

Corso di  
**Proprietà di Biopolimeri**

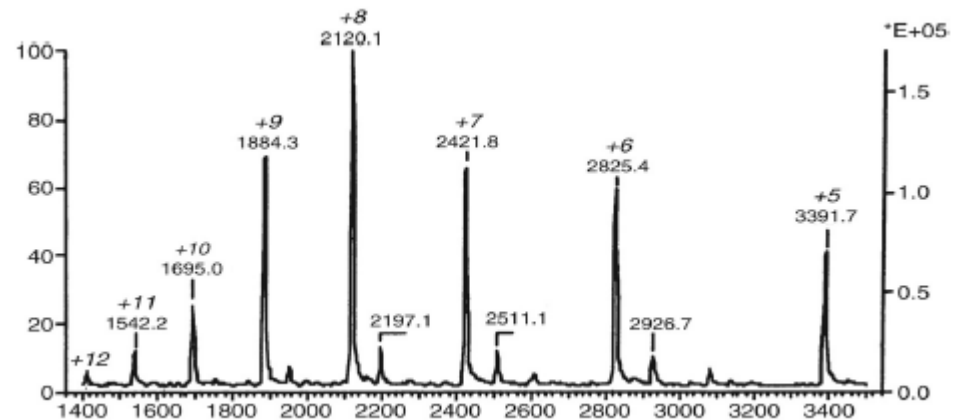
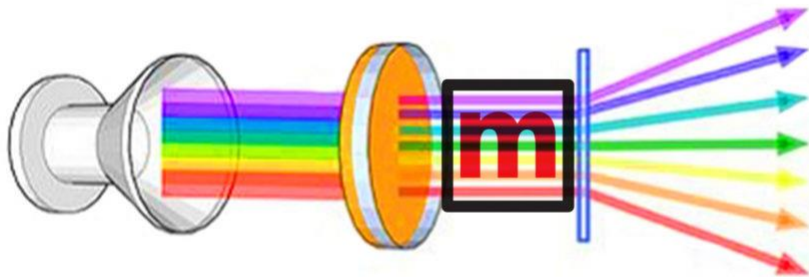
*Prof. R. URBANI*

**Dipartimento di Scienze Chimiche e Farmaceutiche**

Tel 040 558 3684  
e-mail: [rurbani@units.it](mailto:rurbani@units.it)

**a.a. 2021-2022**

## Spettrometria di massa di proteine



70 anni fa circa le innovazioni riguardanti la spettrometria di massa ed in particolare la ionizzazione di biopolimeri [**electrospray (ESI)** e **matrix-assisted laser desorption/ionization (MALDI)**] hanno portato la spettrometria di massa quale tecnica di eccellenza nel campo della emergente **proteomica** (*PROTEins expressed by genOME*).

Con questi strumenti, gli spettrometristi di massa potevano per la prima volta, facilmente e in modo robusto, convertire i peptidi in fase condensata ed anche le proteine intere in ioni di fase gassosa **intatti**.

La «**Peptide Mass Fingerprinting**» è stata quella tra le prime tecniche a guadagnare popolarità. Con questo metodo, una proteina isolata viene digerita enzimaticamente ed i risultanti pesi molecolari dei peptidi vengono misurati con un MALDI o con un MALDI-TOF

Le masse peptidiche vengono poi **confrontate** con quelle di un database di proteine che sono state «**digerite**» *in silico*.

#### **Riferimenti bibliografici «storici»**

**Fenn, J.B., M. Mann, C.K. Meng, S.F. Wong, and C.M. Whitehouse. 1989.** Electrospray ionization for mass-spectrometry of large biomolecules. *Science* 246:64-71.

**Karas, M. and F. Hillenkamp. 1988.** Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons. *Anal. Chem.* 60:2301-2303.

# Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10 000 Daltons

Michael Karas and Franz Hillenkamp

*Anal. Chem.* 1988, 60, 2301-2303

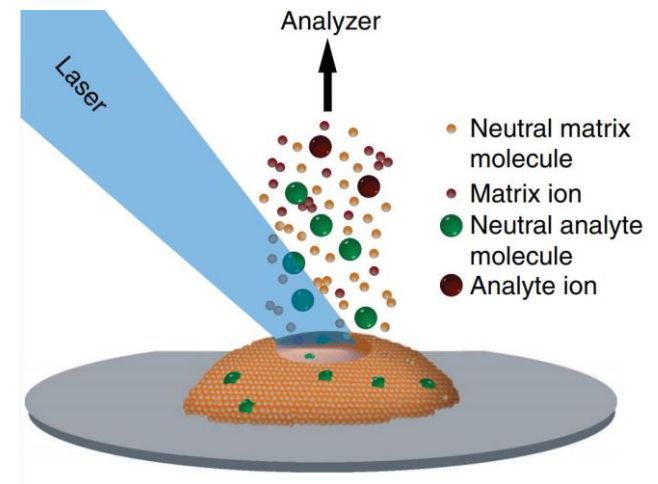
Fast atom bombardment or liquid secondary ion mass spectrometry (SIMS) data on compounds above 10 000 molecular weight show weak signal intensities and poor signal-to-noise (S/N) ratios.

The only exception is the results reported on the analysis of small proteins in the range of 10-24000 daltons.

Sono riportati i primi risultati sul desorbimento con luce laser ultravioletta (LD) di composti bioorganici nel range di massa **superiore a 10000 dalton**.

Un raggio laser è focalizzato sulla camera del campione. Il campione è incorporato in una matrice che assorbe l'energia del laser e poi ne trasmette una parte al campione. Il campione si ionizza e poi viene inviato al rivelatore TOF.

Si chiama  
**Matrix-Assisted Laser Desorption Ionization (MALDI).**



# Electrospray Ionization for Mass Spectrometry of Large Biomolecules

Science 246:64-71 (1989)

JOHN B. FENN, MATTHIAS MANN, CHIN KAI MENG, SHEK FU WONG,  
CRAIG M. WHITEHOUSE

La ionizzazione elettrospray (ESI) è emersa recentemente come una potente tecnica per la produzione di **ioni intatti** nel vuoto da specie grandi e complesse in soluzione.

In una misura maggiore di quanto sia stato possibile in precedenza con i più familiari metodi di ionizzazione (FAB, LD), questa tecnica "soft" offre la potenza e l'eleganza dell'analisi MS applicabile alle grandi e fragili molecole polari che svolgono un ruolo così vitale nei sistemi biologici.

Le caratteristiche degli spettri elettrospray per le grandi molecole sono le sequenze di picchi i cui ioni componenti sono multipli di carica, ossia in cui gli ioni corrispondenti ad ogni picco differiscono **di una carica** da quelli dei vicini adiacenti nella sequenza.

Si sono ottenuti spettri di biopolimeri tra cui oligonucleotidi e proteine, con pesi molecolari fino a 130,000 Da.

# Peptide Mass Fingerprinting

Se una proteina può essere tagliata in modo controllato, le dimensioni dei frammenti formano **un'impronta digitale** per quella proteina.

Se da un database di sequenze si possono idrolizzare proteine nello stesso modo ma *in silico*, l'impronta digitale può servire ad identificare la proteina.

Questa ipotesi è stata la base dei primi esperimenti, in quello che sarebbe diventato il **Peptide Mass Fingerprinting (PMF)**

La forza motrice primaria nello sviluppo della PMF è stata quella necessaria ad aumentare la velocità dell'analisi delle proteine.

Fino al 1989 la degradazione (automatizzata) di Edman aveva ancora **un tempo di ciclo di quasi un'ora per ogni residuo di amminoacido** (idrolisi, elettroforesi, blotting, staining, separazione delle bande e sequenziamento).

La possibilità di identificare rapidamente le proteine ha consentito un più efficiente e veloce uso del sequenziatore di proteine per l'analisi di nuove strutture primarie.

