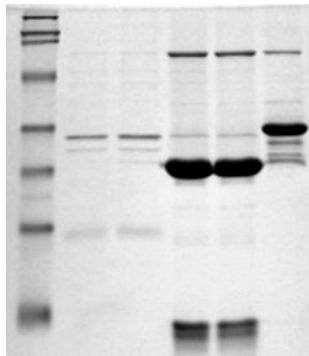


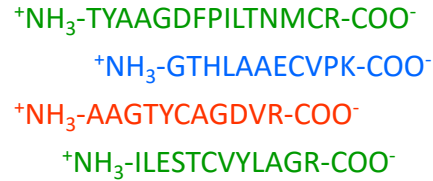
Protein and PTM Identification

IDeA National Resource
Proteomics Workshop
May 24, 2017

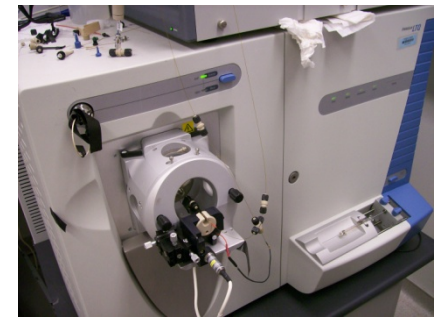
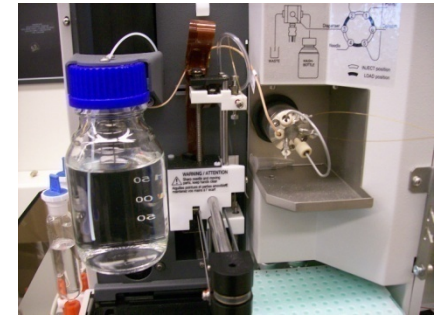
Samuel G. Mackintosh, Ph.D.
Dept. of Biochemistry and Molecular Biology
University of Arkansas for Medical Sciences
sgmackintosh@uams.edu



trypsin



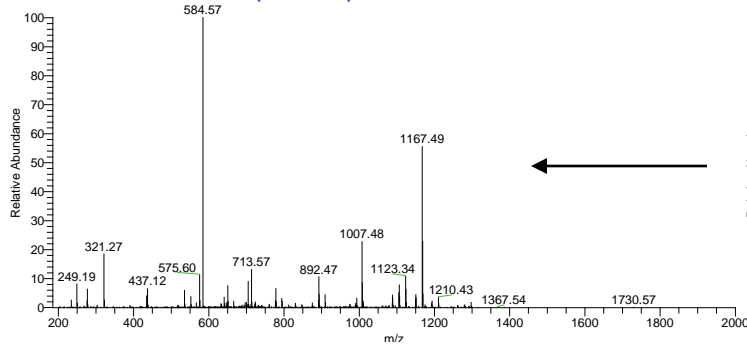
→



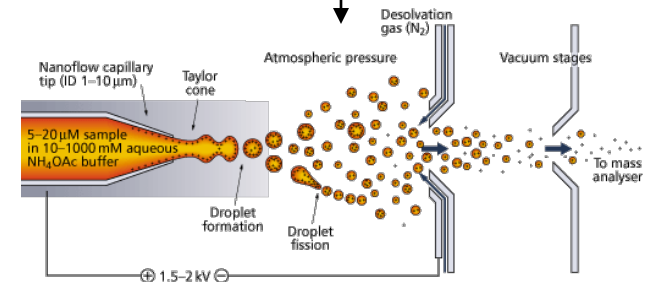
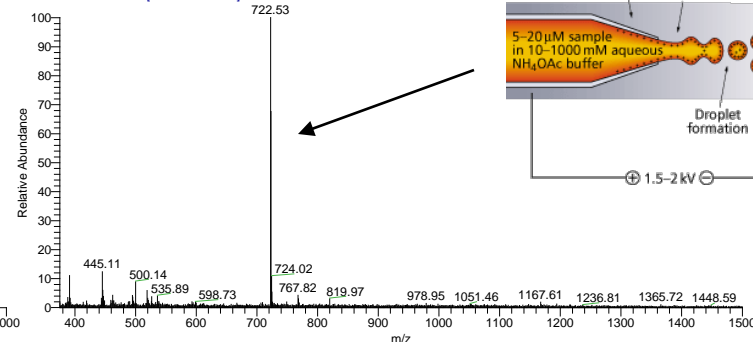
Protein identification by LC-MS/MS:

- in-gel trypsin digestion
- reverse-phase HPLC separation
- electrospray ionization
- peptide ion detection by MS
- peptide fragmentation by MS/MS
- protein identification by database searching

BSA_060908_01 #647 RT: 10.21 AV: 1 NL: 2.71E3
T: ITMS + c ESI d Full ms2 722.65@cid35.00 [185.00-2000.00]



BSA_060908_01 #659 RT: 10.41 AV: 1 NL: 3.02E5
T: ITMS + c ESI Full ms [375.00-1500.00]





Tryptic digestion of BSA:

- cleaves C-terminal of K or R

```

1  MKwvtfislillfssaysrGVFRrDTHKse 30
31  iahrFKdlgeehfkGLVLI AFSQYLQQCPF 60
61  DEHVKlvneltefakTCVADESHAGCEKsl 90
91  htlfgdclckVASLRetygmadcceKQEP 120
121 ERnecflshkDDSPDLPKlkpdpntlcd ef 150
151 kADEKkFWGKylyeiarRhp yfyapellyy 180
181 ankYNGVFQECCQAEDKgacllpkIETMR e 210
211 kVLASSARqrLRcasiqkFGERalkAWSVA 240
241 RlsqkFPKae fvevtkLVTDLTkv hKECCH 270
271 GDLLECADDRadlakYICDNQDTISSKIkE 300
301 CCDKPLLEKshciaev ekDAIPENLPPLTA 330
331 DFAEDKdvcKNYQEAKdaf lgsflyeysrR 360
361 hpeyavsvllrLAKeyeatle eccakDDPH 390
391 ACYSTVFDKlkHLVDEPQNLIKqncdqf ek 420
421 LGEYGFQNALIVRytrKvpqvstptlvevs 450
451 rSLGKvgtrCCTKPESERmpctedy lslil 480
481 nrLCVLHEKtpvsekVTKccteslvnrRPC 510
511 FSALTPDETYVPKafdekLFTFHADICTLP 540
541 DTEKqikKqtalvellkHKPKateeq lktV 570
571 MENFVAFVDKccaaddkEACFAVEGPKlvv 600
601 stqta la 607

```

MW range
of peptide products:

