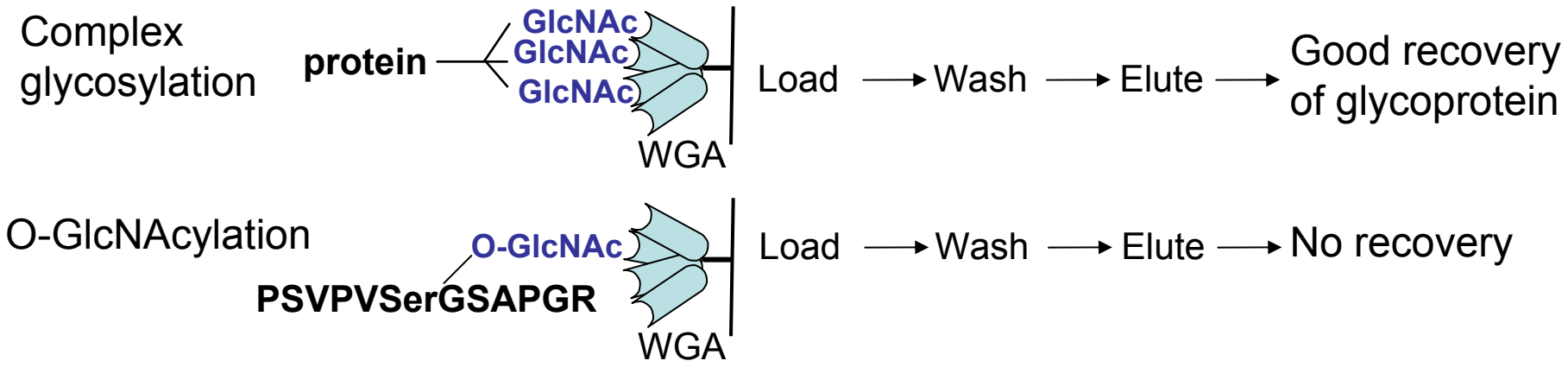
CID fragmentation of O-linked glycopeptides

O-linked GlcNAc

- Regulatory modification of nuclear and cytoplasmic proteins
- Poorly understood due to lack of effective methods for enrichment and detection.

The Enrichment Problem

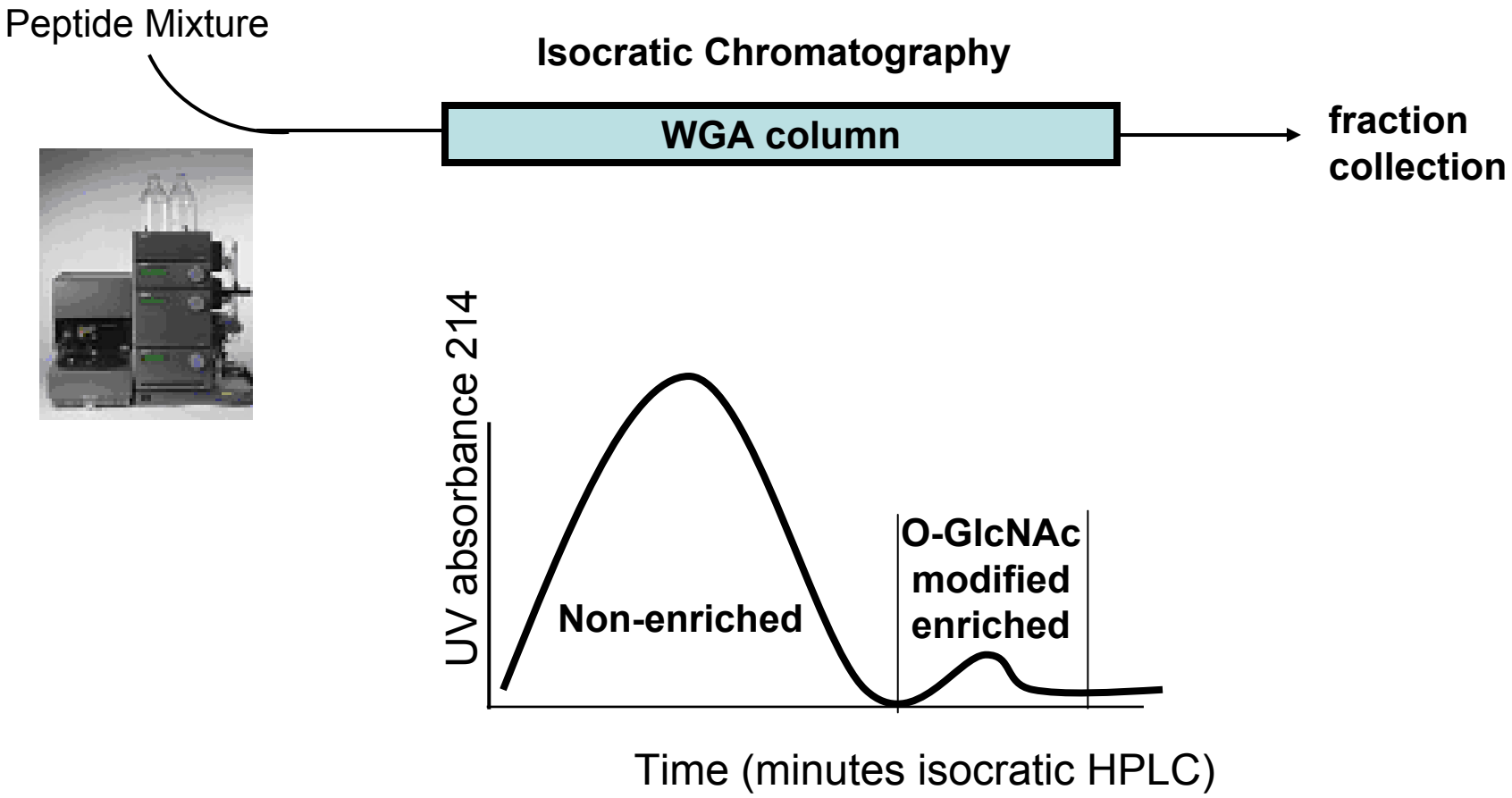
- WGA lectin has affinity for GlcNAc, but affinity to a single GlcNAc moiety is low: millimolar[1].



Ohlson S et al. *Bioseparation* (1998) **7** p.101

O-GlcNAc Enrichment

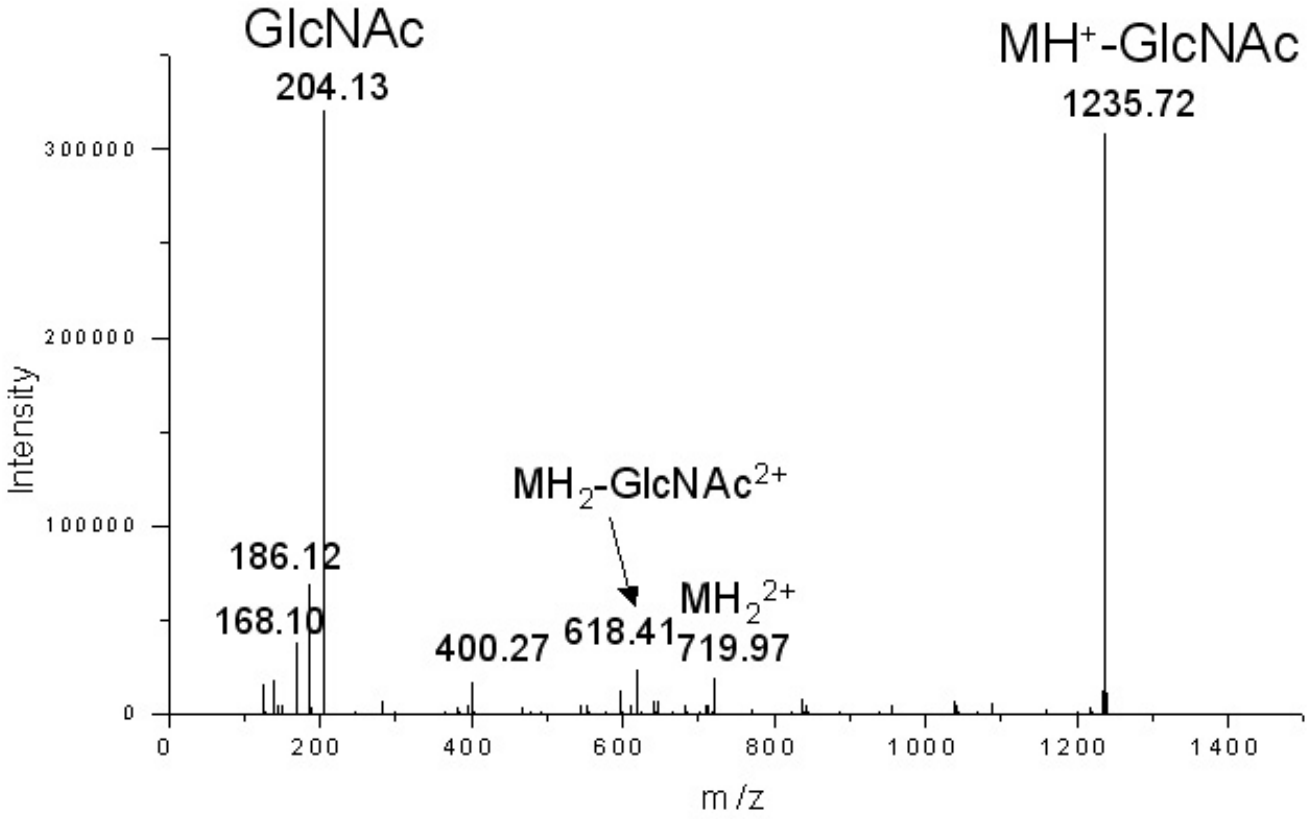
•Selective Enrichment of O-GlcNAc modified peptides using lectin weak affinity chromatography[1].



Vosseller, K. et al. *Mol Cell Proteomics* (2006) **5** 5: p.923-934

CID Analysis of O-GlcNAc-Modified Peptides

- O-glycosidic link is significantly more labile under CID conditions than peptide backbone.
  - Modification site identification using CID often not possible.

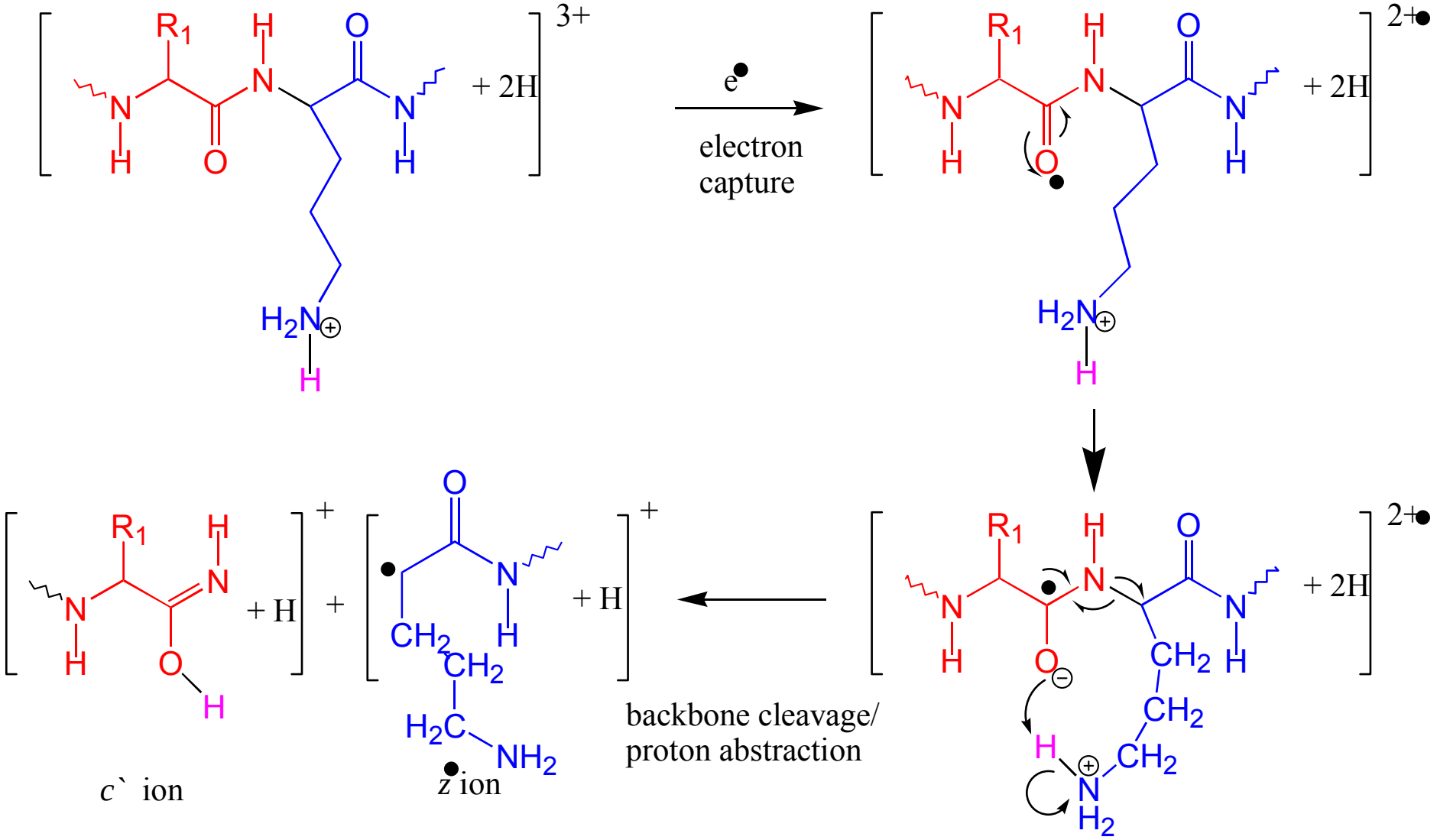


Chalkley, R. J. and Burlingame, A. L. *J. Am. Soc. Mass Spectrom.* (2001) **12** p.1106-1113

A bit about MS/MS alternatives

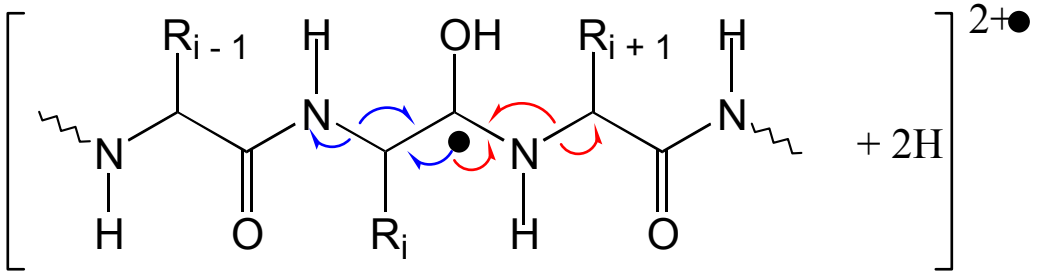
- ECD (electron-capture dissociation) – multiply charged ions meet electron beam in FT-ICR – larger the charge state larger the capture's efficiency
- ETD (electron-transfer dissociation) – multiply charged ions meet stable anion (fluoranthene) in ion traps
- radical ion is formed, different mechanism $\rightarrow$ mostly backbone cleavages