# Peptide Mass Fingerprinting

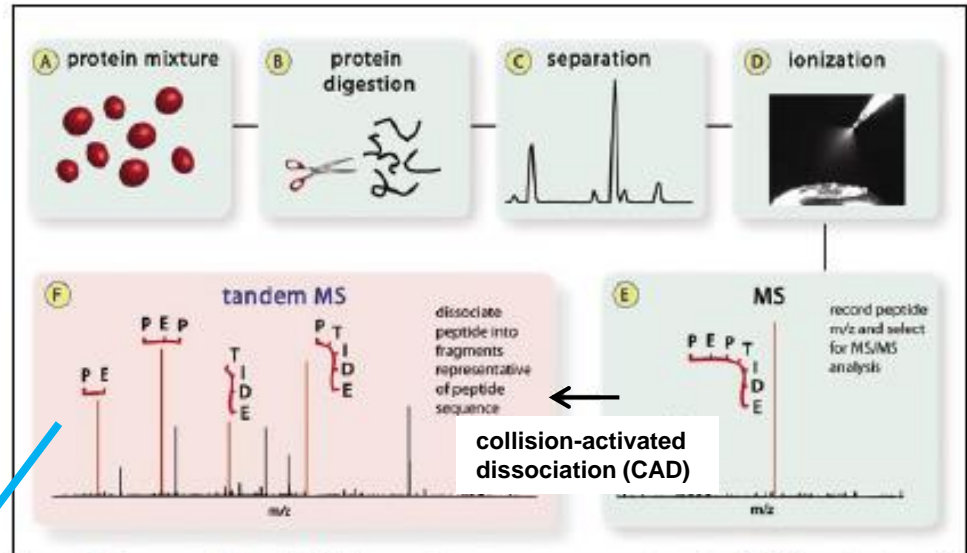
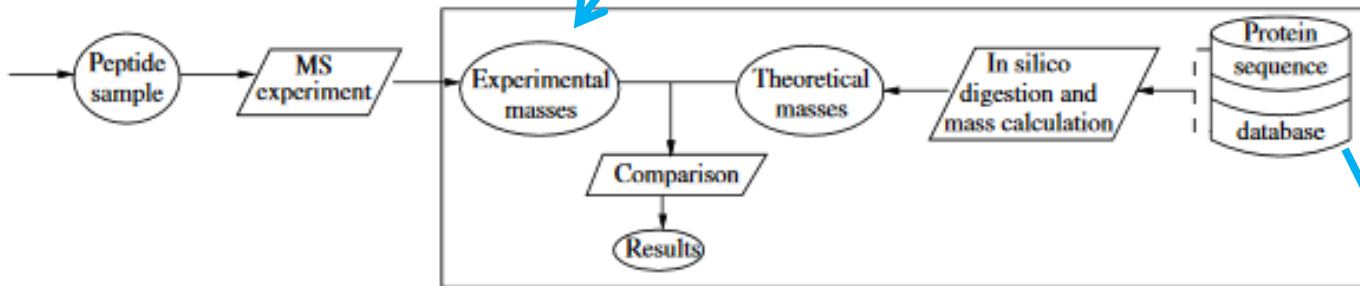


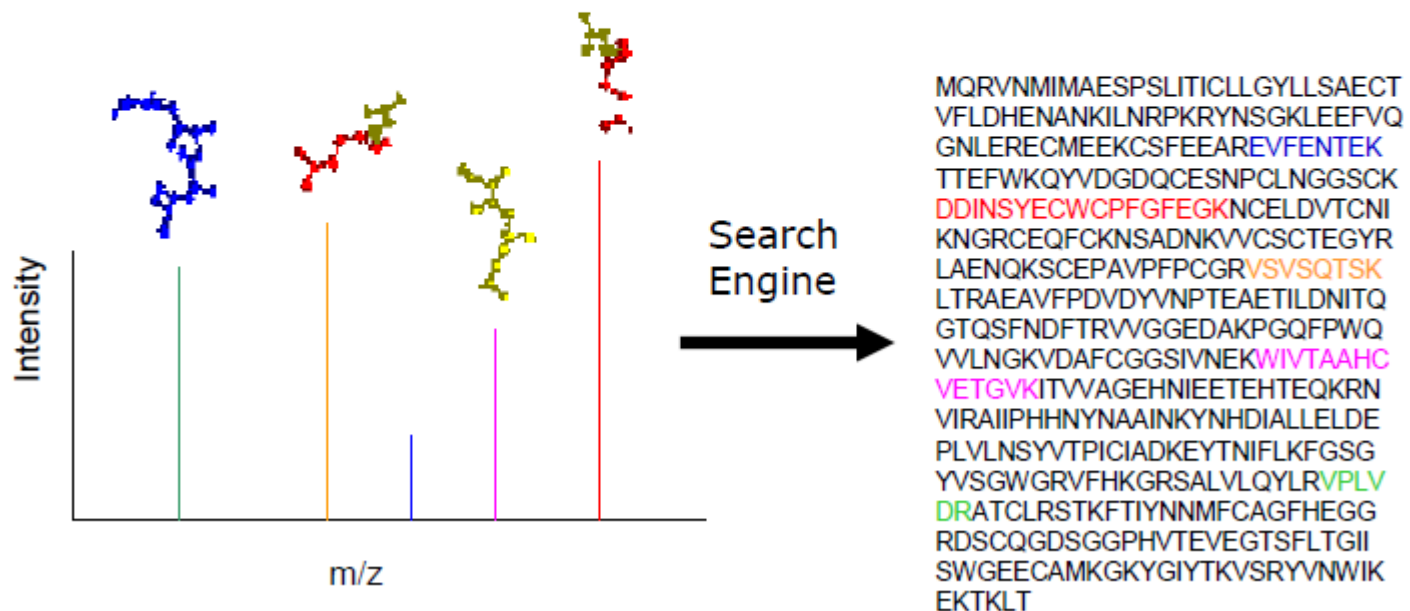
Figure 1. Sequence of events defining a contemporary mass spectrometry (MS)-based proteomics experiment. (A) The process begins with a mixture of proteins from any number of sources. (B) Enzymes



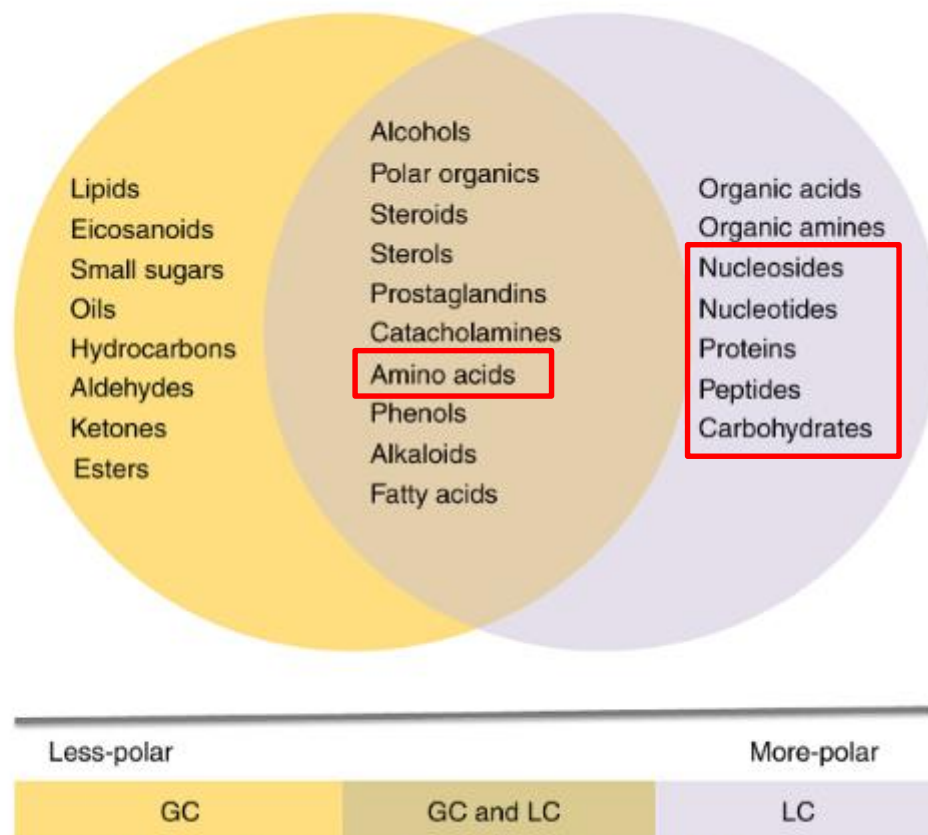
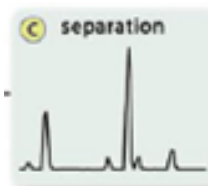
L'obiettivo è quello di trovare una proteina nel database delle sequenze che possa spiegare i dati sperimentali, per quanto possibile

# Identification of a Protein via Peptide Mass Fingerprinting

Direct analysis of the **entire peptide mixture**



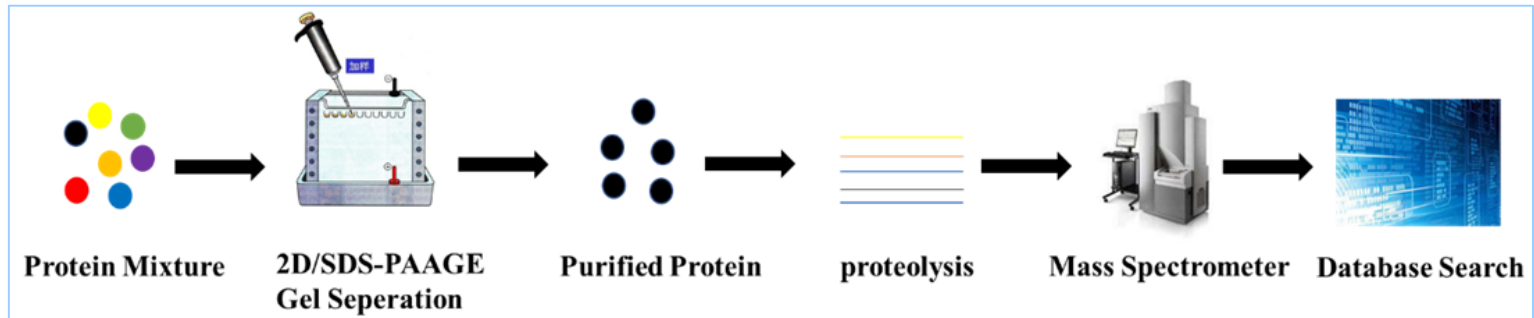
# Separazione dei peptidi



**FIGURE 2.1** The Classes of Biomolecules That are Analyzed by GC/MS and LC/MS



## Peptide Mass Fingerprint



Identificazioni affidabili richiedono esperimenti MS di alta qualità.

