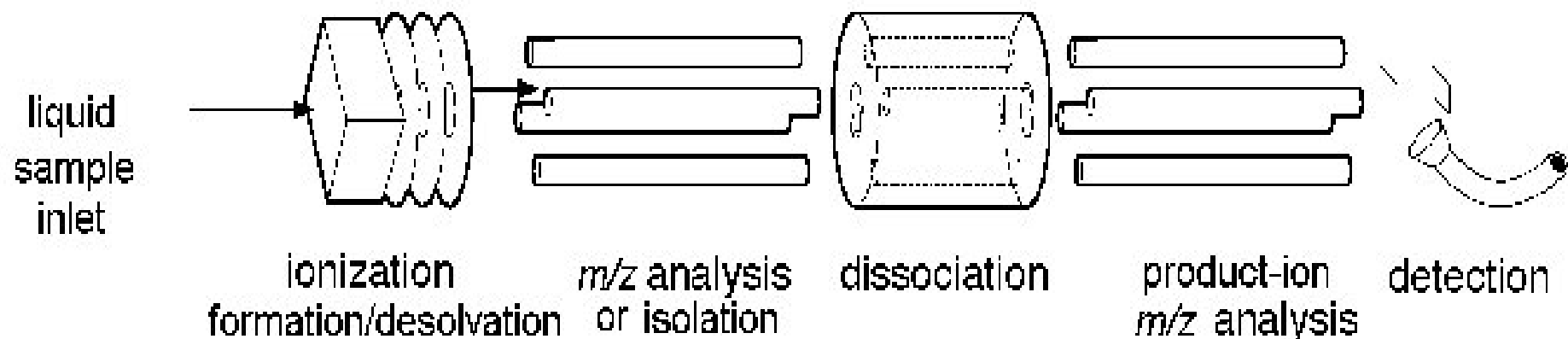
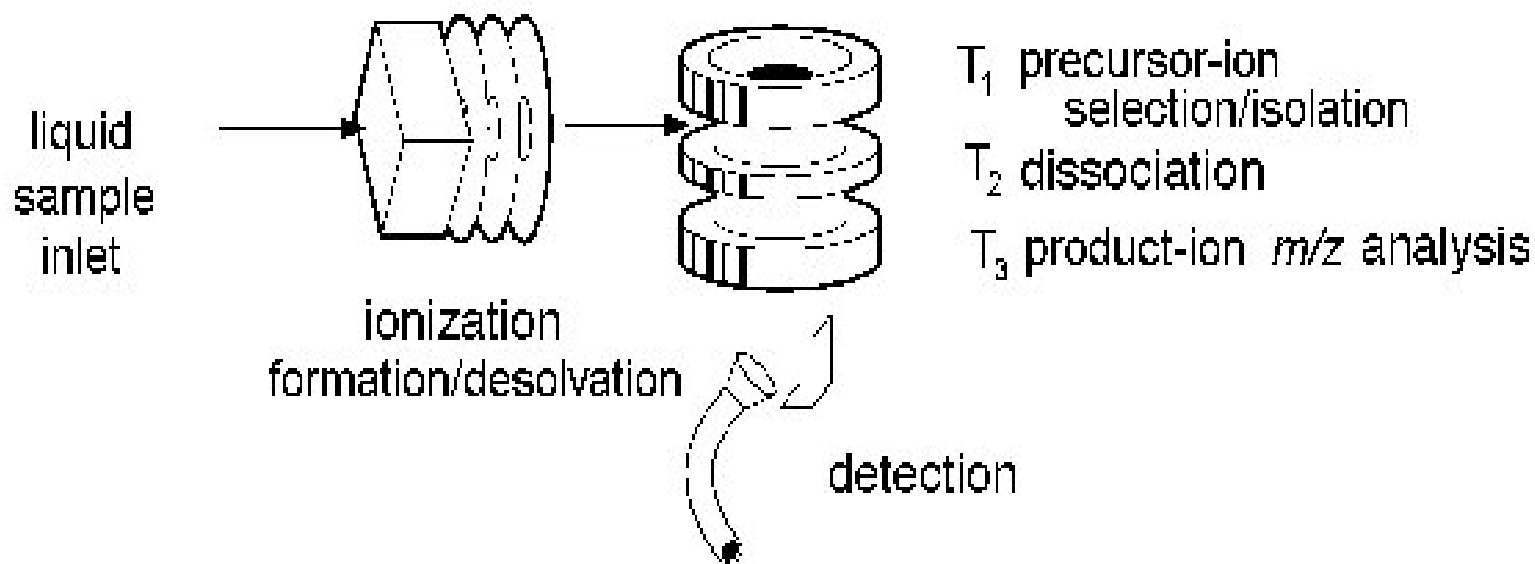


- **WHEN PEPTIDE MASS FINGERPRINTING IS NOT ENOUGH TO IDENTIFY A PROTEIN, AMINOACID SEQUENCE INFORMATION IS NEEDED**
- **SEQUENCE INFORMATION IS PRODUCED USING TANDEM MASS SPECTROMETRY (MS/MS)**
- **SELECTED PEPTIDES ARE GIVEN ENERGY AND COLLIDED WITH GAS MOLECULES INSIDE THE MASS SPECTROMETER PRODUCING FRAGMENT IONS**
- **THIS PROCESS IS KNOWN AS COLLISION INDUCED DISSOCIATION (CID)**