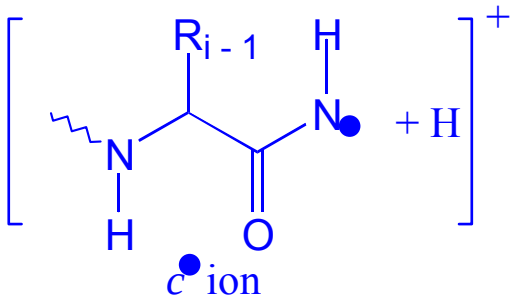
Mechanism of ECD I

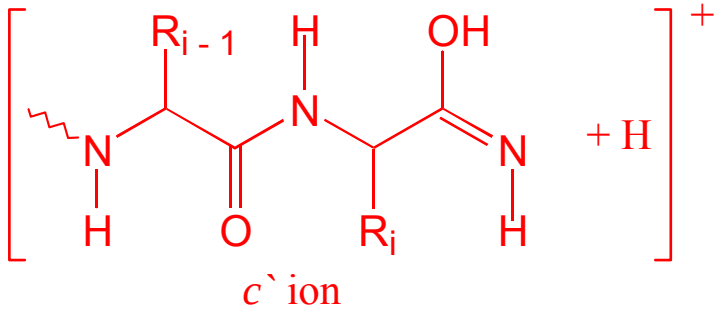


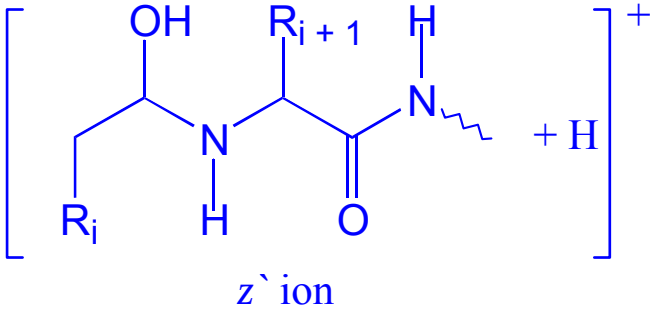
Bakken, V.; Helgaker, T.; Uggerud, E., *Eur. J. Mass Spectrom*. 2004, **10,** 625-638

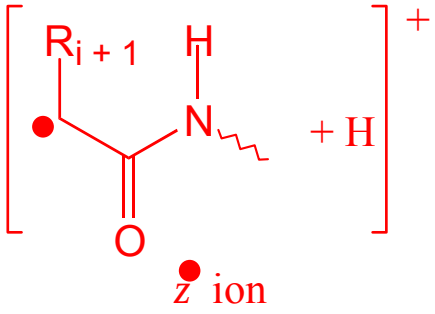
Mechanism of ECD II



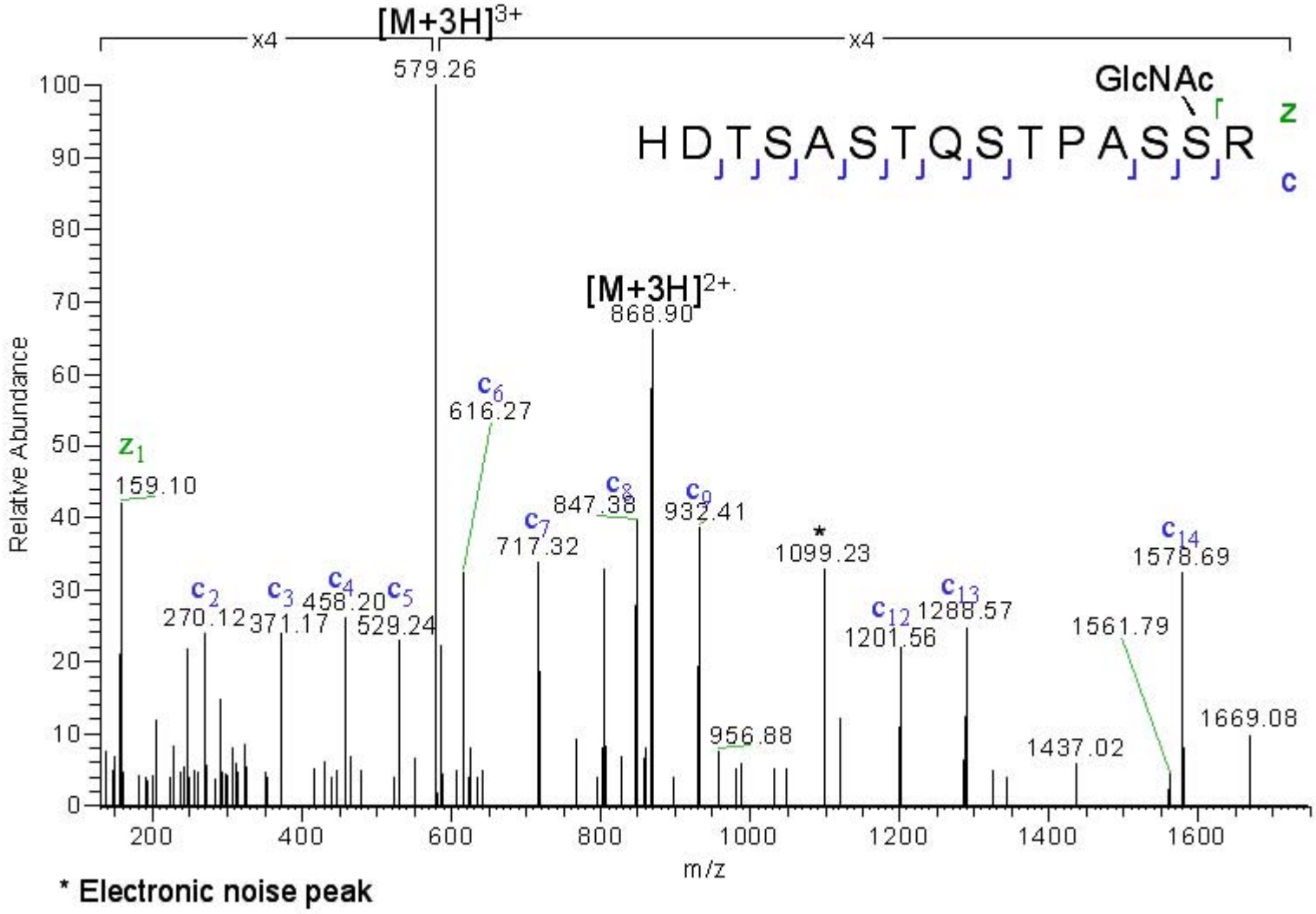








ECD MS Spectrum of GlcNAc-modified Peptide from Spectrin



Vosseller, K et al. *Mol Cell Proteomics* (2006) **5** 5: p.923-924

Phosphorylation
Biological significance

One of the most important regulatory events:

- turns proteins on and off
- induces or prevents other post-translational modifications in the same protein
- signaling pathways : phosphorylation cascades

Difficulties

a) Dynamic process : kinase vs. phosphatase

   - *both* must be blocked during isolation

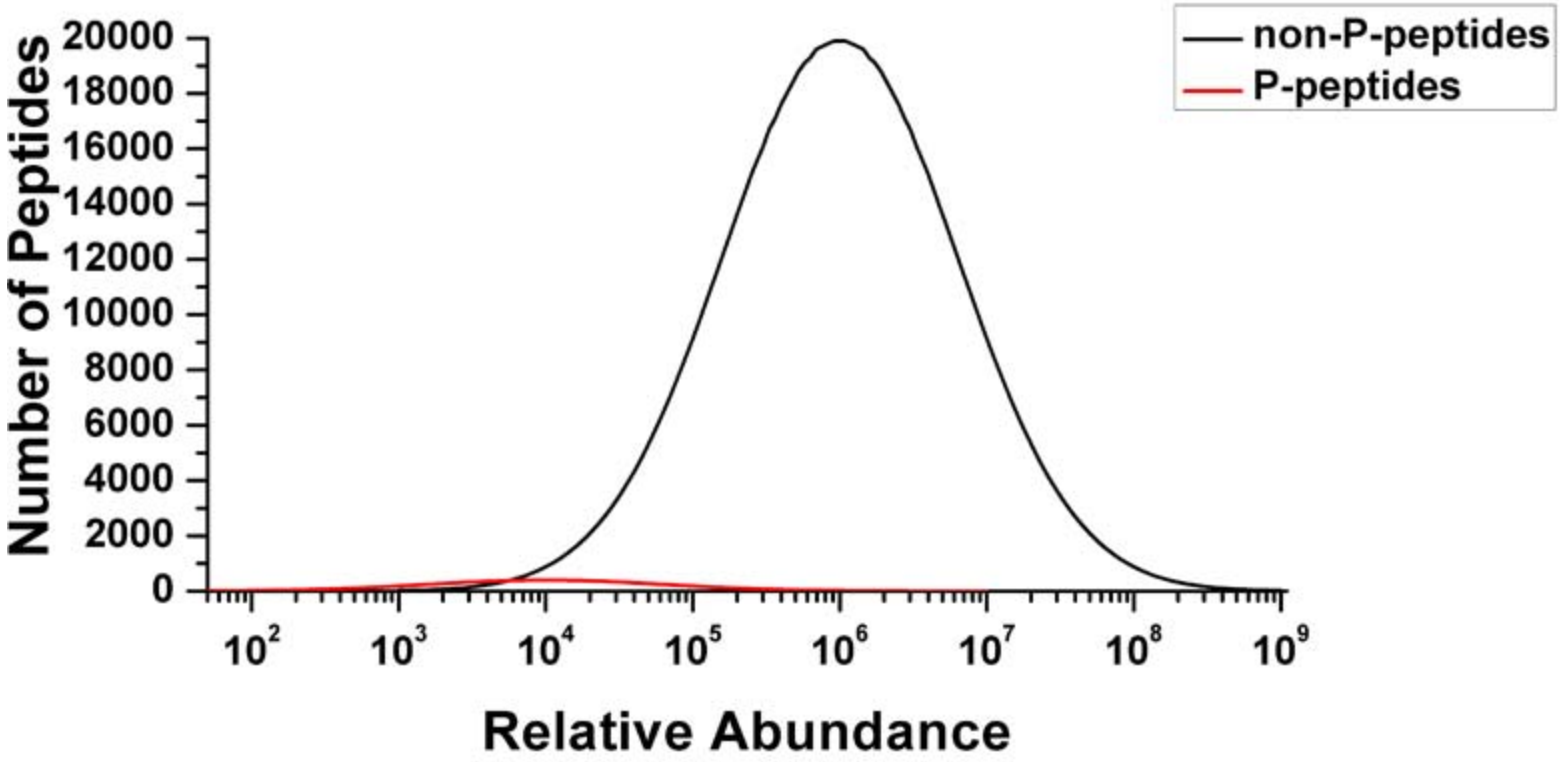
b) Phosphorylation often @ low level (<5%)