Ciò significa che gli spettri devono contenere **picchi chiari e ben risolti** del campione di proteine e, idealmente, nessuno altro picco da altre molecole.

D'altra parte, **le masse teoriche** si ottengono utilizzando programmi per computer che traducono il genoma noto in proteine o le proteine nel database, quindi **tagliano teoricamente le proteine** in peptidi e calcolano le masse assolute dei peptidi.

Si esegue pertanto la **digestione *in silico*** su proteine di database **con lo stesso enzima** utilizzato nella digestione sperimentale e **viene generata una lista teorica dei picchi**.

Mascot, **MS-Fit** e Profound sono alcuni di questi software per il PMF.

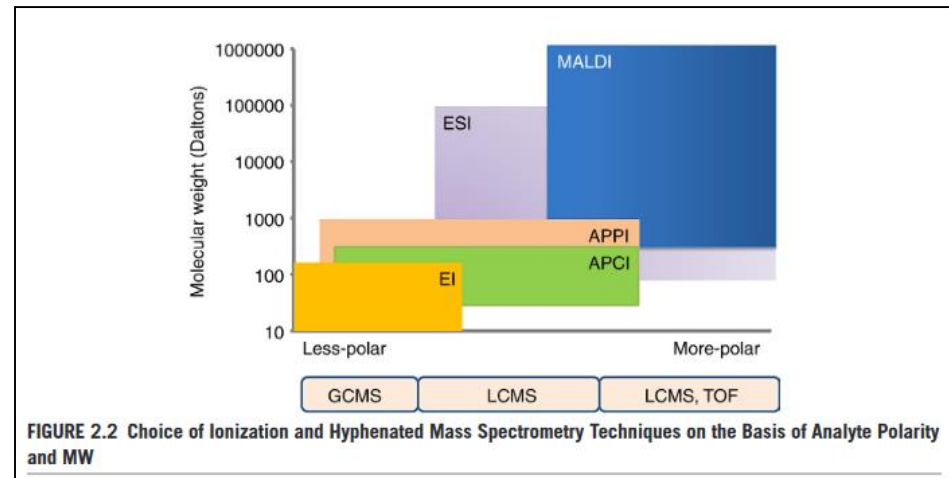
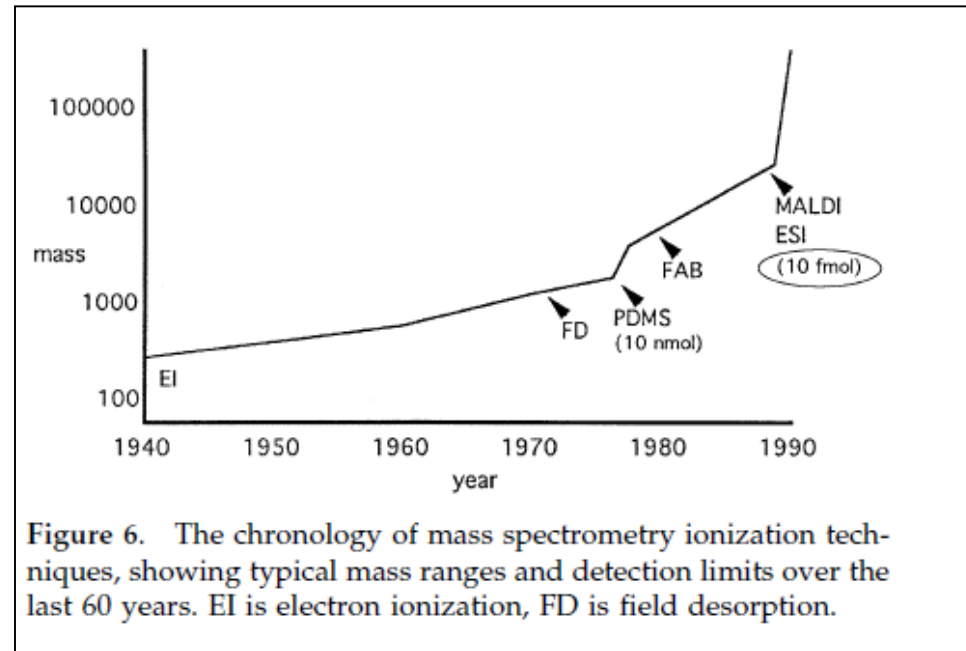
Le masse peptidiche ottenute sperimentalmente vengono confrontate con la massa peptidica teorica. I risultati sono analizzati statisticamente per trovare il miglior «**matching**» tra le strutture.

## Peptide Mass Fingerprinting

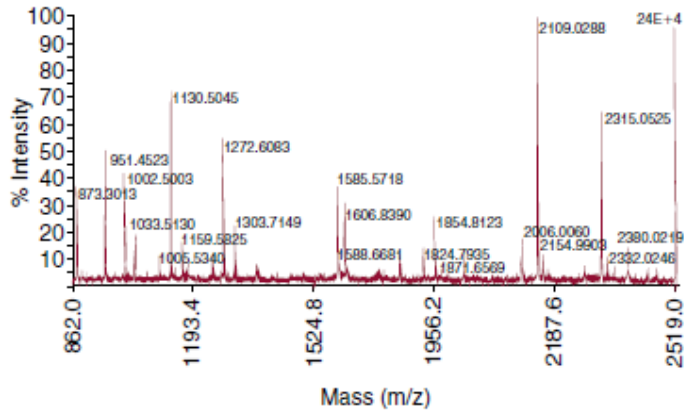
Le due nuove tecniche di ionizzazione, (ESI e MALDI) hanno superato rapidamente le prestazioni di FAB e PDMS (Plasma desorption-mass spectrometry).

Ognuna di queste nuove tecniche presentavano limiti di rilevamento dell'ordine delle fmoli e un range delle masse superiore a 100,000 Da.

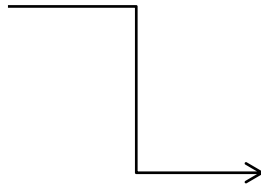
Il grande aumento della sensibilità del MALDI rispetto al metodo FAB ha incoraggiato il **PMF** come strumento per l'identificazione delle proteine



Spec #1⇒BC[BP = 2139.0, 23588]



**FIGURE 2.18** Example of a matrix-assisted laser desorption ionization–time-of-flight spectrum showing peptides generated in a tryptic digest of a spot cored from a two-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis. The 16 most abundant  $m/z$  values were submitted to the MS-Fit database for searching against the nonredundant database. The results for this search are shown in Table 2.1.



**TABLE 2.1** Example of Printout of Bacterial Identification Through Peptide Mass Fingerprinting Using Matrix-Assisted Laser Desorption Ionization MALDI-Time of Flight\*

Rank	Mowse Score	# (%) Masses Matched	Protein Mw (Da)/pI	Species	NCBI Inr. 81602 Accession #	Protein Name
1	1.07e+008	14/16 (87%)	101754.9/9.15	<i>Saccharomyces cerevisiae</i>	6321275	(Z72685) ORF YGL163c
1. 14/16 matches (87%). 101754.9 Da, pI = 9.15. Acc. #6321275. <i>Saccharomyces cerevisiae</i> . (Z72685) ORF YGL163c.						
$m/z$ Submitted	MH <sup>+</sup> Matched	Delta ppm	Start	End	Peptide Sequence (Click for Fragment Ions)	Modifications
870.4746	870.4797	-5.8732	598	606	(K) GVGGSOPLR(A)	
873.3981	873.3929	5.9793	774	779	(K) DCFIYR(F)	C <sup>2</sup> H <sup>2</sup> O <sup>2</sup>
951.4901	951.4900	0.1050	814	821	(R) LFSSDNLK(Q)	
1002.5385	1002.5373	1.2224	515	522	(K) NFENPILR(G)	
1033.5513	1033.5543	-2.8793	46	55	(K) NTHIPPAAGR(I)	
1130.6349	1130.6322	2.4037	120	128	(R) LSHIQYTLR(R)	
1130.6349	1130.6322	2.4037	514	522	(R) KNFENPILR(G)	
1159.6039	1159.6071	-2.7957	56	67	(R) IATGSDNIVGGR(S)	
1272.6508	1272.6483	1.9865	734	746	(K) AGGCGINLIGANR(L)	C <sup>2</sup> H <sup>2</sup> O <sup>2</sup>
1303.7573	1303.7599	-1.9457	270	280	(K) ILRPHQVEGVR(F)	
1585.7190	1585.7215	-1.5602	446	459	(K) NCVGLMLADEGHR(L)	C <sup>2</sup> H <sup>2</sup> O <sup>2</sup>
1606.8861	1606.9029	-10.4650	22	35	(R) LVPRPINVQDSVNR(L)	
2138.0756	2138.0704	2.4250	747	765	(R) LILMDPDVWNPAAADQQALR(M)	
2315.1093	2315.0951	6.1321	401	423	(K) SSMGGGNTTVSQAIHAWAQAQGR(N)	
2388.0671	2388.0731	-2.5004	293	313	(K) DYLEAEAFNTSSEDPLKSDK(A)	

MH<sup>+</sup>, Ion formed by attachment of a proton to molecule M; MOWSE, MOlecular Weight SEarch method; MW, molecular weight;  $m/z$ , mass-to-charge ratio

\*A generated mass list is compared with theoretical tryptic masses for proteins in a database. Match quality is used for pathogen identification.