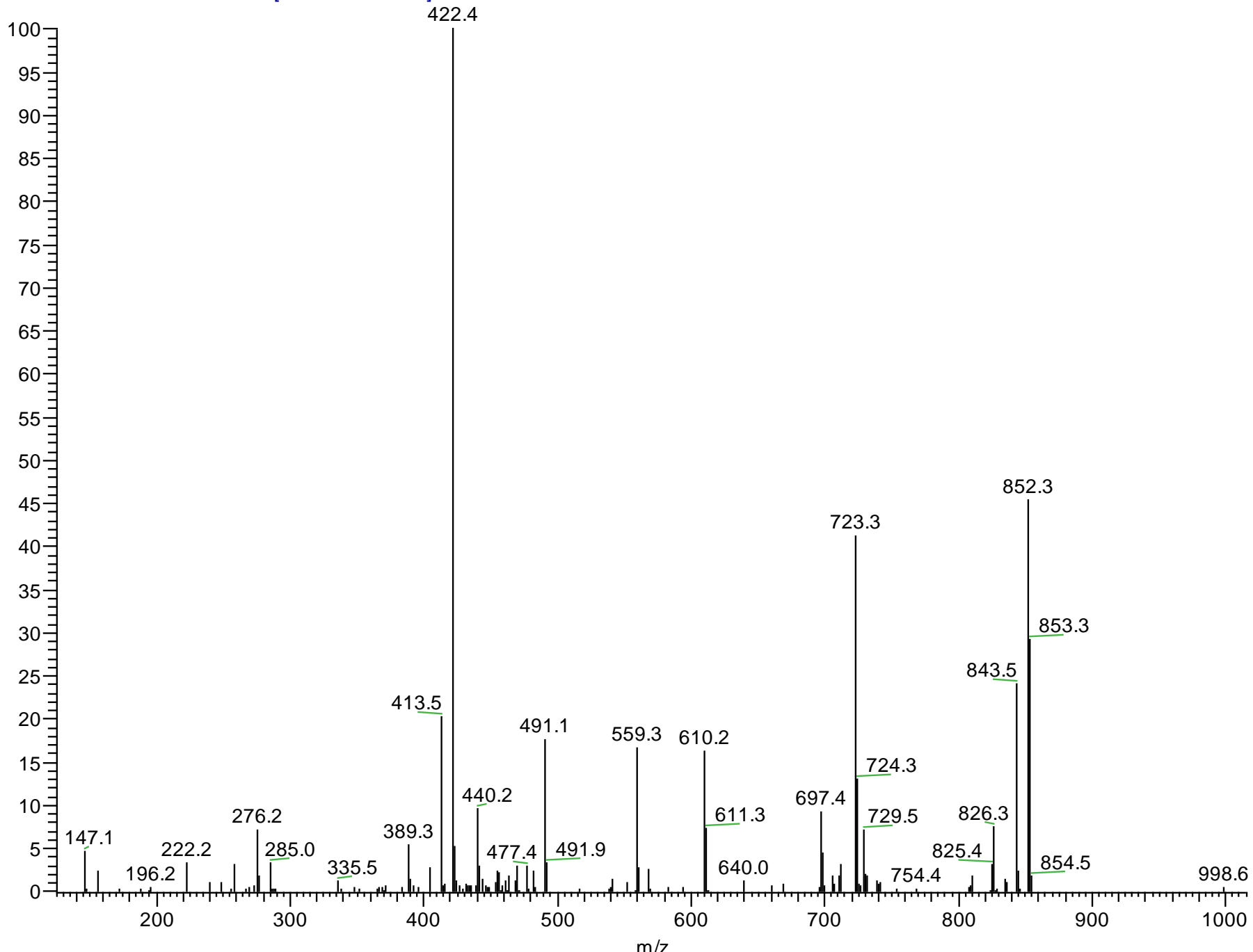
LC / MS / MS Tandem-in-Space



Tandem-in-Time

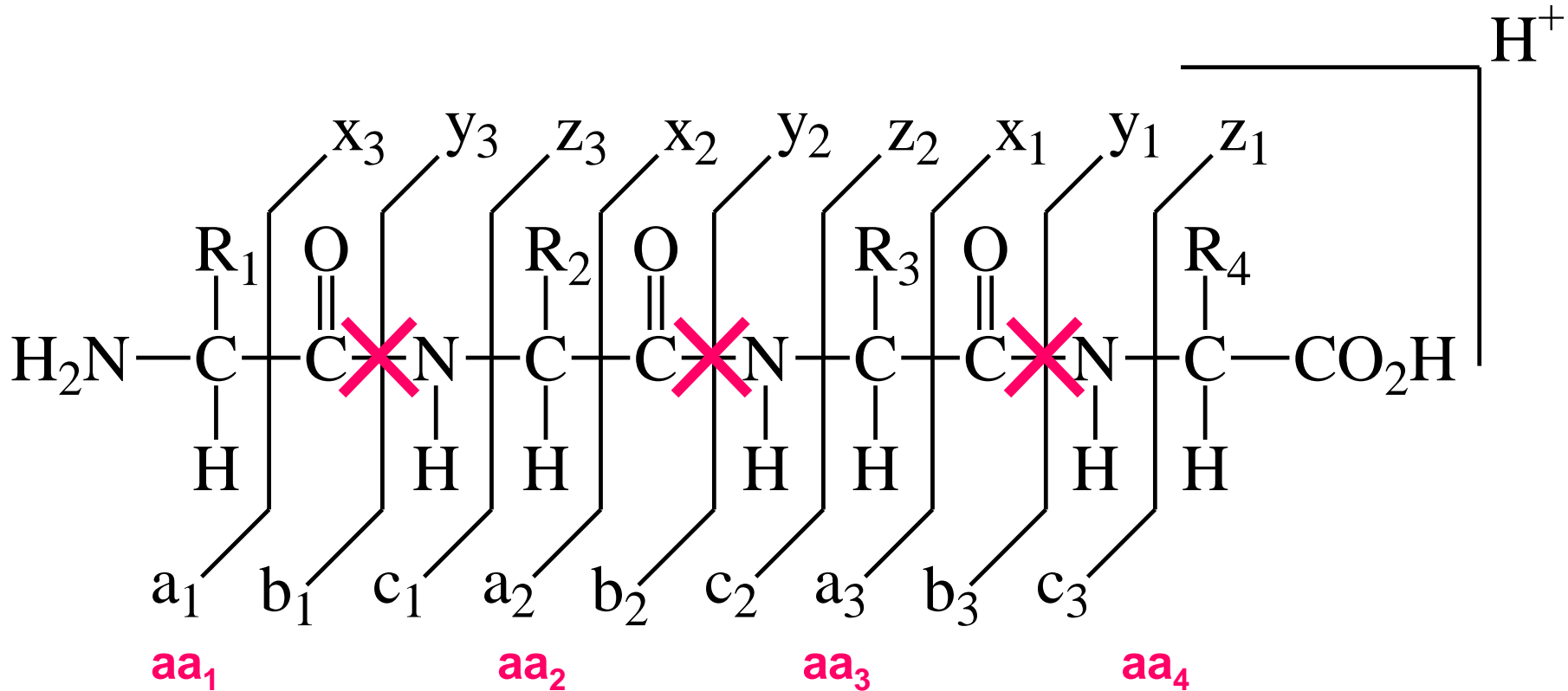


flow051202 #892 RT: 23.91 AV: 1 NL: 4.14E6
T: + c d Full ms2 500.10@35.00 [125.00-1015.00]



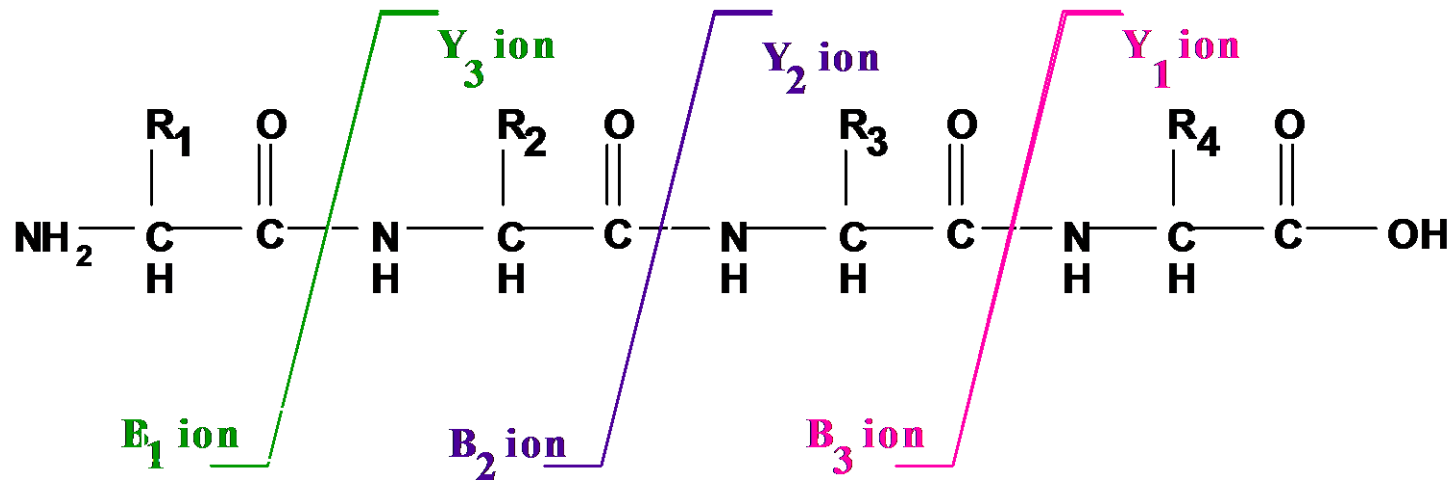
Peptide Fragmentation Nomenclature

b- and **y-**fragments most common in ion traps, q-tof, triple quads



B and Y Fragment Ions

Y-ions from C to N terminus



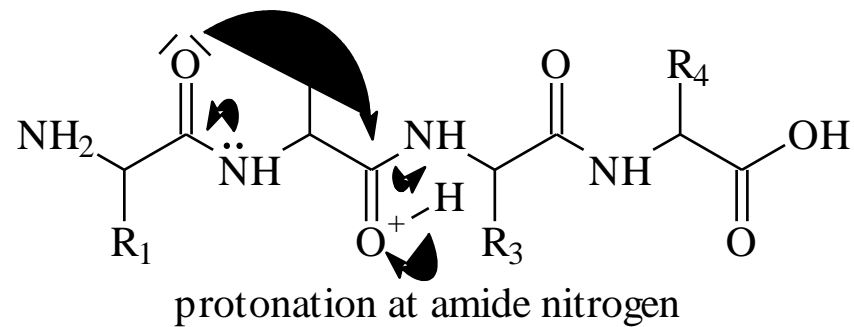
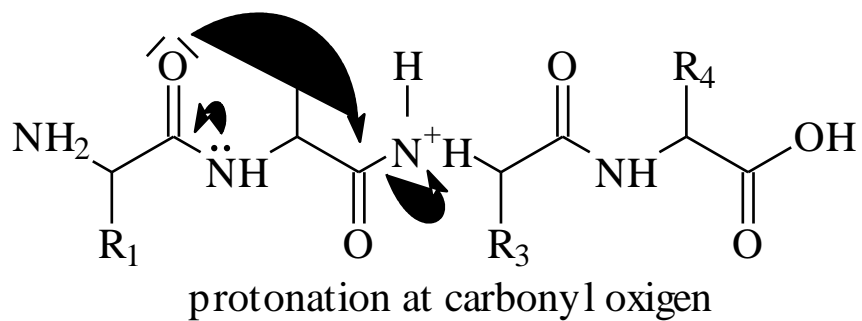
B-ions from N to C terminus

- **IT AND QTOF USED FOR MOST OF THE FRAGMENTATION STUDIES IN PROTEOMICS, SO LOW ENERGY AND CHARGE DIRECTED FRAGMENTATION APPLIES**
- **PEPTIDE FRAGMENTATION IS A GAS PHASE CHEMICAL REACTION: UNIMOLECULAR FRACTURE**
- **PEPTIDE FRAGMENTATION IS PRODUCED BY AN INTRAMOLECULAR NUCLEOPHILIC ATTACK**

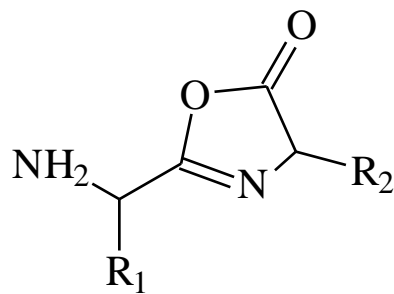
DAUGHTER IONS

- Produced by CID fragmentation of the precursor ion
- Frequently 2+ ions are fragmented
- Different types of fragmentation are possible. At low collision energy, fragmentation is produced mainly at the peptide bonds.
- Most CID peptide fragments keep either the N- or the C-Terminus, being classified accordingly.