c) Lower ionization efficiency – signal of phosphopeptides suppressed

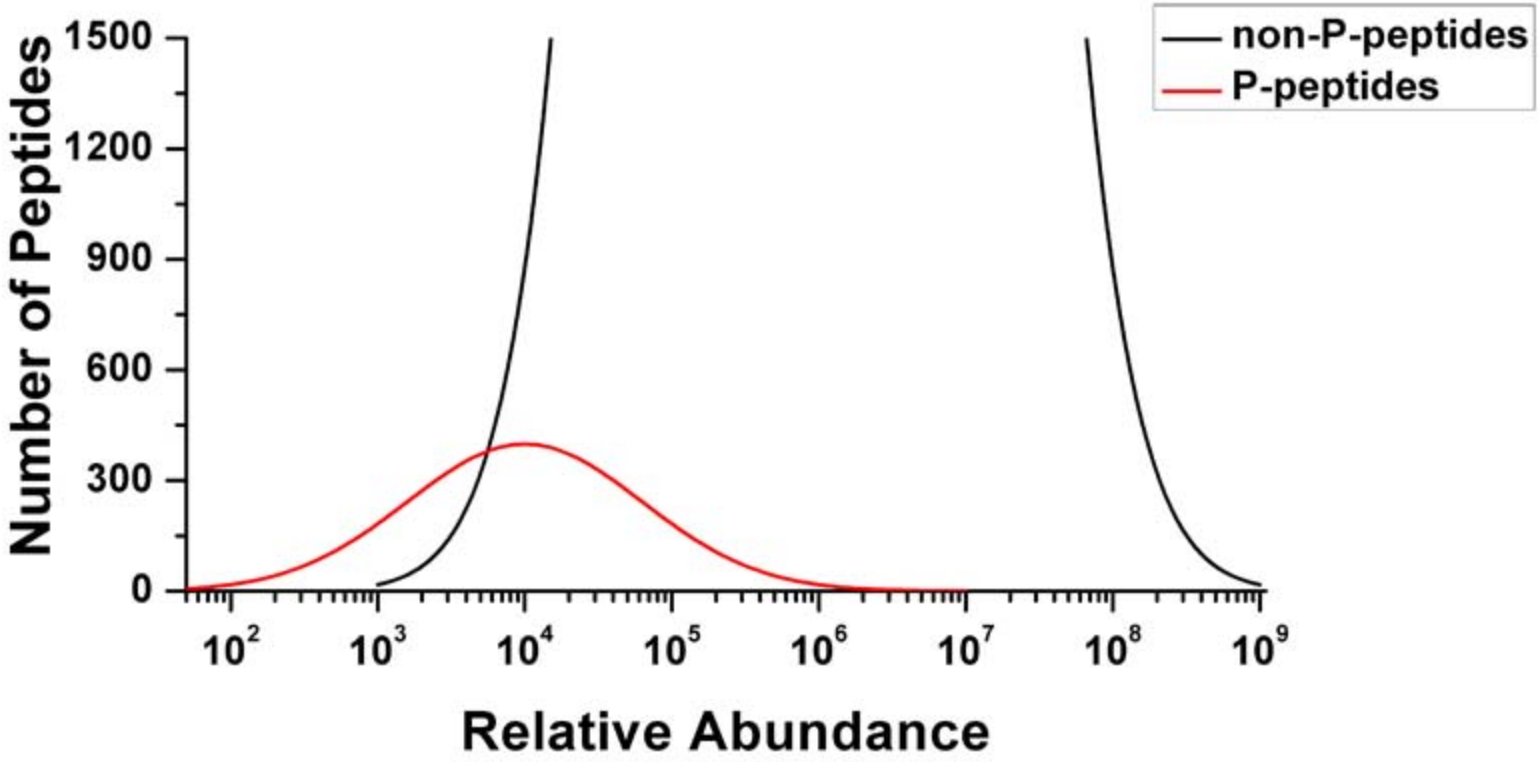
→ Enrichment is a must at protein level

at peptide level

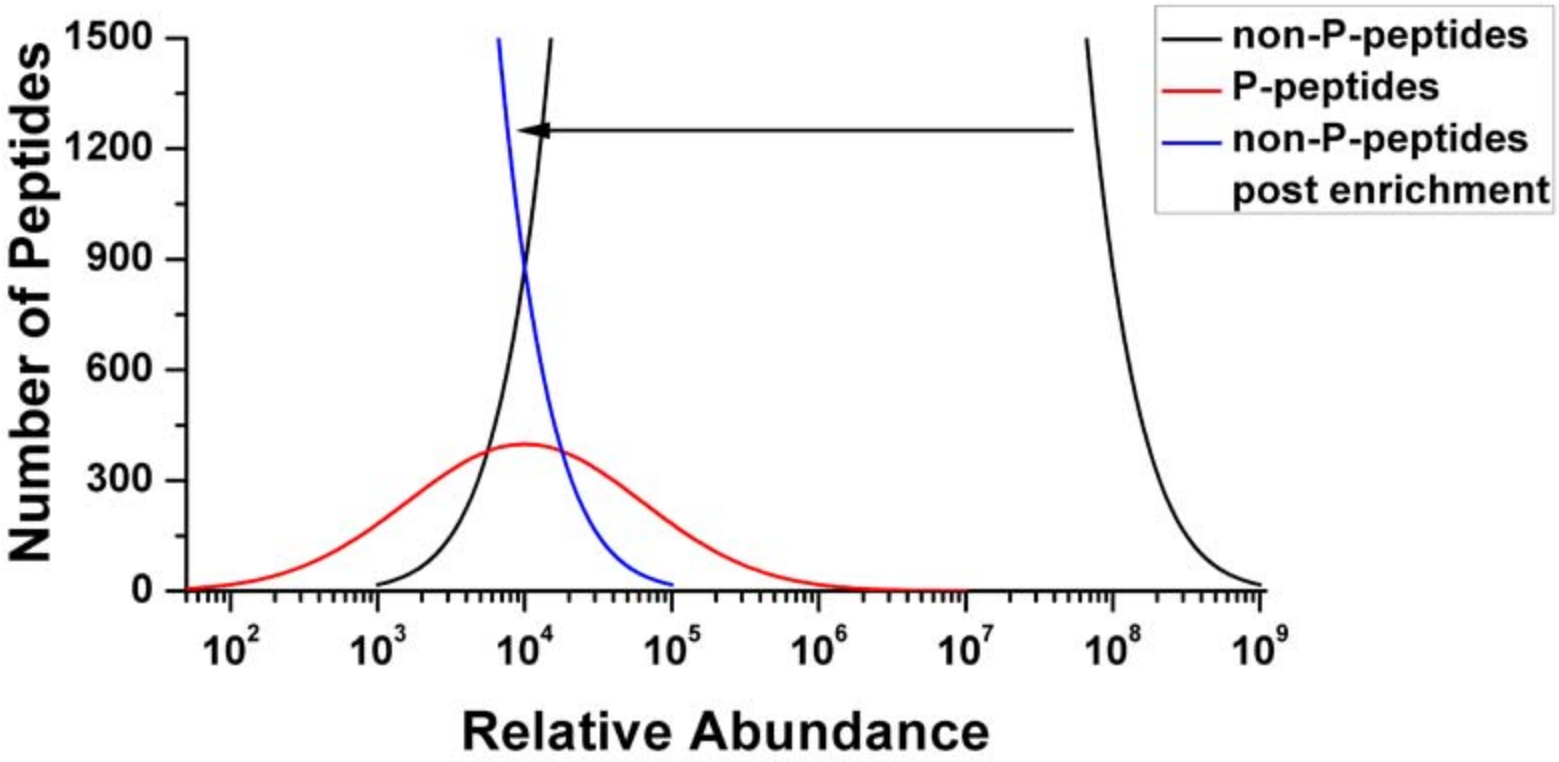
A whole cell lysate with 20,000 sites of phosphorylation at 1% stoichiometry



A whole cell lysate with 20,000 sites of phosphorylation at 1% stoichiometry



phosphopeptide relative distribution after a 10,000 fold enrichment



Confirming the presence of phosphorylation

- Western blot
  - ✓ pTyr large enough for sequence independent recognition - works well
  - ✓ pSer, pThr – not reliable
- Dyes – questionable reliability
- Phosphatase treatment + isoelectric focusing – pI shift

- *In vitro/in vivo* assay with radioactive phosphate

Determining site of phosphorylation

- Mikrosequencing
- Mutation studies
- Mass spectrometry
  - MS spectrum : + 80 Da shift
  - MS/MS fragmentation :
    - pSer, pThr - H_3PO_4 loss : -98 Da
    - pTyr – always retains the modification immonium ion at m/z 216 Da

Enrichment methods

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

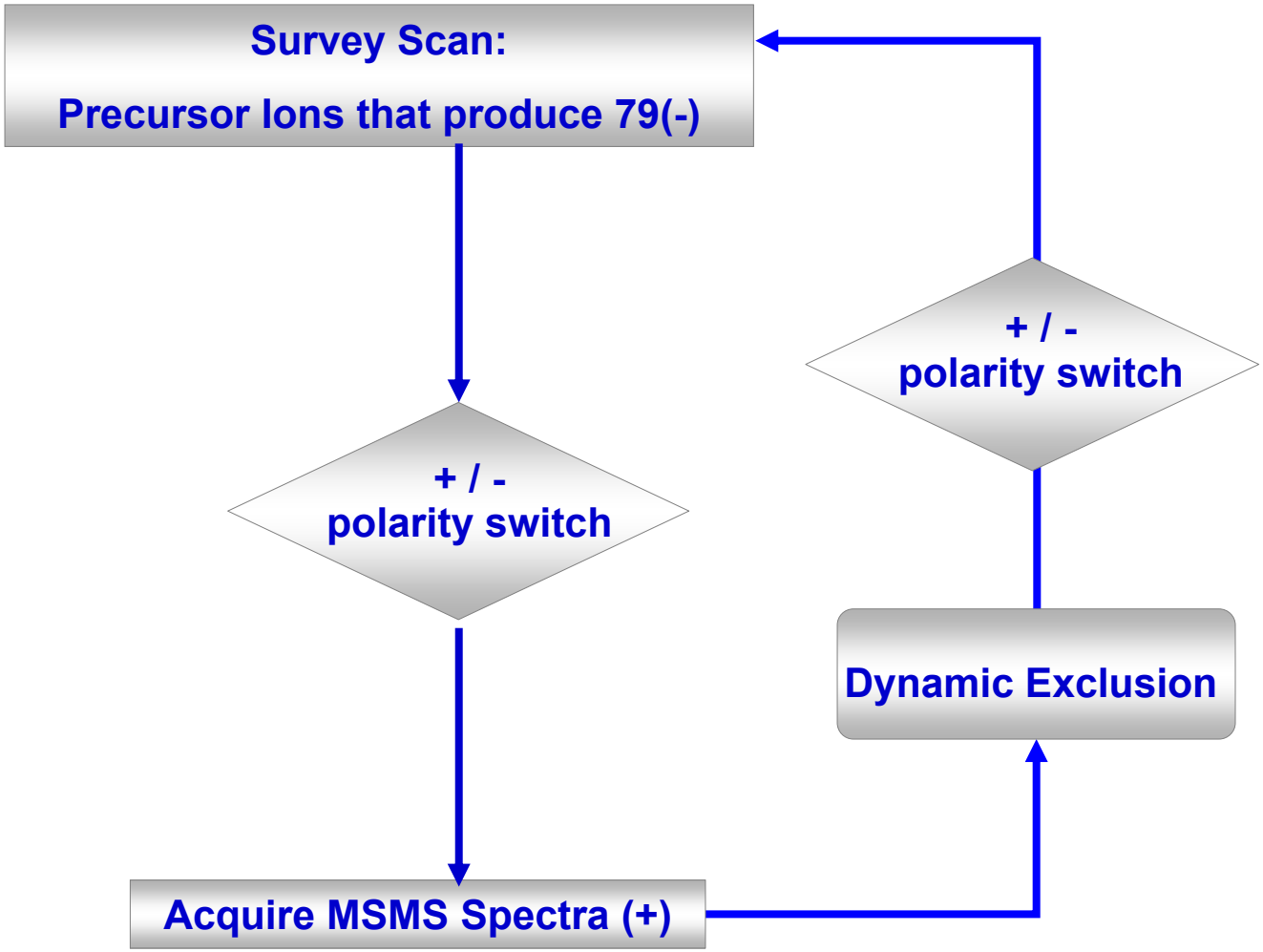
- Immunoprecipitation (only for pTyr)

Large scale vs. 1 protein

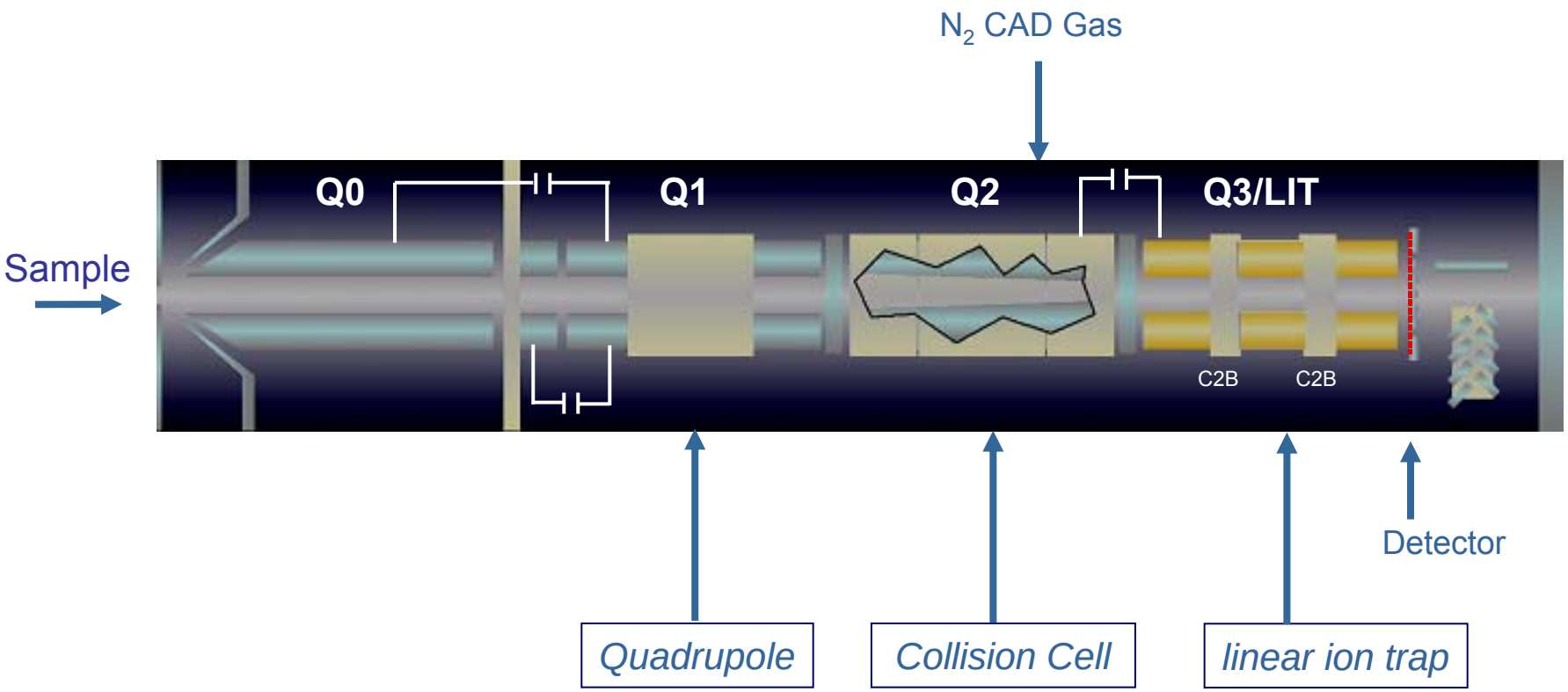
- large scale phosphorylation studies:
  - ✓ large amounts of a complex mixture is analyzed compensating for losses during sample preparation
  - ✓ leads to more PTM identification *but*
  - ✓ low end results are incidental

  - ✓ single protein samples:
  - ✓ sample amount is usually limited
  - ✓ more challenging
  - ✓ combination of techniques may be necessary

