

La Spettrometria di Massa nello Studio di Proteine e Peptidi



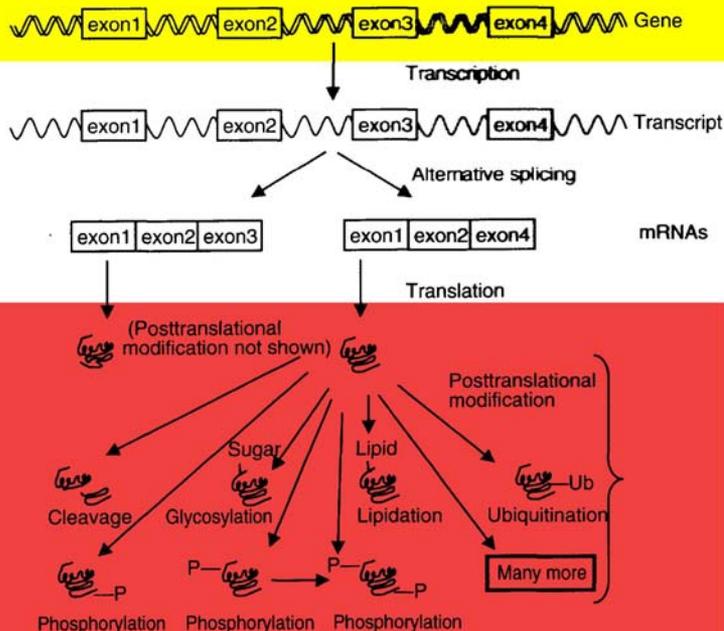
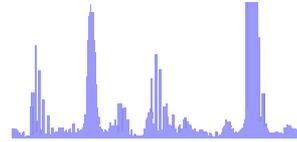
Gianluca Giorgi

Università degli Studi di Siena
Dipartimento di Biotecnologie, Chimica
e Farmacia

via Aldo Moro
53100 Siena

e-mail: gianluca.giorgi@unisi.it

Tel. 0577-234241



Challenges for mass spectrometry in proteome and genome analysis

Challenges	Requirements
mass spectrometry can handle only simple analyte mixtures, adduction complicates mass spectra	<ul style="list-style-type: none">• sample preparation and purification using high-resolution separation methods• high-resolution mass analyzers
low amounts of sample available	<ul style="list-style-type: none">• low detection limits• miniaturization
many samples to be screened	<ul style="list-style-type: none">• fast scanning rate• multiplexing
Identification of compounds	<ul style="list-style-type: none">• high mass accuracy• MS/MS capability• Spectral libraries• database search

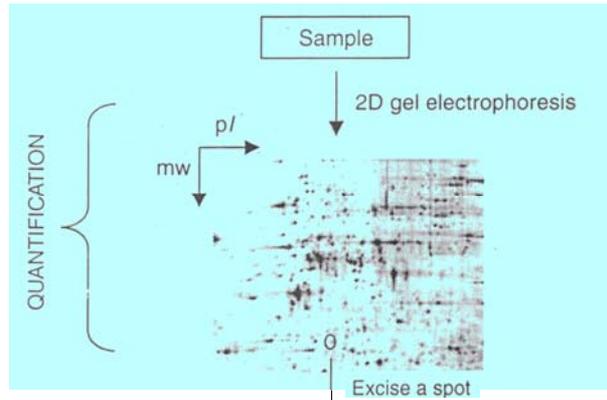
La Spettrometria di Massa nello studio di Proteine e Peptidi

1. Studio di Proteine

- i) Estrazione
- ii) Separazione
- iii) Determinazione del PM
 - i) MALDI-TOF: spettro, risoluzione
 - ii) ESI: ioni multicarica, risoluzione
 - iii) Quale massa?
 - iv) Ricerca in banca dati
- iv) Determinazione della struttura (Sequenza AA): MSⁿ
 - i) Proteina *in toto*: ECD, ETD
 - ii) Digestione enzimatica → **Peptidi**
 - i) MALDI: peptide mass fingerprinting
 - ii) HPLC-ESI-MS e MSⁿ
 - iii) Ricerca in banca dati

1. Studio di Proteine

- i) Estrazione
- ii) Separazione



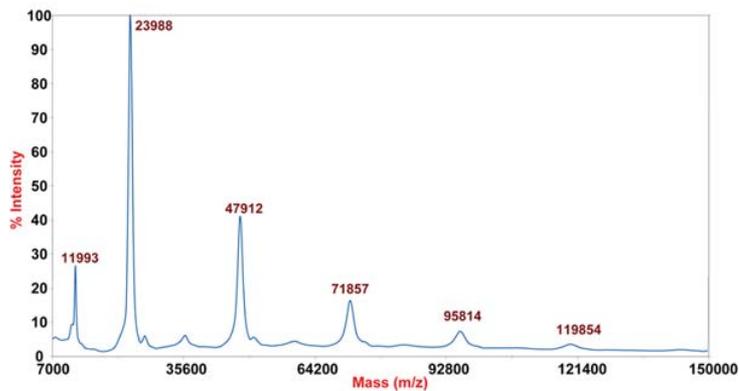
iii) Determinazione del peso molecolare

MALDI

ESI

Proteine - Determinazione del PM

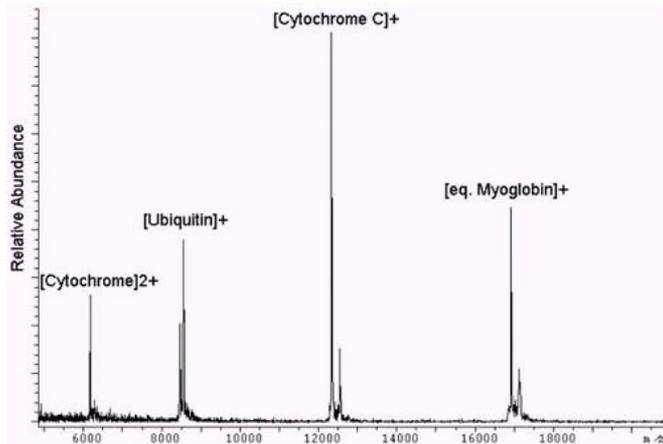
MALDI-TOF: spettro, risoluzione



MALDI-TOF linear mass spectrum of intact protein: Hg-Papain oligomerization

Proteine - Determinazione del PM

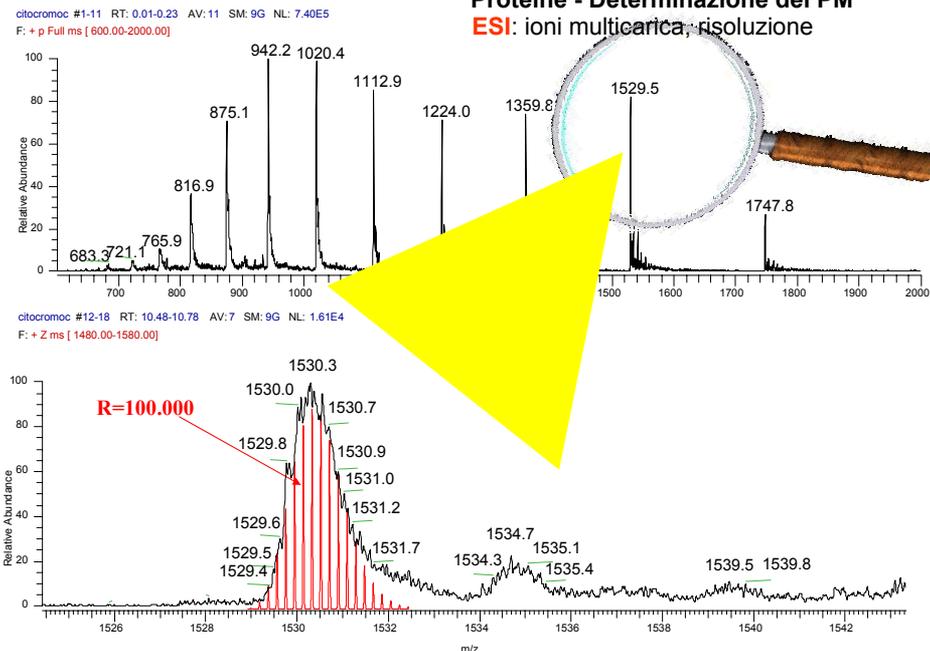
MALDI-TOF: spettro, risoluzione



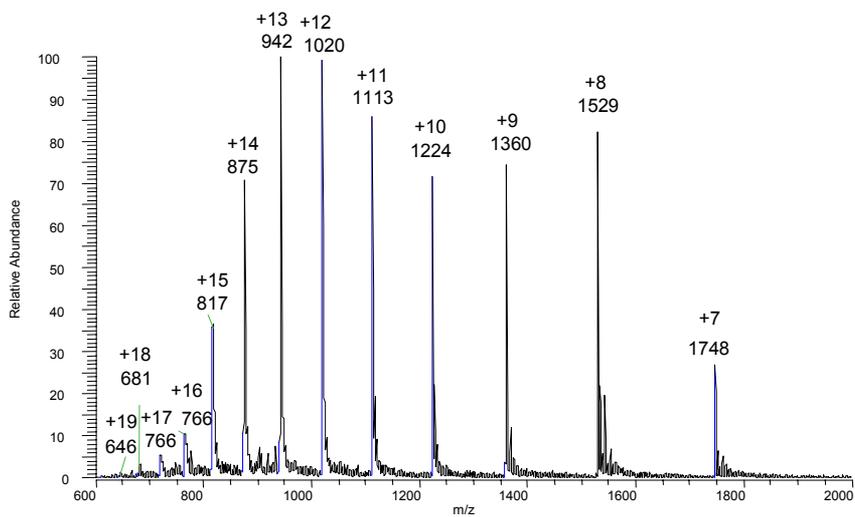
MALDI TOF mass spectrum of a mixture of ubiquitin, cytochrome C and equine myoglobin using 2,5-dihydroxybenzoic acid (DHB) as the matrix.

Proteine - Determinazione del PM

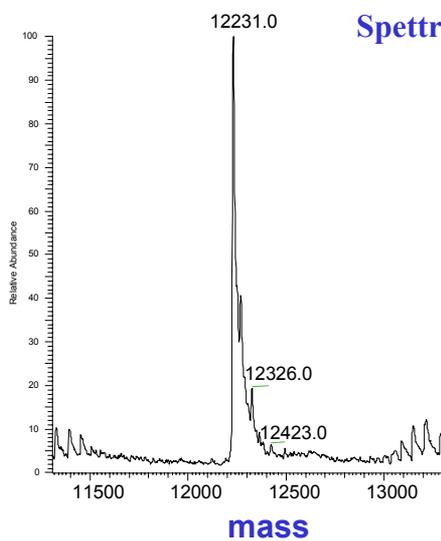
ESI: ioni multicarica, risoluzione



Assegnamento delle cariche



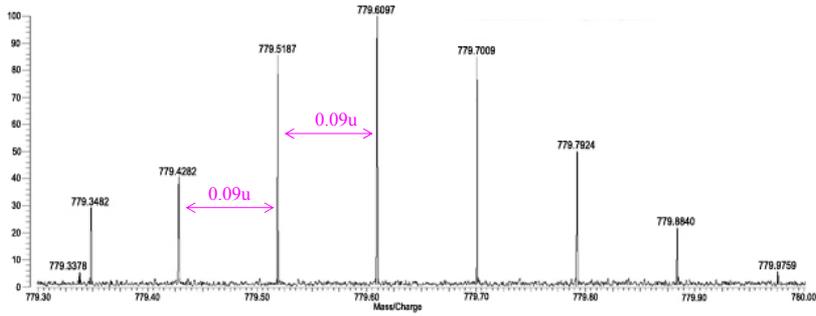
Spettro deconvoluto *in silico*



FT-ICR

ESI (+) of Ubiquitin

1,040,000 Resolution at m/z 779



Proteine & Peptidi- Determinazione del PM

Spettrometria di **massa**

Quale massa?