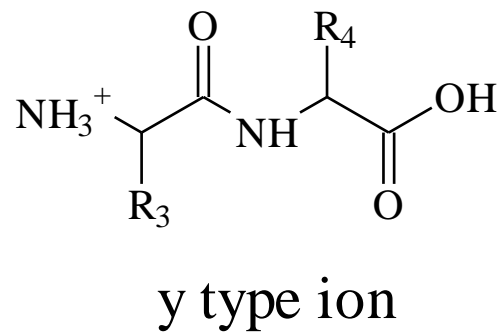
Formation of y ions



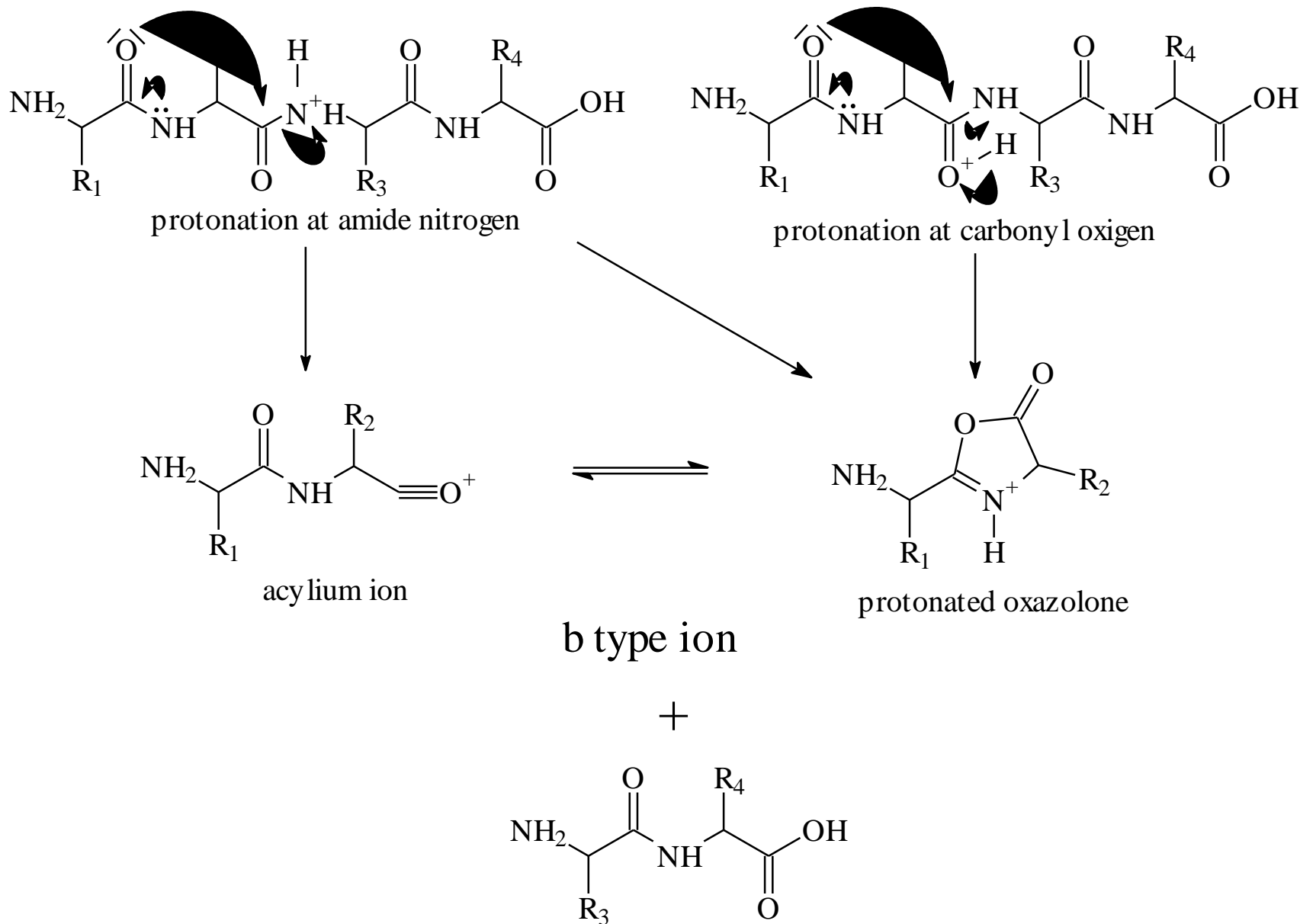
H⁺ transfer



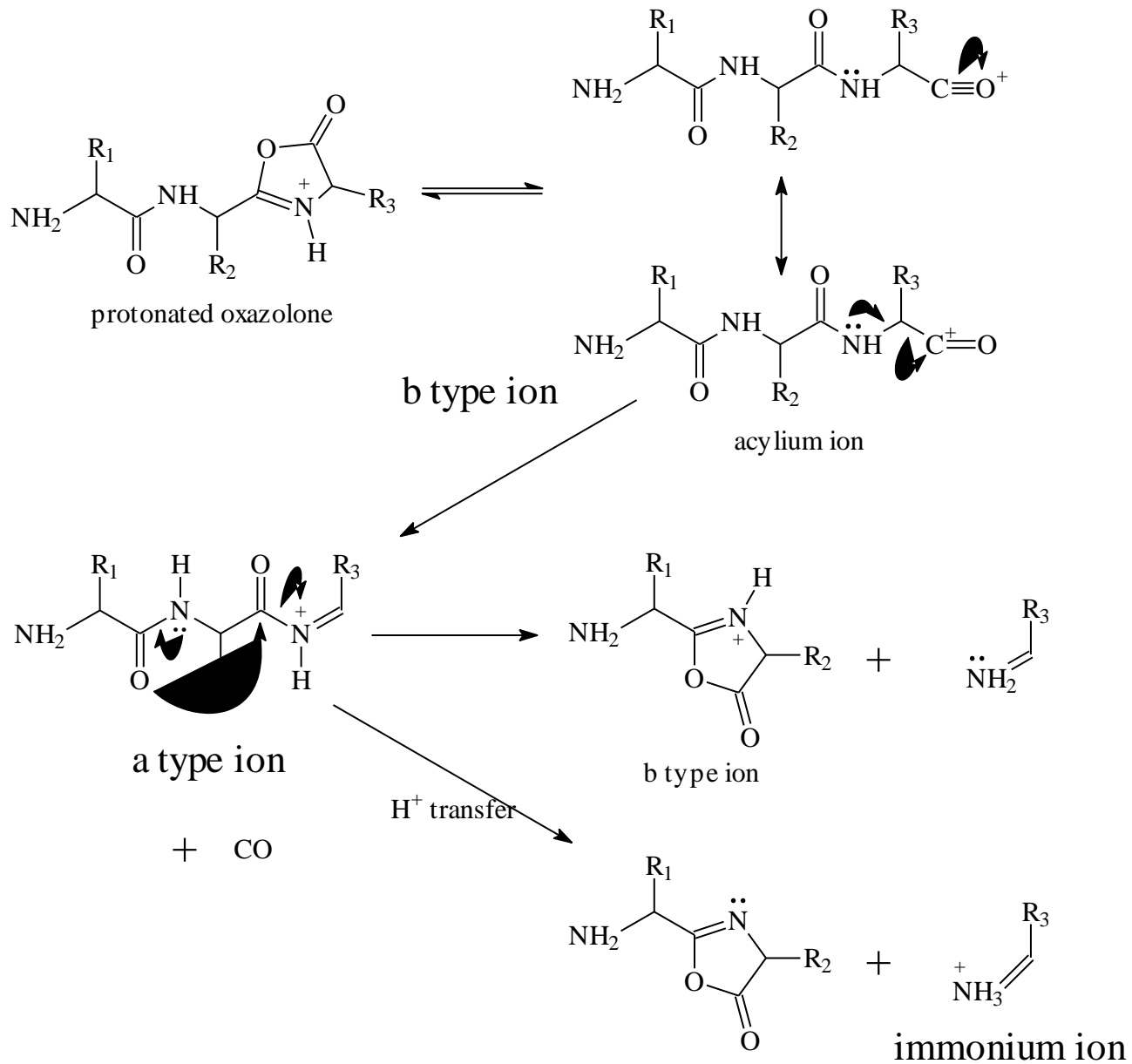
+

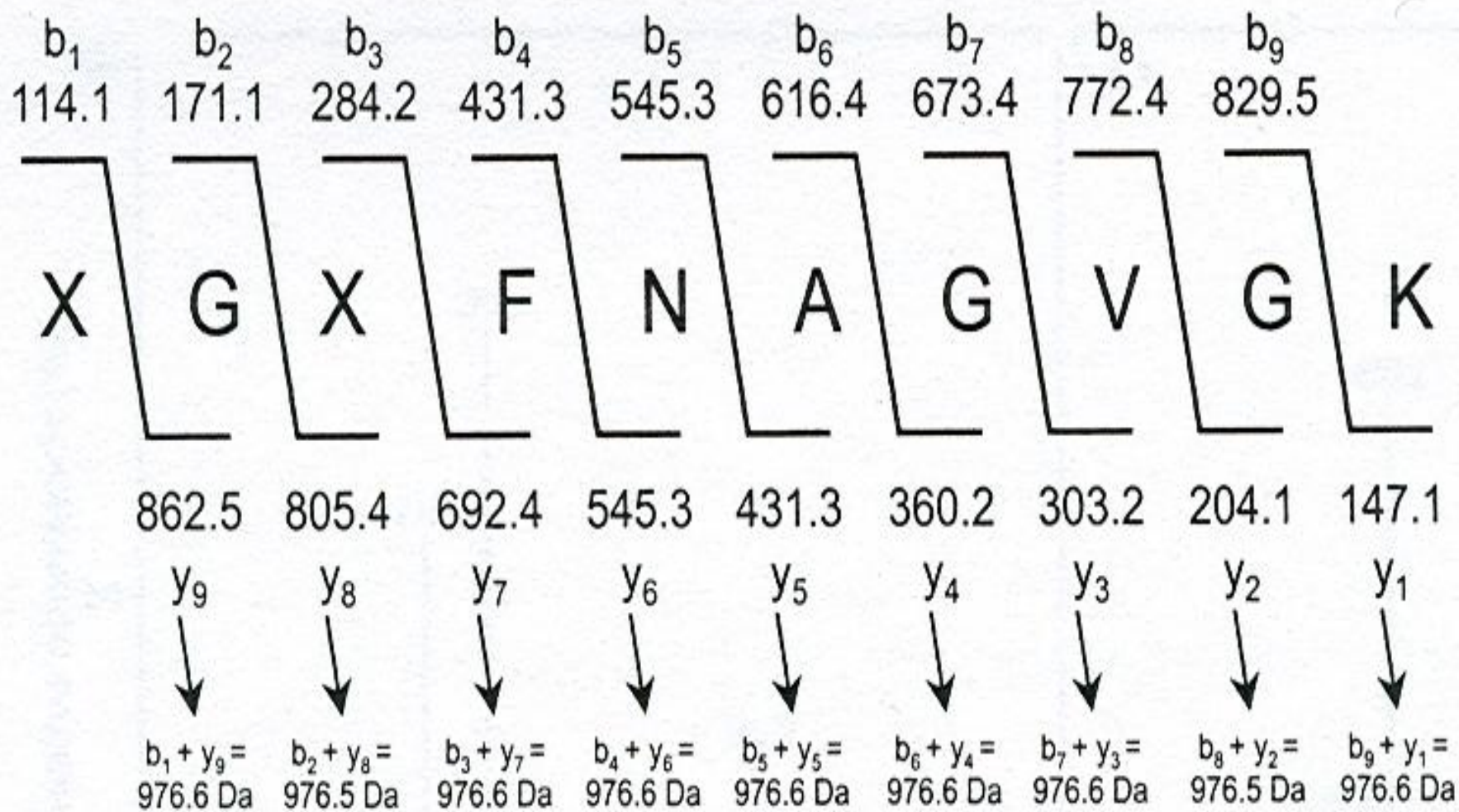


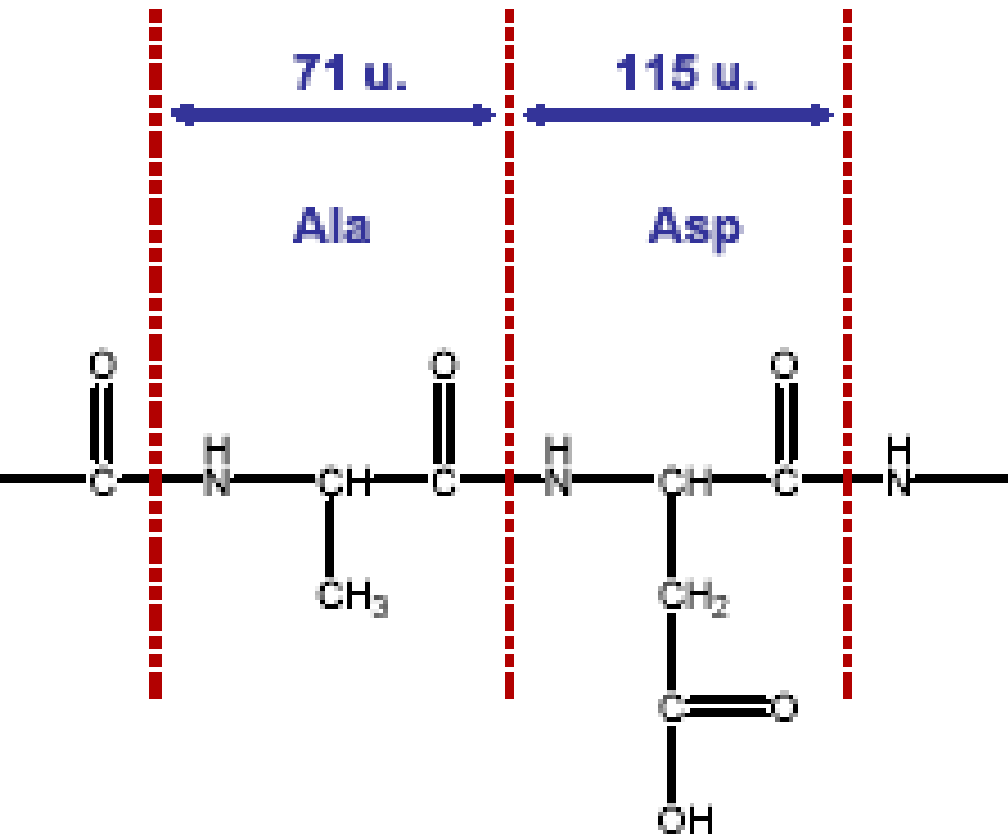
Formation of b ions



Formation of a and immonium ions







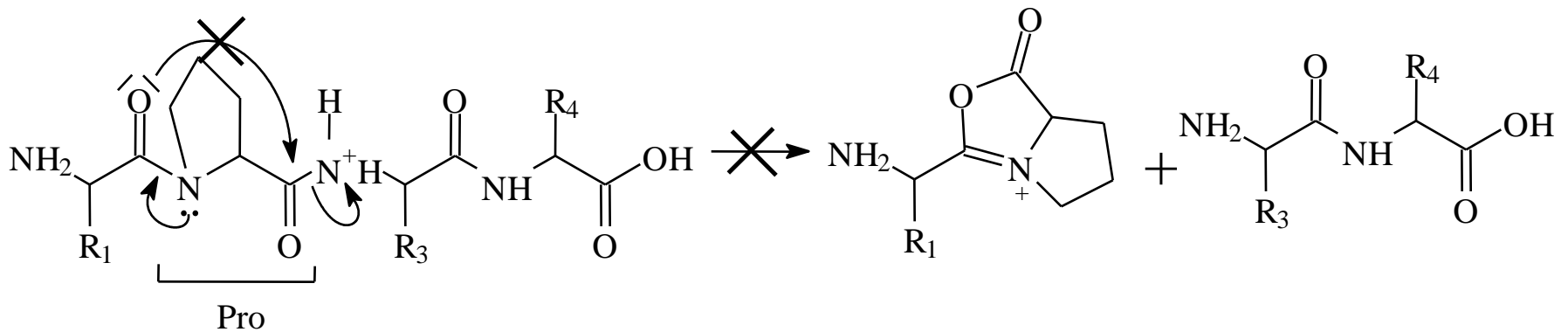
amino acid			mass
Alanine	ALA	A	71.09
Arginine	ARG	R	156.19
Aspartic Acid	ASP	D	115.09
Asparagine	ASN	N	114.11
Cysteine	CYS	C	103.15
Glutamic Acid	GLU	E	129.12
Glutamine	GLN	Q	128.14
Glycine	GLY	G	57.05
Histidine	HIS	H	137.14
Isoleucine	ILE	I	113.16
Leucine	LEU	L	113.16
Lysine	LYS	K	128.17
Methionine	MET	M	131.19
Phenylalanine	PHE	F	147.18
Proline	PRO	P	97.12
Serine	SER	S	87.08
Threonine	THR	T	101.11
Tryptophan	TRP	W	186.12
Tyrosine	TYR	Y	163.18
Valine	VAL	V	99.14

'Neutral Losses'

- Peptides lose water (-18 mass units) y^0 , b^0 , a^0
 - Most prominent losses from Ser, Thr side chains and from Glu N-terminus
 - Also Glu, Asp can lose water
- Peptides lose ammonia : Asn, Gln, Arg, Lys (-17 mass units) y^* , b^* , a^*