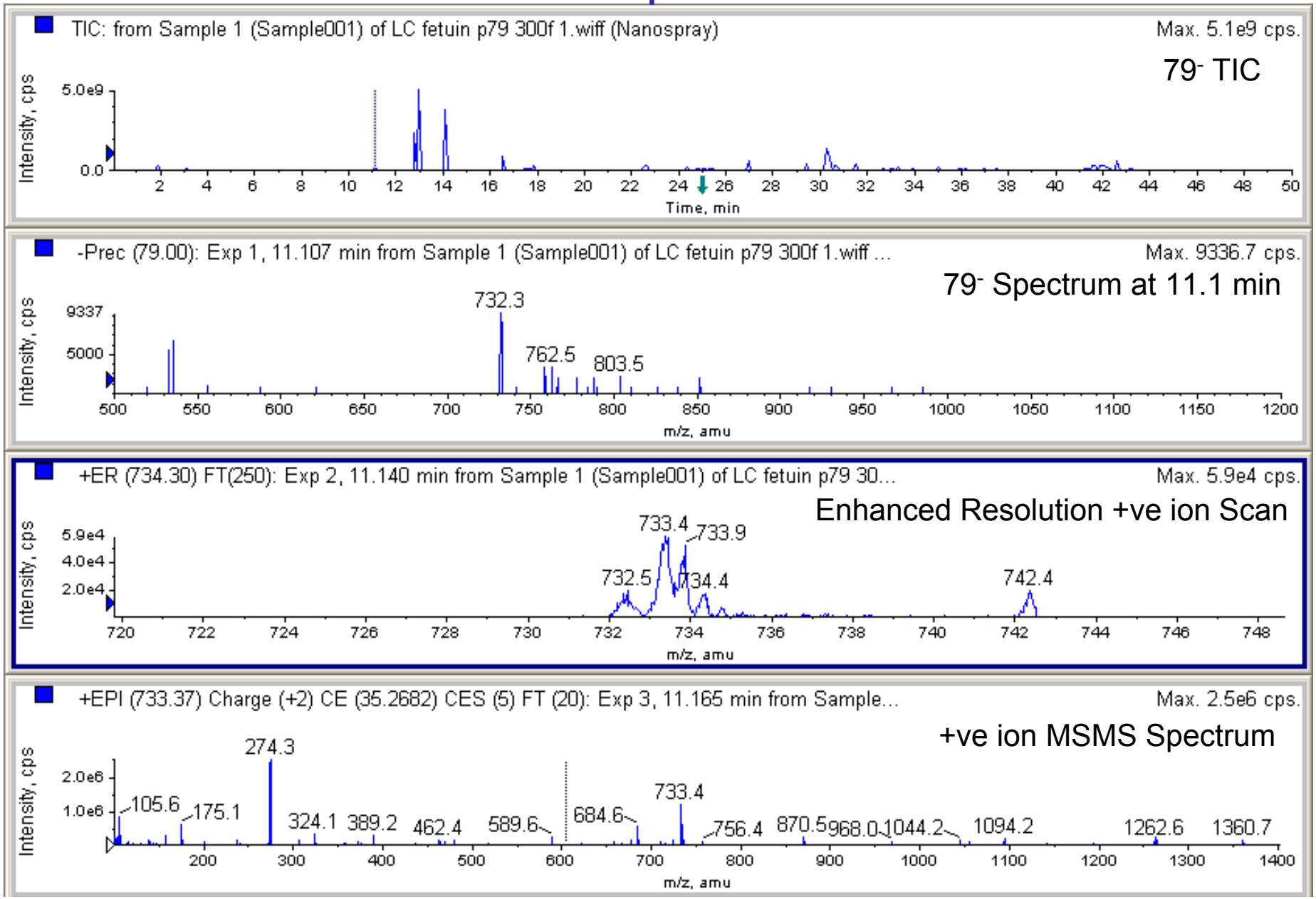
Phosphorylation Discovery Workflow



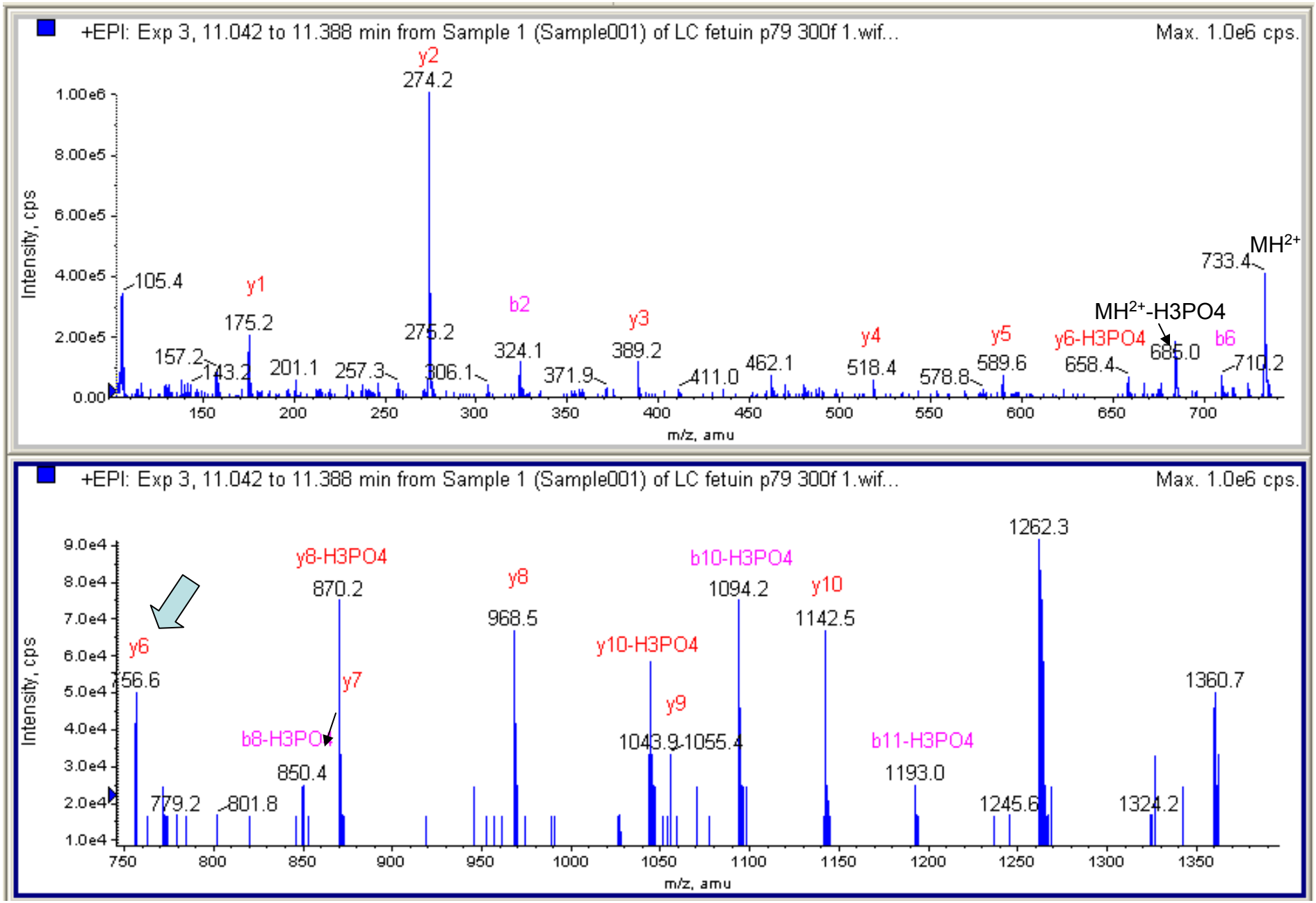
4000 Q TRAP


N_2 CAD Gas
Sample
Q0
Q1
Q2
Q3/LIT
C2B
C2B
Detector
Quadrupole
Collision Cell
linear ion trap

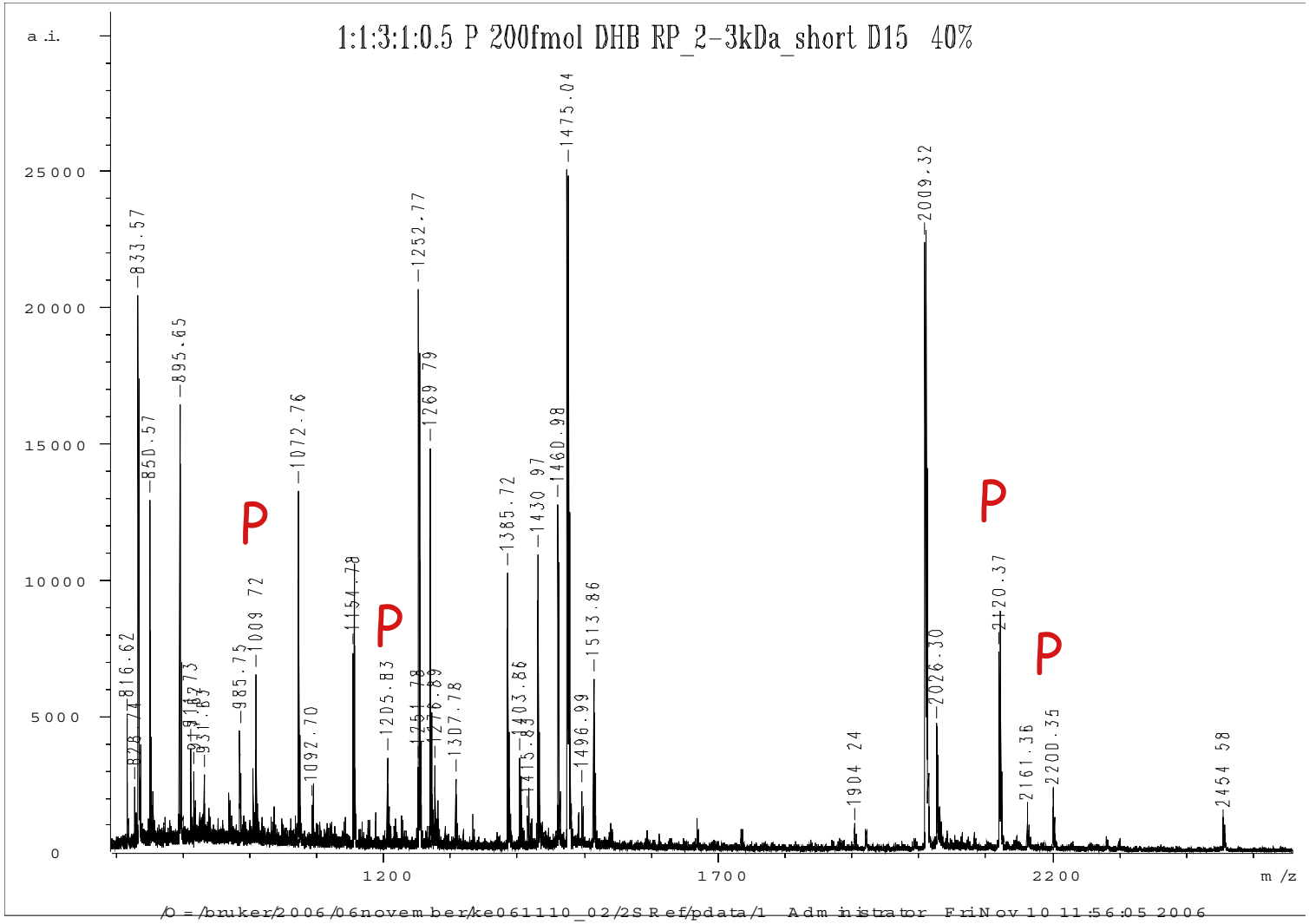
LCMS 300fmole Fetuin: Discovery and Identification of CDSSPDpSAEDVR



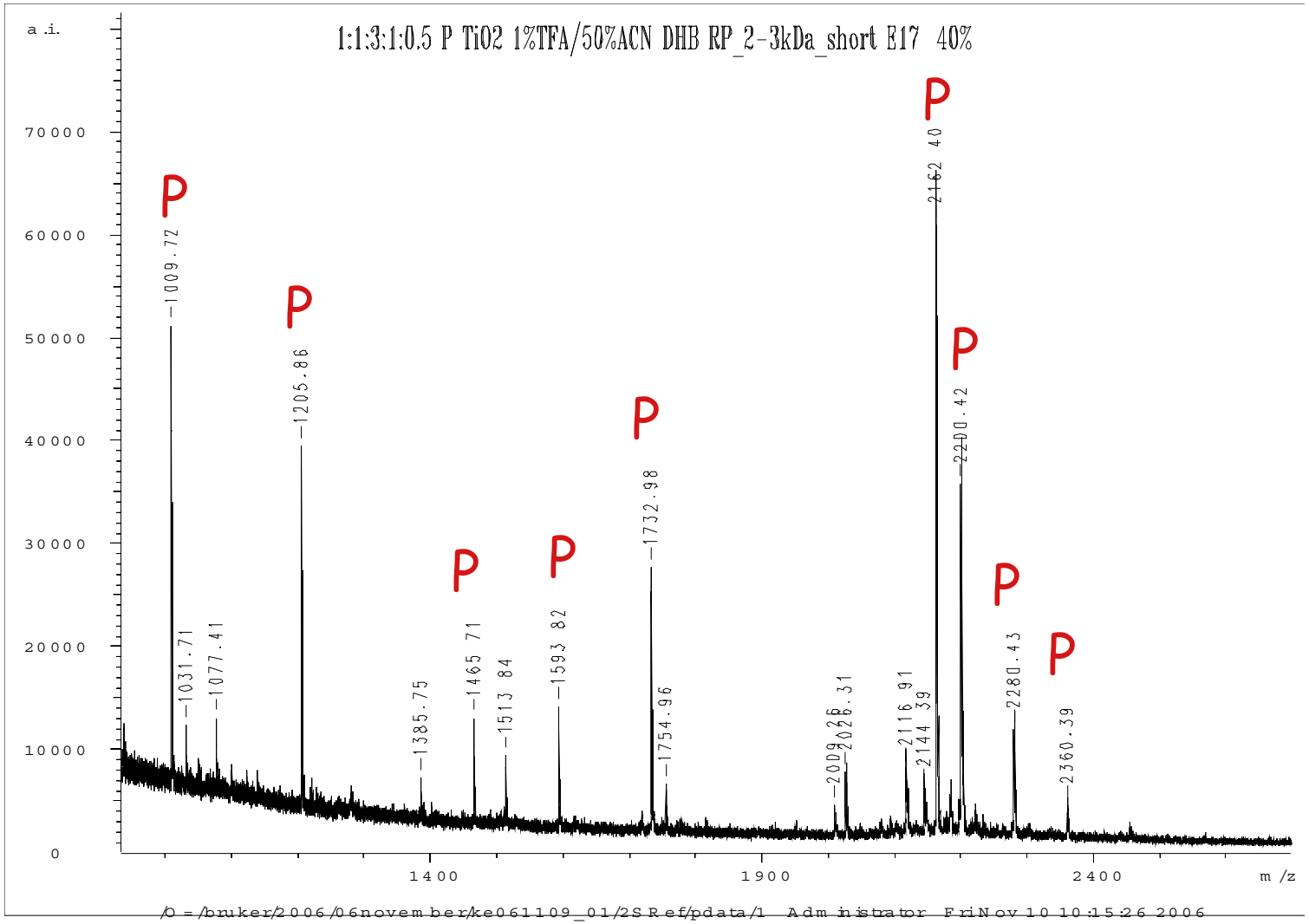
CDSSPDpSAEDVR



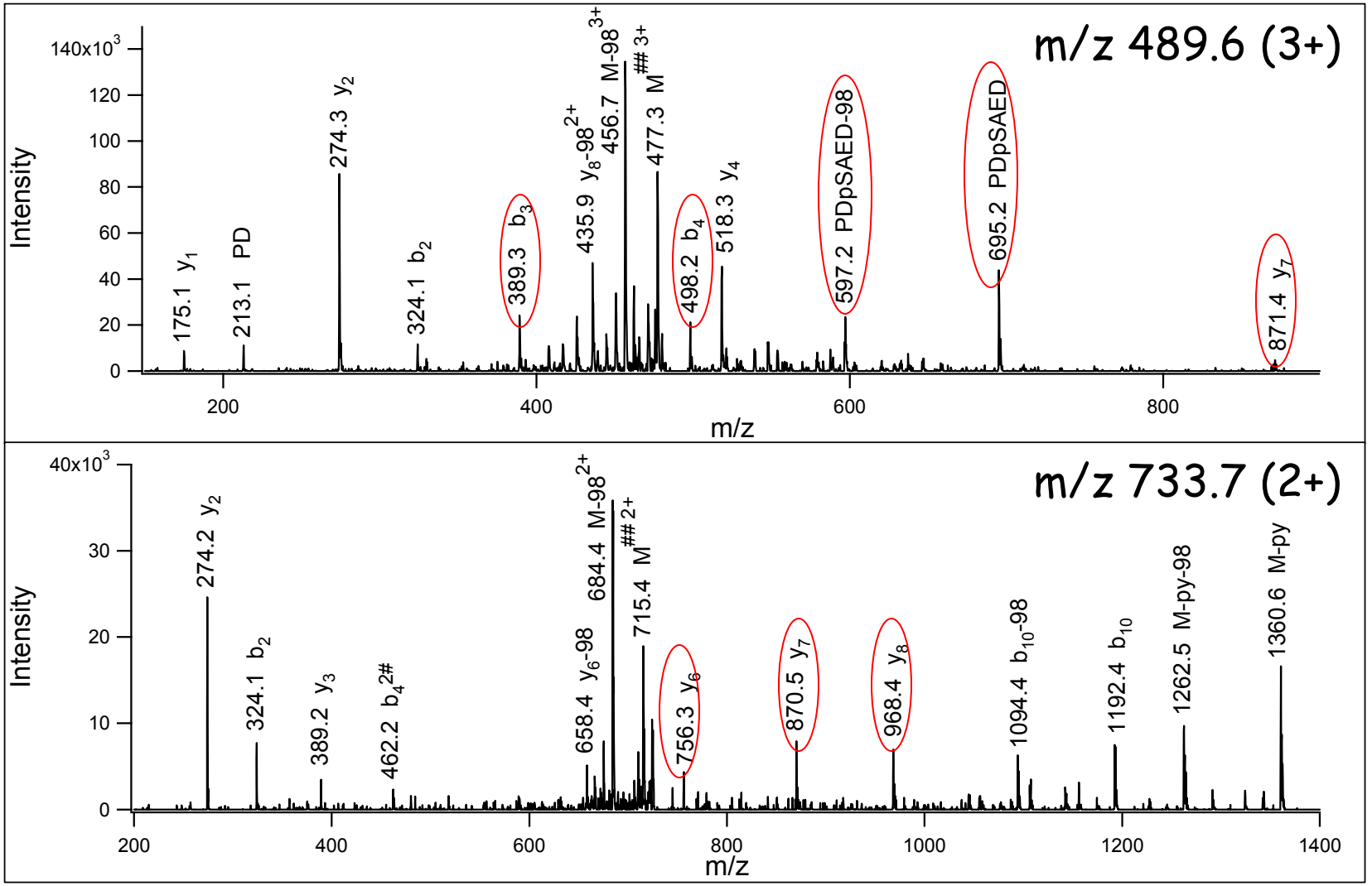
MALDI-TOF MS without enrichment



Enrichment with TiO_2



CID of CDSSPDpSAEDVR



+80 Da: phosphate? sulfate?