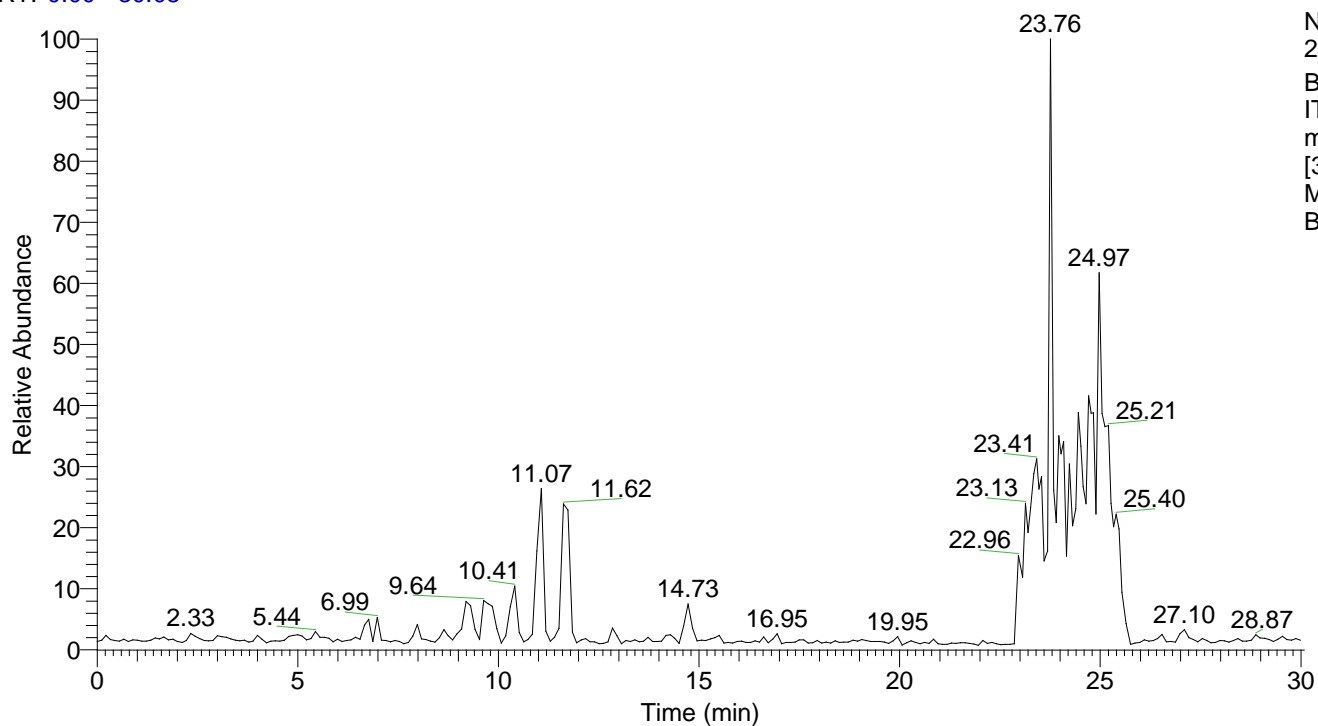
146.19 – 2435.82

- not all peptides will be ionized and detected
- there will be gaps in sequence coverage