L'accuratezza di massa diminuisce all'aumentare della grandezza dello ione



Poiché una proteina è una molecola grande, occorre **elevatissima risoluzione e notevole accuratezza di massa (FT-ICR)** per poter avere informazioni univoche sulla sua formula bruta

La Spettrometria di Massa nello studio di Proteine e Peptidi

1. Studio di Proteine

i) Estrazione

ii) Separazione

iii) Determinazione del PM

- i) MALDI-TOF: spettro, risoluzione
- ii) ESI: ioni multicarica, risoluzione
- iii) Quale massa?
- iv) Ricerca in banca dati

iv) Determinazione della struttura (Sequenza AA): MSⁿ

i) Proteina *in toto*: ECD, ETD

ii) Digestione enzimatica → **Peptidi**

- i) MALDI: peptide mass fingerprinting
- ii) HPLC-ESI-MS e MSⁿ

iii) Ricerca in banca dati

Determinazione della struttura: MSⁿ

i) Proteina *in toto*:

CID non riesce a frammentare ioni prodotti da molecole grandi (PM > 3000 ca)

La decomposizione di ioni prodotti da molecole grandi può avvenire solo attraverso l'interazione con elettroni:

ECD = electron capture dissociation (FT-ICR)

ETD = electron transfer dissociation (2D e 3D IT)

Determinazione della struttura: MSⁿ

i) Proteina *in toto*:

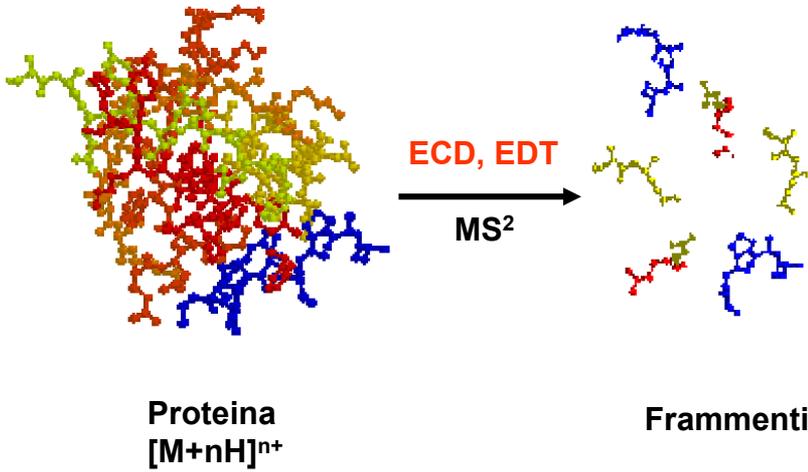
ECD = electron capture dissociation (FT-ICR)

ETD = electron transfer dissociation (Linear IT)

Approccio **Top Down**

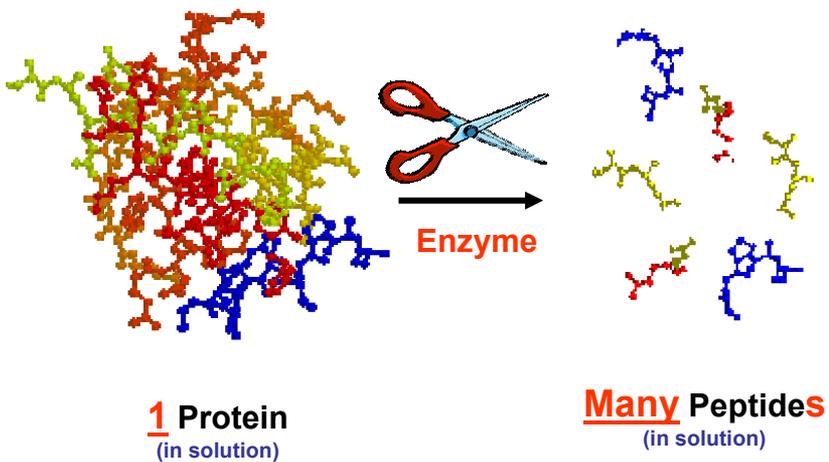
Determinazione della struttura

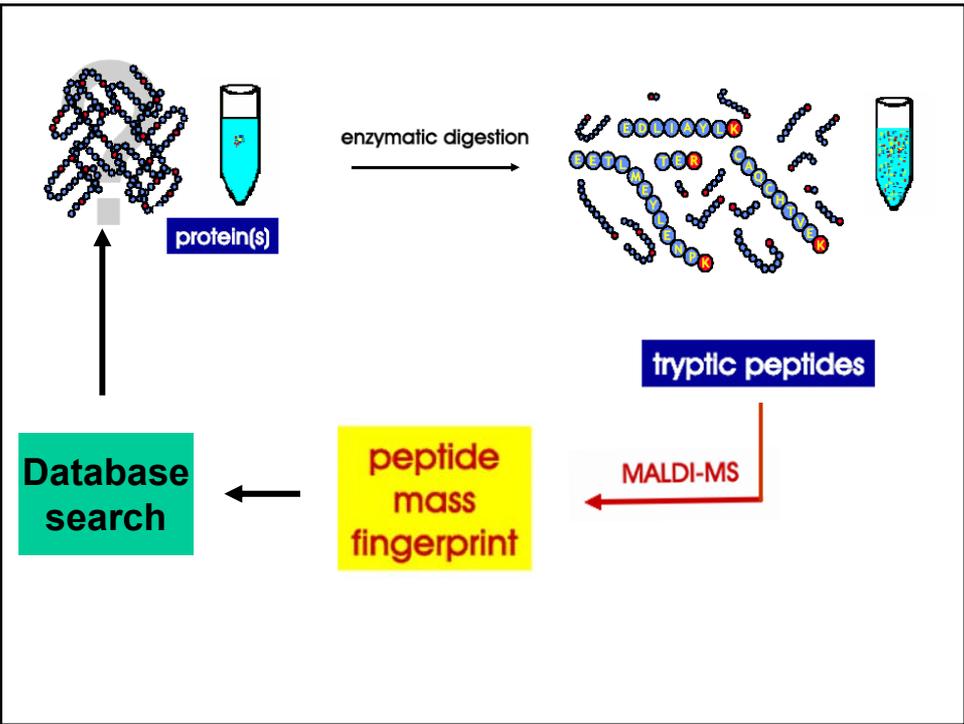
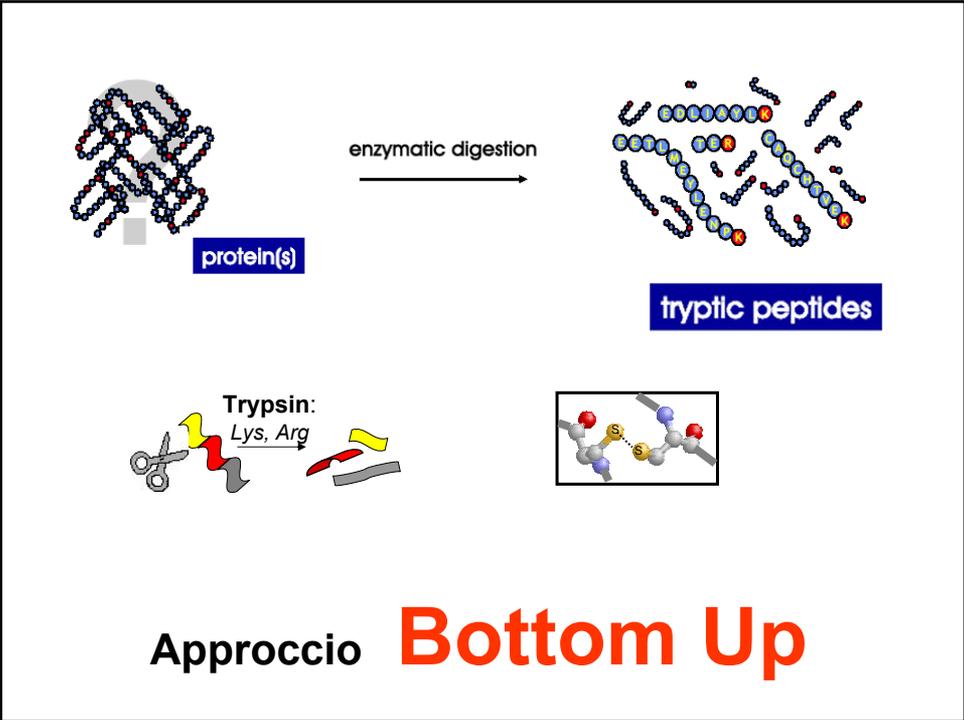
Proteina *in toto*



Determinazione della struttura

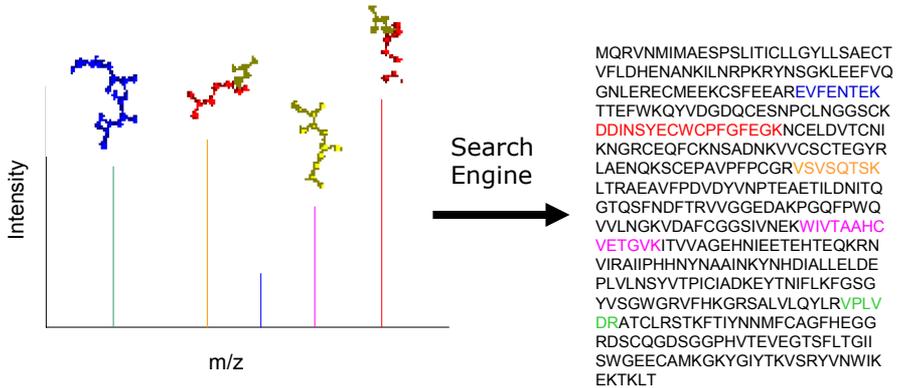
ii) Digestione enzimatica → Peptidi



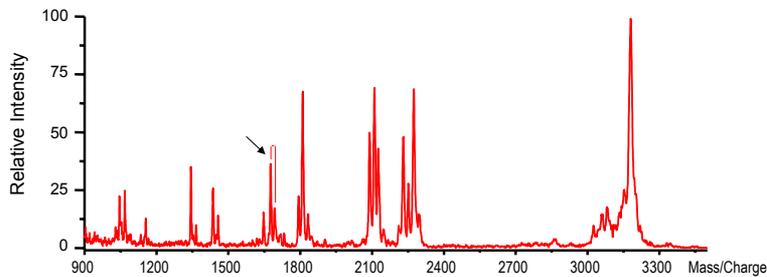


Identification of a Protein via Peptide Mass Fingerprinting

Direct analysis of the **entire peptide mixture**



Protein identification by peptide mapping



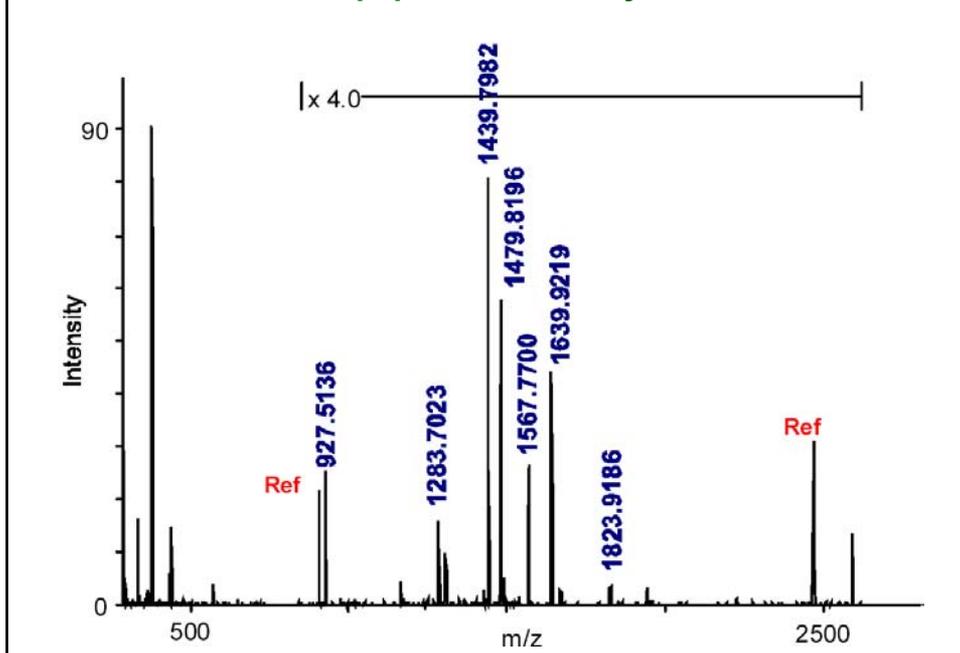
Database search (www.expasy.ch)

Aldolase A

```

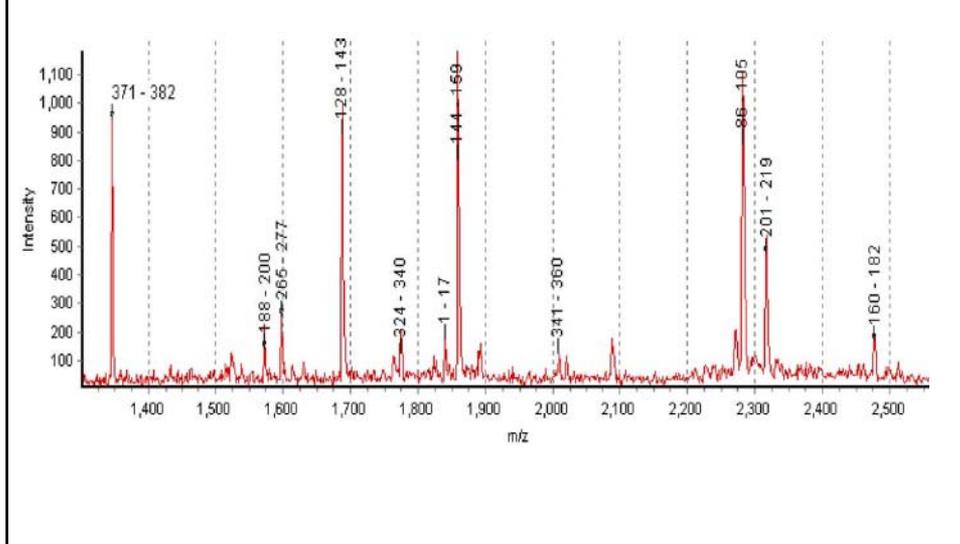
1  PYQYPALTPPEQKKELSDIAHRIVAPGKILAADESTGSIAKRLQSIGTEN
51  TEENRRFYRQLLTADDRVNPCIGGVILFHETLYQKADDGRFPQVIKSK
101  GGVVGIKVDKGVVPLAGTNGETTQQLDGLSERCAQYKKGADFAKWRVCV
151  LKIGHETPSALAIMENANVLRARYASICQQNGIVPIVEPEILPDGDHDLKR
201  CQYVTEKVLAAVYKALSDHHIYLEGTLKPNMVTPGHACTQKFSHEEIAM
251  ATVTALRRTVPPAVTGITFLSGGQSEEBEASINLNAINKCPLLKPWALFIS
301  YGRALQASALKAWGGKKENLKAQEEYVKRALANSLACQGYTPSQQAGA
351  AASESLFVSNHAY
    
```

MS of a peptide mixture by MALDI-TOF



Peptide mass fingerprinting

MALDI-TOF



Identification of a spot based on peptide mapping

Selected protein

Rank: 1 Expectation: 0.000 Coverage: 48.2 pt 5.2 Mass: 42.86 kDa Matched peptides: 11 Measured peptides: 18

Protein information: gi|808974|emb|CAA23716.1 - ovalbumin [Gallus gallus]