Phosphorylation

- Tyr, Ser, Thr, (His)
- Phosphopeptides are stable in MS (except His)
- Tyr – no phosphate loss in CID
- Ser, Thr – H_3PO_4 (98 Da) loss in CID

Sulfation

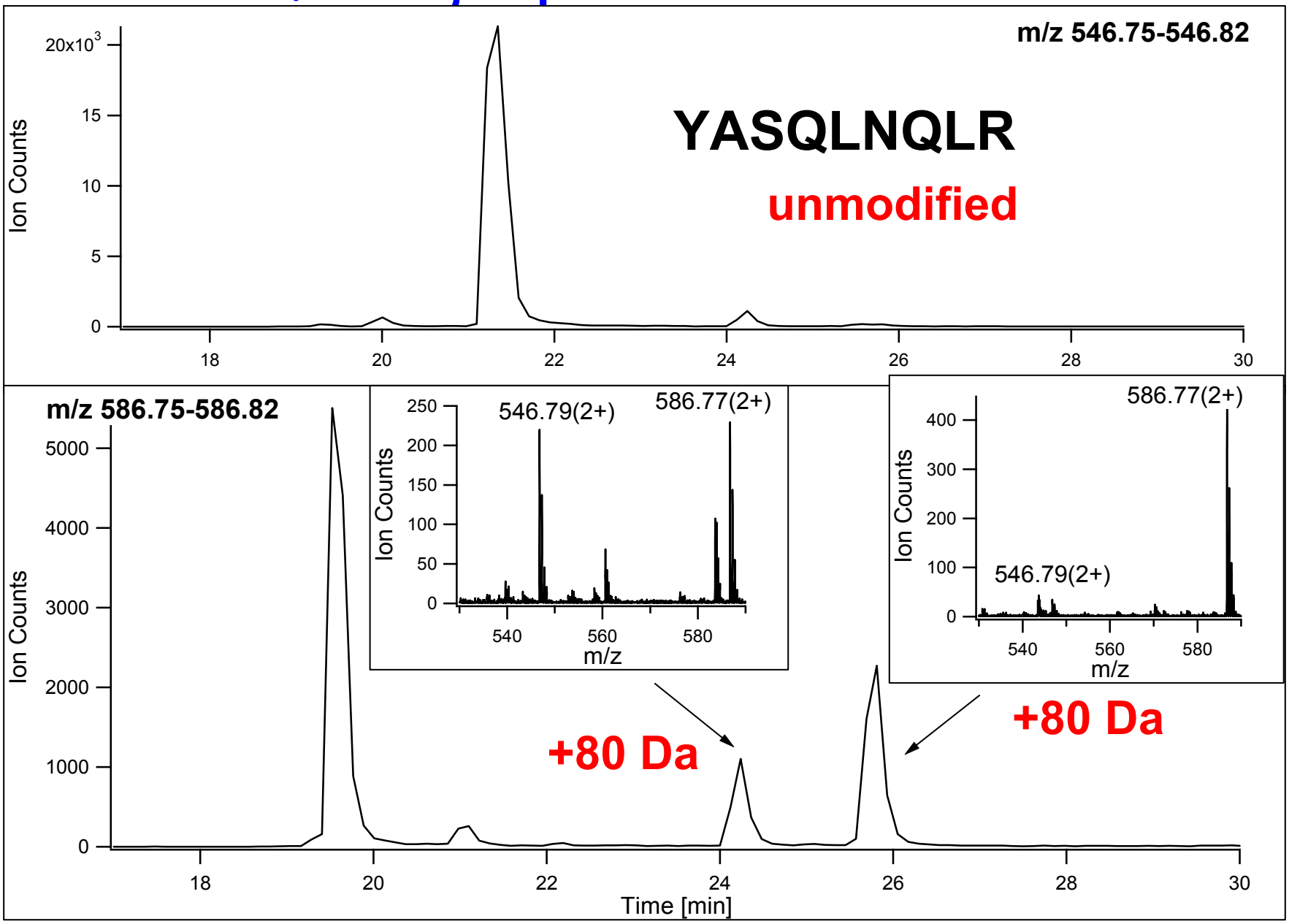
- Tyr only
- Significant SO_3 (80 Da) loss even in MS

FTMS accurate mass measurement will tell the difference

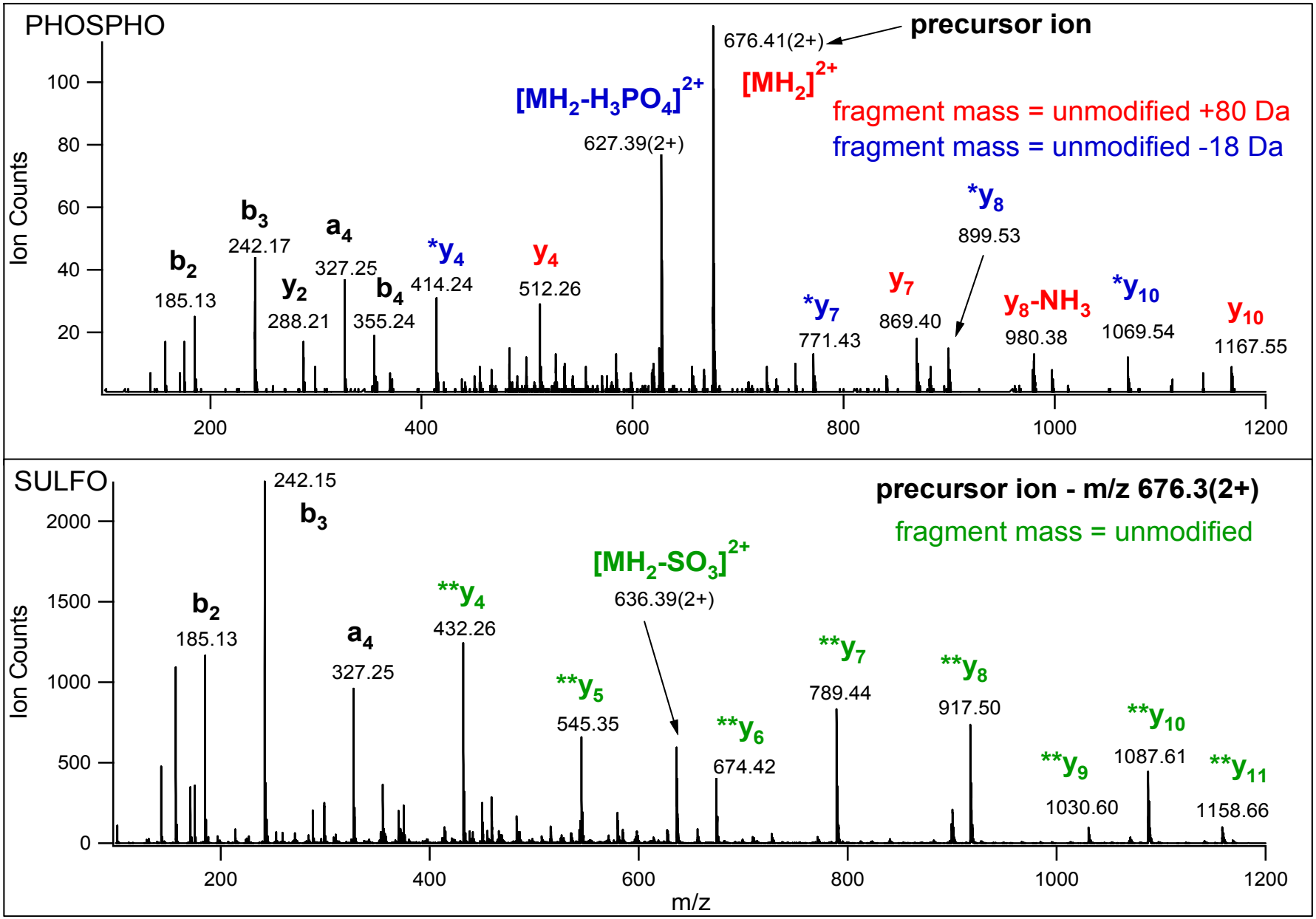
Ion detected	MH^+ determined	sequence	MH^+ calculated	Δ[ppm]
501.27454(3+)	1501.808	GKFDYNTFVGI/LI/LK	1501.8055	+1.6
512.26184(4+)	2046.0239	AcAEEKQGRHTTNVI/LSMFR	2046.0191	+2.3
516.26027(4+)	2062.0176	AcAEEKQGRHTTNVI/LSM(O)FR	2062.0140	+1.7
603.31016(2+)	1205.6125	HTTNVI/LSMFR	1205.6101	+1.9
611.30791(2+)	1221.608	HTTNVI/LSM(O)FR	1221.6050	+2.4
643.28905(2+)	1285.5703	sulfo-HTTNVI/LSMFR	1285.5669	+2.6
		phospho-HTTNVI/LSMFR	1285.5764	-4.7

myosin regulatory light chain, *Lymnaea stagnalis*

Chromatographic and MS-behavior



CID comparison of phospho- and sulfo-LAGLQDEIGSLR



Sulfopeptides easily can be identified as phosphopeptides!

- + 80 Da, 9 mmu difference only
- "identical" behavior by ESIMS, chromatography and under basic conditions.
- different CID fragmentation

K.F. Medzihradszky et al.,
Mol. Cell. Proteomics, May 2004; 3: 429 - 440.

An exciting new cofactor

Reza Ghiladi
Paul Ortiz de Montellano

Catalase-peroxidase

Removes harmful peroxide molecules

$$2\ H_2O_2 \rightarrow H_2O + O_2$$

Catalase

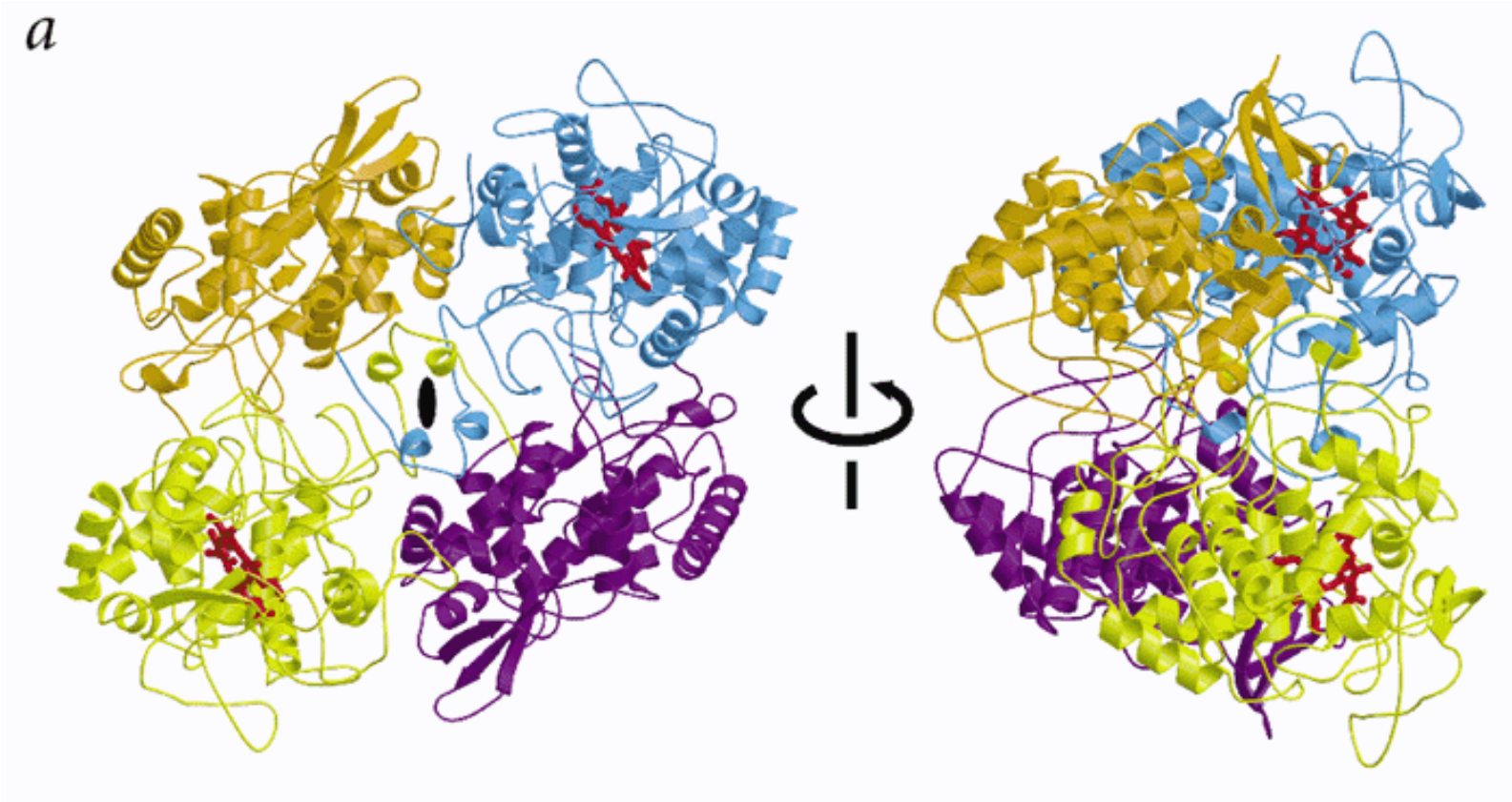
H_2O_2 alternately oxidizes/reduces the heme Fe

Peroxidase

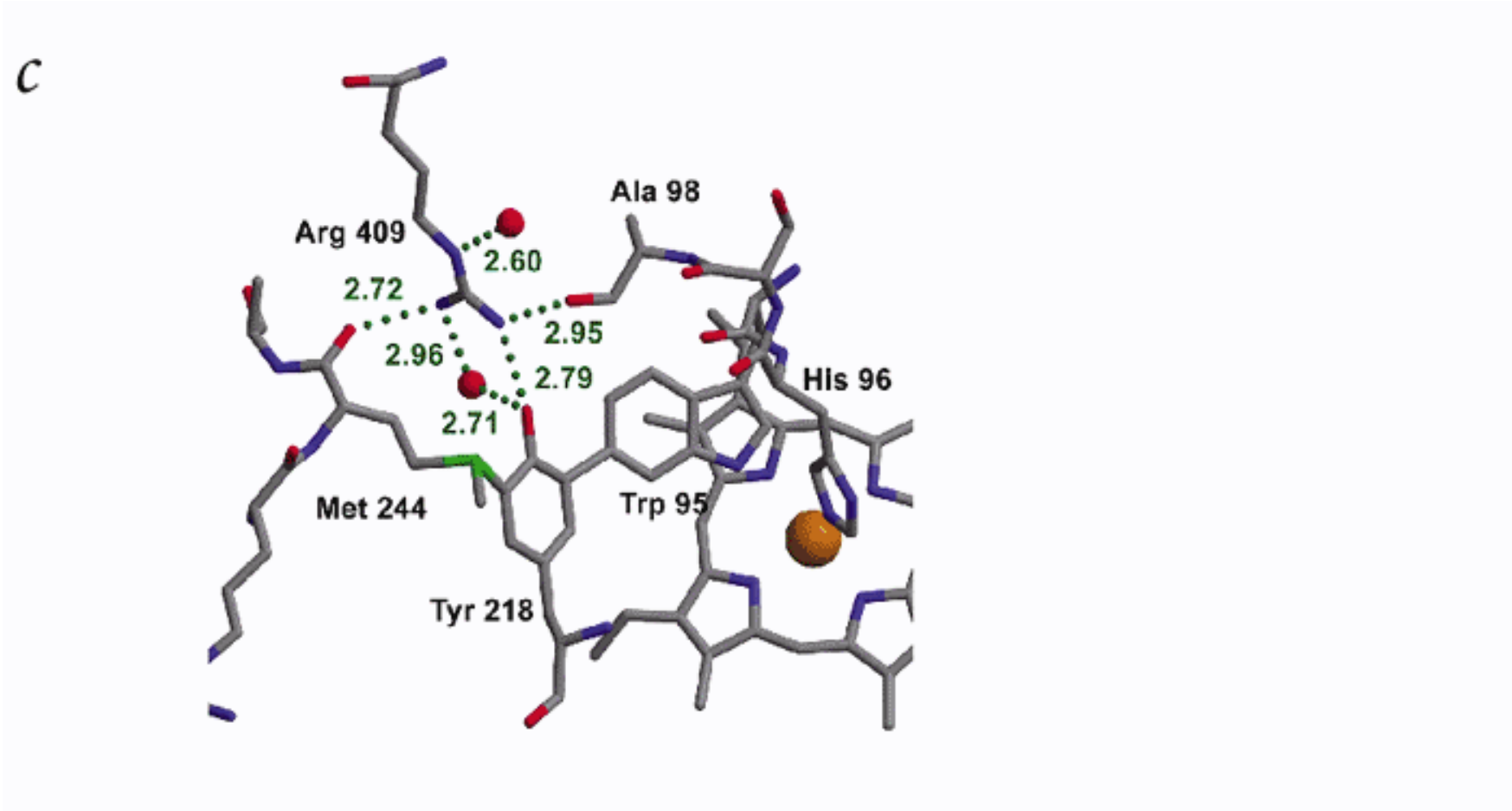
heme is oxidized by H_2O_2;
reduction involves H-donor, such as NADH