LC-MS of 20 fmol of BSA

H₂O/acetonitrile gradient \longrightarrow

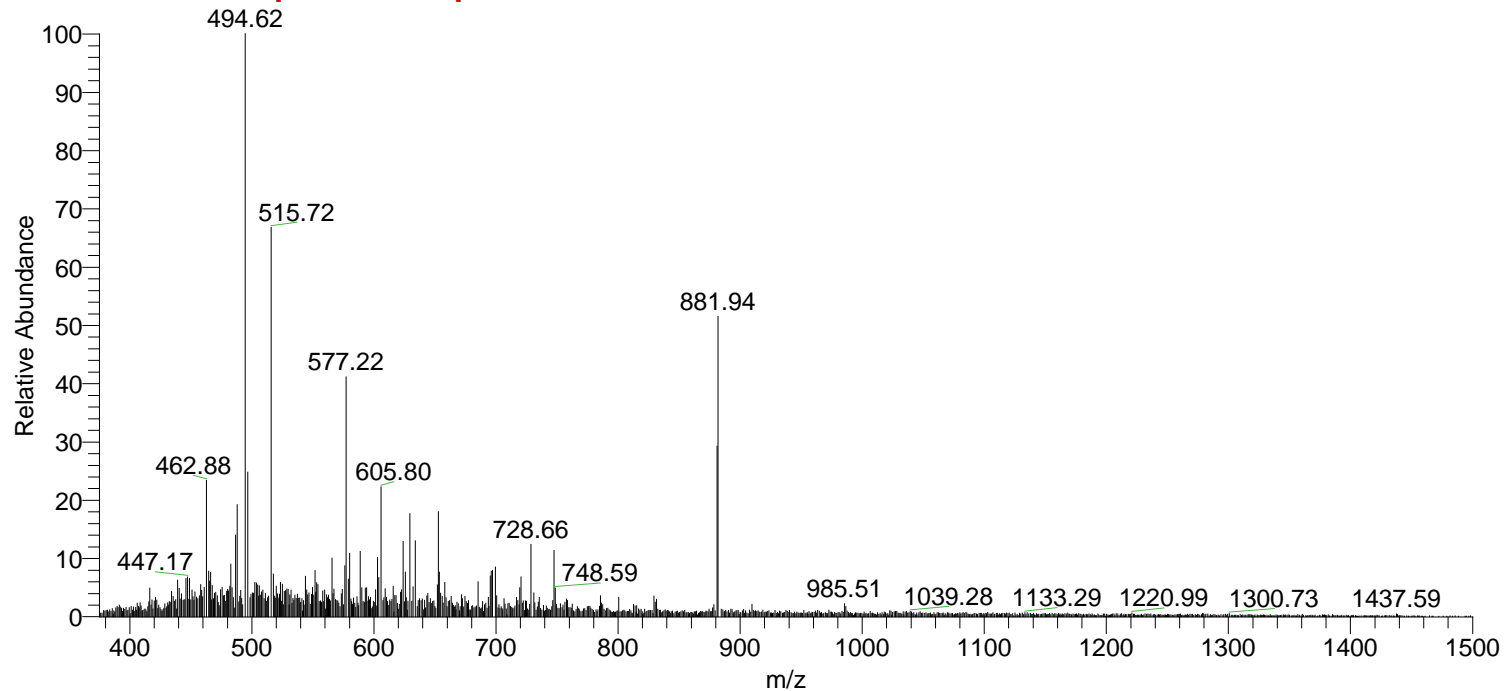
RT: 0.00 - 30.08



NL:
2.89E6
Base Peak F:
ITMS + c ESI Full
ms
[375.00-1500.00]
MS
BSA_060908_01

MS scan – determines masses of intact peptides

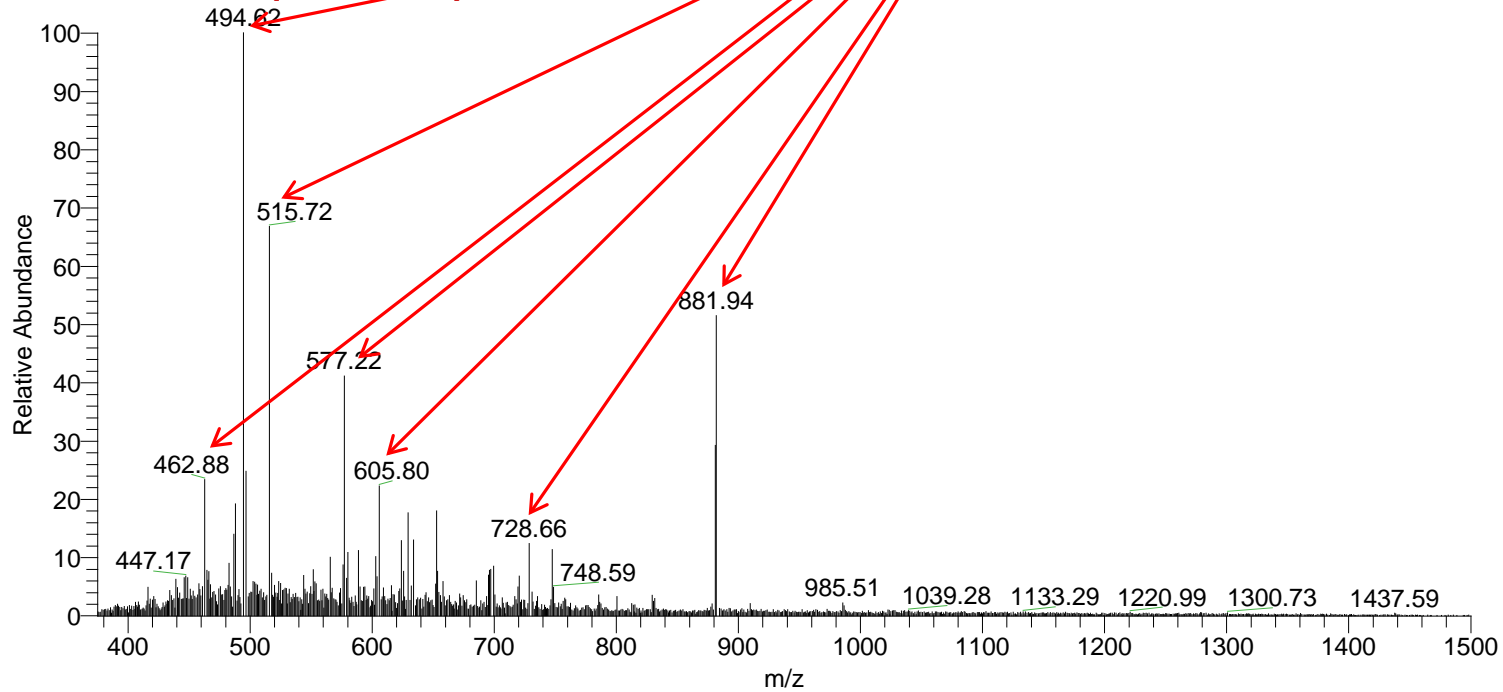
Ecoli_041708_01 #1310 RT: 14.94 AV: 1 NL: 8.31E4
F: ITMS + c ESI Full ms [375.00-1500.00]



Data-Dependent Acquisition (DDA)

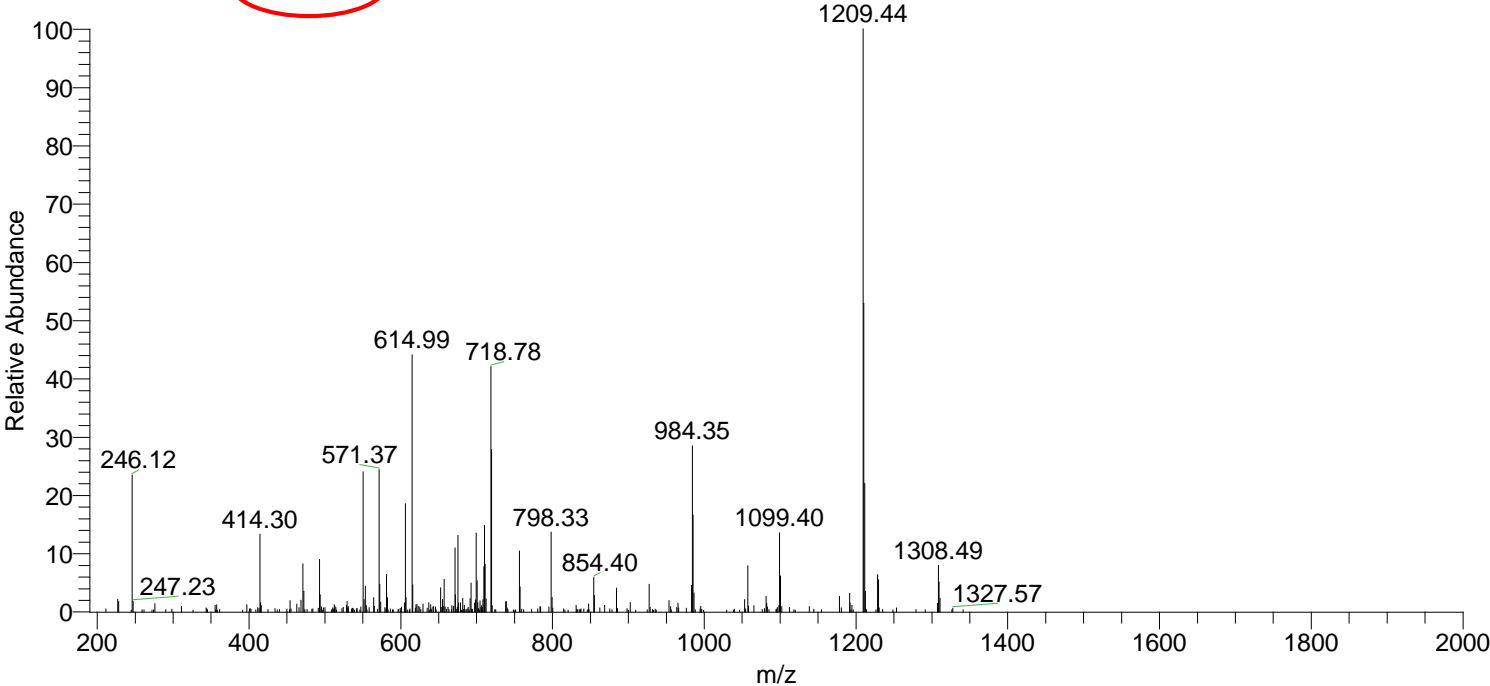
most intense peptide ions are selected for fragmentation

Ecoli_041708_01 #1310 RT: 14.94 AV: 1 NL: 8.31E4
F: ITMS + c ESI Full ms [375.00-1500.00]