- The water/ammonia loss can be very prominent
- B ions lose CO (-28 mass units): **a ions**
- Neutral loss of CH_3SOH (-64 Da) from oxidized Met

Effect of Pro

- Proline is an unusual amino acid
- Secondary amine used in peptide bond formation
- Fragmentation on C-term of Pro is very unfavourable

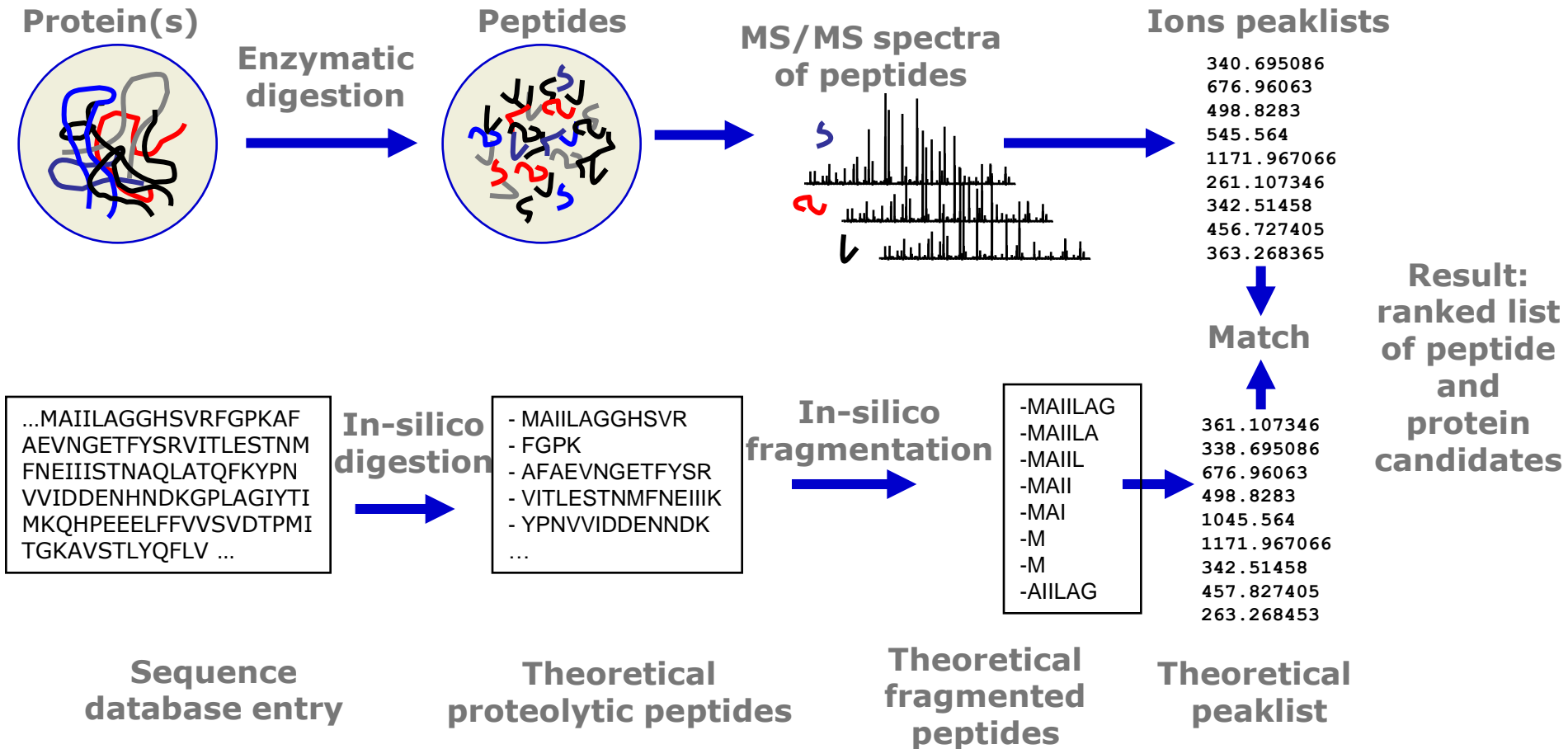


strained bicyclic ion

Resume of fragmentation rules

- Ions from y'' and b series predominate
- Neutral loss of CO (-28 Da) from b ions (a ions)
- Neutral loss of H₂O (-18 Da) from Cys, Ser, Thr, Glu
- Neutral loss of NH₃ (-17 Da) from Gln, Asn, Lys, Arg
- Immonium ions at low m/z range characteristic of aminoacids
- Neutral loss of CH₃SOH (-64 Da) from oxidized Met
- Neutral loss of PO₄H₃ (-98 Da) from phosphoSer and phosphoThr
- Hindered Pro-Xxx
- If mobile proton is available (n . charges > n . Arg residues):
 - Enhanced Xxx-Pro (and production of internal Pro series)
 - Enhanced His-Xxx (and production of internal His series)
- If mobile proton is not available (sequestering by Arg)
 - Enhanced Asp-Xxx and Glu-Xxx
 - Enhanced Asp-Pro