---

## Description, Instructions, and Tips for [MS-Fit](#)

---

### Purpose

This document provides instructions for [MS-Fit](#).

---

### Contents of this document:

- [Introduction and Background](#)
  - [AA Substitutions \(homology mode\)](#)
  - [Report Homologous Proteins](#)
  - [Minimum Number of Peptides Required to Match](#)
  - [Ranking / Scoring of Results](#)
  - [Multiply-charged ions](#)
  - [Monoisotopic/Average Flags](#)
  - [Searching for Mixtures](#)
  - [Looking for Peptides with Non-Specific Cleavages](#)
  - [Hit Statistics](#)
- 

### Links to topics in the general instructions:

- [Search Times](#)
  - [Stopping / Cancelling a Search](#)
  - [Saving Hits](#) from one Protein Prospector program, searching them with another
  - [Databases](#)
  - [Species Filtering](#)
  - [Species Code Filtering](#)
  - [Intact Protein MW Filtering](#)
  - [Intact Protein pI Filtering](#)
  - [Enzyme specificity / Missed cleavages](#)
-

# MS-Digest

<b>Database</b> User Protein <input type="text"/>	<b>Retrieve Entry by</b> Accession Number <input type="text"/>	
<b>Output</b> HTML <input type="text"/> <b>Hits to file</b> <input type="checkbox"/> <b>Name</b> lastres <input type="text"/>	<b>N Term AA Limit</b> <input type="text"/>	
<b>Digest</b> Trypsin <input type="text"/> <b>Max. Missed Cleavages</b> 1 <input type="text"/>	<b>List of Entries</b> ( <a href="#">see instructions</a> ) <div style="border: 1px solid black; padding: 5px; min-height: 100px;">P15497</div>	
<b>[+] End Terminus Parameters</b> <div style="border: 1px solid gray; padding: 5px;"><ul style="list-style-type: none"><li>Abu (N-term)</li><li>Acetonydrazide (C-term)</li><li>Acetonydrazide (DE)</li><li>Acetyl (K)</li><li>Acetyl (N-term)</li><li>Acetyl (R)</li></ul></div>		
<b>Constant Mods</b> <div style="border: 1px solid gray; padding: 5px;"><ul style="list-style-type: none"><li>Acetyl (K)</li><li>Acetyl (N-term)</li><li>Acetyl (R)</li></ul></div>	<b>Peptide Mass</b> <input type="text" value="800.0"/> to <input type="text" value="4000.0"/> <b>Min Peptide Length</b> <input type="text" value="5"/> <b>Hide Protein Sequences</b> <input type="checkbox"/> <b>Hide HTML Links</b> <input type="checkbox"/> <b>Report Multiple Charges</b> <input type="checkbox"/> <b>Bull Breese Indices</b> <input type="checkbox"/> <b>HPLC Indices</b> <input type="checkbox"/> <b>Separate Proteins</b> <input type="checkbox"/>	
<b>[+] Present Amino Acids</b> ( <a href="#">see instructions</a> )		<b>Perform Digest</b> <input type="button" value="Perform Digest"/>
<b>Variable Mods</b> <div style="border: 1px solid gray; padding: 5px;"><ul style="list-style-type: none"><li>Acetyl (Protein N-term)</li><li>Gln-&gt;pyro-Glu (N-term Q)</li><li>Met-loss (Protein N-term M)</li><li>Met-loss+Acetyl (Protein N-term M)</li><li>Oxidation (M)</li></ul></div>		
<b>Add (+) or remove (-) mods using menus &amp; buttons below</b> <div style="border: 1px solid gray; padding: 5px;"><b>Frequent</b> <input type="text"/> <div style="border: 1px solid gray; padding: 5px;"><ul style="list-style-type: none"><li>Acetyl (K)</li><li>Acetyl (Protein N-term)</li><li>Acetyl (Uncleaved K)</li><li>Acetyl+Oxidation (Protein N-term M)</li></ul></div></div>		
<b>[+] Set Contents</b> <input type="button" value="+"/> <input type="button" value="-"/>		
<b>For digestion of a user supplied sequence select User Protein above.</b>		
<b>User Protein Sequence</b> <pre>MKAVVLT LAVLFLTGSQARHFWQQDDPQSSWDRVKDFATVYVEAIKDSGRDYVAQFEASALGKQLNLKLLDNWDTLASTL SKVREQLGVPVTQEFWDNLEKETASLRQEMHKDLEEVKQKVPYLDFQKKWHEEVEIYRQKVAPLGEEFREGARQKVQEL QDKLSPLAQELRDRARAHVETLRQQLAPYSDDLRLQRLTARLEALKEGGGSLAEYHAKASEQLKALGEKAKPVLEDLRQGL LPVLES LKVSILAAIDEASKKLNAAQ</pre>		

# MS-Digest Search Results

## [+] Parameters

Index Number: **1**

pI of Protein: **5.7**

Protein MW: **30277**

Amino Acid Composition: **A27 D17 E27 F7 G11 H5 I4 K23 L38 M2 N4 P8 Q23 R16 S15 T9 V18 W5 Y6**

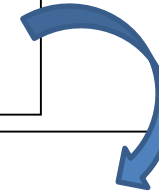
1 MKAVVLT<sup>1</sup>LAV LFLTGSQ<sup>2</sup>ARH FWQQDDP<sup>3</sup>QSS WDRV<sup>4</sup>KDFATV YVEAI<sup>5</sup>KDSGR DYVAQ<sup>6</sup>FEASA LGKQL<sup>7</sup>NLKL<sup>8</sup>LN DNWD<sup>9</sup>TLASTL  
 81 SKVRE<sup>10</sup>QLGPV TQEFWD<sup>11</sup>NLEK ETASLR<sup>12</sup>QEMH KDLEEV<sup>13</sup>KQKV QPYLDEF<sup>14</sup>QKK WHEEVEI<sup>15</sup>YRQ KVAPL<sup>16</sup>GEEFR EGARQ<sup>17</sup>KVQEL  
 161 QDKLS<sup>18</sup>PLAQE LRDRAR<sup>19</sup>AHVE TLRQQL<sup>20</sup>APYS DDLRQ<sup>21</sup>RLTAR LEALKE<sup>22</sup>GGGS LAEYHAK<sup>23</sup>ASE QLKAL<sup>24</sup>GEKAK PVLED<sup>25</sup>LRQGL  
 241 LPVLES<sup>26</sup>LKVS ILAAIDE<sup>27</sup>ASK KLNA<sup>28</sup>Q

Number	m/z (mi)	m/z (av)	Modifications	Start End	Missed Cleavages	Sequence
1	825.4577	825.9485		177 183	0	(R)AHVETLR(Q)
1	859.4520	859.9602		157 163	0	(K)VQELQDK(L)
1	988.5310	989.1201		112 119	1	(K)DLEEVKQK(V)
1	1014.6306	1015.2485		197 205	1	(R)LTARLEALK(E)
1	1017.5364	1018.1645		142 150	0	(K)VAPLGEEFR(E)
1	1026.5942	1027.2160		164 172	0	(K)LSPLAQELR(D)
1	1040.6099	1041.2430		229 237	0	(K)AKPVLEDLR(Q)
1	1052.5960	1053.2165		175 183	1	(R)ARAHVETLR(Q)
1	1098.5790	1099.2363	1Gln->pyro-Glu	155 163	1	(R)QKVQELQDK(L)
1	1115.6055	1116.2670		155 163	1	(R)QKVQELQDK(L)
1	1173.6474	1174.3475		218 228	1	(K)ASEQLKALGEK(A)
1	1179.6984	1180.4387	1Gln->pyro-Glu	238 248	0	(R)QGILLPVLES <sup>26</sup> LK(V)
1	1196.7249	1197.4694		238 248	0	(R)QGILLPVLES <sup>26</sup> LK(V)
1	1216.6783	1217.4134		249 260	0	(K)VSILAAIDEASK(K)
1	1218.5749	1219.3040		206 217	0	(K)EGGGSLAEYHAK(A)
1	1255.6569	1256.4500		36 46	0	(K)DFATVYVEAIK(D)
1	1256.6634	1257.4406	1Gln->pyro-Glu	140 150	1	(R)QKVAPLGEEFR(E)
1	1260.6008	1261.3879		131 139	0	(K)WHEEVEIYR(Q)
1	1266.6365	1267.4328		120 129	0	(K)VOPYI DFFOK(K)

## 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

FT-ICR=Fourier transform-ion  
cyclotron resonance.