MS/MS scan – determines masses of peptide fragments

Ecoli_041708_01 #1315 RT: 15.02 AV: 1 NL: 1.44E2
T: ITMS + c ESI d Full ms2 728.66@cid35.00 [190.00-2000.00]



MS/MS fragmentation methods:

Collision-induced dissociation (CID)
is most likely to occur at peptide bonds

b ions: charge is retained on N-terminus

y ions: charge is retained on C-terminus

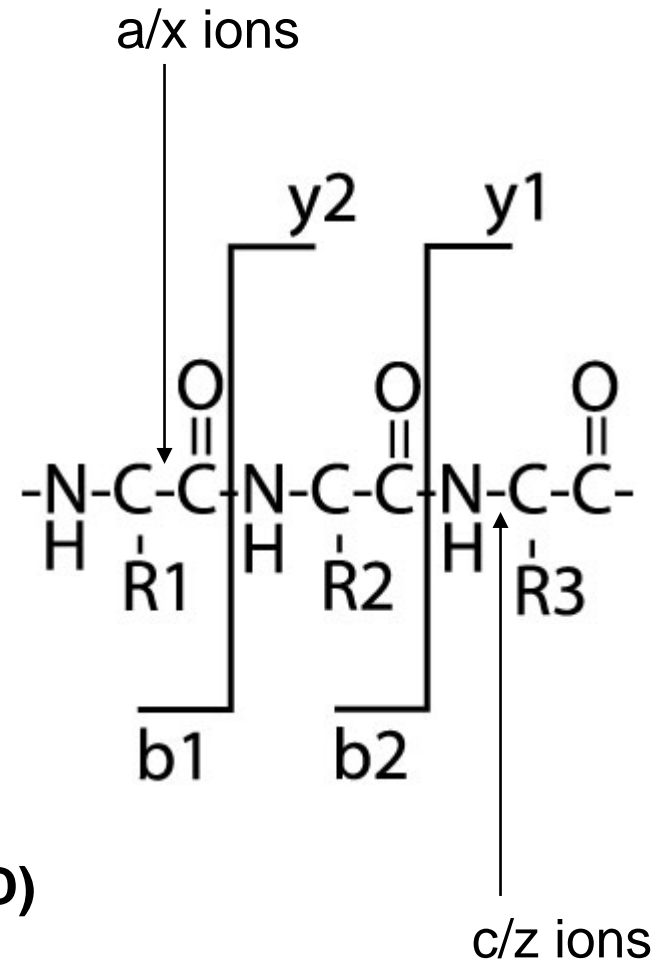
Electron-transfer dissociation (ETD)
is more likely to occur at N-C α bonds

c ions: charge is retained on N-terminus

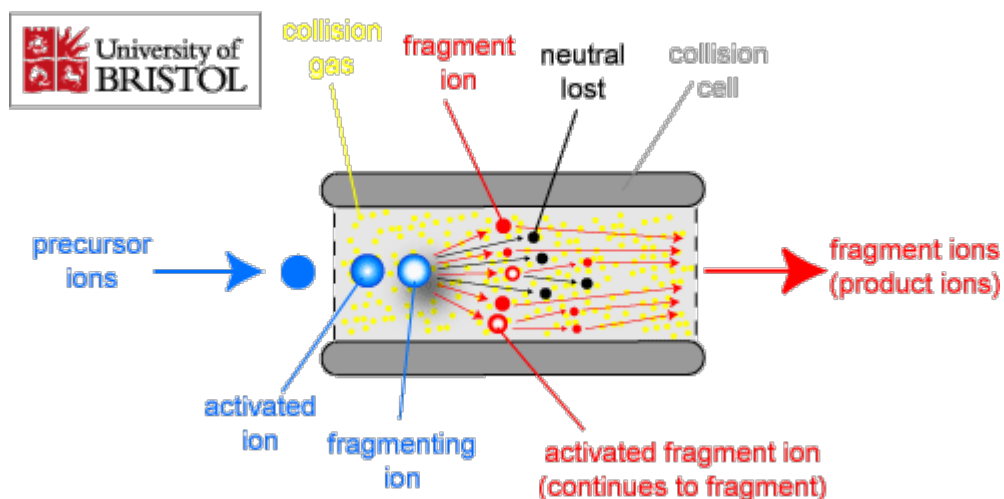
z ions: charge is retained on C-terminus

High-energy collisional dissociation (HCD)

produces b and y ions



CID – collision-induced dissociation



takes place in ion trap in presence of inert collision gas (He or Ar)

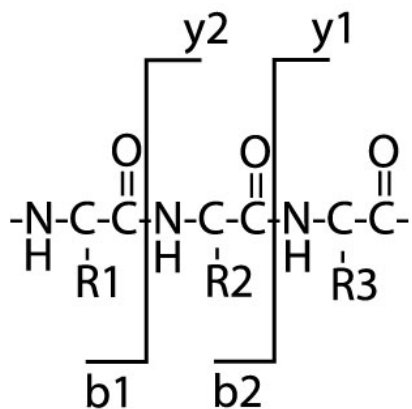
CID fragmentation of peptide YICDNQDTISSK

b ion series:

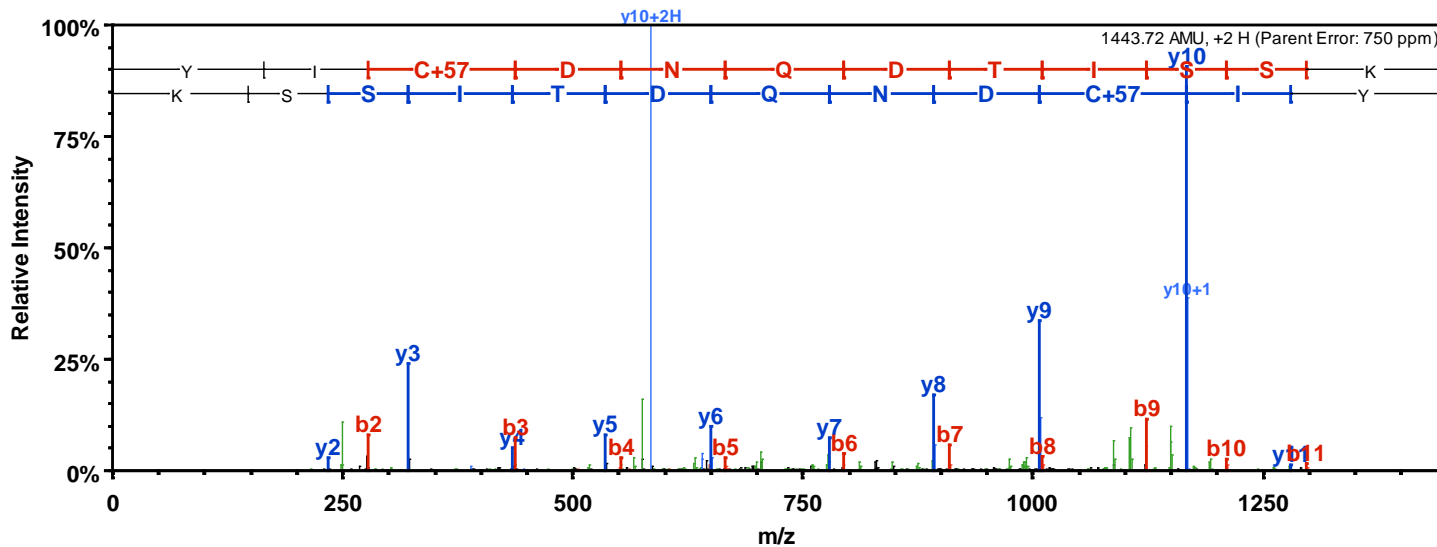
b1: Y-
b2: YI-
b3: YIC-
b4: YICD-
b5: YICDN-
b6: YICDNQ-
b7: YICDNQD-
b8: YICDNQDT-
b9: YICDNQDTI-
b10: YICDNQDTIS-
b11: YICDNQDTISS-

y ion series:

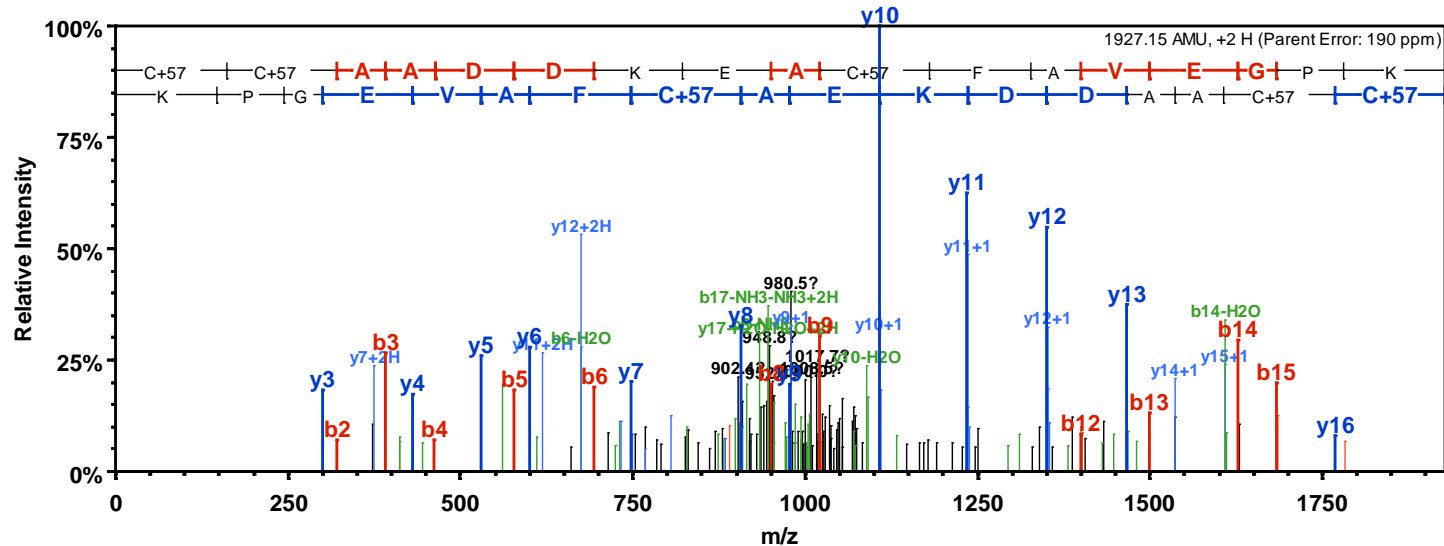
y1: K-
y2: KS-
y3: KSS-
y4: KSSI-
y5: KSSIT-
y6: KSSITD-
y7: KSSITDQ-
y8: KSSITDQN-
y9: KSSITDQND-
y10: KSSITDQNDC-
y11: KSSITDQNDCI-



Peptide sequence determination by MS/MS fragmentation



Peptide/protein identification by database searching



- perform a theoretical trypsin digestion of all known proteins
- determine a theoretical MS/MS fragmentation for each peptide
- compare experimental spectrum to each theoretical spectrum

Mascot peptide scoring

Mascot Search Results: Peptide View - Microsoft Internet Explorer

Address: http://acrc300811/mascot/cgi/peptide_view.pl?file=../data/20080916/F005597.dat&query=1069&hit=1&index=ALBU_BOVIN&px=1§ion=5&ave_thresh=46

Google

Monoisotopic mass of neutral peptide Mr(calc): 1442.6347
 Fixed modifications: Carbamidomethyl (C)
 Ions Score: 71 Expect: 0.00021
 Matches (Bold Red): 42/116 fragment ions using 55 most intense peaks

#	b	b ⁺⁺	b [*]	b ^{*++}	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y [*]	y ^{*++}	y ⁰	y ⁰⁺⁺	#
1	164.0706	82.5389					Y							12
2	277.1547	139.0810					I	1280.5787	640.7930	1263.5522	632.2797	1262.5681	631.7877	11
3	437.1853	219.0963					C	1167.4946	584.2510	1150.4681	575.7377	1149.4841	575.2457	10
4	552.2123	276.6098			534.2017	267.6045	D	1007.4640	504.2356	990.4374	495.7224	989.4534	495.2304	9
5	666.2552	333.6312	649.2286	325.1180	648.2446	324.6259	N	892.4371	446.7222	875.4105	438.2089	874.4265	437.7169	8
6	794.3138	397.6605	777.2872	389.1472	776.3032	388.6552	Q	778.3941	389.7007	761.3676	381.1874	760.3836	380.6954	7
7	909.3407	455.1740	892.3142	446.6607	891.3301	446.1687	D	650.3355	325.6714	633.3090	317.1581	632.3250	316.6661	6
8	1010.3884	505.6978	993.3618	497.1846	992.3778	496.6925	T	535.3086	268.1579	518.2821	259.6447	517.2980	259.1527	5
9	1123.4725	562.2399	1106.4459	553.7266	1105.4619	553.2346	I	434.2609	217.6341	417.2344	209.1208	416.2504	208.6288	4
10	1210.5045	605.7559	1193.4779	597.2426	1192.4939	596.7506	S	321.1769	161.0921	304.1503	152.5788	303.1663	152.0868	3
11	1297.5365	649.2719	1280.5100	640.7586	1279.5259	640.2666	S	234.1448	117.5761	217.1183	109.0628	216.1343	108.5708	2
12							K	147.1128	74.0600	130.0863	65.5468			1

Mass error 467 ppm

Mass error 467 ppm

NCBI BLAST search of [YICDNQDTISSK](#)
 (Parameters: blastp, nr protein database, expect=20000, no filter, PAM30)
 Other BLAST [web gateways](#)