Peptide fragmentation fingerprinting = PFF = ion search MS/MS database matching



Instruments

Q-TOF

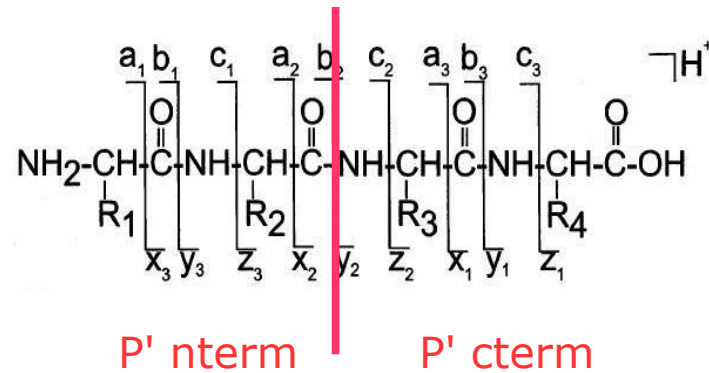
- easier to interpret since the spectra are less crowded (cleaner spectra)
- higher accuracy and resolution -which compensates for low frequency of a and b ions

3D Ion Traps

- a, b and y ions- less easier to interpret but more significant since more ions to confirm the sequence (noisy spectra)

Ion-types

Ion Type	Ion Mass	offset
a	[N]+[M]-CO	-28
a*	a-NH ₃	-45
a ^o	a-H ₂ O	-46
a ⁺⁺	(a+H)/2	
b	[N]+[M]	0
b*	b-NH ₃	-17
b ^o	b-H ₂ O	-18
b ⁺⁺	(b+H)/2	
c	[N]+[M]+NH ₃	+17
d	a-partial side chain	
v	y-complete side chain	
w	z-partial side chain	
x	[C]+[M]+CO	+28
y	[C]+[M]+H ₂	+ 2
y*	y-NH ₃	-15
y ^o	y-H ₂ O	-16
y ⁺⁺	(y+H)/2	
z	[C]+[M]-NH	-15



It is very important to know the ionic series produced by a spectrometer, otherwise potential matches will be missed.

In the other hand, if an ion-type not present in the original spectrum is taken into account, it will contribute to get false positive matches.

[N] is the mass of the N-term group

[M] is the mass of the sum of the neutral amino acid residue masses

MS/MS based identification tools

- Tag search- Tools that search peptides based on a MS/MS Sequence Tag
 - MS-Tag and MS-Seq, PeptideSearch
- Ion search or PFF - Tools that match MS/MS experimental spectra with “theoretical spectra” obtained via in-silico fragmentation of peptides generated from a sequence database
 - Phenyx, Mascot, Sequest, X!Tandem, OMSSA, ProID, ...
- *de novo* sequencing - Tools that directly interpret MS/MS spectra and try to deduce a sequence
 - Convolution/alignment (PEDENTA)
 - De-novo sequencing followed by sequence matching (Peaks, Lutefisk, Sherenga, PeptideSearch)
 - Guided Sequencing (Popitam)

In all cases, the output is a peptide structure per MS/MS spectrum

Some PFF tools

Same principle of a PMF, but using MS/MS spectra

Software	Source website
InsPecT	peptide.ucsd.edu/inspect.py
Mascot	www.matrixscience.com/search_form_select.html
MS-Tag and MS-Seq	prospector.ucsf.edu
PepFrag	prowl.rockefeller.edu/prowl/pepfragch.html
Phenyx	phenyx.vital-it.ch
Popitam	www.expasy.org/tools/popitam
ProID (download)	sashimi.sourceforge.net/software_mi.html
Sequest*	fields.scripps.edu/sequest/index.html
Sonar	65.219.84.5/service/prowl/sonar.html
SpectrumMill*	www.home.agilent.com
VEMS	www.bio.aau.dk/en/biotechnology/vems.htm
X!Tandem (download)	www.thegpm.org/TANDEM

*Commercialized

Non exhaustive list!

Searches with MS/MS Data

- In a similar fashion to peptide mass fingerprinting, the predicted fragment ion mass from each peptide of a database sequence are calculated
- The calculated and observed ion masses are compared and given a score
- Individual peptide scores are combined to give a protein score

Mascot: MS/MS Ions Search

[Your name](#) [Email](#)

[Search title](#)

[Database](#)

[Taxonomy](#)

[Enzyme](#) [Allow up to](#) missed cleavages

[Fixed modifications](#)

- Acetyl (K)
- Acetyl (N-term)
- Amide (C-term)
- Biotinylated (K)
- Biotinylated (N-term)

[Variable modifications](#)

- Oxidation (HW)
- Oxidation (M)
- PEO Biotin (C)
- Phospho (ST)
- Phospho (Y)

[Protein mass](#) kDa [ICAT](#)

[Peptide tol. ±](#) Da [MS/MS tol. ±](#) Da

[Peptide charge](#) [Monoisotopic](#) [Average](#)

[Data file](#)

[Data format](#) [Precursor](#) m/z

[Instrument](#)

[Overview](#) [Report top](#) hits

[HOME](#)[MASCOT](#)[HELP](#)[WHAT'S NEW](#)[PRODUCTS](#)[SUPPORT](#)[SITE SEARCH](#)[LINKS](#)[EMPLOYMENT](#)[CONTACT US](#)

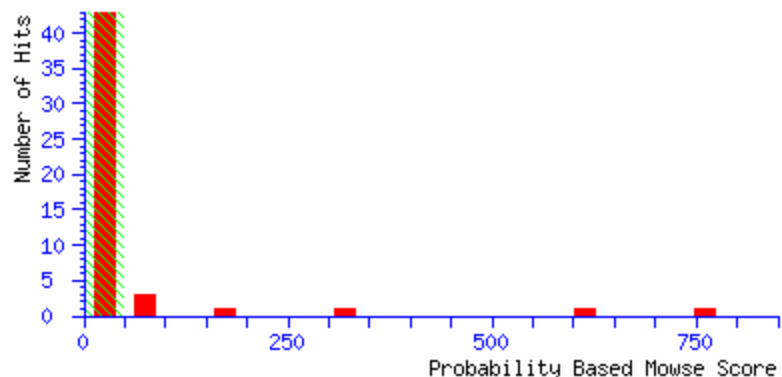
Mascot Search Results

[HOME](#)[MASCOT](#)[HELP](#)[WHAT'S NEW](#)[PRODUCTS](#)[SUPPORT](#)[SITE SEARCH](#)[LINKS](#)[EMPLOYMENT](#)[CONTACT US](#)

User : bcanas
Email : bcanasmon@yahoo.co.uk
Search title :
MS data file : C:\WINNT\Profiles\PS1\Desktop\BSA004.pkl
Database : NCBI nr 20020220 (883557 sequences; 276611002 residues)
Timestamp : 15 Mar 2002 at 11:09:00 GMT
Significant hits: [gi|2190337](#) (X58989) serum albumin [Bos taurus]
[gi|229552](#) albumin [Bos taurus]
[gi|3319897](#) (Y17737) albumin [Canis familiaris]
[gi|1351908](#) Serum albumin precursor (Allergen Fel d 2)
[gi|28590](#) (V00494) reading frame HSA [Homo sapiens]
[gi|11277085](#) serum albumin - mouse (fragment)
[gi|18309363](#) (NC_003366) conserved hypothetical protein [Clostridium perfringens]

Probability Based Mowse Score

Score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event. Individual ions scores > 48 indicate identity or extensive homology ($p < 0.05$).