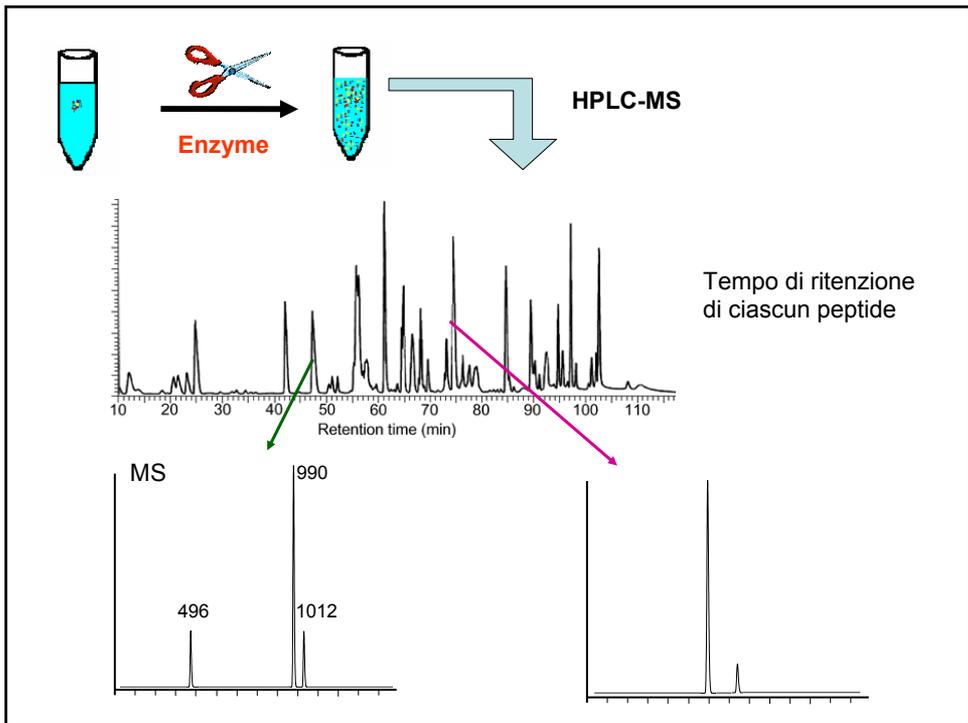
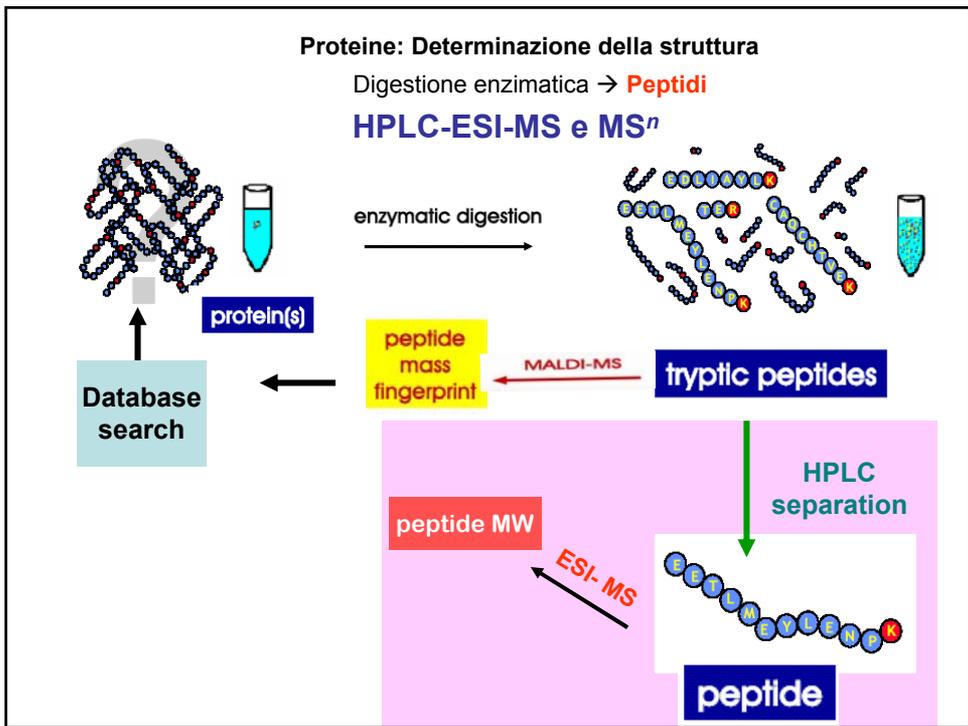
Matched peptides

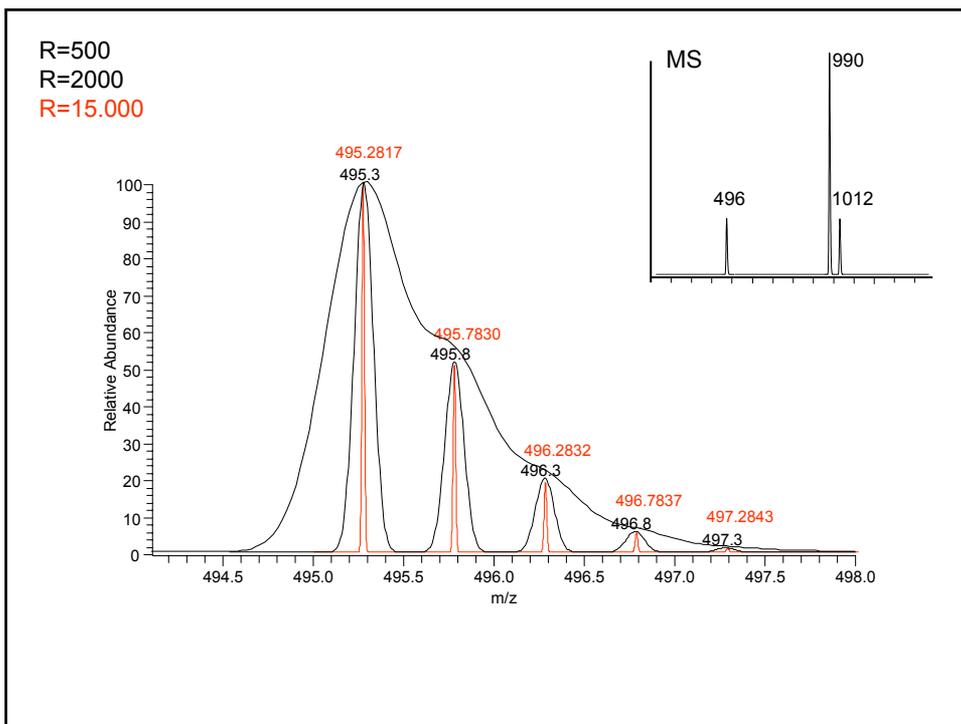
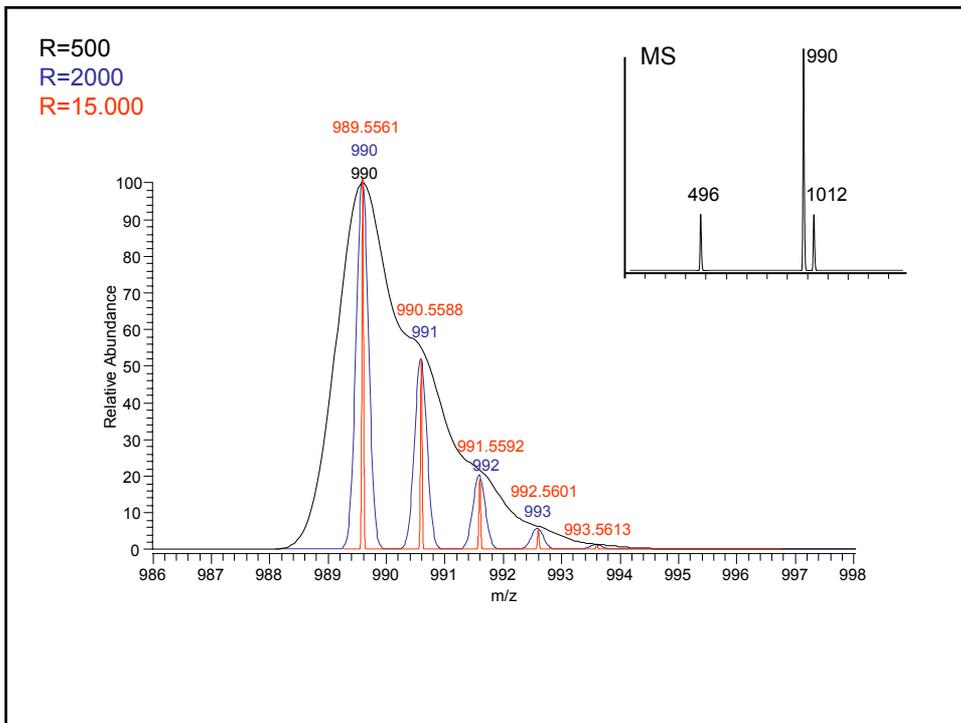
	Meas. mass	Ave/isotopic	Comp. mass	Error (ppm)	Res. start	Res. end	M. cut	Peptide sequence
	1344.706	Mono	1344.730	-18.1465	371	382	0	HIAITNAVLFFFGR
▶	1570.734	Mono	1570.709	16.1735	188	200	1	AFKDEDTQAMPFR 1)+O@M
	1596.663	Mono	1596.709	-28.5483	265	277	0	LTEWTSNNVMEER 1)+O@M
	1686.774	Mono	1686.833	-34.9025	128	143	0	GGLEPINFQTAADQAR
	1772.802	Mono	1772.892	-50.8825	324	340	0	ISQAVHAAHAEINEAGR
	1838.630	Mono	1838.804	-94.6184	1	17	0	MGSIGAASMEFCDFVK
	1857.842	Mono	1857.958	-62.5625	144	159	0	ELINSWVESQTNGIIR
	2007.771	Mono	2007.939	-83.3058	341	360	0	EVVGSAAEAGVDAASVSEEFR

Protein identifications were made by peptide mass mapping (peptide mass fingerprinting) using an unspecified search program and an unspecified database.

Se il TOF ha alta risoluzione, è possibile determinare la massa accurata e la formula bruta di ciascun peptide

Se proteine diverse, digerite con lo stesso enzima nelle stesse condizioni sperimentali, producono lo stesso peptide mass fingerprint, le due proteine sono uguali



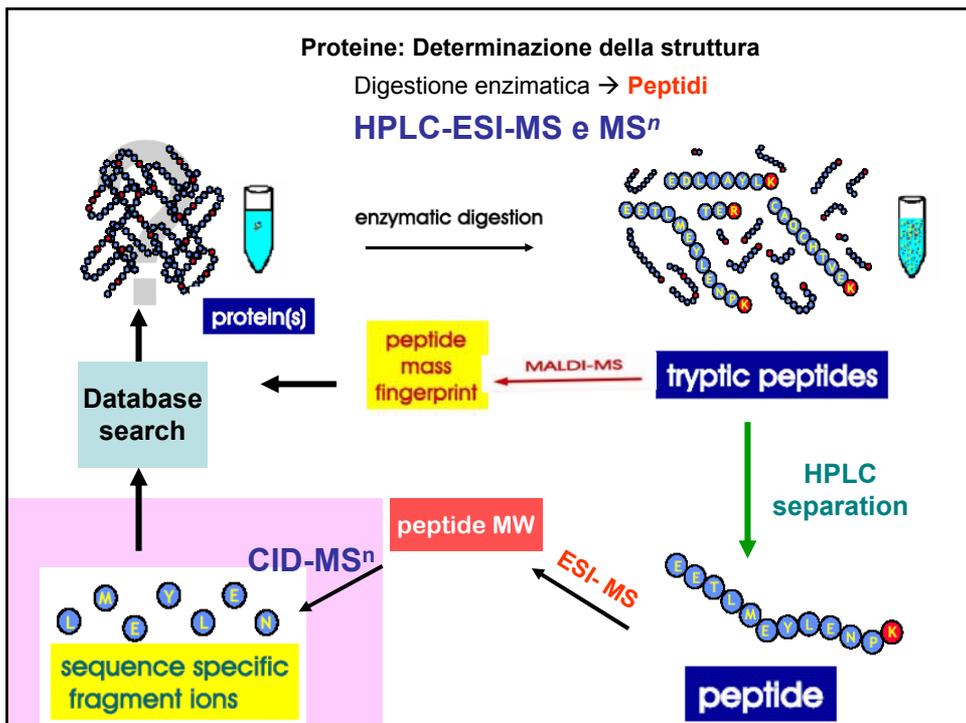


Se si dispone di un analizzatore con alta risoluzione e massa accurata

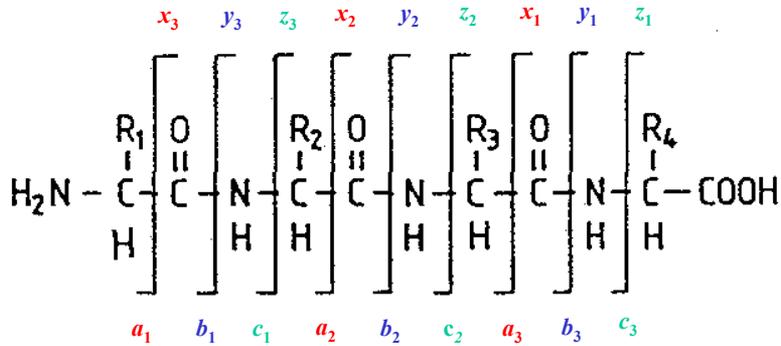
settori, ToF, Orbitrap, FT-ICR

è possibile misurare la massa accurata di ciascun peptide e quindi la formula bruta

Poiché le tecniche di ionizzazione utilizzate sono *soft* non si ha frammentazione in sorgente e quindi non si hanno informazioni strutturali, ovvero non è possibile determinare la sequenza AA