All matches to this query

Score	Mr(calc):	Delta	Sequence
70.8	1442.6347	1.0839	YICDNQDTISSK
22.2	1442.6347	1.0839	YICENQDSISSK
16.2	1442.7558	0.9629	YLKEIFVDNFR
14.8	1441.7962	1.9224	ANPAVVNELLMKK

Done

Local intranet

Start | Inboxes - Micr... | Seminar_10... | Core Facility ... | 2 Adobe R... | Matrix Scien... | Peptide Su... | Mascot Se... | Scaffold Vie... | Microsoft Po... | 3:53 PM

Compilation of database search results

Scaffold Viewer - BSAcontrol

File Edit View Experiment Export Window Help

Min Protein: 95.0% Min # Peptides: 1 Min Peptide: 0%

Serum albumin precursor (Allergen Bos d 6) (BS... All MSMS Samples

Sequence Coverage	Protein	Category	Bio Sample	MSMS Sam...	Prob	%St
	Serum album...	Uncategoriz...	BSA	BSA_09160...	100%	2.2

Good?	Sequence	Prob	Masco...	Masco...	NTT	Modifications
<input checked="" type="checkbox"/>	(K)ECCKPLLEK(S)	95%	54.6	47.1	2	Carbamidomethyl ...
<input checked="" type="checkbox"/>	(R)NECFLSHKDDSPDLPK(L)	48%	27.6	46.5	2	Carbamidomethyl ...
<input checked="" type="checkbox"/>	(R)RHPYAVSVLLR(L)	95%	73.6	46.9	2	
<input checked="" type="checkbox"/>	(K)LVTDLTK(Y)	80%	34.1	48.4	2	
<input checked="" type="checkbox"/>	(K)LVTDLTK(V)	95%	61.4	48.1	2	
<input checked="" type="checkbox"/>	(K)DDSPDLPK(L)	67%	31.9	48.2	2	
<input checked="" type="checkbox"/>	(K)DDSPDLPK(L)	95%	54.1	47.8	2	
<input checked="" type="checkbox"/>	(K)DLGEEHFK(G)	95%	43.1	48.0	2	
<input checked="" type="checkbox"/>	(K)AEFVEVTK(L)	94%	37.5	46.9	2	
<input type="checkbox"/>	(K)AEFVEVTK(L)	0%	11.1	47.0	2	
<input type="checkbox"/>	(K)AEFVEVTK(L)	7%	20.7	47.0	2	
<input checked="" type="checkbox"/>	(K)AEFVEVTK(L)	95%	38.4	47.0	2	
<input type="checkbox"/>	(K)AEFVEVTK(L)	5%	19.8	47.0	2	
<input checked="" type="checkbox"/>	(K)AEFVEVTK(L)	95%	51.9	46.5	2	
<input checked="" type="checkbox"/>	(K)EACFAVEGPK(L)	94%	37.3	47.0	2	Carbamidomethyl ...
<input checked="" type="checkbox"/>	(K)CTESLVNR(R)	95%	58.3	46.7	2	Carbamidomethyl ...
<input checked="" type="checkbox"/>	(K)CTESLVNR(R)	69%	30.7	46.8	2	Carbamidomethyl ...
<input checked="" type="checkbox"/>	(K)DDPHACYSTVFDK(L)	95%	53.0	47.0	2	Carbamidomethyl ...
<input type="checkbox"/>	(K)DDPHACYSTVFDK(L)	11%	22.1	46.8	2	Carbamidomethyl ...
<input checked="" type="checkbox"/>	(K)DDPHACYSTVFDK(L)	67%	30.5	46.9	2	Carbamidomethyl ...
<input checked="" type="checkbox"/>	(K)ATEEQLK(T)	95%	48.9	48.6	2	
<input checked="" type="checkbox"/>	(K)KGTALVELLK(H)	95%	72.1	47.5	2	
<input checked="" type="checkbox"/>	(K)KGTALVELLK(H)	80%	33.5	47.6	2	
<input checked="" type="checkbox"/>	(R)HPEYAVSVLLR(L)	95%	46.1	47.0	2	
<input checked="" type="checkbox"/>	(K)YLYEIAR(R)	95%	44.8	47.3	2	

Protein Sequence Similar Proteins Spectrum Spectrum/Model Error Fragmentation Table

ALBU_BOVIN (100%), 69294.2 Da
Serum albumin precursor (Allergen Bos d 6) (BSA) - Bos taurus (Bovine)
28 unique peptides, 45 unique spectra, 47 total spectra, 251/607 amino acids (41% coverage)

```

M K W V T F I S L L L L F S S A Y S R G V F R R D T H K S E I A H R F K D L G E E H F K G L V L I A
F S Q Y L Q Q C P F D E H V K L V N E L T E F A K T C V A D E S H A G C E K S L H T L F G D E L C K
V A S L R E T Y G D M A D C C E K Q E P E R N E C F L S H K D D S P D L P K L K P D P N T L C D E F
K A D E K K F W G K Y L Y E I A R R H P Y F Y A P E L L Y Y A N K Y N G V F Q E C C Q A E D K G A C
L L P K I E T M R E K V L A S S A R Q R L R C A S I Q K F G E R A L K A W S V A R L S Q K F P K A E
F V E V T K L V T D L T K V H K E C C H G D L L E C A D D R A D L A K Y I C D N Q D T I S S K L K E
C C D K P L L E K S H C I A E V E K D A I P E N L P P L T A D F A E D K D V C K N Y Q E A K D A F L
G S F L Y E Y S R R H P E Y A V S V L L R L A K E Y E A T L E E C C A K D D P H A C Y S T V F D K L
K H L V D E P Q N L I K Q N C D Q F E K L G E Y G F Q N A L I V R Y T R K V P Q V S T P T L V E V S
R S L G K V G T R C C T K P E S E R M P C T E D Y L S L I L N R L C V L H E K T P V S E K V T K C C
T E S L V N R R P C F S A L T P D E T Y V P K A F D E K L F T F H A D I C T L P D T E K Q I K K Q T
A L V E L L K H K P K A T E E Q L K T V M E N F V A F V D K C C A A D D K E A C F A V E G P K L V V
S T Q T A L A
    
```

Start | Inboxes - Microsoft Out... | Seminar_102908 | Core Facility 2008 | Microsoft PowerPoint ... | Scaffold Viewer - B... | 7:27 AM