# Determinazione della struttura: MS<sup>n</sup>

## i) Proteina *in toto*:

collision-induced dissociation

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)

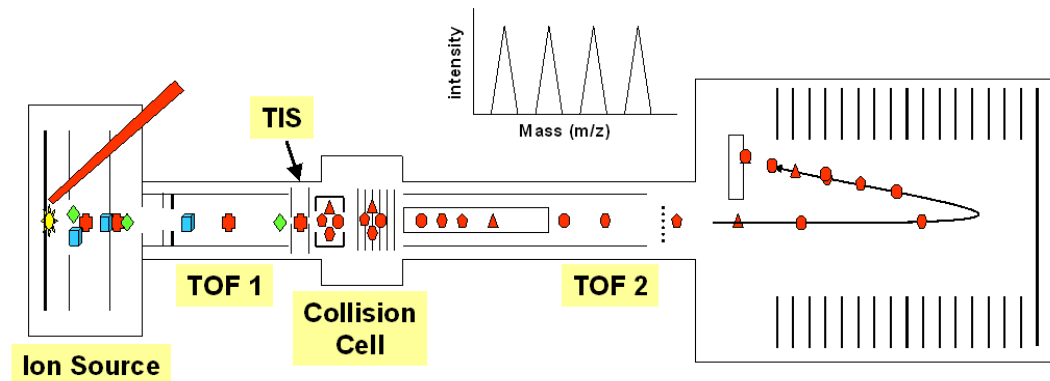
Approccio **Top Down**



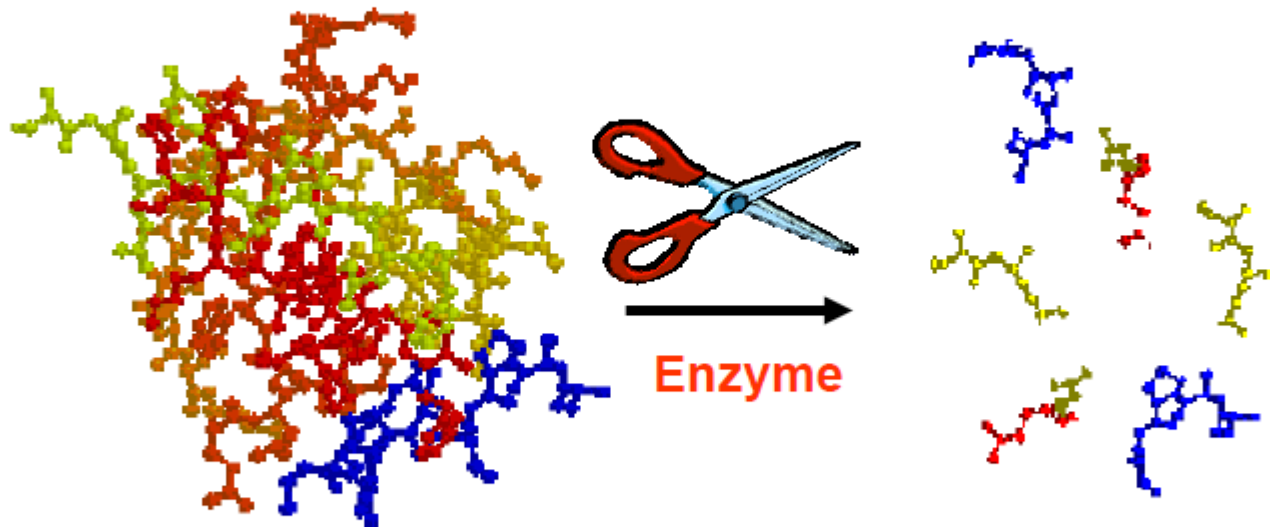
L'obiettivo chiave nell'analisi «**TOP-DOWN**» è quello di frammentare (blandamente) in modo efficiente le proteine intatte per ottenere informazioni specifiche della sequenza per la ricerca nel database e l'identificazione delle proteine

L'emergere delle nuove tecniche MALDI TOF-TOF permette di ottenere sequenza di informazioni specifiche da proteine intatte **senza trattamento proteolitico**.

## MALDI TOF-TOF: *MS/MS Mode*



ii) Digestione enzimatica → Peptidi

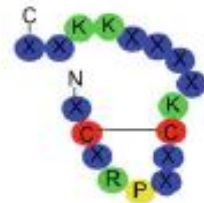


1 Protein  
(in solution)

Many Peptides  
(in solution)

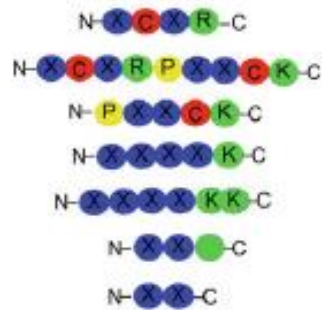
Approccio **Bottom Up**

Bottom up

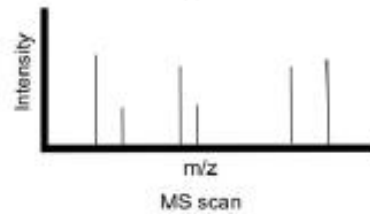


Protein(s)

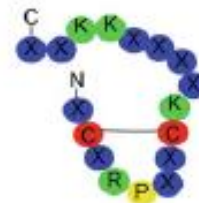
↓ Digestion and alkylation



↓ MS

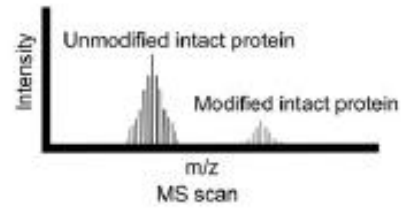


Top down

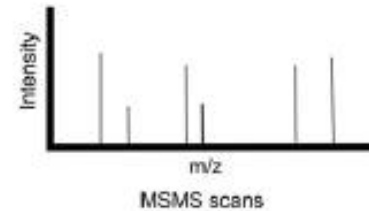


Protein

↓ MS



↓ Gas phase dissociation MS



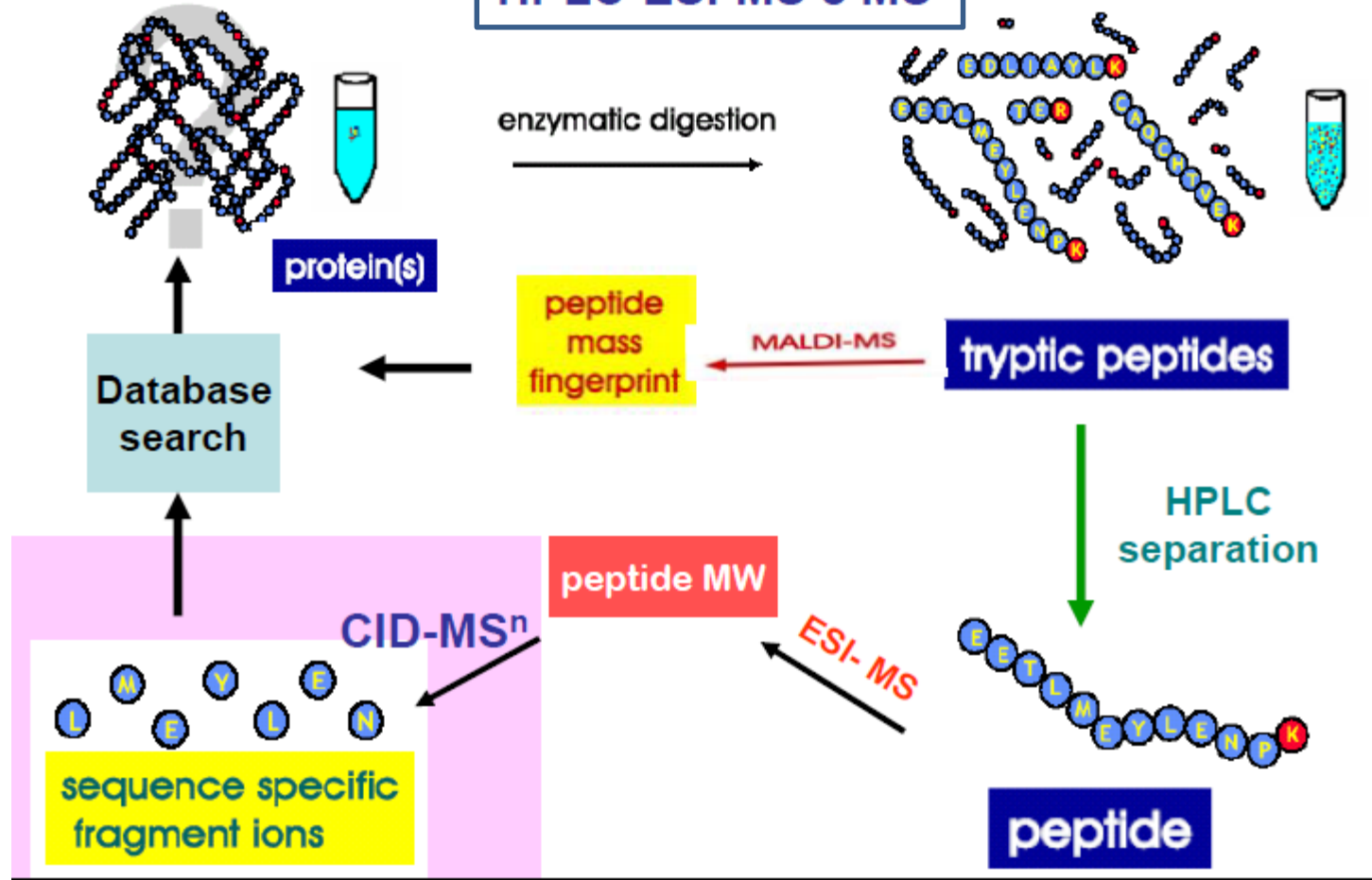
Se il TOF ha alta risoluzione, è possibile determinare la massa accurata e la formula bruta di ciascun peptide