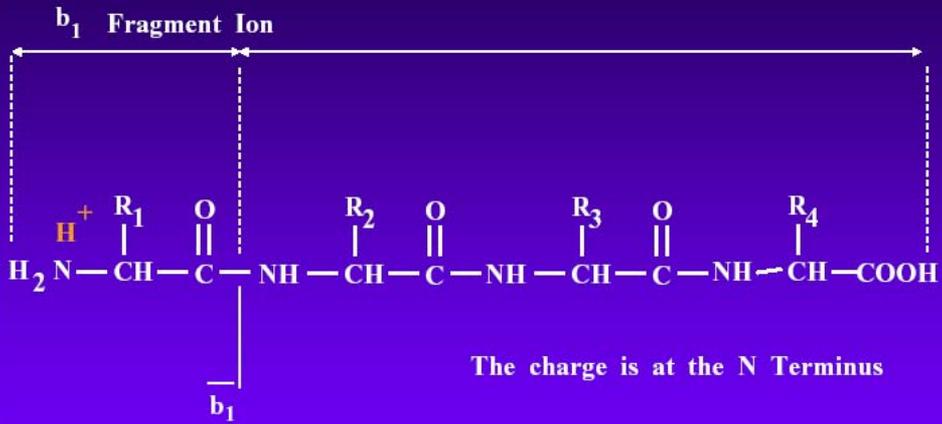
CID-MSⁿ di peptidi



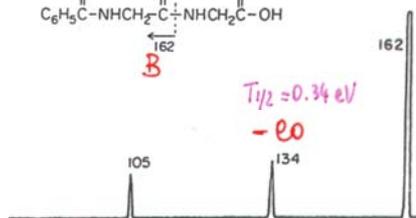
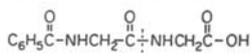
Ioni C-terminali:
 y, x, z

Ioni N-terminali:
 b, a, c

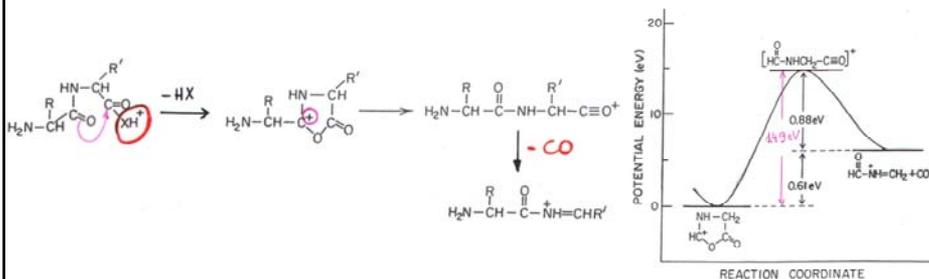
FRAGMENT ION b



Why *b* ions are very stable species?

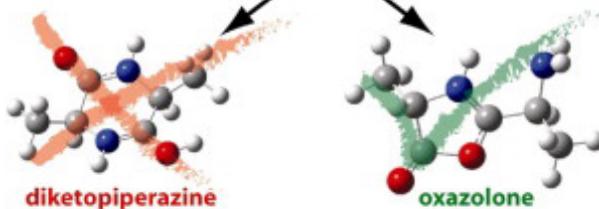


CID spectra (45 eV) of B ion derived from *N*-benzoyl- Gly-Gly-OH

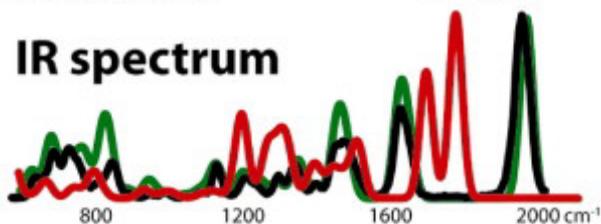


Calculated potential energy profile for fragmentation of B ion.

Spectroscopic Evidence for an Oxazolone Structure of the b_2 Fragment Ion from Protonated Tri-Alanine



IR spectrum

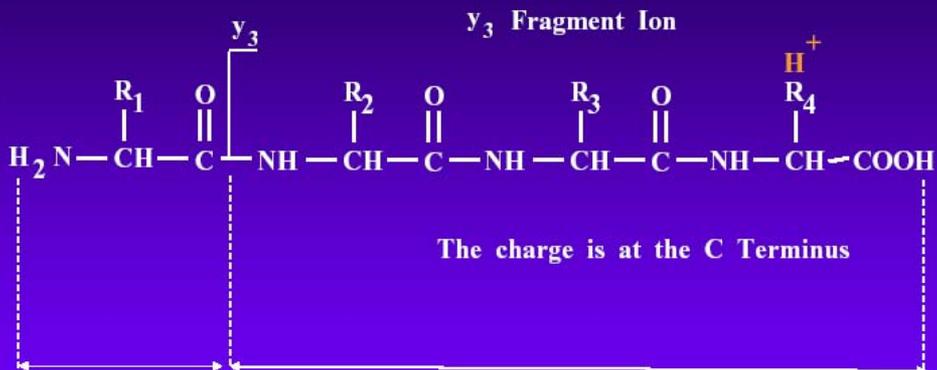


Whenever you identify a **b** ion look for an **a** ion at -28u.

This gives some assurance that your assignment is correct.

Also look for ammonia and water losses, -17 and -18u respectively.

FRAGMENT ION y



R_4 = Arginine or Lysine

Attribuire la giusta sequenza al peptide

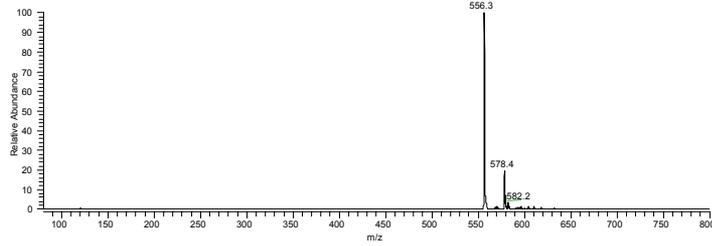
Gly Phe Tyr Gly Leu/Ile

Residui AA:

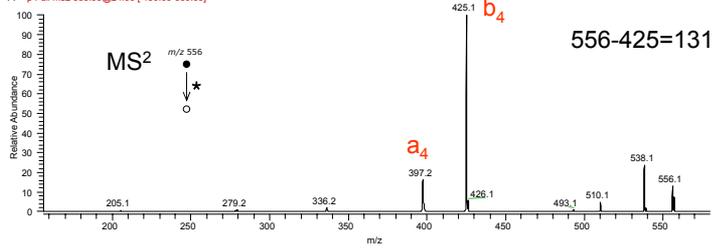
Gly 57 Phe 147 Tyr 163 Leu/Ile 113

C:\CIADSL\...peptide-ms7\peptide99

peptide99 #21-22 RT: 0.31-4.05 AV: 2 NL: 2.41E6
F: + p Full ms [80.00-800.00]



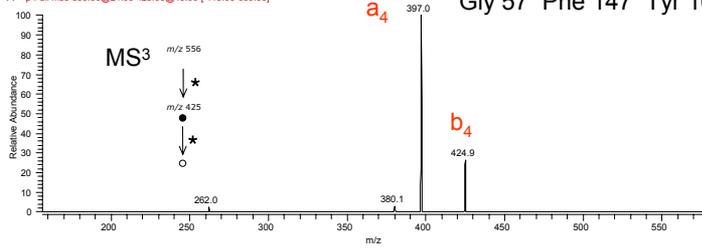
peptide99 #27-45 RT: 18.17-18.38 AV: 19 NL: 6.73E5
F: + p Full ms2 556.00@24.00 [150.00-650.00]



Residui AA:

Gly 57 Phe 147 Tyr 163 Leu/Ile 113

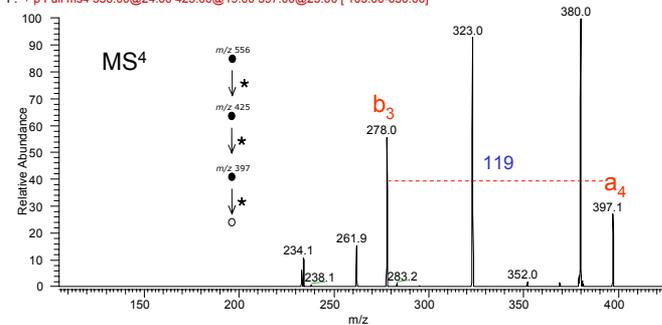
peptide99 #46-79 RT: 43.66-44.05 AV: 24 NL: 1.54E5
F: + p Full ms3 556.00@24.00 425.00@19.00 [115.00-650.00]



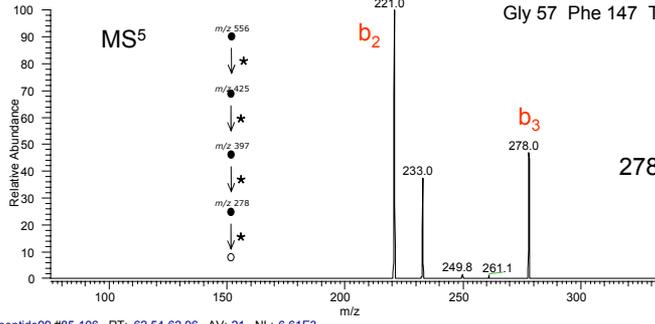
C:\CIADSL\...peptide-ms7\peptide99

21/02/2007 12.23.21

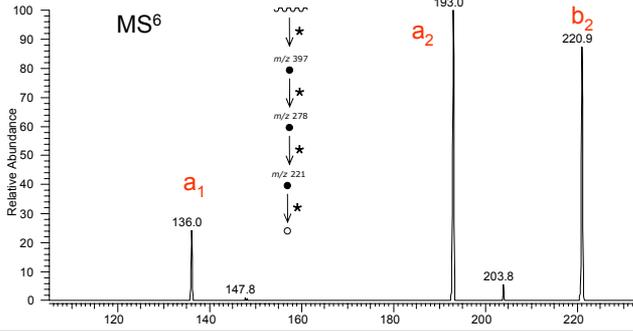
peptide99 #69-79 RT: 48.00-48.20 AV: 10 NL: 6.20E3
F: + p Full ms4 556.00@24.00 425.00@19.00 397.00@25.00 [105.00-650.00]



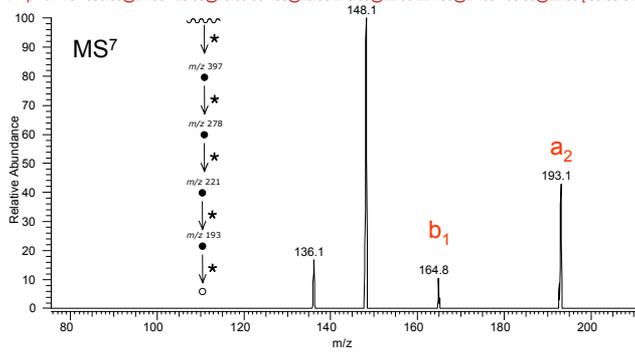
peptide99 #80-106 RT: 57.64-57.75 AV: 6 NL: 1.11E4
 F: + p Full ms5 556.00@24.00 425.00@19.00 397.00@25.00 278.00@22.00 [75.00-650.00] Residui AA:
 Gly 57 Phe 147 Tyr 163 Leu/Ile 113



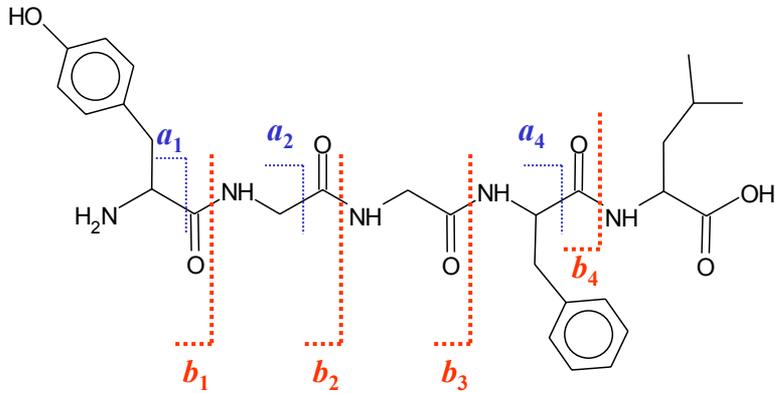
peptide99 #85-106 RT: 62.54-62.96 AV: 21 NL: 6.61E3
 F: + p Full ms6 556.00@24.00 425.00@19.00 397.00@25.00 278.00@22.00 221.00@21.00 [60.00-650.00]



peptide99 #106-144 RT: 72.78-73.65 AV: 38 NL: 7.87E2
 F: + p Full ms7 556.00@24.00 425.00@19.00 397.00@25.00 278.00@22.00 221.00@21.00 193.00@22.00 [50.00-6 ...