Crystal structure of catalase-peroxidase

Haloarcula marismortui (Yamada *et al. Nature structural biology)*



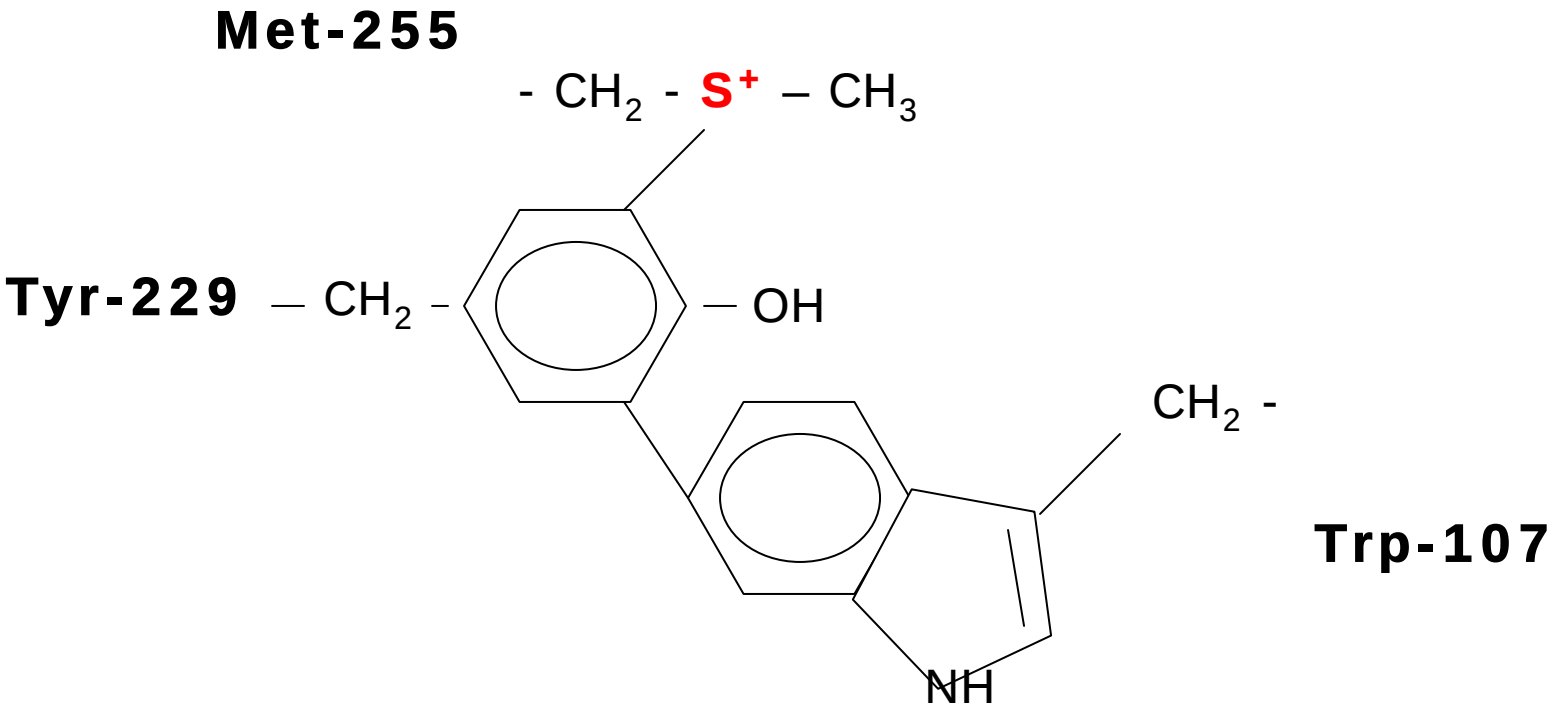
Active site structure



Active site structure

- We studied the *Mycobacterium tuberculosis* enzyme

1) Ghiladi RA,et al., *J Biol Chem*. **280,** 22651-63 (2005).
2) Ghiladi RA, et al, *Biochemistry*, **44**, 15093-105 (2005).
3) Ghiladi RA, et al., *J Am Chem Soc*. **127,** 13428-42 (2005).



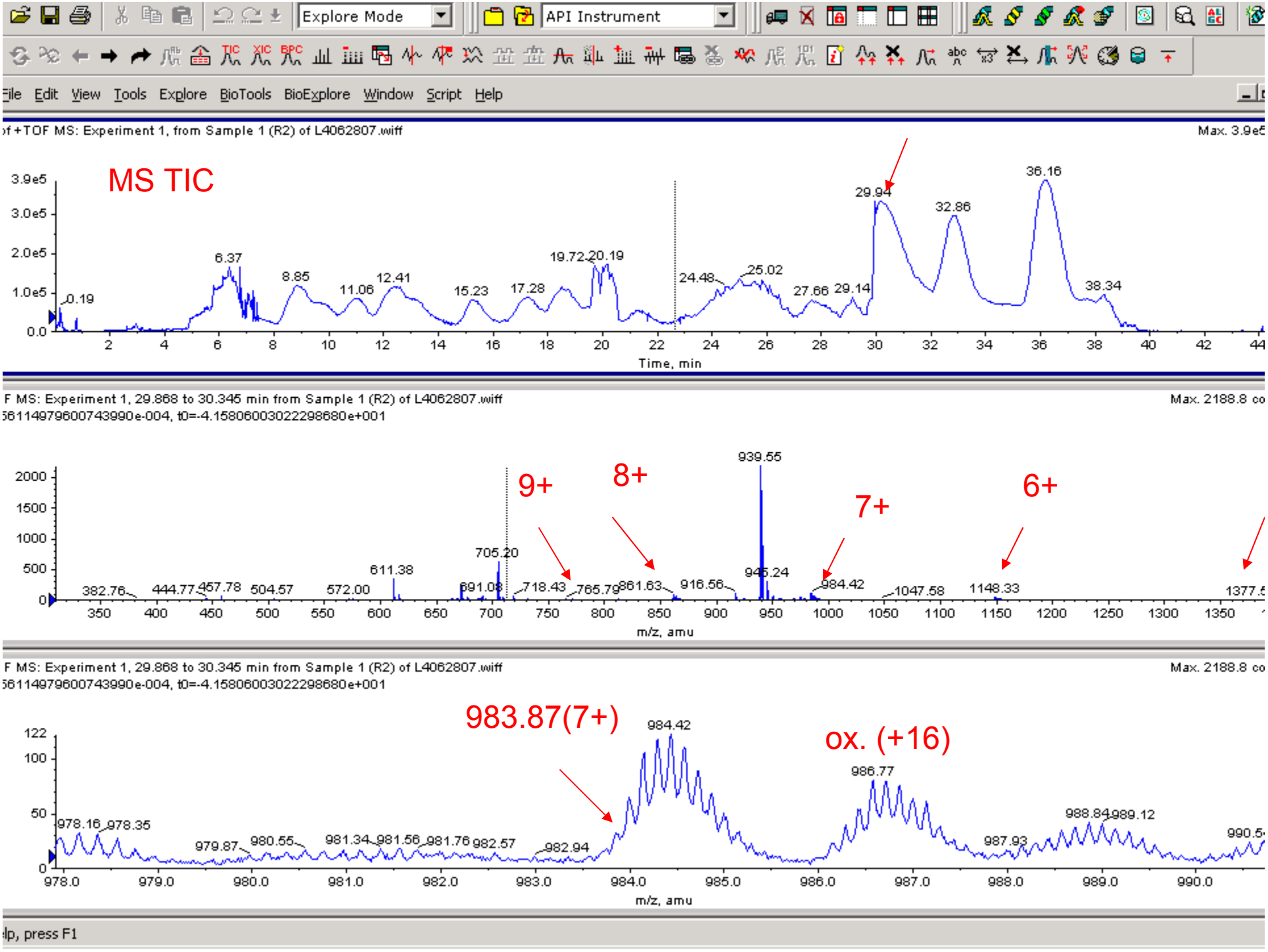
What are we looking for?

* Cross-linked "tri"-peptide

  calculated neutral (zwitter ionic)
  mass: 6880.31 Da

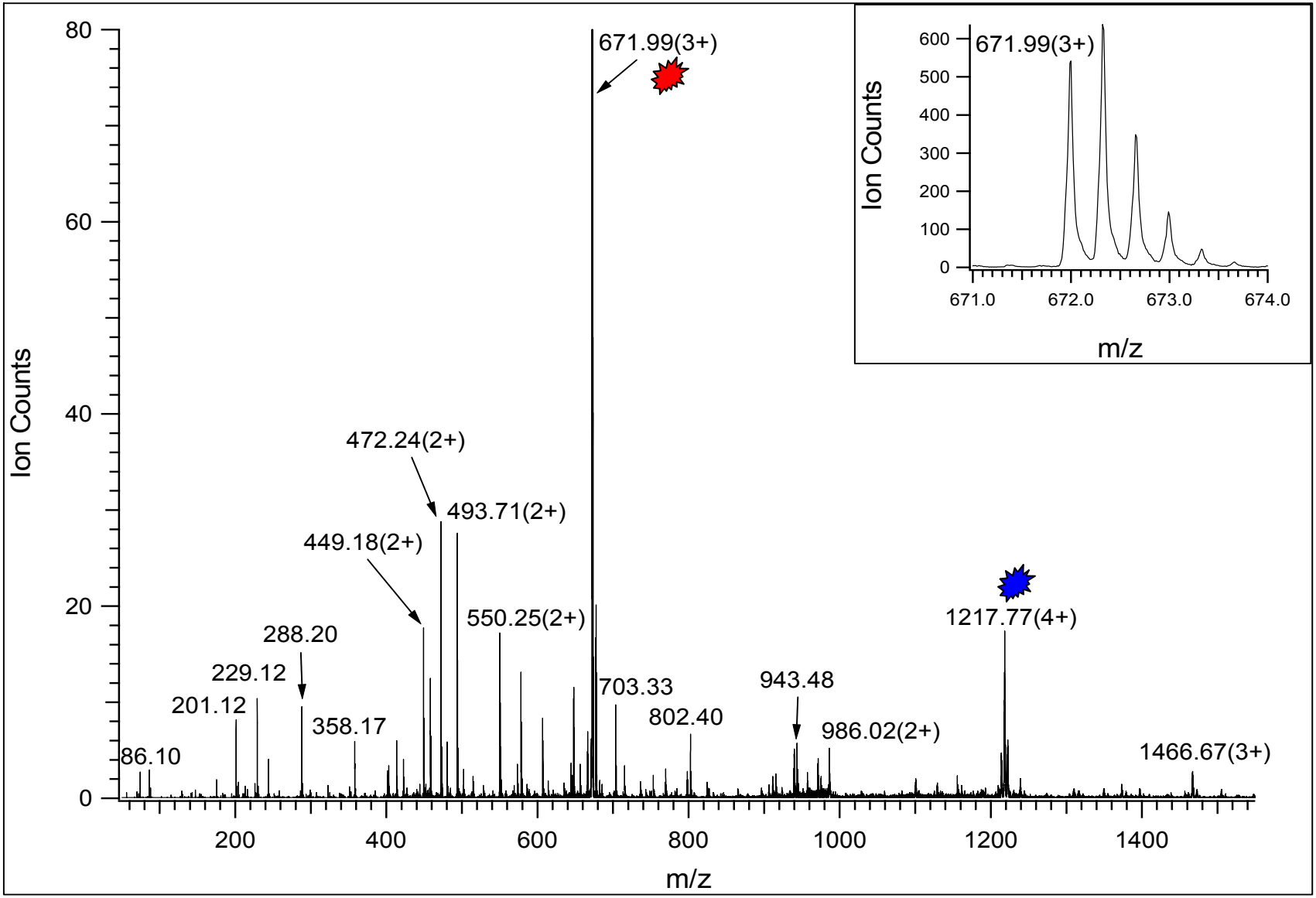
  "measured"

  mass: 6879.99 Da

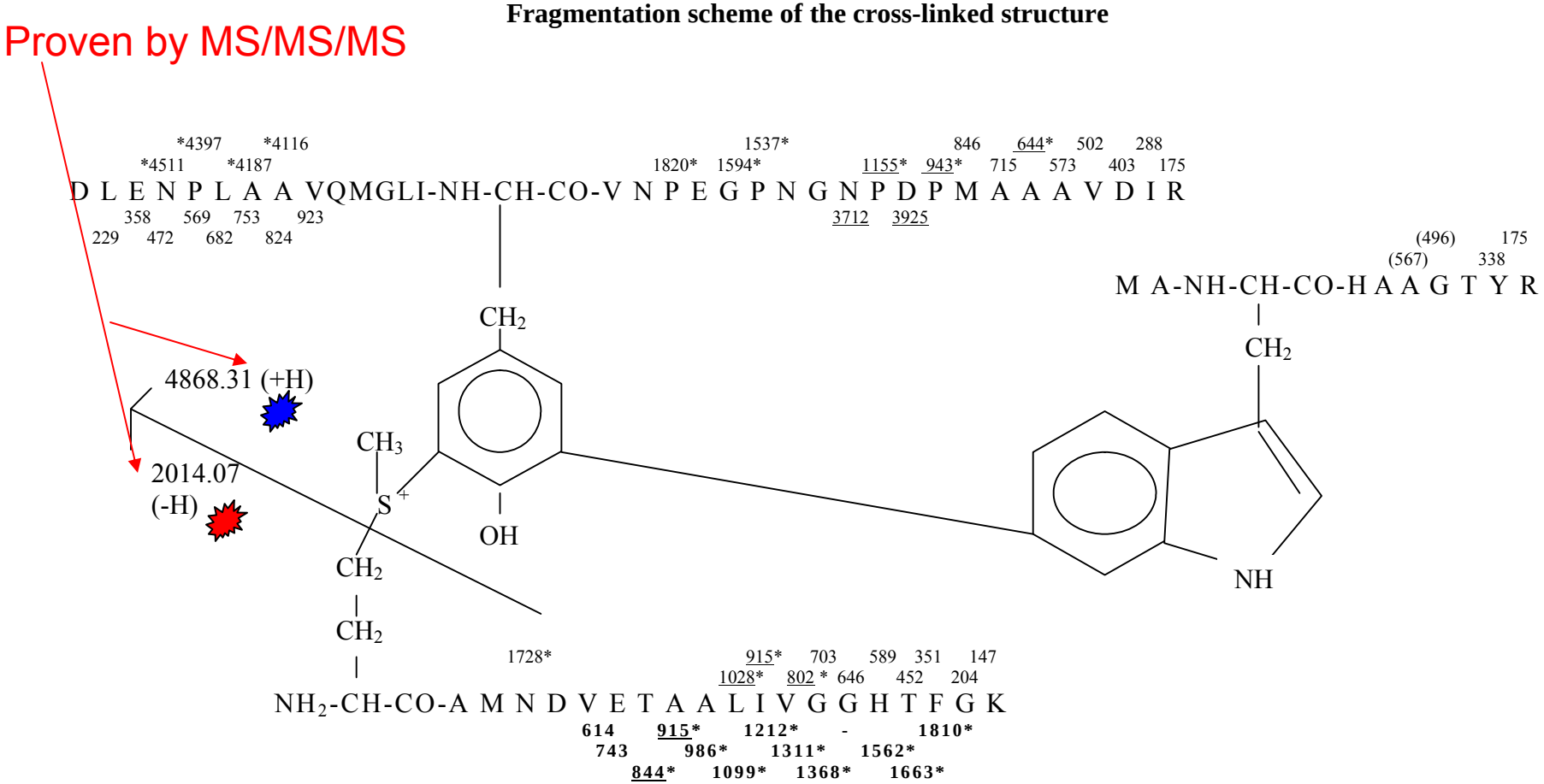

Explore Mode
API Instrument
File Edit View Tools Explore BioTools BioExplore Window Script Help
of +TOF MS: Experiment 1, from Sample 1 (R2) of L4062807.wiff
Max. 3.9e5
MS TIC
3.9e5
3.0e5
2.0e5
1.0e5
0.0
0.19
6.37
8.85
11.06
12.41
15.23
17.28
19.72
20.19
24.48
25.02
27.66
29.14
29.94
32.86
36.16
38.34
Time, min
F MS: Experiment 1, 29.868 to 30.345 min from Sample 1 (R2) of L4062807.wiff
Max. 2188.8 co
561149796007439 90e-004, t0=-4.1580600302229868 0e+001
9+
8+
7+
6+
939.55
705.20
611.38
382.76
444.77
457.78
504.57
572.00
691.05
718.43
765.79
861.63
916.56
945.24
984.42
1047.58
1148.33
m/z, amu
F MS: Experiment 1, 29.868 to 30.345 min from Sample 1 (R2) of L4062807.wiff
Max. 2188.8 co
561149796007439 90e-004, t0=-4.1580600302229868 0e+001
983.87(7+)
ox. (+16)
984.42
986.77
978.16
978.35
979.87
980.55
981.34
981.56
981.76
982.57
982.94
987.93
988.84
989.12
990.5
m/z, amu
lp, press F1