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

# Proteine: Determinazione della struttura

Digestione enzimatica → Peptidi

HPLC-ESI-MS e MS<sup>n</sup>



Tandem MS-MS

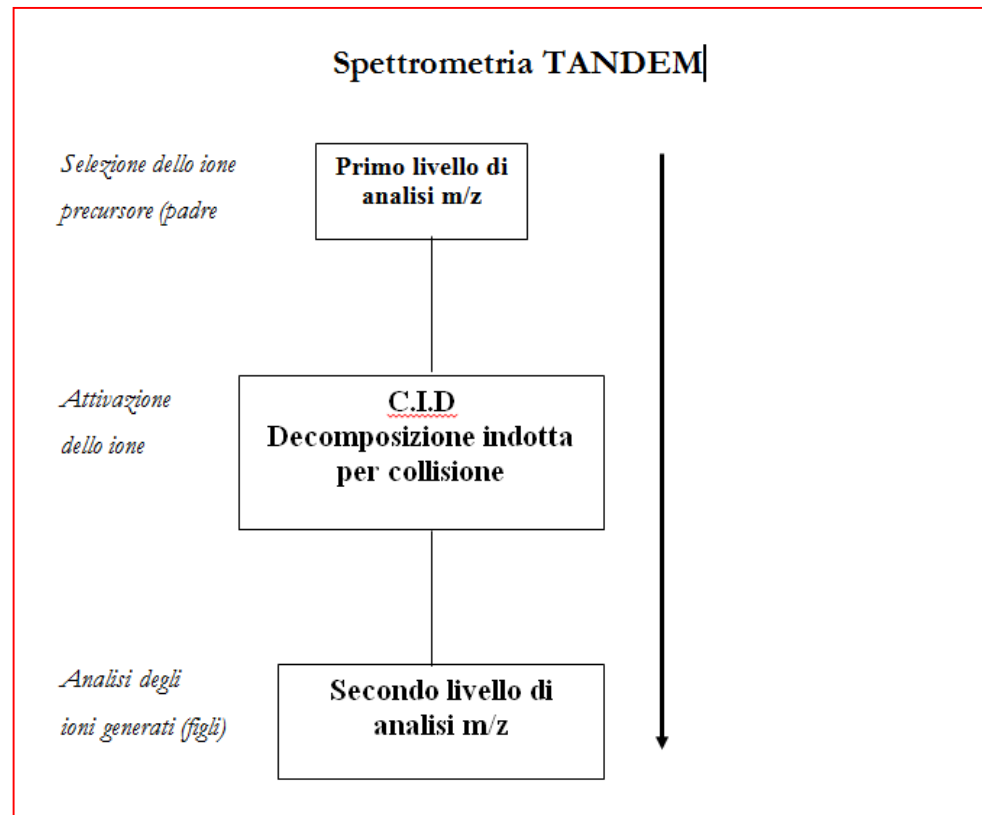
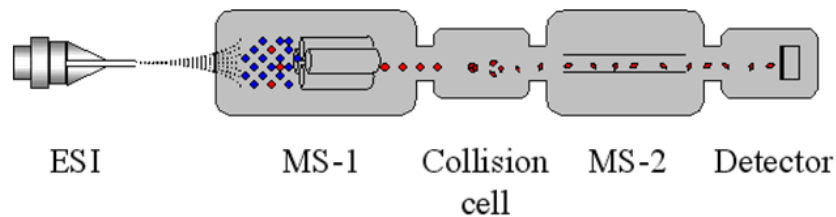
# **Spettrometria di massa tandem (MS-MS)**

## **per ottenere informazioni strutturali e sulla sequenza**

La spettrometria di massa Tandem (MS-MS) è utilizzata per ottenere **informazioni strutturali** di un composto attraverso la frammentazione di ioni specifici prodotta all'interno dello spettrometro ed identificando i frammenti ottenuti.

Da queste informazioni si possono ottenere dati strutturali sulla molecola nativa.

E' stata messa a punto la *Spettrometria di Massa Tandem* (MS/MS) nella quale lo ione molecolare stabile, immediatamente dopo la determinazione della sua massa, viene fatto collidere con molecole di gas neutro (elio od argon), acquista ulteriore energia e la dissipa frammentandosi. Il processo è detto *Dissociazione Indotta da Collisione* (CID). I frammenti ionici prodotti vengono quindi sottoposti ad una seconda analisi.



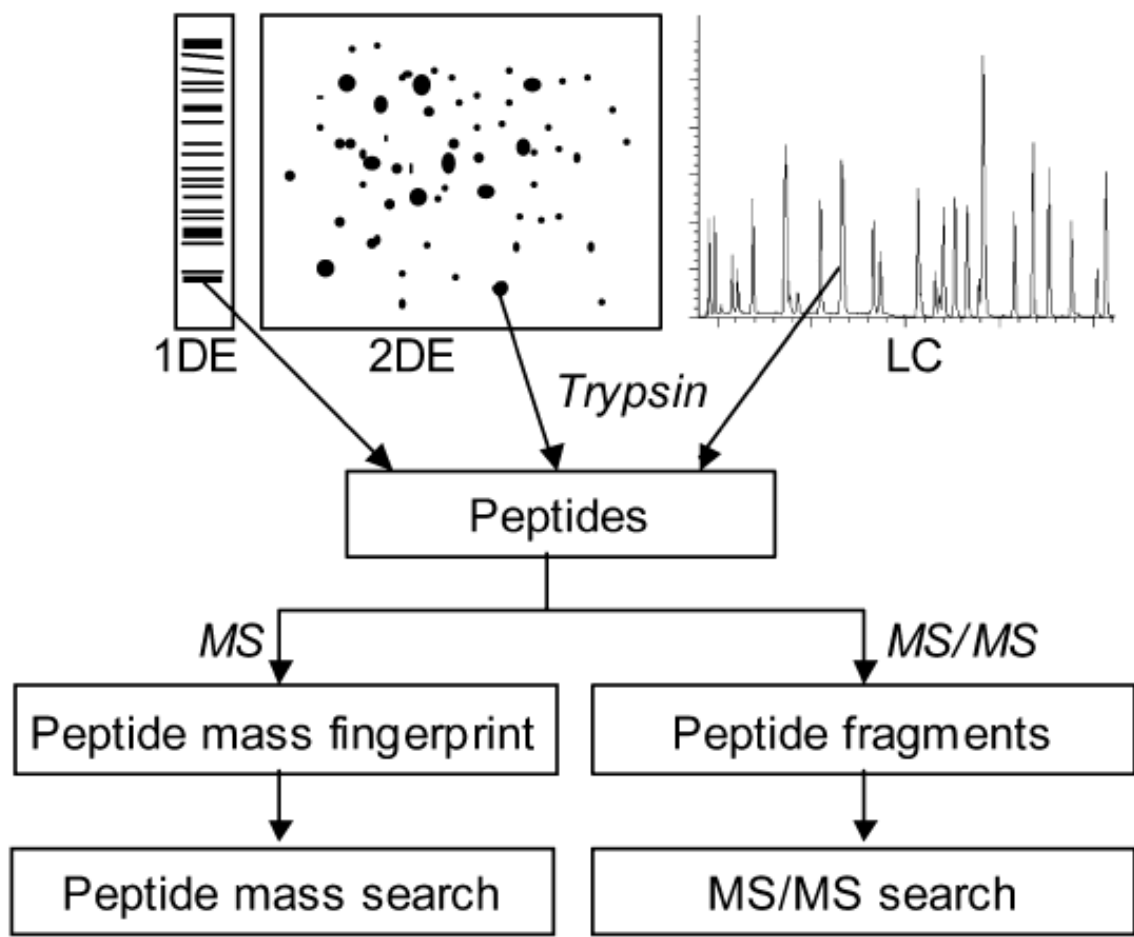


Fig. 1. Common procedures to identify proteins by mass spectrometry.