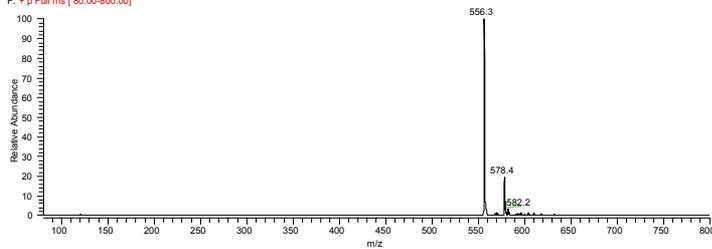
Tyr-Gly-Gly-Phe-Leu



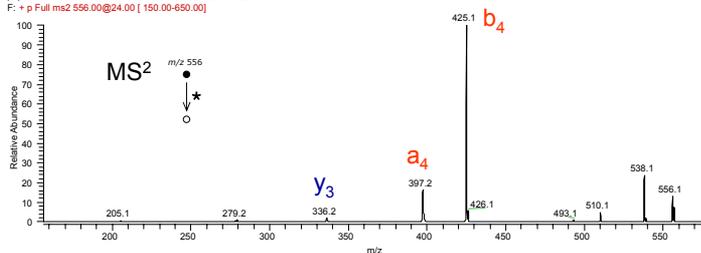
Attribuire la sequenza al peptide

C:\CIADSL...peptide-ms7\peptide99

peptide99 #21-22 RT: 0.31-4.05 AV: 2 NL: 2.41E6
F: + p Full ms [80.00-800.00]



peptide99 #27-45 RT: 18.17-18.38 AV: 19 NL: 6.73E5
F: + p Full ms2 556.00@24.00 [150.00-650.00]



Cyclic peptides:

Naturally occurring in bacteria, fungi and plants

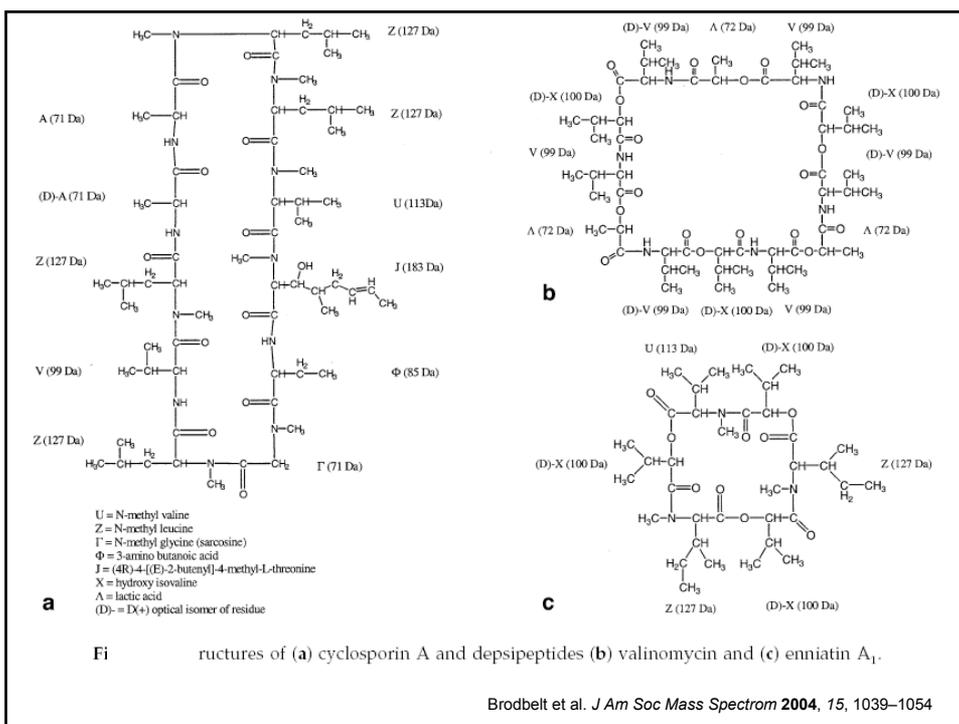
Synthetic compounds

They can have a lot of biological and pharmaceutical properties:

enzyme inhibitors, antifungal and antibacterial agents, immunosuppressant, antibiotics and anticancer drugs

Better bioavailability and higher resistance to proteolytic degradation than their linear analogues

Highly selective metal ion chelators:
carriers of metal ions inside the cells. Use in radiotherapy

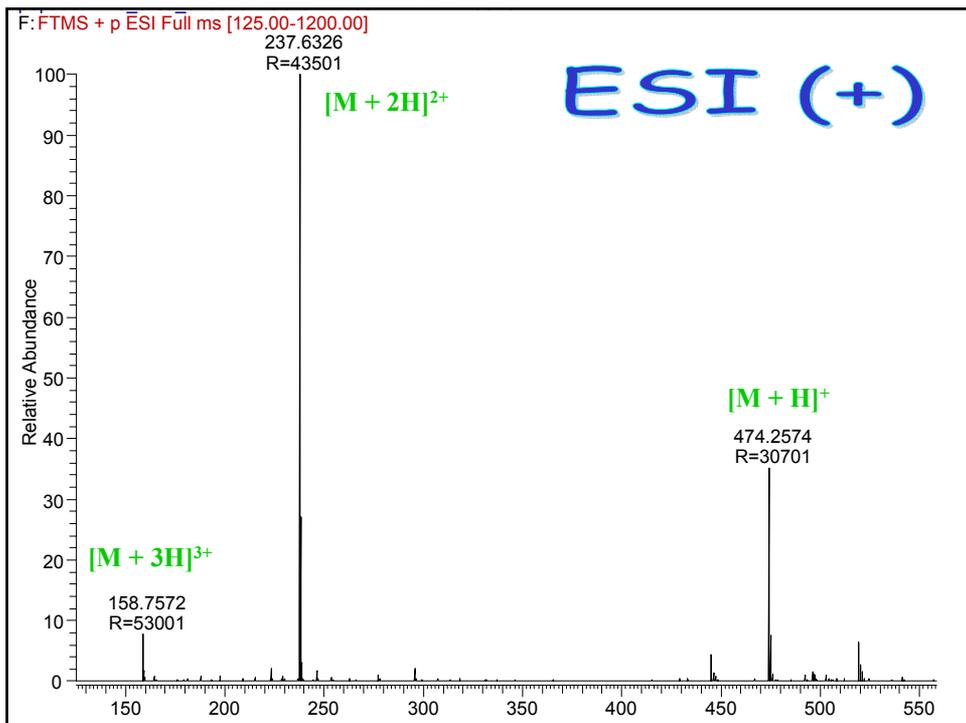


Characterization and sequencing of cyclic peptides represent an important challenge for mass spectrometry.

No definite terminations, as it occurs in their linear analogues.

The elimination of an aminoacidic residue must involve the fission of two bonds of the cycle.

The ring-opening may occur at several backbone positions, meaning that all the fragment ions are not referenced to a single "terminal" position.



Resolving power S :

Orbitrap:

at constant detection time:
resolving power inversely
proportional to $\sqrt{m/z}$

Thus if

$R=100,000$ at m/z 100,

at m/z 1000 it will be $R=100,000(100/1000)^{1/2}=31,646$

Elemental composition search on mass 474.26

$m/z = 469.26-479.26$

m/z	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
474.2573	474.2572	0.18	5.0	C ₂₂ H ₃₈ O ₉ N ₂
474.2572	474.2572	0.20	10.5	C ₂₁ H ₃₂ O ₄ N ₉
474.2585	474.2585	-2.64	10.0	C ₂₃ H ₃₄ O ₅ N ₆
474.2558	474.2558	5.02	5.5	C ₂₀ H ₃₆ O ₈ N ₅
474.2599	474.2599	-5.46	15.0	C ₂₄ H ₃₀ O ₁ N ₁₀
474.2599	474.2599	-5.47	9.5	C ₂₅ H ₃₆ O ₆ N ₃

RDB=double bond/ring equivalents

$$D = 1 + \frac{\sum_i^{i_{max}} N_i (V_i - 2)}{2}$$

For ions with odd charge:

Odd-electron ions = RDB integer

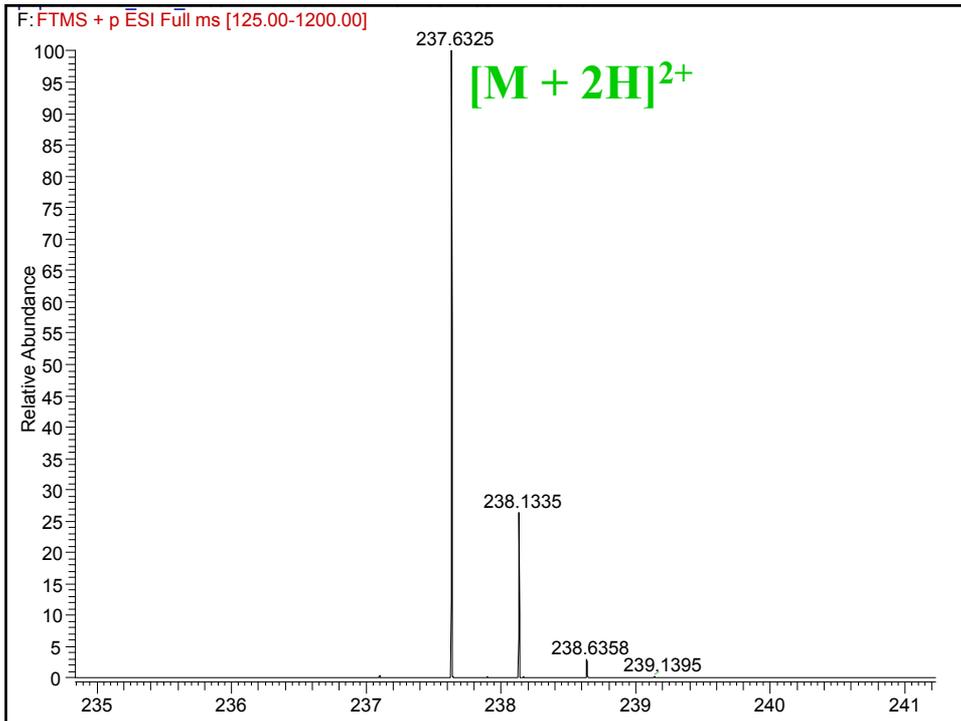
Even-electron ions = RDB x.5

Elemental composition search on mass 475.26

$m/z = 470.26-480.26$

m/z	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
475.2605	475.2605	-0.03	10.5	C ₂₀ ¹³ C ₁ H ₃₂ O ₄ N ₉
475.2605	475.2605	-0.04	5.0	C ₂₁ ¹³ C ₁ H ₃₈ O ₉ N ₂
475.2605	475.2605	0.10	18.5	C ₃₀ H ₃₁ N ₆
475.2610	475.2610	-0.98	0.5	C ₁₇ H ₃₉ O ₁₁ N ₄
475.2596	475.2596	1.85	1.0	C ₁₅ H ₃₇ O ₁₀ N ₇
475.2592	475.2592	2.79	5.5	C ₁₉ ¹³ C ₁ H ₃₆ O ₈ N ₅

¹² C	0÷30
¹³ C	0÷1
¹⁴ N	0÷10
¹⁶ O	0÷15
¹ H	0÷60



Elemental composition search on mass 237.63

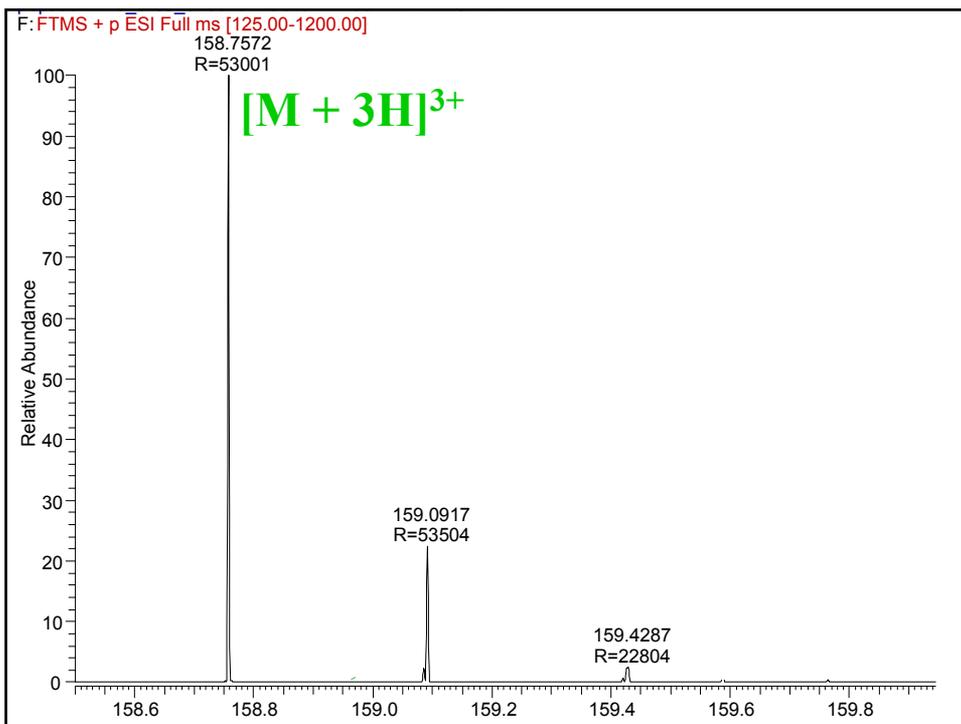
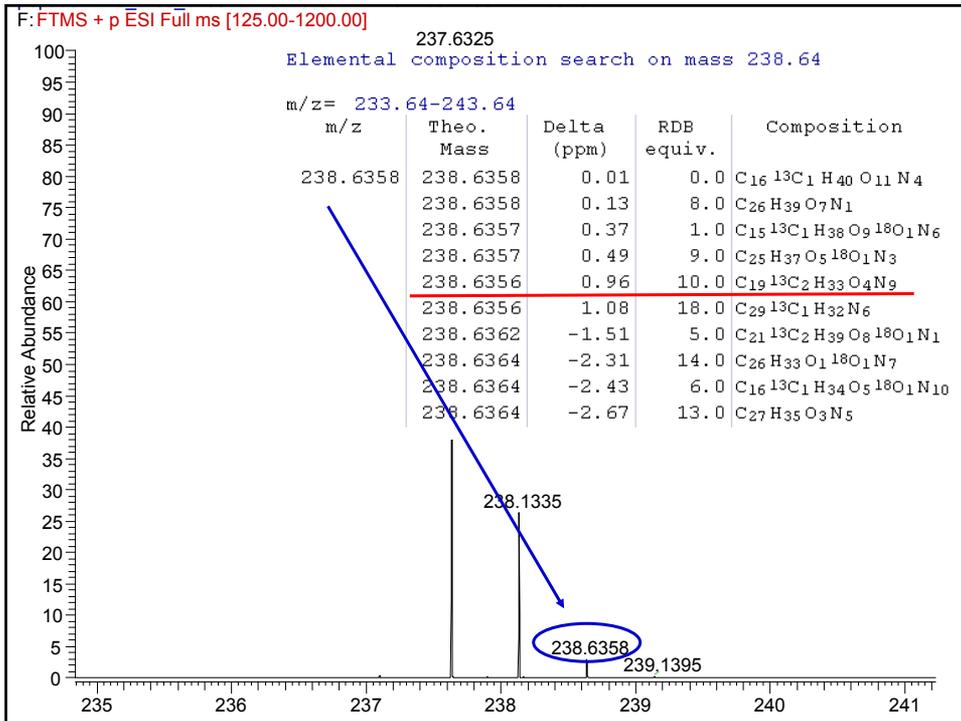
m/z= 232.63-242.63

m/z	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
237.6325	237.6327	-0.78	14.0	C ₂₇ ¹³ C ₁ H ₃₄ O ₃ N ₄
	237.6322	1.11	10.0	C ₂₁ H ₃₃ O ₄ N ₉
	237.6320	2.04	9.0	C ₂₆ ¹³ C ₁ H ₃₈ O ₇
	237.6316	3.92	5.0	C ₂₀ H ₃₇ O ₈ N ₅
	237.6336	-4.54	9.0	C ₂₅ H ₃₇ O ₆ N ₃
	237.6336	-4.67	1.0	C ₁₅ ¹³ C ₁ H ₃₈ O ₁₀ N ₆