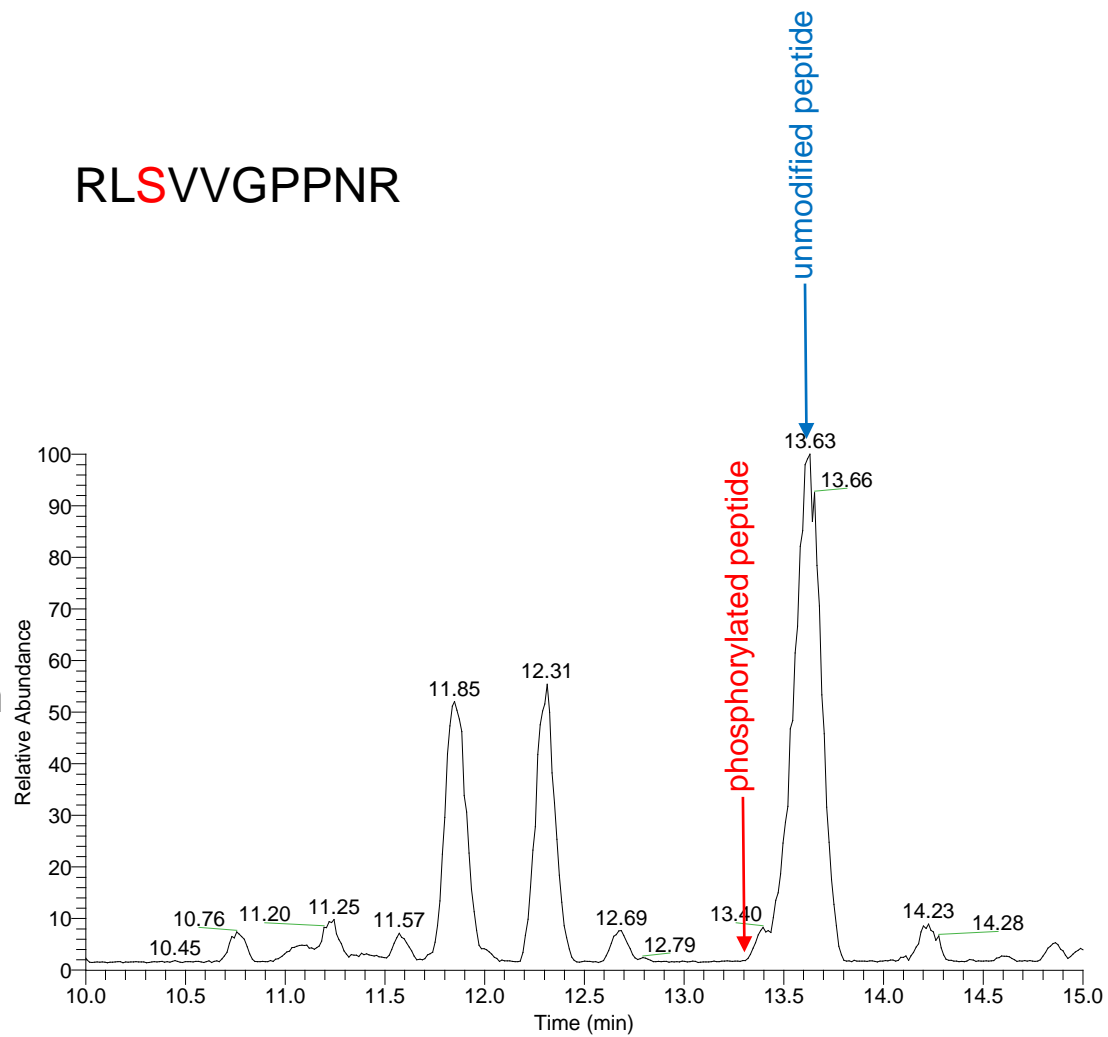


Challenges in mapping post-translational modifications

- Stoichiometry is often unfavorable
- Sequence coverage gaps
- Relatively unstable modifications are often lost during peptide fragmentation

RLSVVGPPNR

chromatogram



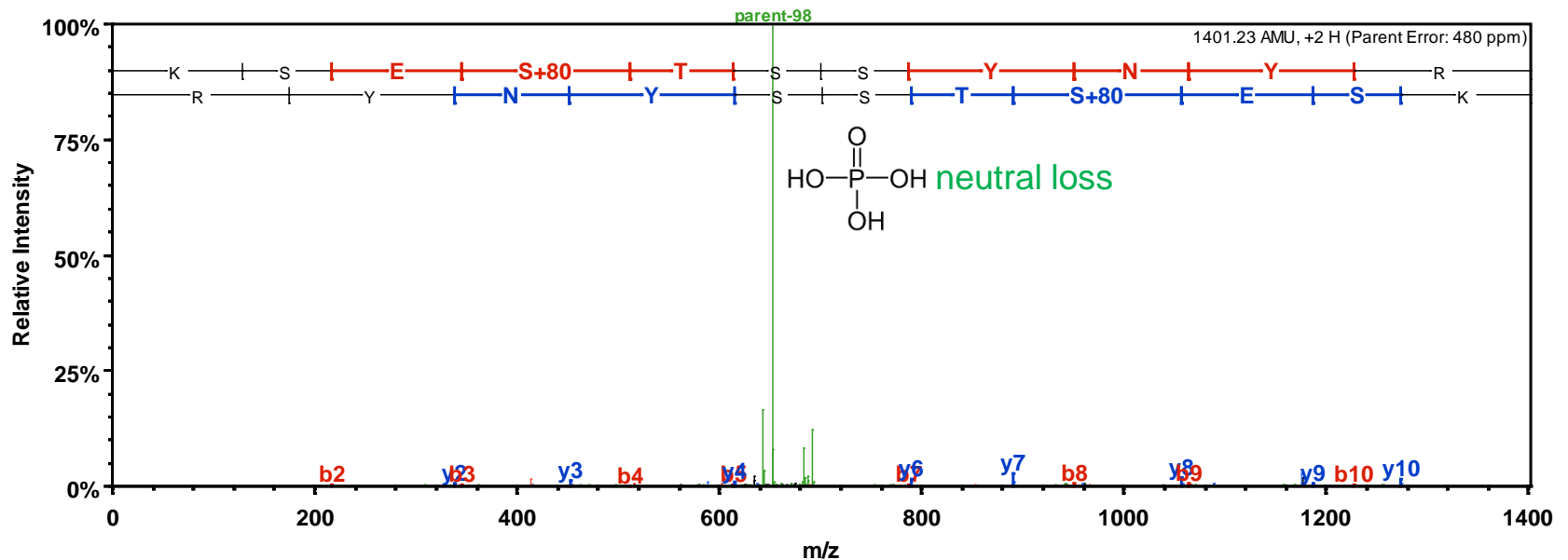
Protease digestion of proteins for MS analysis