Circa 4  $\mu\text{L}$  di soluzione bastano per l'analisi della miscela triptica che corrisponde ad una concentrazione di proteina originale pari a 1-10  $\text{pmol}/\mu\text{L}$ .

Quindi il sequenziamento mediante MS/MS è una tecnica sensibile che consuma poco campione.

La quantità di informazione estraibile varia da peptide a peptide. Alcuni peptidi possono generare l'informazione sufficiente per determinare l'intera sequenza, altri solo sequenze di 4 o 5 amminoacidi.

Alcuni procedimenti di analisi possono essere automatizzati da software specifici, ma in generale l'analisi di tutta l'informazione contenuta è un processo che richiede tempo ed accuratezza se si vogliono ottenere dati credibili.

## Tagli specifici con metodi chimici o enzimatici

### Reagente

### Sito di taglio

#### Taglio chimico

Bromuro di cianogeno  
O-iodobenzoato  
Idrossilammia  
2-Nitro-5-tiocianobenzoato

Lato carbossilico di residui Met  
Lato carbossilico di residui Trp  
Legami Asp-Gly  
Lato aminico di residui di Cys

#### Taglio enzimatico

Tripsina  
Clostripaina  
Proteasi dello Stafilococco

Lato carbossilico di residui Lys e Arg  
Lato carbossilico di residui Arg  
Lato carbossilico di residui Asn e Gln

## Sequenziamento di peptidi:



I peptidi frammentano in un modo ben conosciuto.

Le molecole protonate frammentano **in catena principale e a volte anche sulle catene laterali.**

Ci sono tre diversi tipi di legami che possono essere rotti in catena principale:

**NH-CH**

**CH-CO**

**CO-NH.**

La carica può risiedere su l'uno o l'altro dei due frammenti in dipendenza della chimica e dell'affinità delle due specie per il protone.

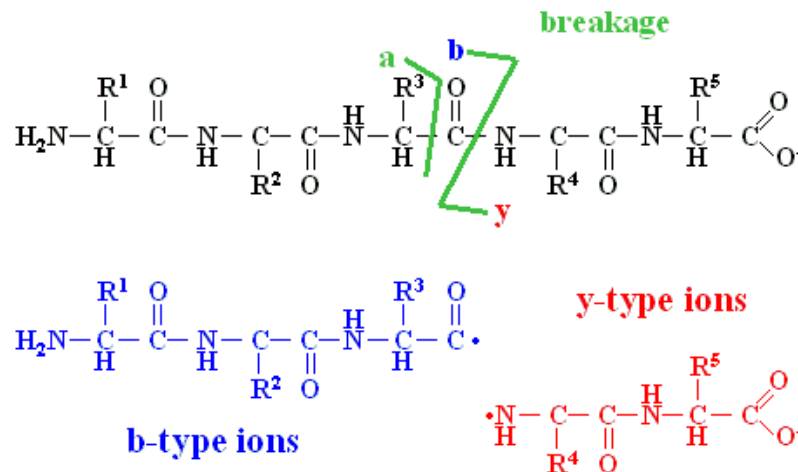
Quindi per ogni residuo di aminoacido possono essere ottenuti sei tipi di frammenti e sono marcati con:

**a, b e c** gli ioni che hanno la carica sul frammento ammino-terminale

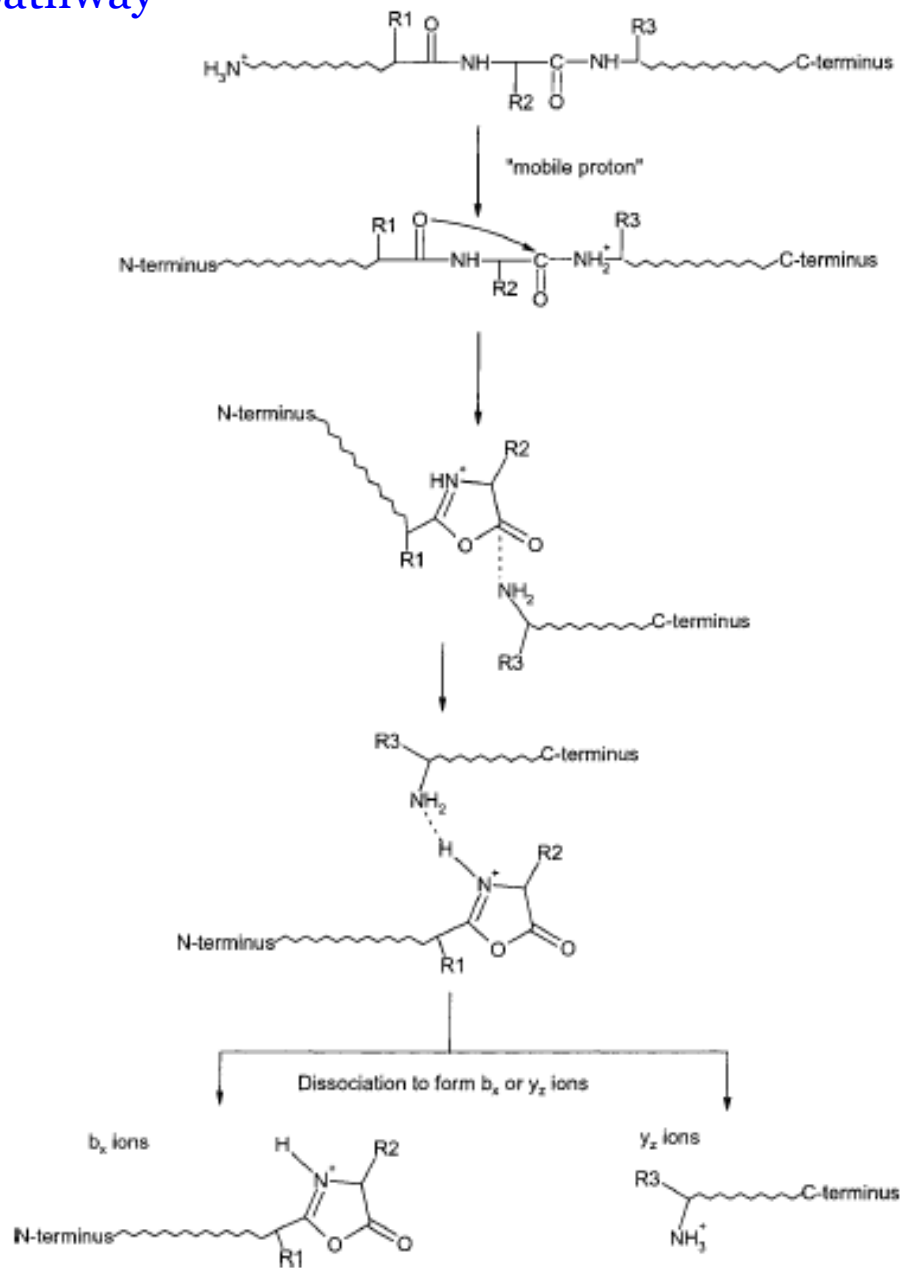
**x, y e z** gli ioni che hanno la carica sul frammento C-terminale.

La rottura più comune è quella sul legame CO-NH che dà luogo a ioni **b** e/o **y**.

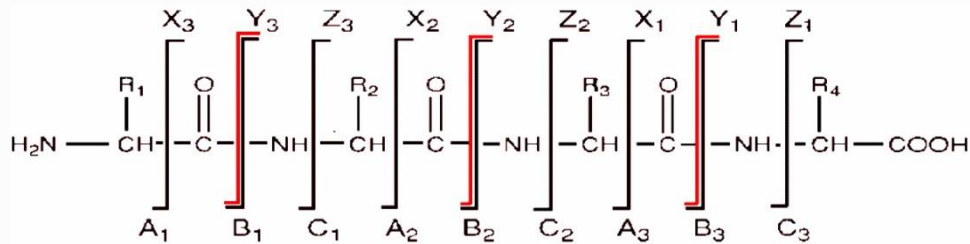
La differenza di massa tra due ioni **b** (o **y**) adiacenti indica la presenza di un certo residuo.