Elemental composition search on mass 238.13

m/z= 233.13-243.13

m/z	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
238.1335	238.1332	1.11	5.0	C ₁₉ ¹³ C ₁ H ₃₇ O ₈ N ₅
	238.1332	1.24	13.0	C ₂₉ H ₃₆ O ₄ N ₂
	238.1339	-1.57	18.0	C ₃₀ H ₃₂ N ₆
	238.1339	-1.70	10.0	C ₂₀ ¹³ C ₁ H ₃₃ O ₄ N ₉
	238.1341	-2.65	0.0	C ₁₇ H ₄₀ O ₁₁ N ₄
	238.1328	2.99	1.0	C ₁₃ H ₃₆ O ₉ N ₁₀



Elemental composition search on mass 158.76

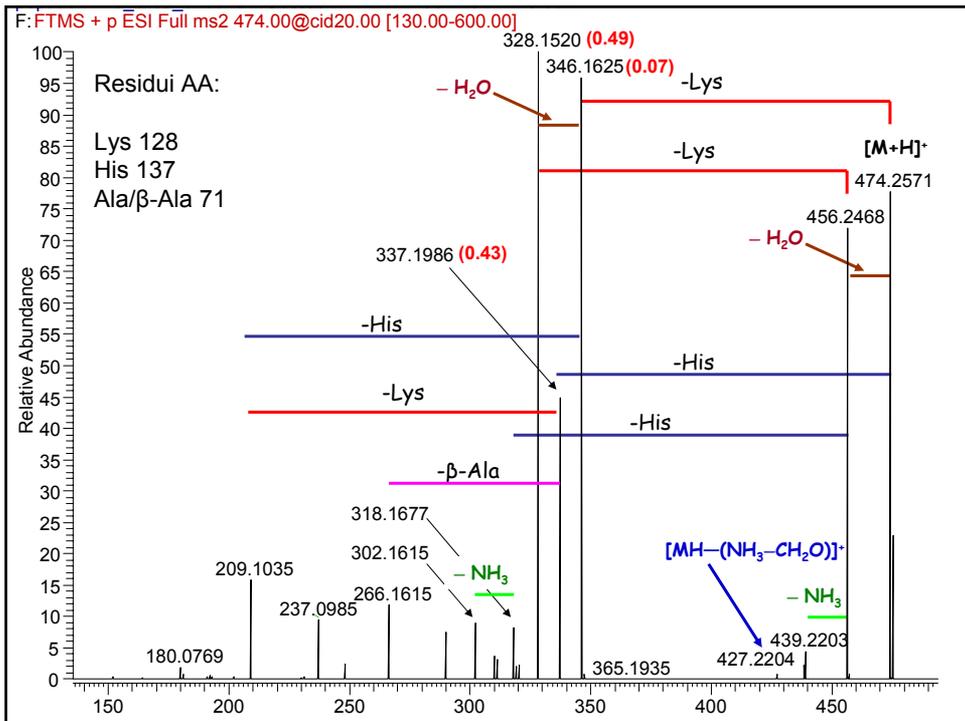
m/z= 153.76-163.76

m/z	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
158.7572	158.7572	-0.34	9.5	C ₂₁ H ₃₄ O ₄ N ₉
	158.7572	-0.35	4.0	C ₂₂ H ₄₀ O ₉ N ₂
	158.7571	0.59	8.5	C ₂₆ ¹³ C ₁ H ₃₉ O ₇
	158.7571	0.60	14.0	C ₂₅ ¹³ C ₁ H ₃₃ O ₂ N ₇
	158.7575	-2.22	13.5	C ₂₇ ¹³ C ₁ H ₃₅ O ₃ N ₄
	158.7568	2.47	4.5	C ₂₀ H ₃₈ O ₈ N ₅

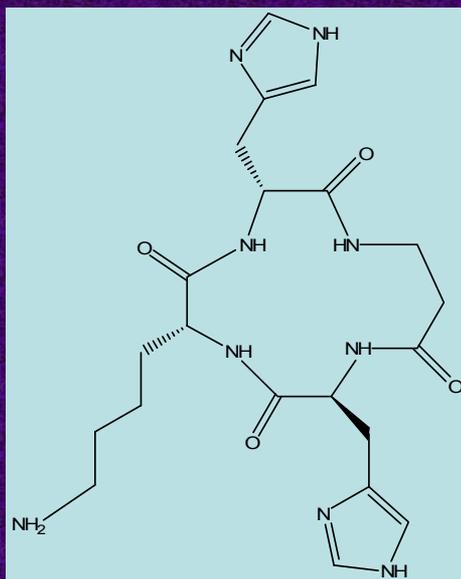
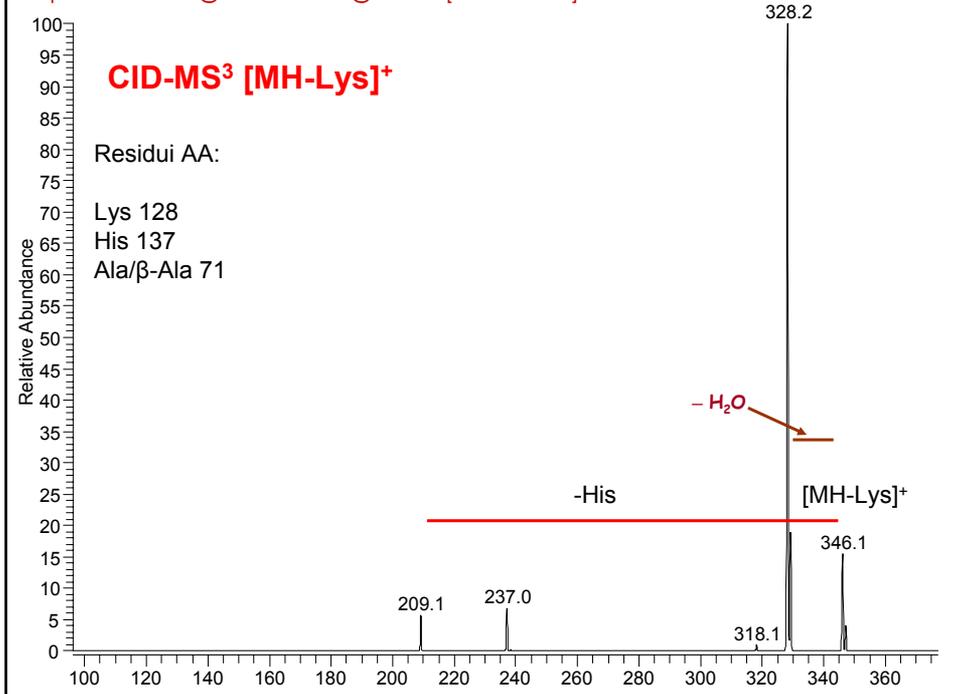
Elemental composition search on mass 159.09

m/z= 154.09-164.09

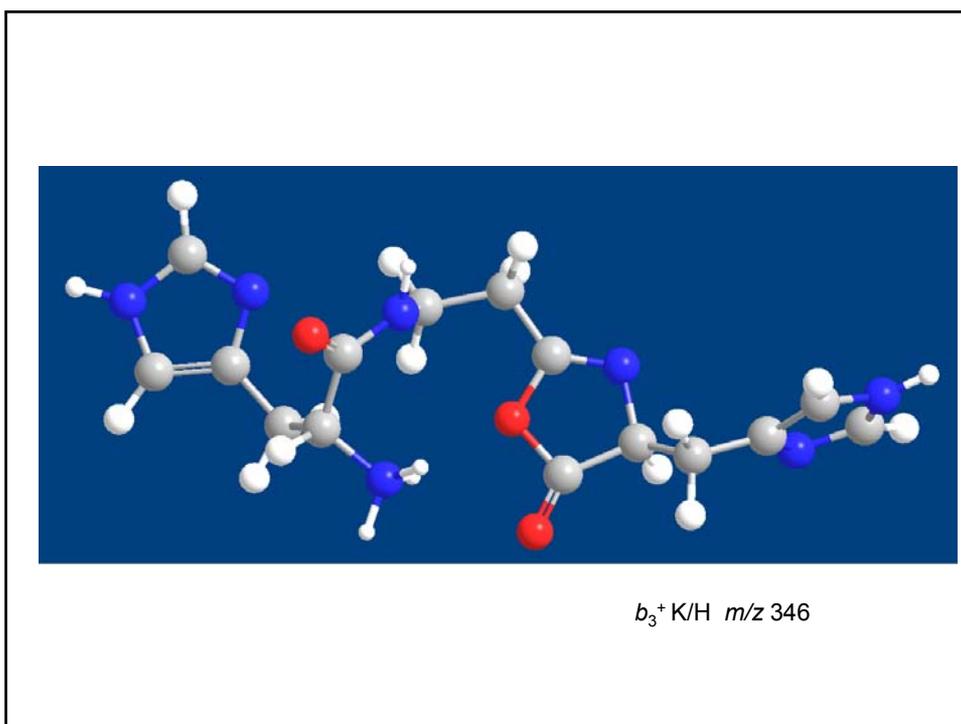
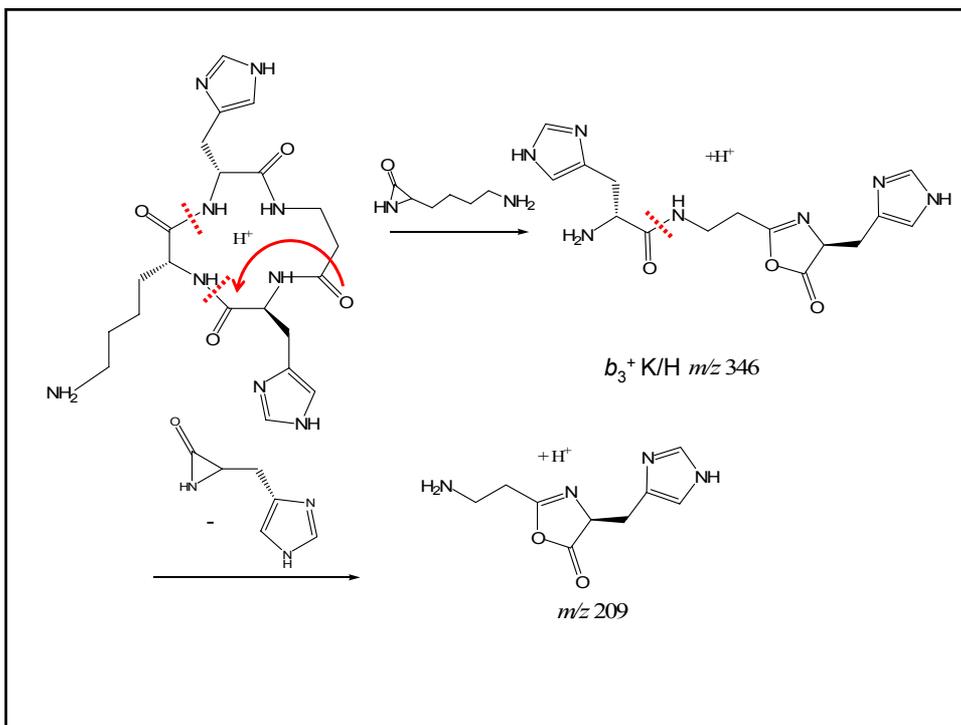
m/z	Theo. Mass	Delta (ppm)	RDB equiv.	Composition
159.0916	159.0917	-0.28	17.5	C ₃₀ H ₃₃ N ₆
	159.0917	-0.41	9.5	C ₂₀ ¹³ C ₁ H ₃₄ O ₄ N ₉
	159.0917	-0.42	4.0	C ₂₁ ¹³ C ₁ H ₄₀ O ₉ N ₂
	159.0918	-1.36	-0.5	C ₁₇ H ₄₁ O ₁₁ N ₄
	159.0914	1.45	0.0	C ₁₅ H ₃₉ O ₁₀ N ₇
	159.0912	2.39	4.5	C ₁₉ ¹³ C ₁ H ₃₈ O ₈ N ₅