Commonly used alternate proteases

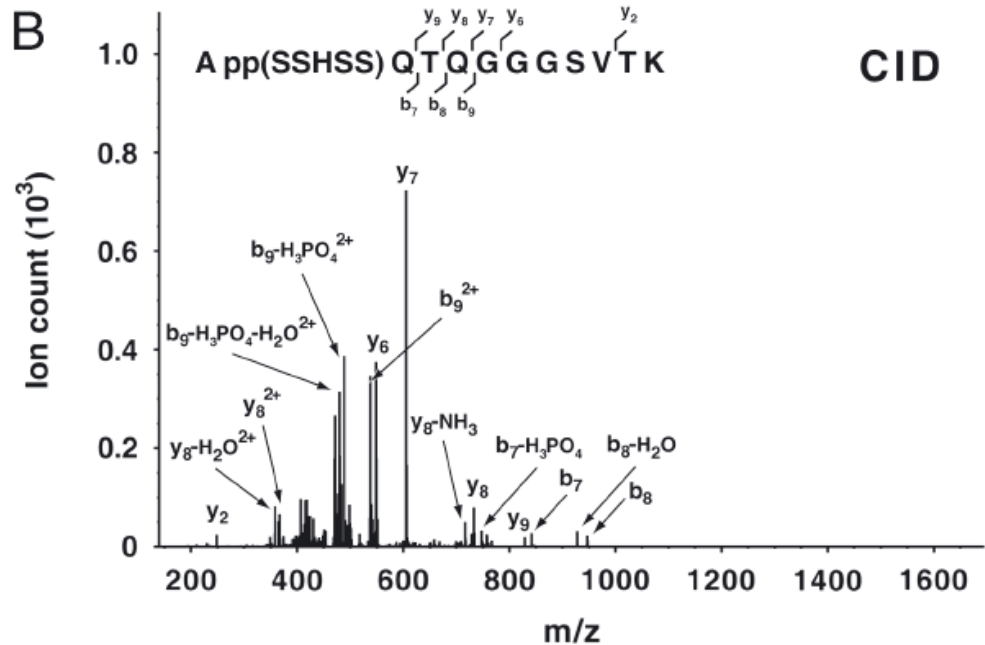
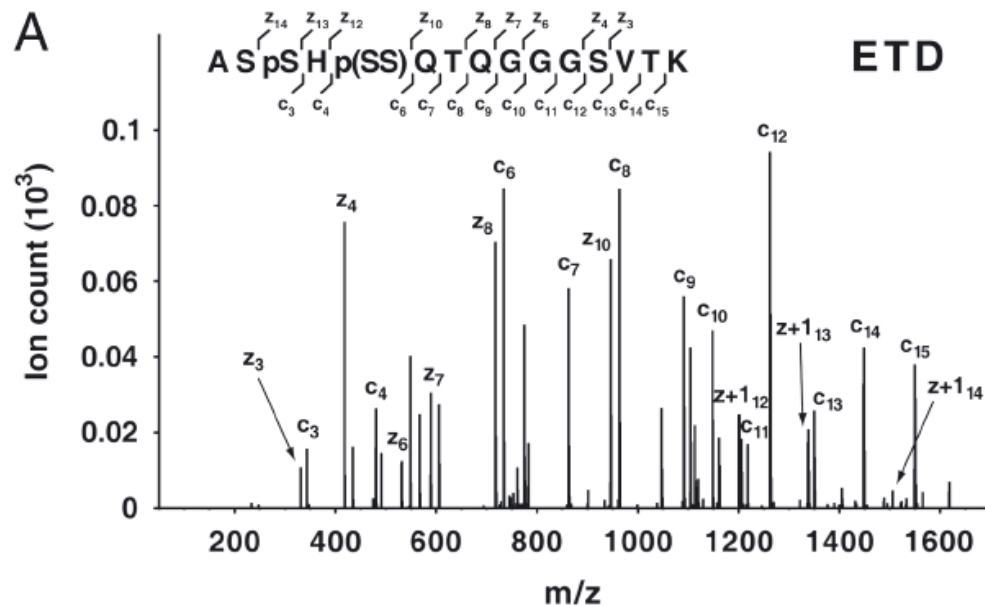
- AspN – cleaves on N-terminal side of Asp residues
- GluC – cleaves on C-terminal side of Glu residues
- LysC – cleaves on C-terminal side of Lys residues
- ArgC – cleaves on C-terminal side of Arg residues

- chymotrypsin – cleaves on C-terminal side of hydrophobic residues
- proteinase K – not sequence specific; used for limited digestion

“Typical” phosphopeptide spectrum:



Comparison of ETD and CID for phosphopeptide MS/MS





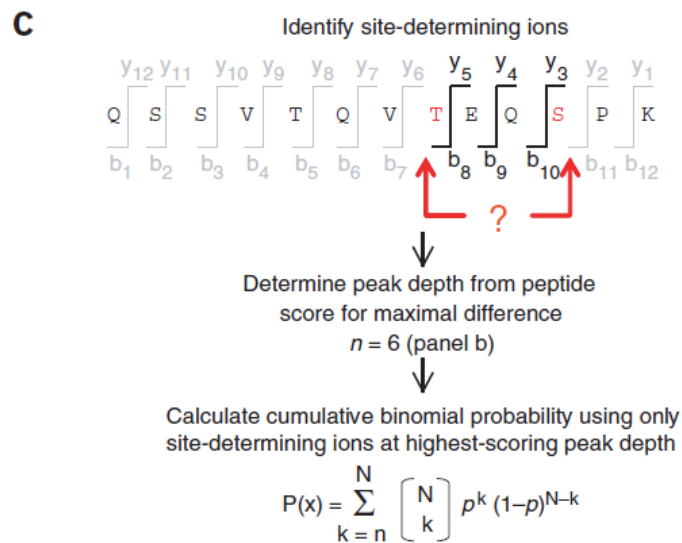
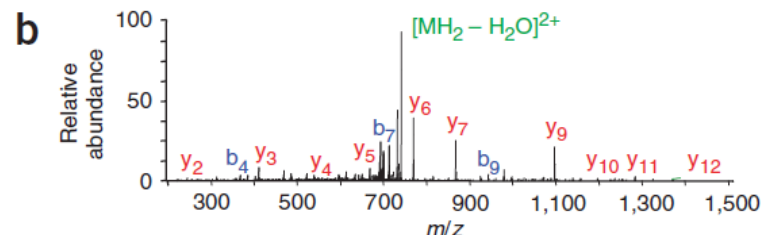
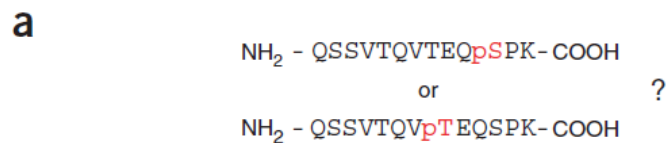
Mapping of post-translational modifications requires

1) Identification of modified peptides

AND

2) Localization of modified sites

PTM site localization using Ascore



Phosphopeptide	QSSVTQVTE Q p SPK	QSSVTQ V p TEQSPK
Trials (N)	6 (y ₃ , y ₄ , y ₅ , b ₈ , b ₉ , b ₁₀)	6 (y ₃ , y ₄ , y ₅ , b ₈ , b ₉ , b ₁₀)
Successes (n)	5 (y ₃ , y ₄ , y ₅ , b ₉ , b ₁₀)	0
p (6 peaks / 100 m/z)	0.06	0.06
P	0.0000044	1.0
Score [-10 × log(P)]	53.57	0
Ascore = ambiguity score (difference of the top two candidates)	53.57 - 0 = 53.57	