Select All

Select None

Search Selected

 Error tolerant[HOME](#)[MASCOT](#)[HELP](#)[WHAT'S NEW](#)[PRODUCTS](#)[SUPPORT](#)[SITE SEARCH](#)[LINKS](#)[EMPLOYMENT](#)[CONTACT US](#)

1. [gi|2190337](#) **Mass:** 69278 **Total score:** 760 **Peptides matched:** 19
(X58989) serum albumin [Bos taurus]

Check to include this hit in error tolerant search

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Rank	Peptide
<input checked="" type="checkbox"/> 2	464.20	926.38	926.49	-0.10	0	40	1	YLYEIAR
<input checked="" type="checkbox"/> 5	507.74	1013.46	1013.61	-0.15	0	67	1	QTALVELLK
<input checked="" type="checkbox"/> 6	507.76	1013.50	1013.61	-0.11	0	(48)	1	QTALVELLK
<input checked="" type="checkbox"/> 7	571.79	1141.55	1141.71	-0.15	1	76	1	KQTALVELLK
<input checked="" type="checkbox"/> 8	582.25	1162.48	1162.62	-0.15	0	70	1	LVNELTEFAK
<input checked="" type="checkbox"/> 9	582.25	1162.49	1162.62	-0.14	0	(70)	1	LVNELTEFAK
<input checked="" type="checkbox"/> 10	642.28	1282.54	1282.70	-0.16	0	44	1	HPEYAVSVLLR
<input checked="" type="checkbox"/> 11	642.28	1282.55	1282.70	-0.15	0	(25)	1	HPEYAVSVLLR
<input checked="" type="checkbox"/> 13	653.25	1304.49	1304.71	-0.22	0	55	1	HLVDEPQNLIK
<input checked="" type="checkbox"/> 14	700.27	1398.52	1398.69	-0.17	0	67	1	TVMENFVAFVDK
<input checked="" type="checkbox"/> 15	480.57	1438.67	1438.80	-0.13	1	48	1	RHPEYAVSVLLR
<input checked="" type="checkbox"/> 16	740.33	1478.64	1478.79	-0.15	0	(85)	1	LGEYGFQNALIVR
<input checked="" type="checkbox"/> 17	740.33	1478.64	1478.79	-0.15	0	85	1	LGEYGFQNALIVR
<input checked="" type="checkbox"/> 20	784.27	1566.52	1566.74	-0.21	0	87	1	DAFLGSFLYEYSR
<input checked="" type="checkbox"/> 21	784.32	1566.62	1566.74	-0.12	0	(87)	1	DAFLGSFLYEYSR
<input checked="" type="checkbox"/> 24	820.38	1638.74	1638.93	-0.19	1	53	1	KVPQVSTPTLVEVSR
<input checked="" type="checkbox"/> 27	944.89	1887.76	1887.92	-0.16	0	(32)	1	HPYFYAPELLYYANK
<input checked="" type="checkbox"/> 28	944.89	1887.76	1887.92	-0.16	0	32	1	HPYFYAPELLYYANK
<input checked="" type="checkbox"/> 30	978.40	1954.78	1954.95	-0.17	0	40	1	DAIPENLPPLTADFAEDK

Proteins matching the same set of peptides:

[gi|418694](#) **Mass:** 69225 **Total score:** 760 **Peptides matched:** 19

Mascot Search Results

[HOME](#)[MASCOT](#)[HELP](#)[WHAT'S NEW](#)[PRODUCTS](#)[SUPPORT](#)[SITE SEARCH](#)[LINKS](#)[EMPLOYMENT](#)[CONTACT US](#)

Protein View

Match to: **gi|2190337**; Score: **760****(X58989) serum albumin [Bos taurus]**

Found in search of C:\WINNT\Profiles\PS1\Desktop\BSA004.pk1

Nominal mass (M_r): **69278**; Calculated pI value: **5.82**NCBI BLAST search of [gi|2190337](#) against nrUnformatted [sequence string](#) for pasting into other applicationsTaxonomy: [Bos taurus](#)

Links to retrieve other entries containing this sequence from NCBI Entrez:

[gi|3336842](#) from [Bos taurus](#)

Cleavage by Trypsin: cuts C-term side of KR unless next residue is P

Sequence Coverage: **22%**Matched peptides shown in **Bold Red**

```

1  MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA
51  FSQYLQCCPF DEHVKLVNEL TEFAKTCVAD ESHAGCEKSL HTLFGDELCK
101 VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKPK PDPNTLCDEF
151 KADEKKFWGK YLYEIARRHP YFYAPPELLYY ANKYNGVFQE CCQAEDKGAC
201 LLPKIETMRE KVLTSARQR LRCASIQKFG ERLKAWVA RLSQKFPKAE
251 FVEVTKLVTD LTKVHKECCH GDLLCADDR ADLAKYICDN QDTISSKLKE
301 CCDKPLLEKS HCIAEVEKDA IPENLPLTA DFAEDKDVCK NYQEAKDAFL
351 GSFLYEYSRR HPEYAVSVLL RLAKEYEATL EECCAADDPH ACYSTVFDKL
401 KHLVDEPQNL IKQNCQDFEK LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS
451 RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT PVSEKVTKCC
501 TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT
551 ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPKLVV
601 STQTALA

```

MATRIX SCIENCE Mascot Search Results

Peptide View

MS/MS Fragmentation of **LVNELTEFAK**

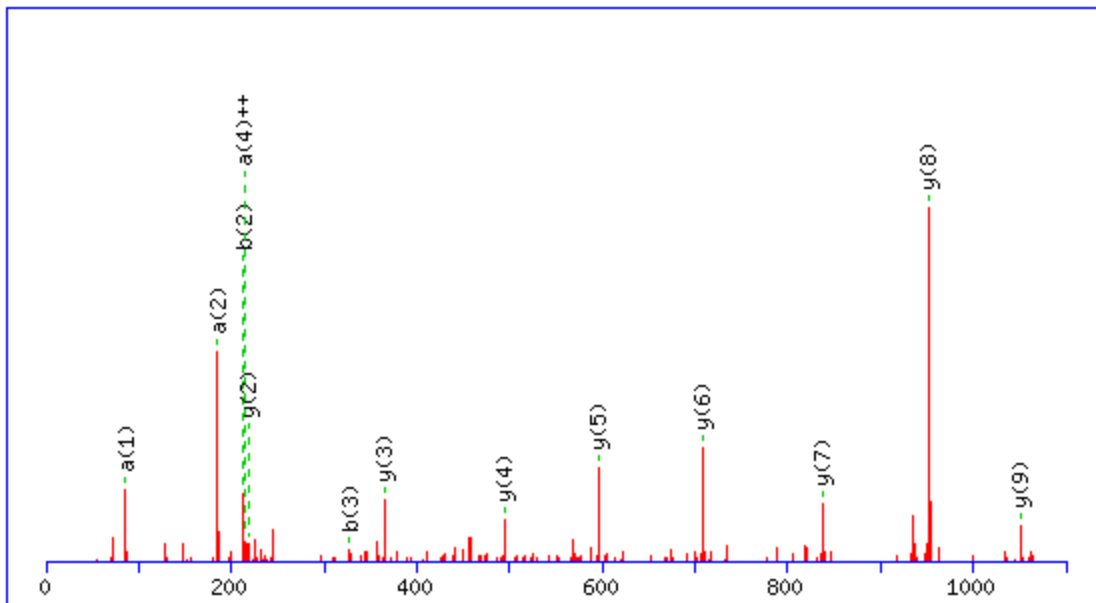
Found in [gi|2190337](#), (X58989) serum albumin [Bos taurus]

Match to Query 8 (582.25,2+)

From data file C:\WINNT\Profiles\PS1\Desktop\BSA004.pkl

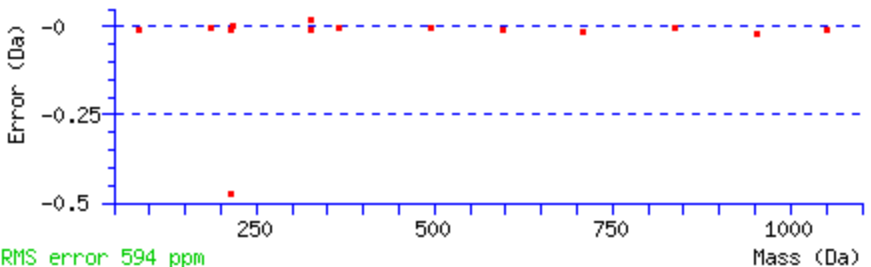
Click mouse within plot area to zoom in by factor of two about that point

Or, Plot from to Da



Monoisotopic mass of neutral peptide (Mr): 1162.62
Ions Score: 70 **Matches (Bold Red):** 14/100 fragment ions using 22 most intense peaks