CID spectrum of the "tri"-peptide

precursor: 984.4(7+); intensities magnified by ~7.5

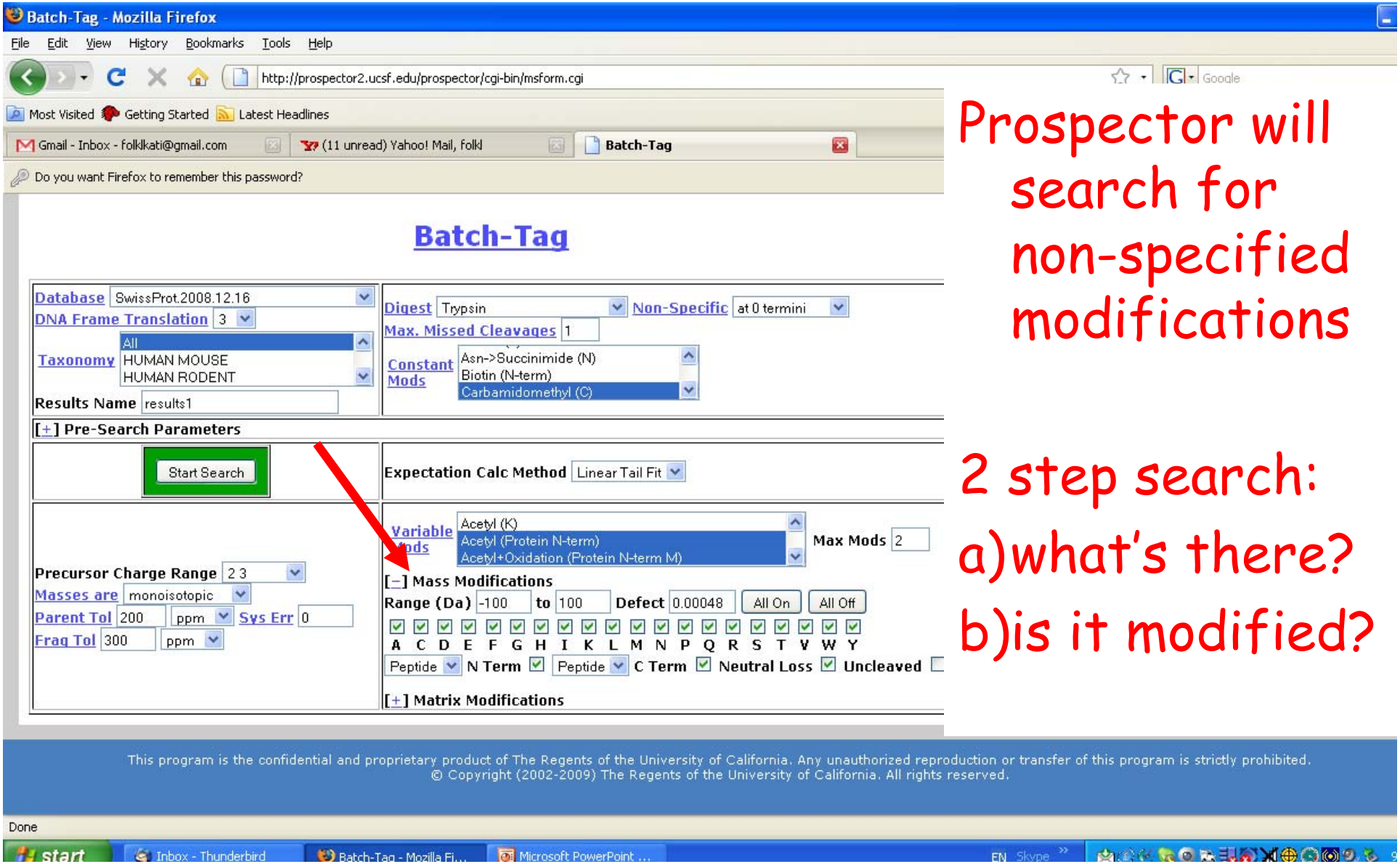


Fragmentation scheme of the cross-linked structure



*4511 – **y**-type ions, cleavage from 4868 – mostly triply charged ions detected
3712 – **b**-type ions, cleavage from 4868
614 – **b**-type ions, derived from 2014
915* - doubly charged ion was detected
915* – singly charged ion was detected too

When we don't have a clue…



Prospector will search for non-specified modifications

2 step search:

a)what's there?

b)is it modified?

Characterization of protein populations

- structural information obtained from peptides

→ conclusions about the proteins

- Analysis of the intact proteins yielding extensive structural information

"top-down" approach

Requirements for Top-Down Analysis

- Reasonable amount of protein – pmoles.
- Protein compatible with ESIMS.
- Relatively simple protein mixture.
- High resolution and mass accuracy instrument - FTMS
- Efficient protein fragmentation: ECD > CID

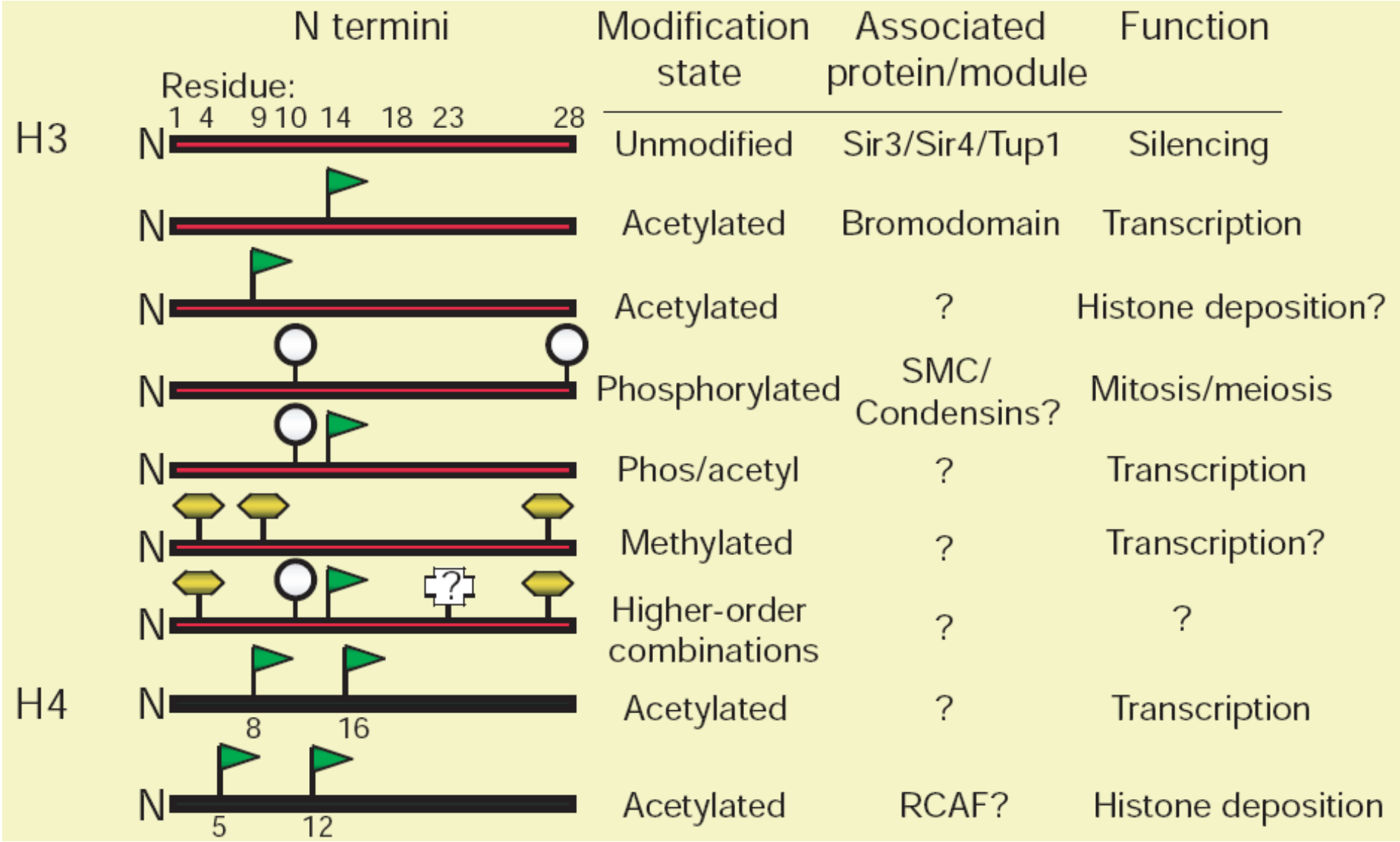
- Good ion statistics and deconvolution software (needs to be able to predict isotope pattern to determine monoisotopic mass).

Histones – Ideal samples?

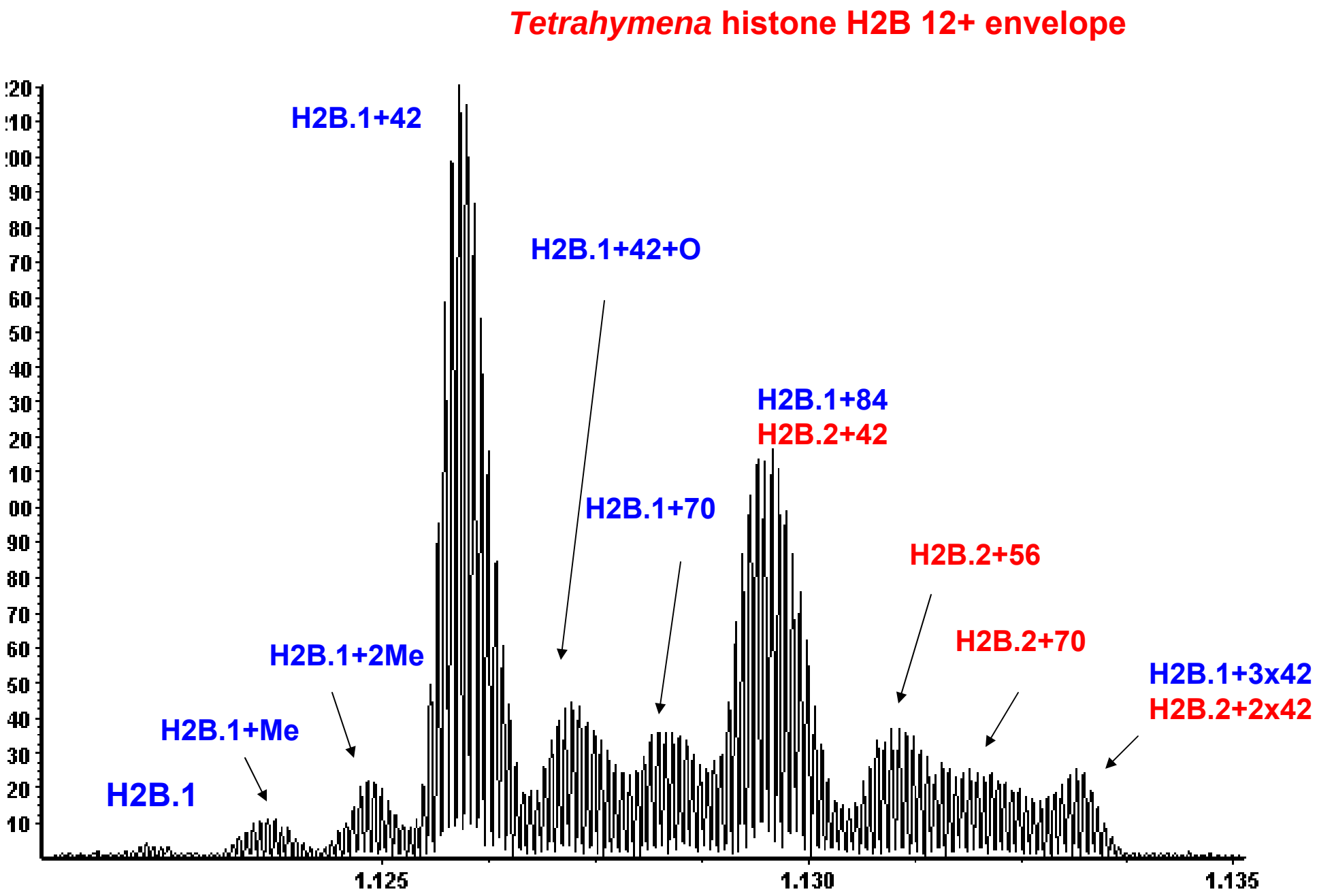
- Small proteins
- Post-translationally extensively modified: phosphorylation, acetylation, methylation, ubiquitination
- Histone code: PTM state regulates gene transcription 'on' and 'off' state