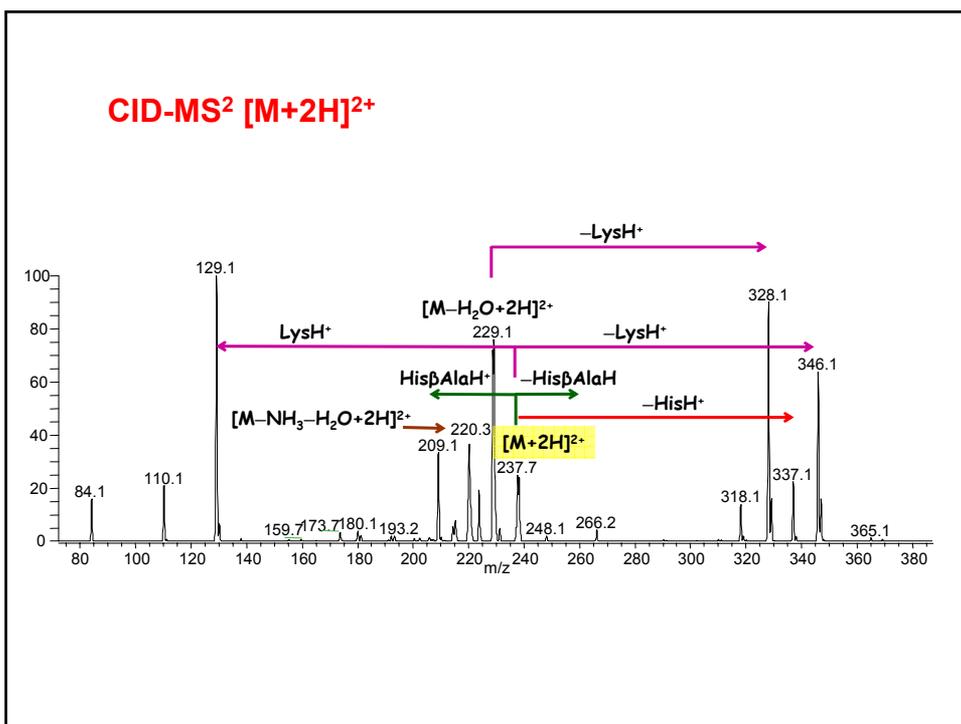
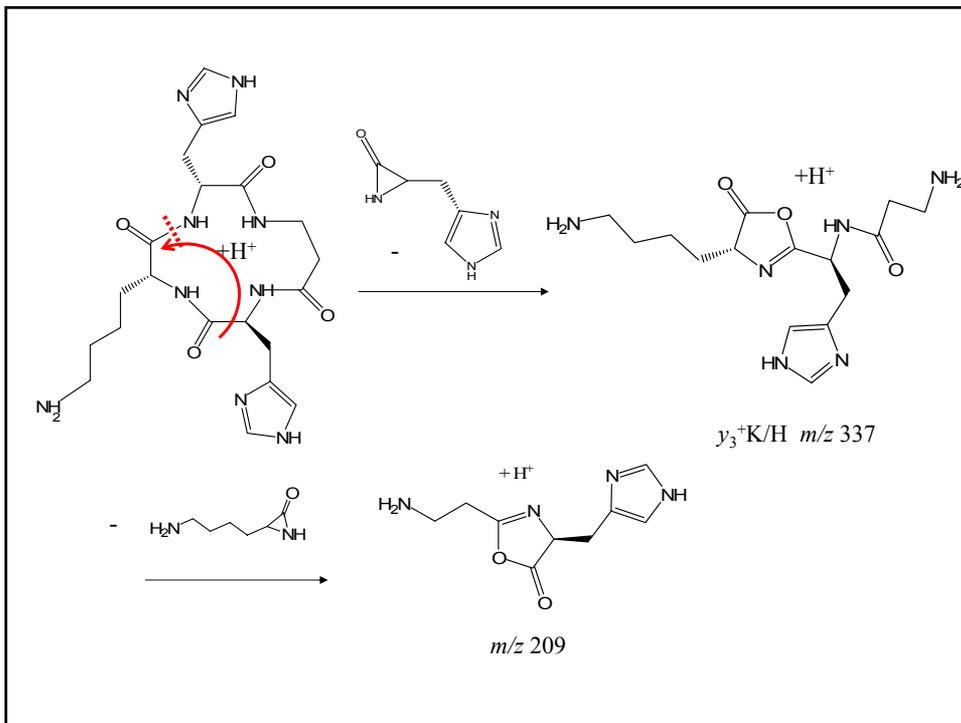


F: + p Full ms3 474.00@cid28.00 346.00@cid22.00 [95.00-1000.00]

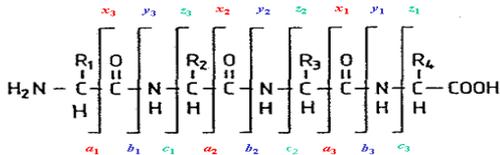
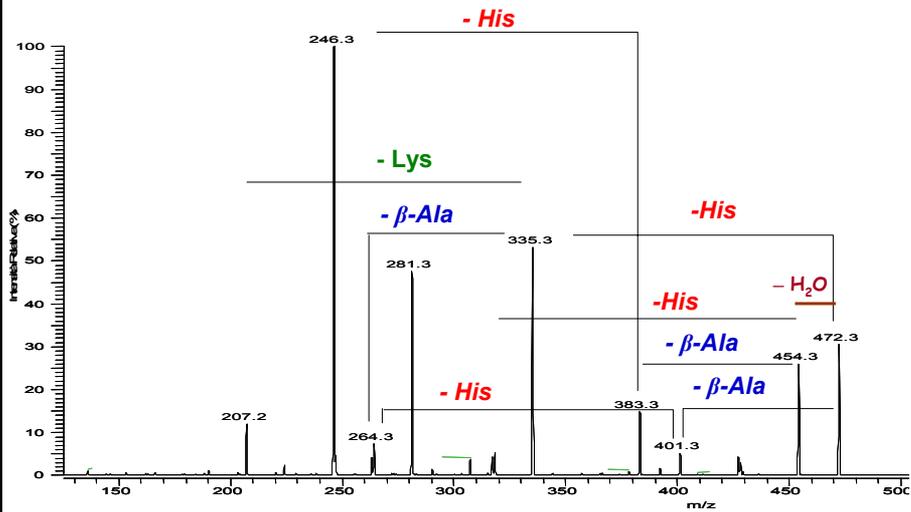


c-(Lys-D/L-His-β-Ala-His)





CID-MS² [M-H]⁻



If you are working with a **tryptic peptide**, look for the arginine or lysine **y₁ ion** at the low end of the spectrum:

$$[M+H]^+ - 156 - 18 = \text{penultimate b ion} \quad \text{Arg is the carboxy AA}$$

$$[M+H]^+ - 128 - 18 = \text{penultimate b ion} \quad \text{Lys is the carboxy AA}$$

This may give you a clue to the **C-terminal residue** of your tryptic peptide.

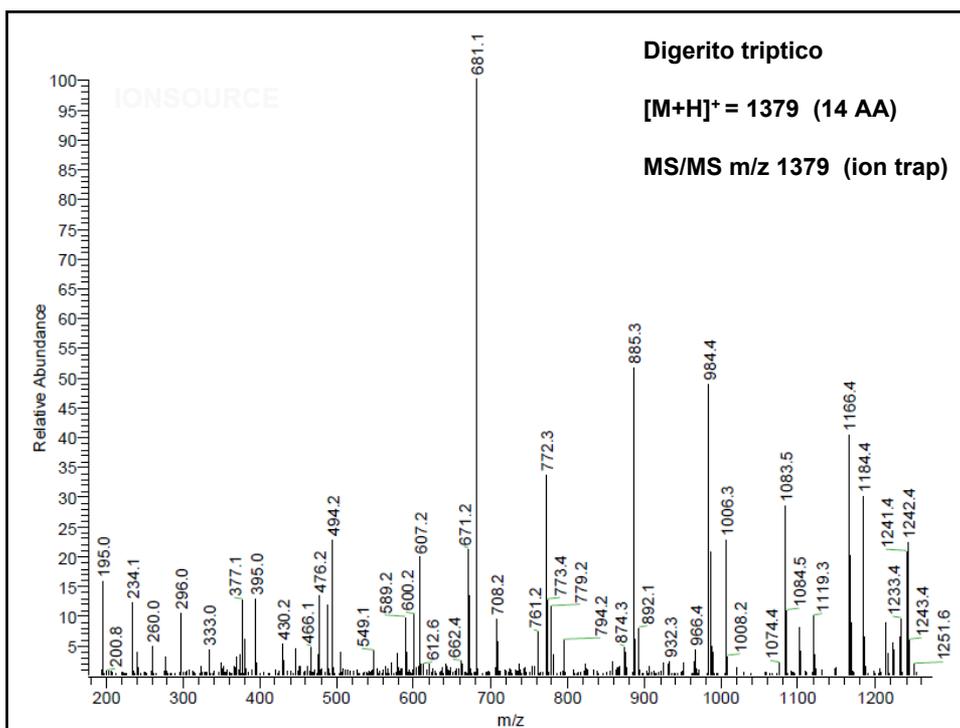
In general: **protonated peptide - AA residue mass - 18 = penultimate b ion**

or you could use the standard formula for calculating the corresponding **b ion** once the **y ion** is known.

$$[M+H]^+ - y_1 + 1 = \text{penultimate b ion}$$

If you are using an ion trap you may not be able to observe the low end of the spectrum, and in this case, you will need to do both of these calculations

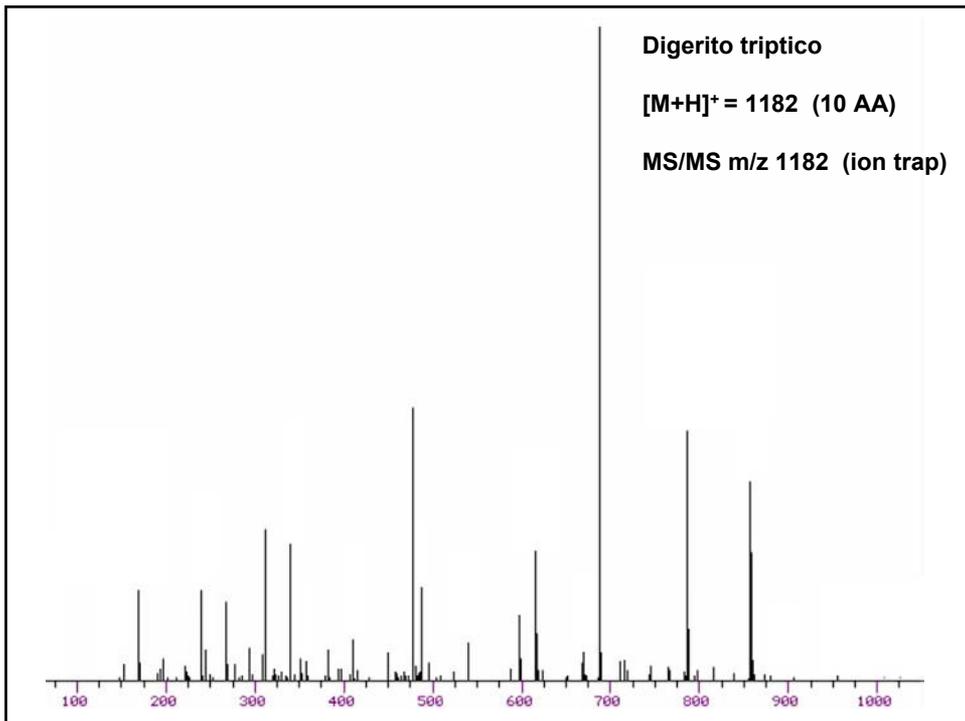
$[M+H]^+$ - observed **b** ion + 1 = corresponding **y** ion.