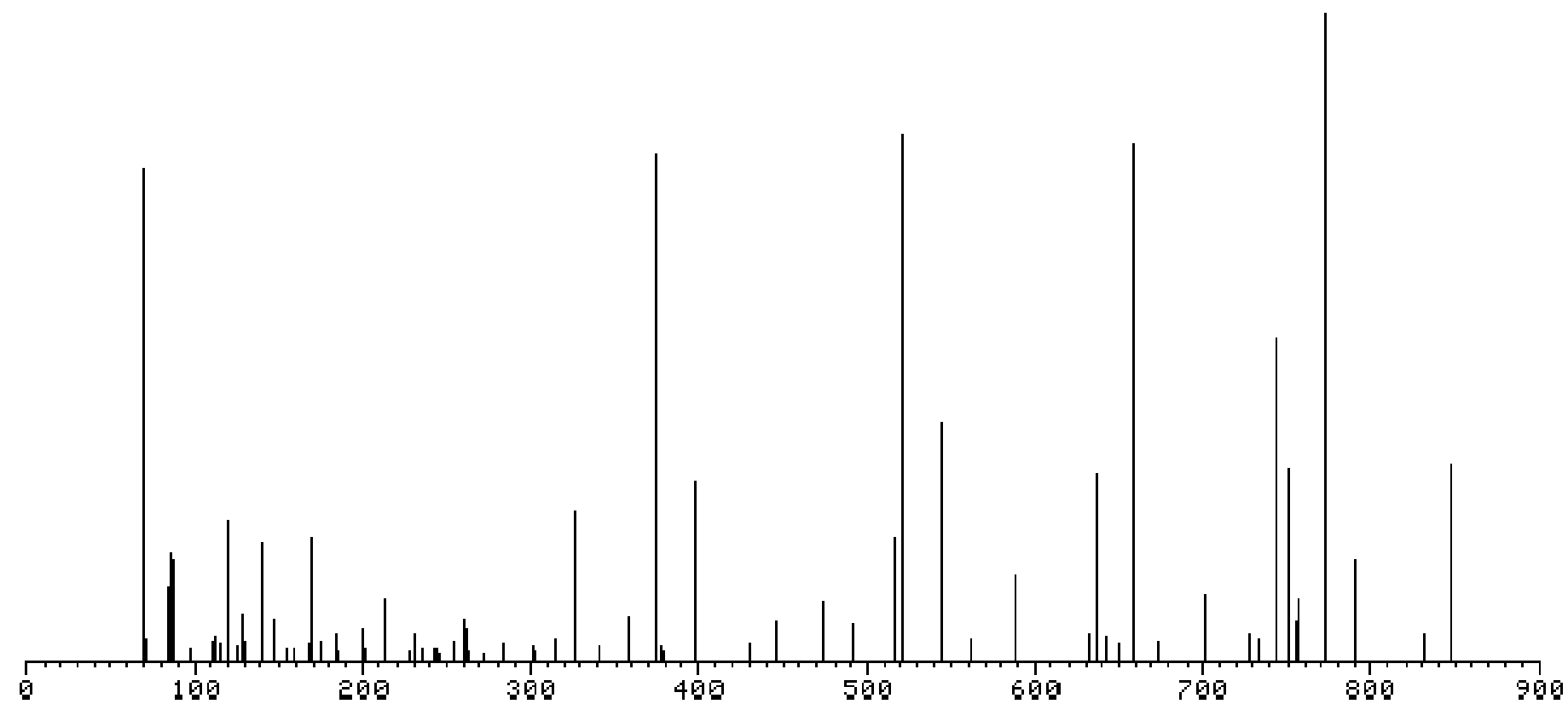
#	a	a ⁺⁺	a [*]	a ⁺⁺⁺	b	b ⁺⁺	b [*]	b ⁺⁺⁺	Seq.	y	y ⁺⁺	y [*]	y ⁺⁺⁺	#
1	86.10	43.55			114.09	57.55			L					10
2	185.17	93.09			213.16	107.08			V	1050.55	525.78	1033.52	517.26	9
3	299.21	150.11	282.18	141.59	327.20	164.11	310.18	155.59	N	951.48	476.24	934.45	467.73	8
4	428.25	214.63	411.22	206.12	456.25	228.63	439.22	220.11	E	837.44	419.22	820.41	410.71	7
5	541.33	271.17	524.31	262.66	569.33	285.17	552.30	276.66	L	708.39	354.70	691.37	346.19	6
6	642.38	321.70	625.36	313.18	670.38	335.69	653.35	327.18	T	595.31	298.16	578.28	289.65	5
7	771.43	386.22	754.40	377.70	799.42	400.21	782.39	391.70	E	494.26	247.63	477.23	239.12	4
8	918.49	459.75	901.47	451.24	946.49	473.75	929.46	465.23	F	365.22	183.11	348.19	174.60	3
9	989.53	495.27	972.50	486.76	1017.53	509.27	1000.50	500.75	A	218.15	109.58	201.12	101.07	2
10									K	147.11	74.06	130.09	65.55	1



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

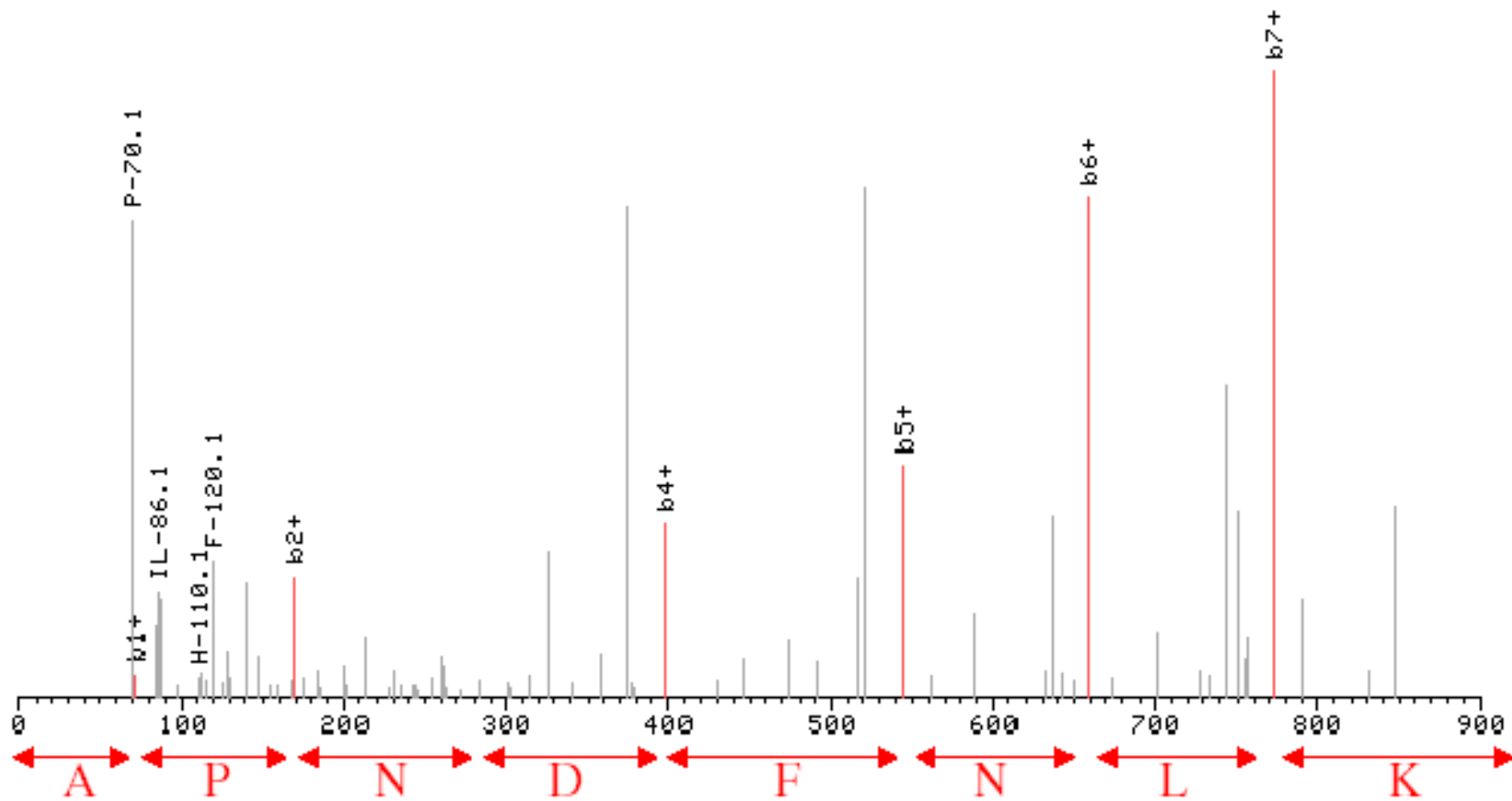
De-Novo Sequencing

- **If the protein is still not identified, the sequence of a peptide has to be reconstructed from the MS/MS data**
- **Very time consuming and demands a great deal of skill, noisy data is very problematic**
- **Sequencing is carried out by finding mass differences between peaks that correspond to amino acid masses**



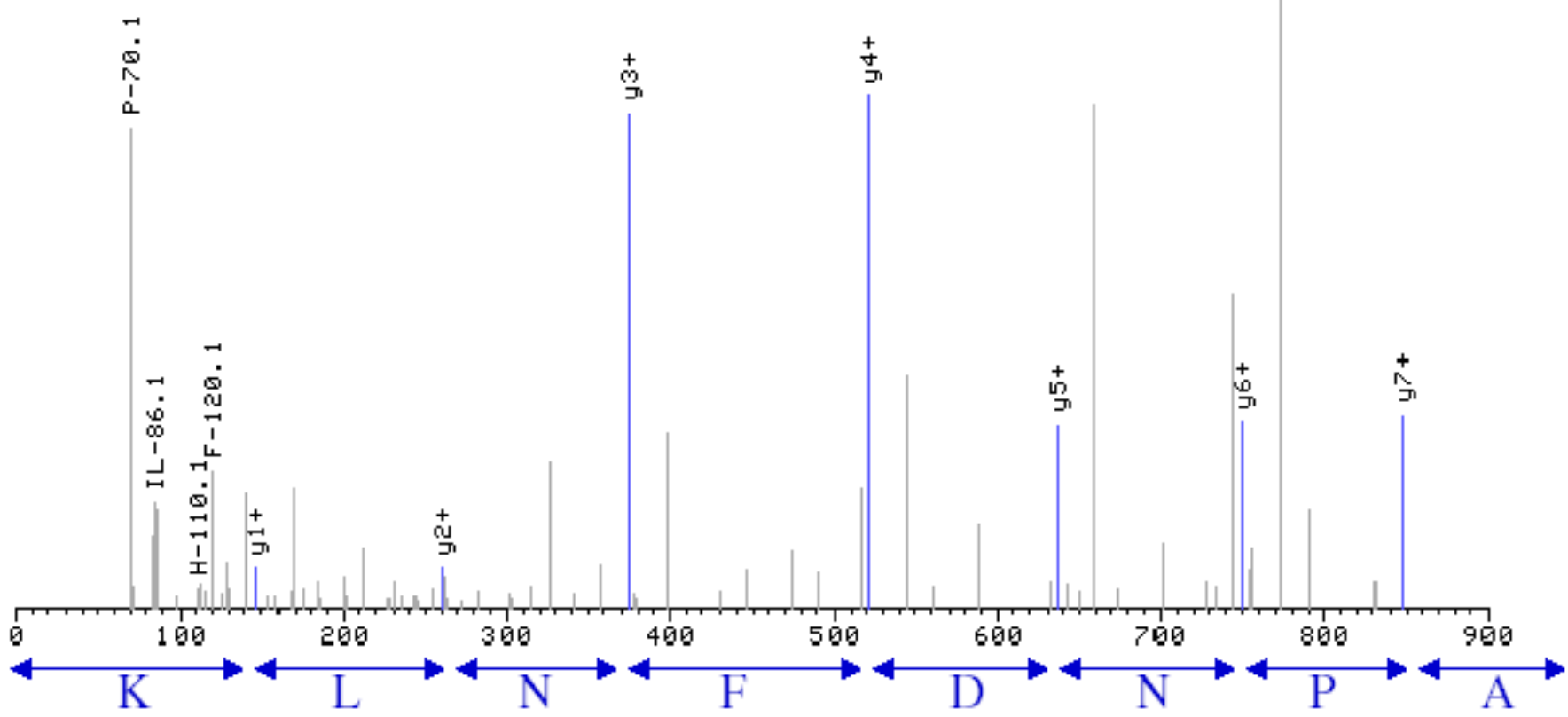
A-P-N-D-F-N-L-K

(MH+ 804.9)