Tetrahymena histone H2B variants studied

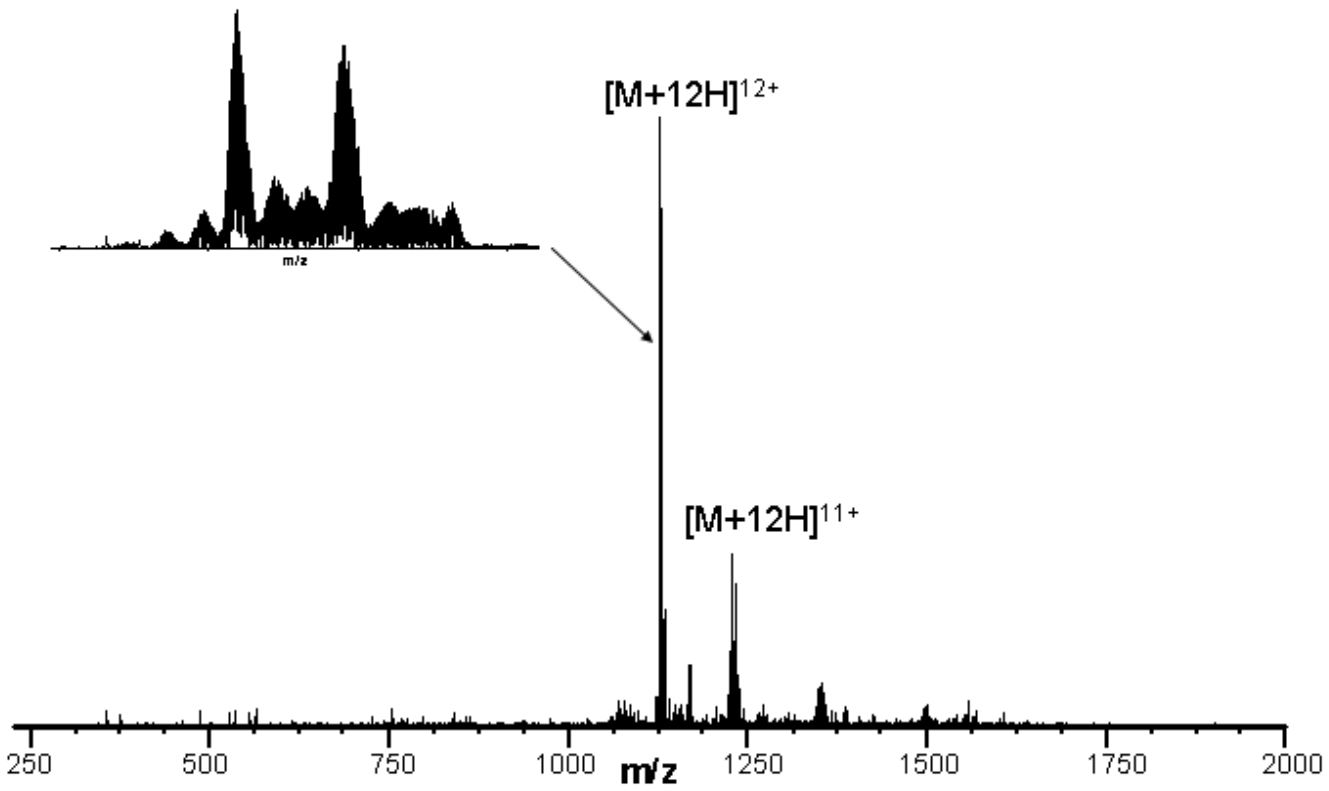
`Histone Code' Hypothesis



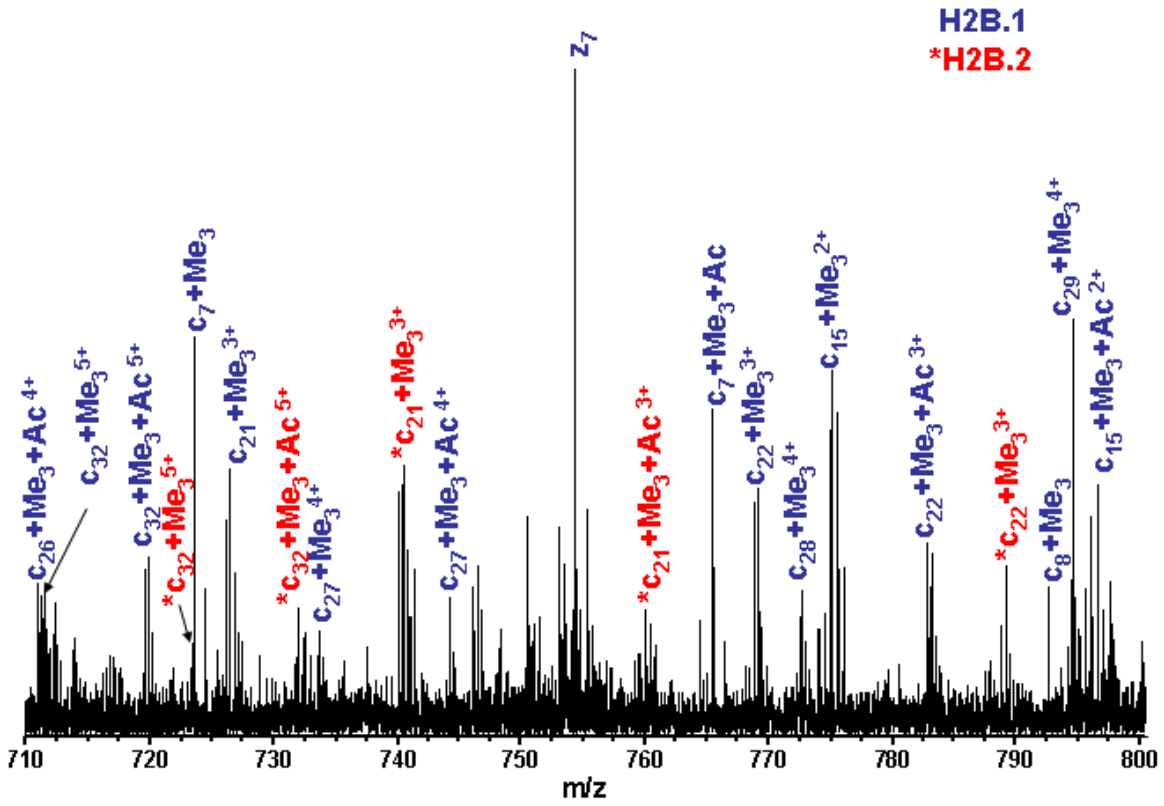
Strahl and Allis, Nature (2000) 403, 41-45.


Tetrahymena histone H2B 12+ envelope
H2B.1+42
H2B.1+42+O
H2B.1+84
H2B.2+42
H2B.1+70
H2B.2+56
H2B.2+70
H2B.1+3x42
H2B.2+2x42
H2B.1+2Me
H2B.1+Me
H2B.1
1.125
1.130
1.135

Electron-capture dissociation spectrum of the H2B mixture



Same ECD spectrum – “zoom-in”

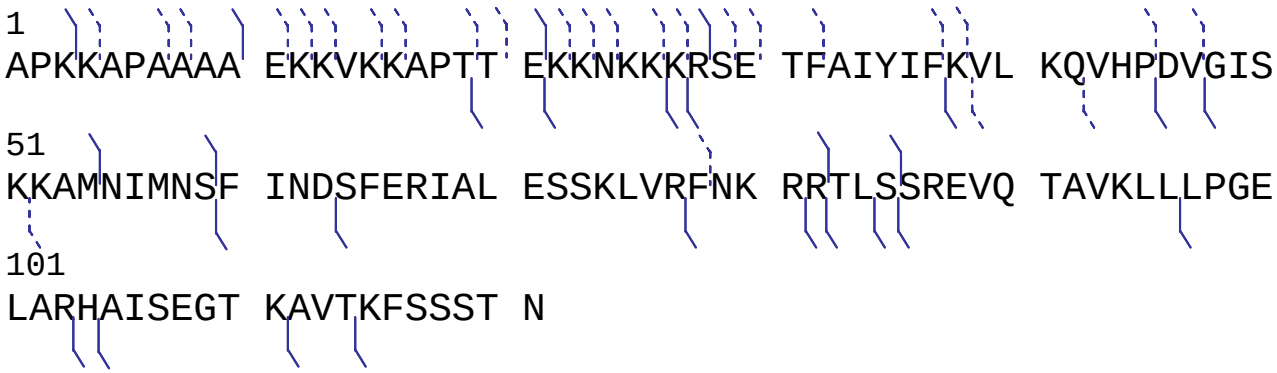


Mass Accuracy is Used to Distinguish between Isobaric Trimethylation and Acetylation

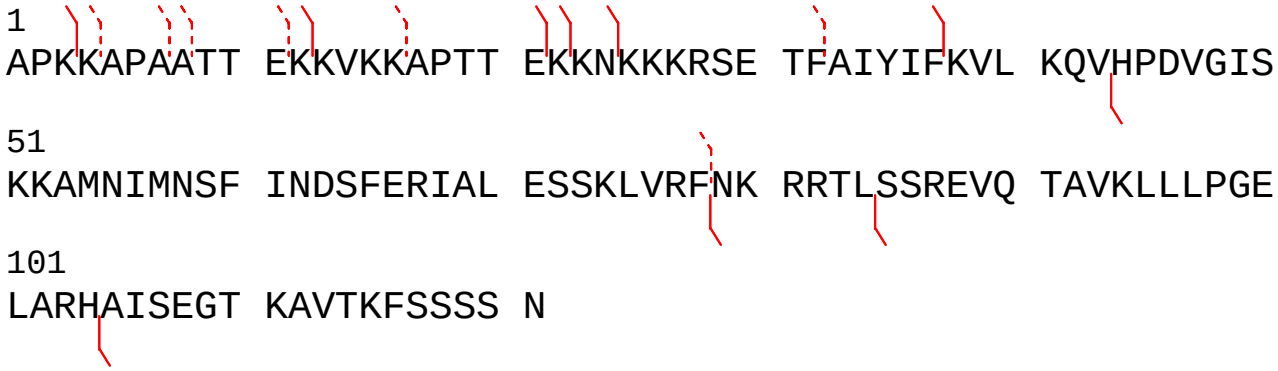
$Mass_{observed}$	**Assignment (Ac/Me_3; $2Ac/AcMe_3/2\ Me_3$) with Mass Calculated**		**Δ[ppm]**
356.2667	c_3+ 42	356.2298 /**356.2661**	+104/ **+1.6**
526.3735	c_4+ 2x42	526.3354/**526.3718**/526.4082	+72/ **+3.2**/ -66
723.4889	c_7+ 42	723.4518/**723.4882**	+51/ **+ 0.9**
754.3535	z_7	54.3497	-5
• • • • •			
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 on 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.

ECD sequence coverage of the two isoforms



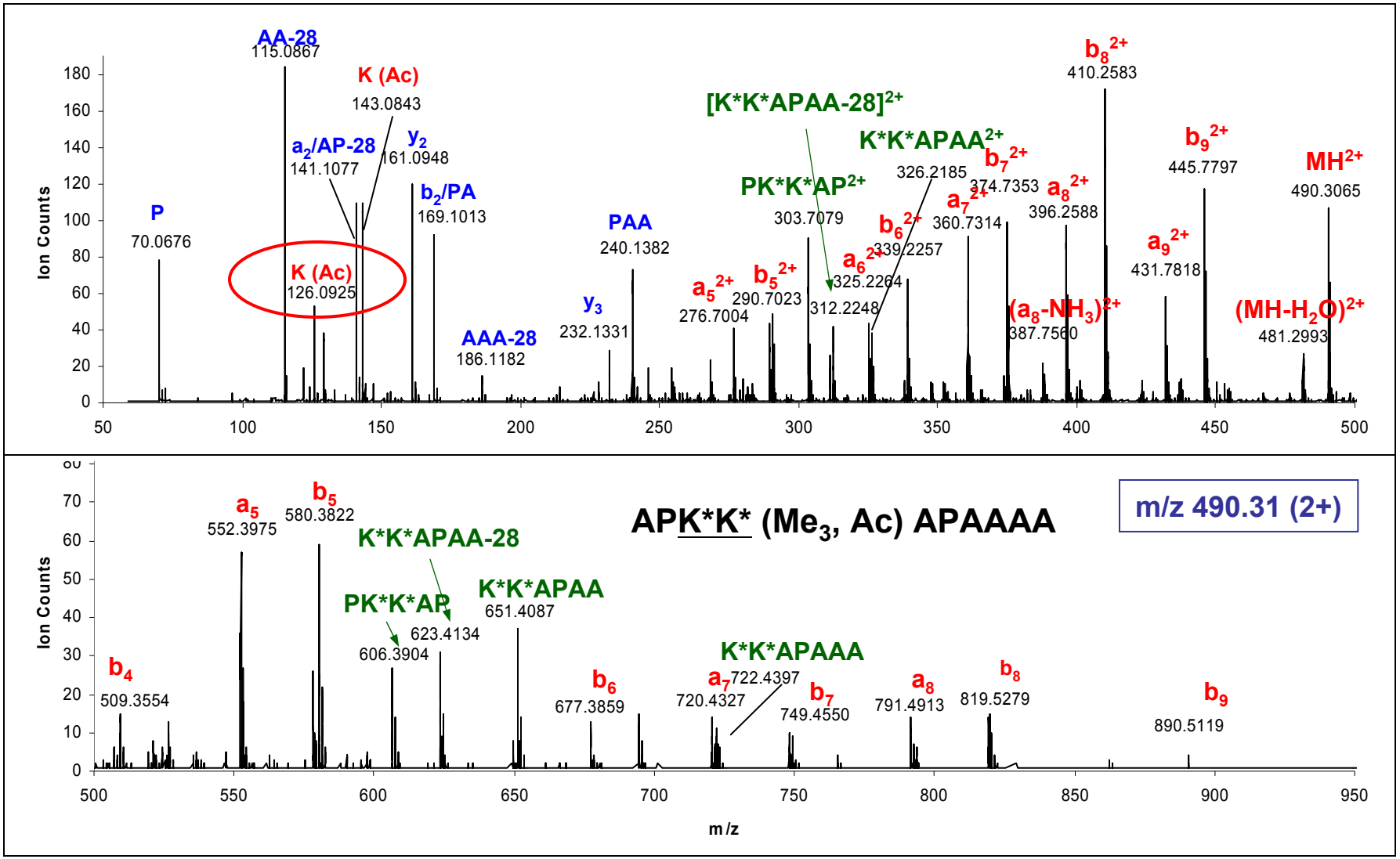
H2B.1



H2B.2

N-terminus or Lys-3 is trimethylated, Lys-4 is acetylated in both

CID Spectrum of an +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.

Mass Accuracy of Fragment Ion Series Determines the +84 Da as a Tri-methylation and an 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 Posttranslational 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

Modified peptides in the tryptic digest

Ion detected	MH^+ calculated	Structure	Δ[ppm]
418.608 (3+)	1253.758	^{4}K(Ac)APAAAAEKKVK^{15}	+40
438.611 (3+)	1313.779	^{4}K(Ac)APAATTEKKVK^{15}	+29
341.978 (4+)	1364.889	APKK (Ac,Me_3)APAAAAEKK^{13}	+44
413.773 (4+)	1652.011	APKK(Ac,Me_3)APAATTEKKVK^{15}	+35
		Me_3APKK(Ac)APAATTEKKVK^{15}	
410.269 (4+)	1637.995	Me_2APKK(Ac) APAATTEKKVK^{15}	+35
398.770 (4+)	1591.990	APKK(Ac,Me_3)APAAAAEKKVK^{15}	
		Me_3APKK (Ac) APAAAAEKKVK^{15}	+42
395.266 (4+)	1577.974	Me_2APKK (Ac) APAAAAEKKVK^{15}	+42
487.961 (3+)	1461.843	^{39}VLK(Ac)QVHPDVGISK^{51}	+17
631.668 (3+)	1892.972	HAISEGTK(Me_2)AVTKFSSSTN^{121}	+9
		^{104}HAISEGTKAVTK(Me_3)FSSSSN^{121}	
	1892.935	^{104}HAISEGTKAVTK(Ac)FSSSSN^{121}	+28
636.323(3+)	1906.950	^{104}HAISEGTK(Me_3)AVTKFSSSTN^{121}	+2

Me_3 @ N-terminus, Lys-3, Lys-111; Ac @ Lys-4; Lys-41
[1-15] peptide always doubly modified!

Why to use both?

"bottom-up" approach:

- More sensitive ~300 fmoles injected
- Reveals more modifications

BUT

- Misses the major protein component!

"top-down" reveals the relative distribution of differently modified populations

Conclusions

•Mass spectrometry is a sensitive, non-biased approach to peptide/protein modification analysis.

•Isolation, MS analysis has to be adjusted to the PTM of interest

•Enrichment methods work much better on large amounts of protein than small.

•For single protein/simple mixture PTM analysis, best approach is to use no enrichment, try to fragment as many components as possible then try to find spectra of modified peptides.

• Peptide-level analysis may not reflect accurately the composition of protein populations