Sequence: HGTVVLTALGGILK, pI: 8.75725

Fragment Ion Table, monoisotopic masses

Seq	#	A	B	C	X	Y	Z	# (+1)
H	1	110.07187	138.06678	155.09333	-	1378.84225	1361.81570	14
G	2	167.09333	195.08825	212.11479	1267.76260	1241.78334	1224.75679	13
T	3	268.14101	296.13592	313.16247	1210.74114	1184.76188	1167.73533	12
V	4	367.20942	395.20434	412.23089	1109.69346	1083.71420	1066.68765	11
V	5	466.27784	494.27275	511.29930	1010.62505	984.64578	967.61924	10
L	6	579.36190	607.35681	624.38336	911.55664	885.57737	868.55082	9
T	7	680.40958	708.40449	725.43104	798.47257	772.49331	755.46676	8
A	8	751.44669	779.44161	796.46815	697.42489	671.44563	654.41908	7
L	9	864.53076	892.52567	909.55222	626.38778	600.40852	583.38197	6
G	10	921.55222	949.54713	966.57368	513.30372	487.32445	470.29790	5
G	11	978.57368	1006.56860	1023.59515	456.28225	430.30299	413.27644	4
I	12	1091.65775	1119.65266	1136.67921	399.26079	373.28152	356.25498	3
L	13	1204.74181	1232.73672	1249.76327	286.17673	260.19746	243.17091	2
K	14	1332.83677	1360.83169	-	173.09266	147.11340	130.08685	1

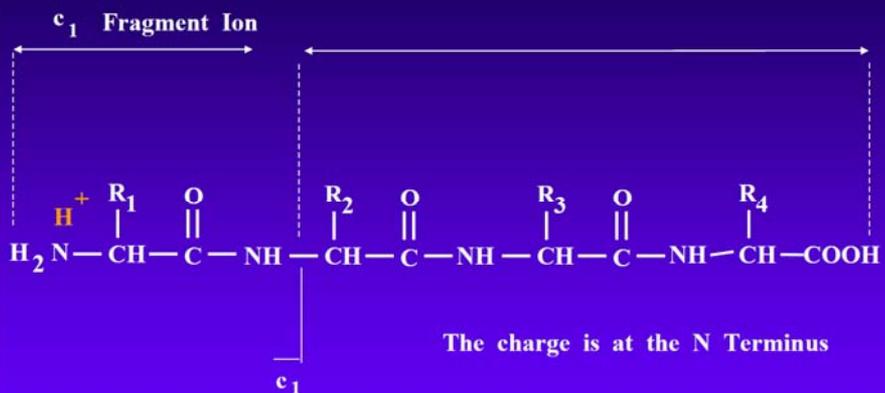


ETD, ECD

High Collision CID

Ions **c**, **z**

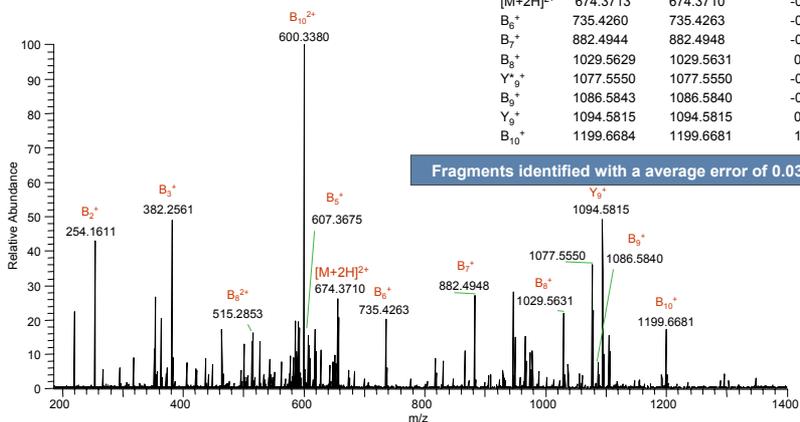
FRAGMENT ION **c**



Single Scan Ion Trap CID with ICR Cell Detection

R P K P Q Q F F G L M

Ion	Theoretical	Measured	ppm
B ₂ ⁺	254.1612	254.1611	0.19
B ₃ ⁺	382.2561	382.2561	-0.04
B ₈ ²⁺	515.2851	515.2853	0.64
B ₁₀ ²⁺	600.3378	600.3380	-0.06
B ₅ ⁺	607.3675	607.3675	0.00
[M+2H] ²⁺	674.3713	674.3710	-0.44
B ₆ ⁺	735.4260	735.4263	-0.31
B ₇ ⁺	882.4944	882.4948	-0.00
B ₈ ⁺	1029.5629	1029.5631	0.52
Y ₉ ⁺	1077.5550	1077.5550	-0.18
B ₉ ⁺	1086.5843	1086.5840	-0.28
Y ₉ ⁺	1094.5815	1094.5815	0.32
B ₁₀ ⁺	1199.6684	1199.6681	1.17

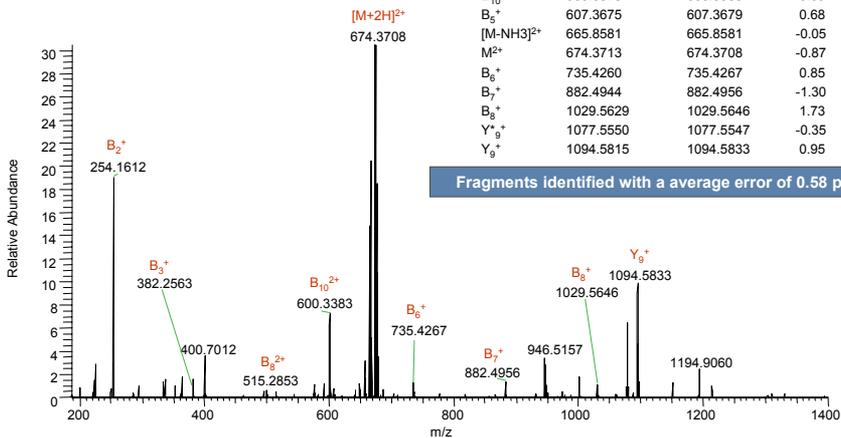


Fragments identified with an average error of 0.03 ppm

Single Scan IRMPD with ICR Cell Detection

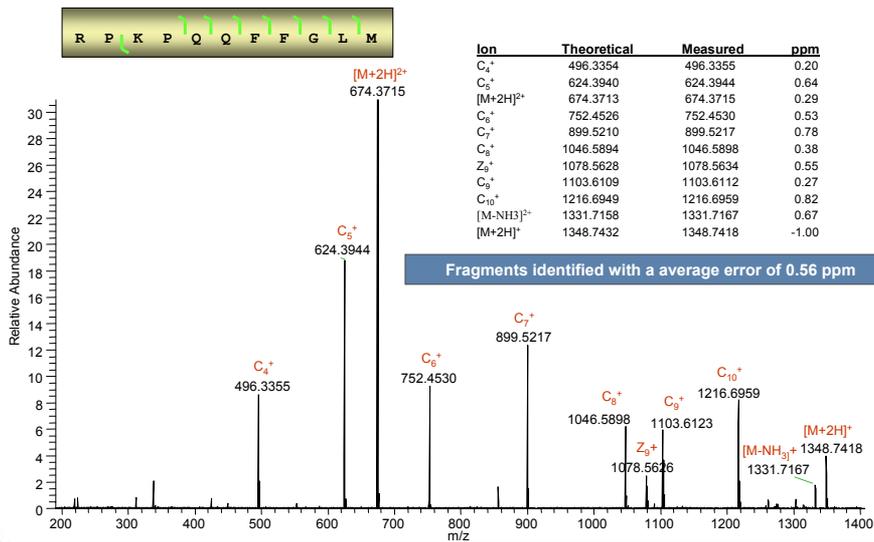
R P K P Q Q F F G L M

Ion	Theoretical	Measured	ppm
B ₂ ⁺	254.1612	254.1612	0.04
B ₃ ⁺	382.2561	382.2563	0.38
B ₈ ²⁺	515.2851	515.2853	0.51
B ₁₀ ²⁺	600.3378	600.3383	0.69
B ₅ ⁺	607.3675	607.3679	0.68
[M-NH3] ²⁺	665.8581	665.8581	-0.05
M ²⁺	674.3713	674.3708	-0.87
B ₆ ⁺	735.4260	735.4267	0.85
B ₇ ⁺	882.4944	882.4956	-1.30
B ₈ ⁺	1029.5629	1029.5646	1.73
Y ₉ ⁺	1077.5550	1077.5547	-0.35
Y ₉ ⁺	1094.5815	1094.5833	0.95



Fragments identified with an average error of 0.58 ppm

Single Scan ECD with ICR Cell Detection



Top down

vs

approaches

Bottom up

for studying large molecules

Determinazione della struttura: MSⁿ

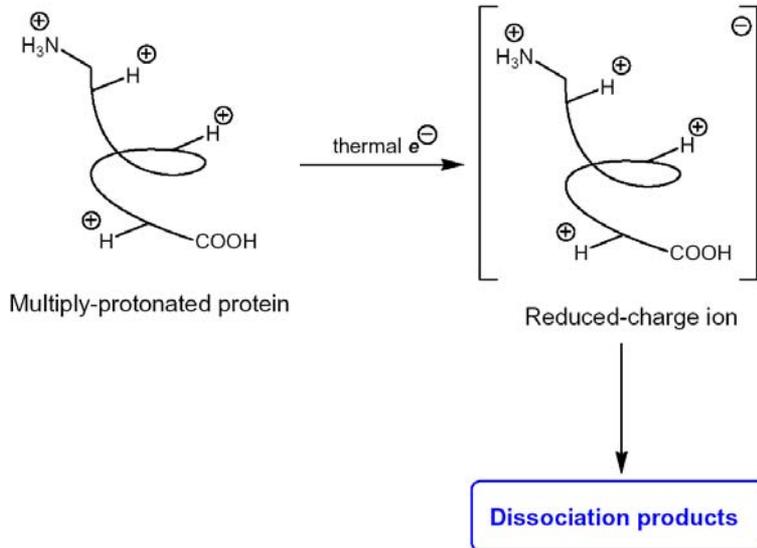
i) Proteina *in toto*:

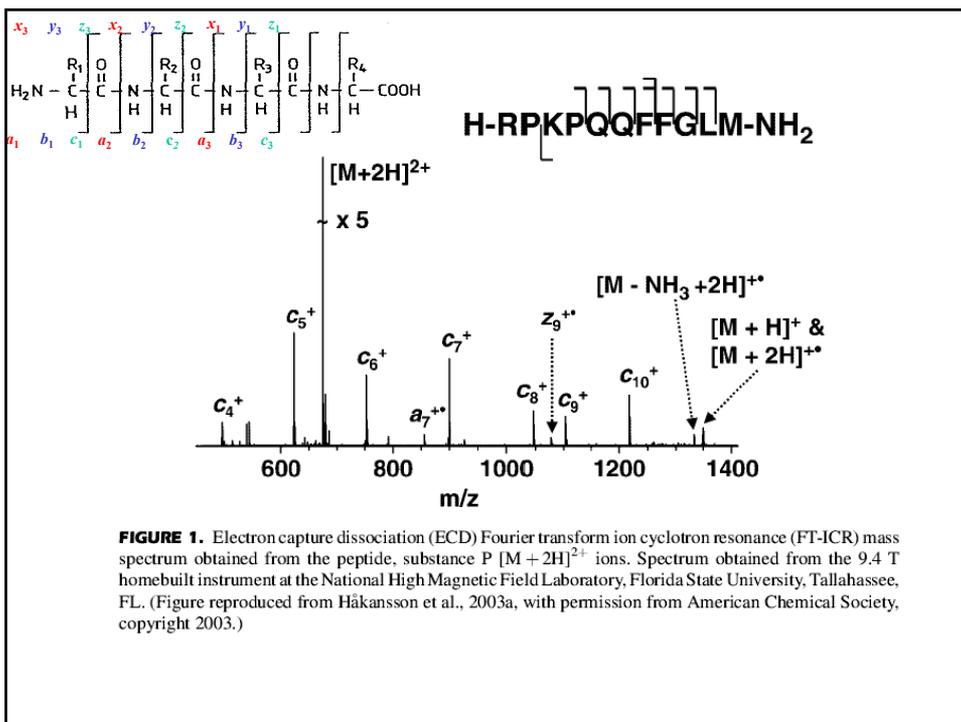
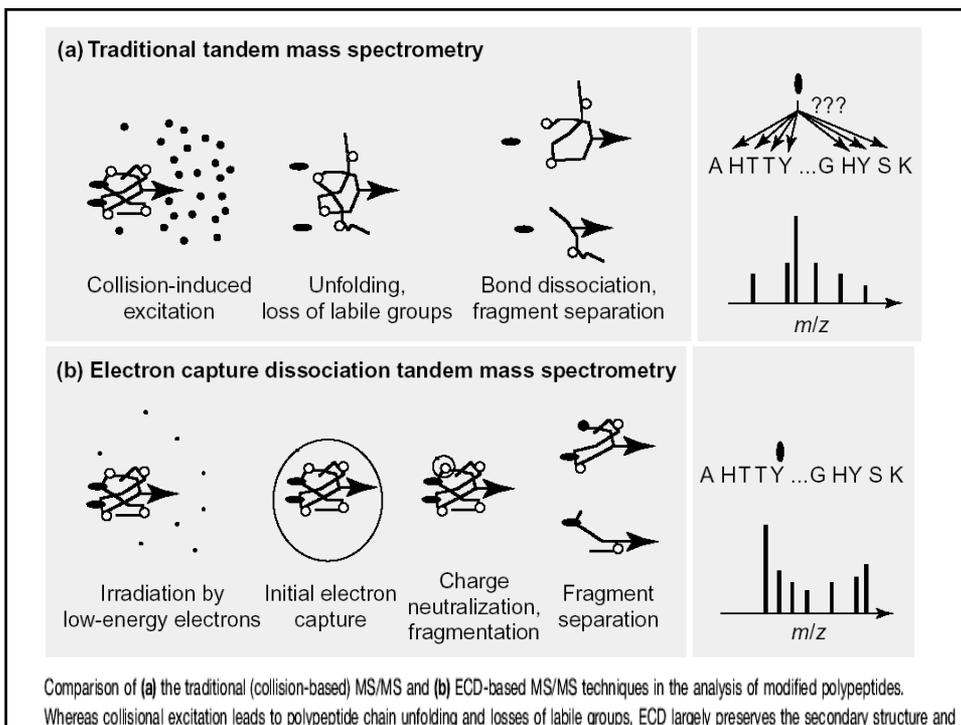
ECD = electron capture dissociation (FT-ICR)

ETD = electron transfer dissociation (Linear IT)

Approccio **Top Down**

Electron Capture Dissociation



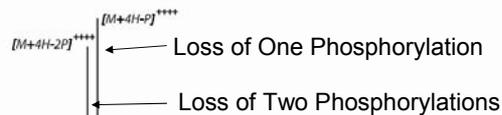


ETD

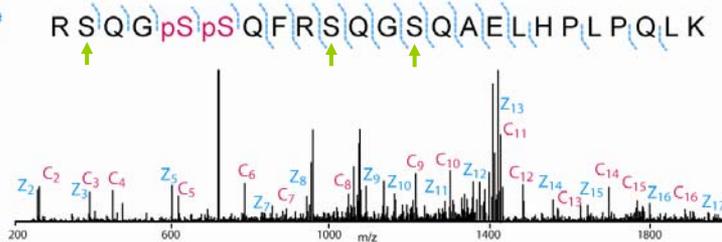
Electron Transfer Dissociation

CID vs ETD for Phosphorylated Peptides

CAD
Scan#
9525



ETD
Scan#
9526



alpha isoform of regulatory subunit B56, serine/threonine protein phosphatase 2A