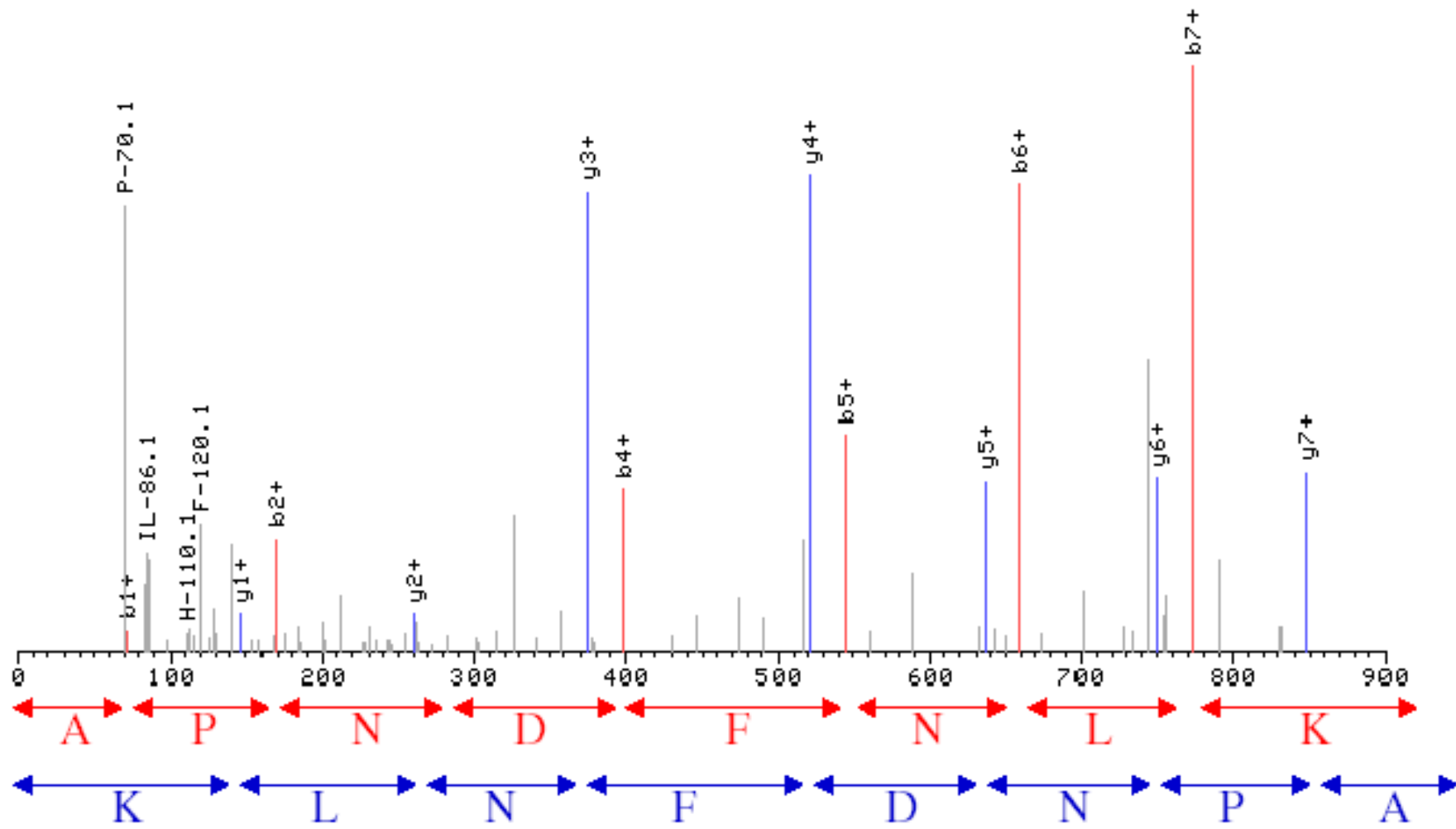


B-ions

7.3e+04



Y-ions

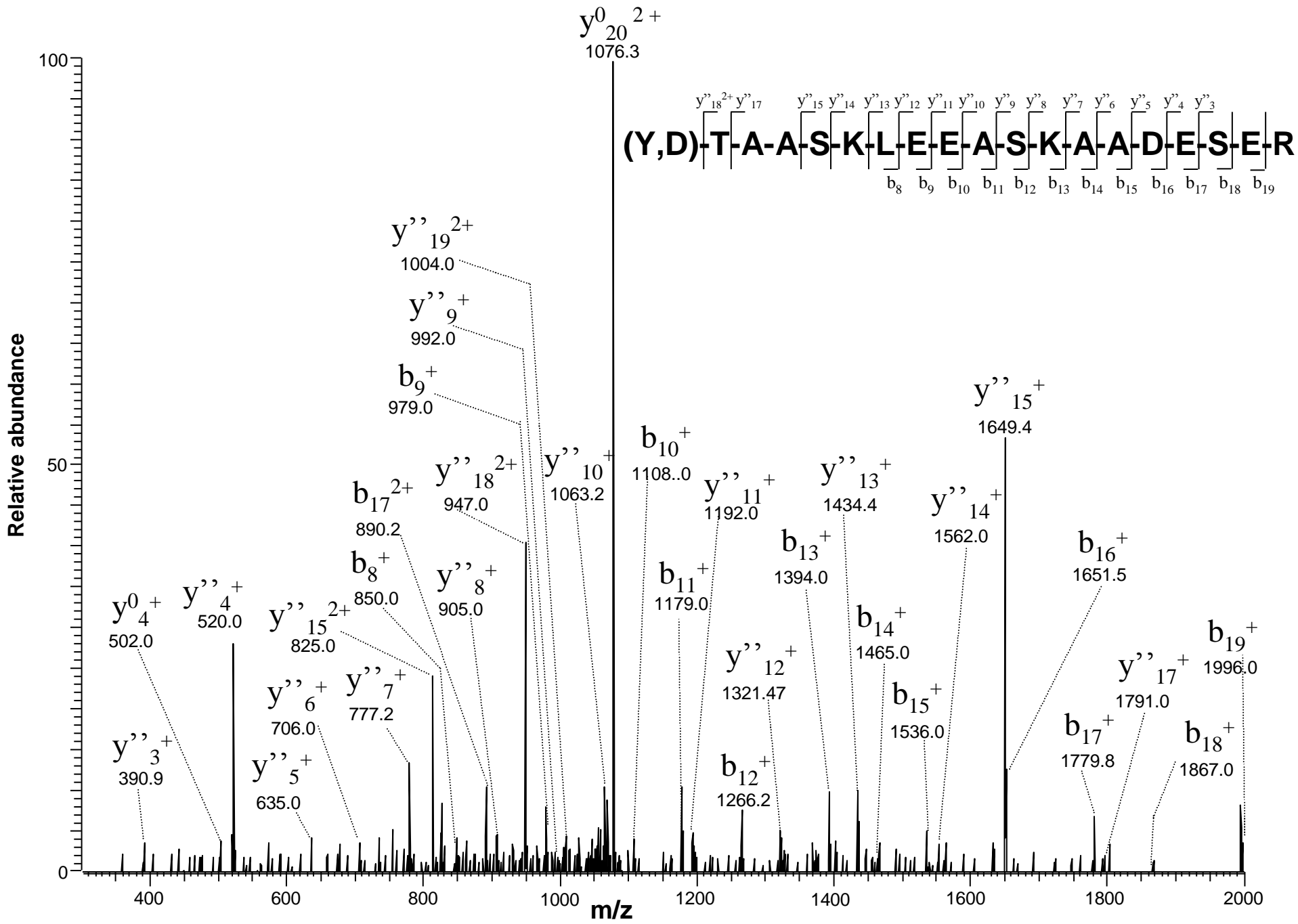


Why de novo sequencing is difficult

1. Leucine and isoleucine have the same mass
2. Glutamine and lysine differ in mass by 0.036Da
3. Phenylalanine and oxidized methionine differ in mass by 0.033Da
4. Cleavages do not occur at every peptide bond (or cannot be observed on the MS-MS)
 - Poor quality spectrum (some fragment ions are below noise level)
 - The C-terminal side of proline is often resistant to cleavage
 - Absence of mobile protons
 - Peptides with free N-termini often lack fragmentation between the first and second amino acids

Why de novo sequencing is difficult (II)

5. Certain amino acids have the same mass as pairs of other amino acids
 - Gly +Gly (114.0429) Asn (114.0429)
 - Ala +Gly (128.0586) Gln (128.0586)
 - Ala +Gly (128.0586) Lys (128.0950)
 - Gly + Val (156.0742) Arg (156.1011)
 - Ala + Asp (186.0641) Trp (186.0793)
 - Ser + Val (186.1005) Trp (186.0793)
6. Directionality of an ion series is not always known (are they b- or y-ions?)



Summary of *de novo* sequencing tools

Software	Source website
PEAKS	www.bioinformaticsolutions.com
SeqMS (download)	www.protein.osaka-u.ac.jp/rcsfp/profiling/SeqMS.html
Sherenga (included in SpectrumMill)	N/A
Lutefisk (download)	www.hairyfatguy.com/Lutefisk
DeNovoX*	www.thermo.com
PepNovo	peptide.ucsd.edu/pepnovo.py
SpectrumMill*	www.home.agilent.com

*Commercialized