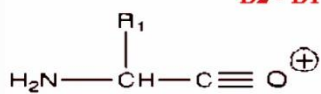
# The $b_x$ and $y_z$ pathway



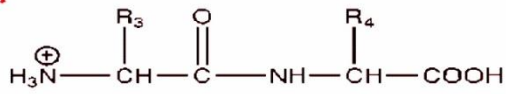
**SCHEME 3.**



$b_2 - b_1 = \text{amino acid}$   
 $B_2 - B_1 = \text{amino acid}$

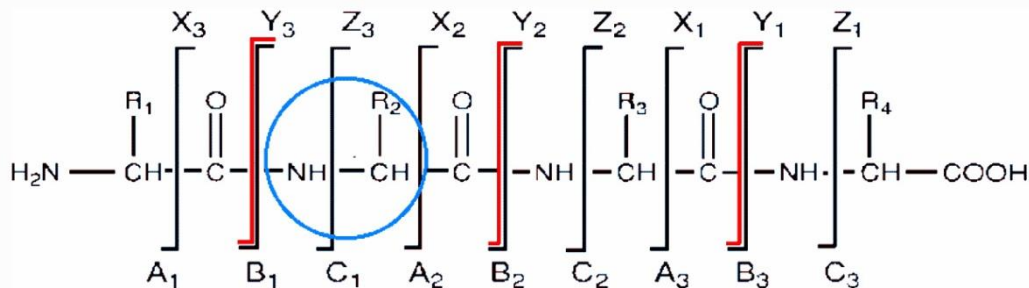


B<sub>1</sub>



Y<sub>2</sub><sup>+</sup>

Ione immonio  $\text{H}_2\text{N}^+=\text{CH}-\text{R}_2$



*Rottura di legami nel sequenziamento mediante massa Tandem*

Gli **ioni immonio** compaiono nella zona a **bassi** m/z dello spettro MS – MS.

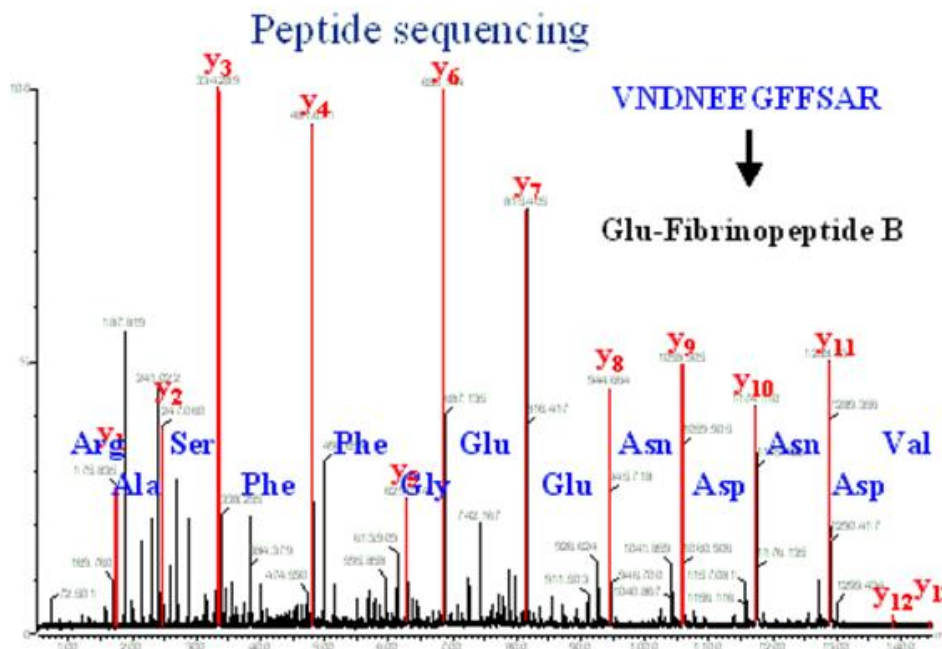
Ciascun residuo di amminoacido porta ad uno **ione immonio diagnostico** ad eccezione delle coppie lisina (K) - glutammina (Q) e leucina (L) – isoleucina (I) che producono ioni immonio con la stessa m/z ( 86 per I e L e 101 per K e Q).

Gli ioni immonio sono utili per confermare la presenza di molti degli amminoacidi in un peptide sebbene non sempre diano informazioni sulla posizione di quei residui nella sequenza.

**Table 17.2:**  
**Monoisotopic residue masses for DNA-encoded Amino Acids**

Amino Acid		Monoisotopic Residue Mass (Da)		Immonium Ion ( $m/z$ )
G	Gly	57	57.0215	30
A	Ala	71	71.0371	44
S	Ser	87	87.0320	60
P	Pro	97	97.0528	70
V	Val	99	99.0684	72
T	Thr	101	101.0477	74
C	Cys	103	103.0092	76
L	Leu	113	113.0841	86
I	Ile	113	113.0841	86
N	Asn	114	114.0429	87
D	Asp	115	115.0269	88
Q	Gln	128	128.0586	101
K	Lys	128	128.0950	101
E	Glu	129	129.0426	102
M	Met	131	131.0405	104
H	His	137	137.0589	110
F	Phe	147	147.0684	120
R	Arg	156	156.1011	129
Y	Tyr	163	163.0633	136
W	Trp	186	186.0793	159





Simbolo	Struttura	Massa (Da)
Ala A	-NH.CH.(CH <sub>3</sub> ).CO-	71.0
Arg R	-NH.CH.((CH <sub>2</sub> ) <sub>3</sub> .NH.C(NH).NH <sub>2</sub> ).CO-	156.1
Asn N	-NH.CH.(CH <sub>2</sub> CONH <sub>2</sub> ).CO-	114.0
Asp D	-NH.CH.(CH <sub>2</sub> COOH).CO-	115.0
Cys C	-NH.CH.(CH <sub>2</sub> SH).CO-	103.0
Gln Q	-NH.CH.(CH <sub>2</sub> CH <sub>2</sub> CONH <sub>2</sub> ).CO-	<u>128.1</u>
Glu E	-NH.CH.(CH <sub>2</sub> CH <sub>2</sub> COOH).CO-	129.0
Gly G	-NH.CH <sub>2</sub> .CO-	57.0
His H	-NH.CH.(CH <sub>2</sub> C <sub>3</sub> H <sub>3</sub> N <sub>2</sub> ).CO-	137.1
Ile I	-NH.CH.[CH.(CH <sub>3</sub> )CH <sub>2</sub> .CH <sub>3</sub> ].CO-	<u>113.1</u>
Leu L	-NH.CH.[CH <sub>2</sub> CH(CH <sub>3</sub> ) <sub>2</sub> ].CO-	<u>113.1</u>
Lys K	-NH.CH.((CH <sub>2</sub> ) <sub>4</sub> NH <sub>2</sub> ).CO-	<u>128.1</u>
Met M	-NH.CH.((CH <sub>2</sub> ) <sub>2</sub> .SCH <sub>3</sub> ).CO-	131.0
Phe F	-NH.CH.(CH <sub>2</sub> Ph).CO-	147.1
Pro P	-NH.(CH <sub>2</sub> ) <sub>3</sub> .CH.CO-	97.1
Ser S	-NH.CH.(CH <sub>2</sub> OH).CO-	87.0
Thr T	-NH.CH.[CH(OH)CH <sub>3</sub> ].CO-	101.0
Trp W	-NH.CH.[CH <sub>2</sub> .C <sub>8</sub> H <sub>6</sub> N].CO-	186.1
Tyr Y	-NH.CH.((CH <sub>2</sub> ).C <sub>6</sub> H <sub>4</sub> .OH).CO-	163.1
Val V	-NH.CH.[CH(CH <sub>3</sub> ) <sub>2</sub> ].CO-	99.1

Il grado di **frammentazione sulle catene laterali** dipende dal tipo di analizzatori utilizzati.

Alcuni analizzatori danno collisioni **ad alta energia** che favoriscono la frammentazione delle catene laterali.

Gli strumenti a quadrupolo–quadrupolo e quadrupolo–TOF sono caratterizzati da frammentazioni **a bassa energia** e quindi danno poche frammentazioni sulle catene laterali.

For a known sequence, the sequence ions of **b-** and **y-type** can be predicted using the equations summarized in the Table

Prediction of peptide sequence ions			
From N-terminal (-NH <sub>2</sub> )		From C-terminal (-COOH)	
<b>b</b> <sub>1</sub>	$M_{\text{residue}} + H$	<b>y</b> <sub>1</sub>	$M_{\text{residue}} + 19$
<b>b</b> <sub>2</sub>	$b_1 + M_{\text{residue}}$	<b>y</b> <sub>2</sub>	$y_1 + M_{\text{residue}}$
<b>b</b> <sub>n-1</sub>	$[M+H]^+ - 18 - M_{\text{residue}}$	<b>y</b> <sub>n-1</sub>	$[M+H]^+ - M_{\text{residue}}$
<b>b</b> <sub>n</sub>	$[M+H]^+ - 18$	<b>y</b> <sub>n</sub>	$[M+H]^+$

**ESEMPIO:**

Qui di seguito viene illustrata l'analisi MS – MS di ioni prodotti da un peptide (**NFESGK**).

La massa molecolare del peptide è misurata mediante spettrometria standard trovando il valore **680.4 Da**.

Nello spettro il valore dominante è a  $m/z$  681.4 che corrisponde allo **ione protonato ( $M+H^+$ )**.

Gli ioni che danno origine a questo frammento sono stati separati dal primo analizzatore, fatti passare per la camera di frammentazione ed analizzati dal secondo analizzatore per produrre lo spettro MS – MS.

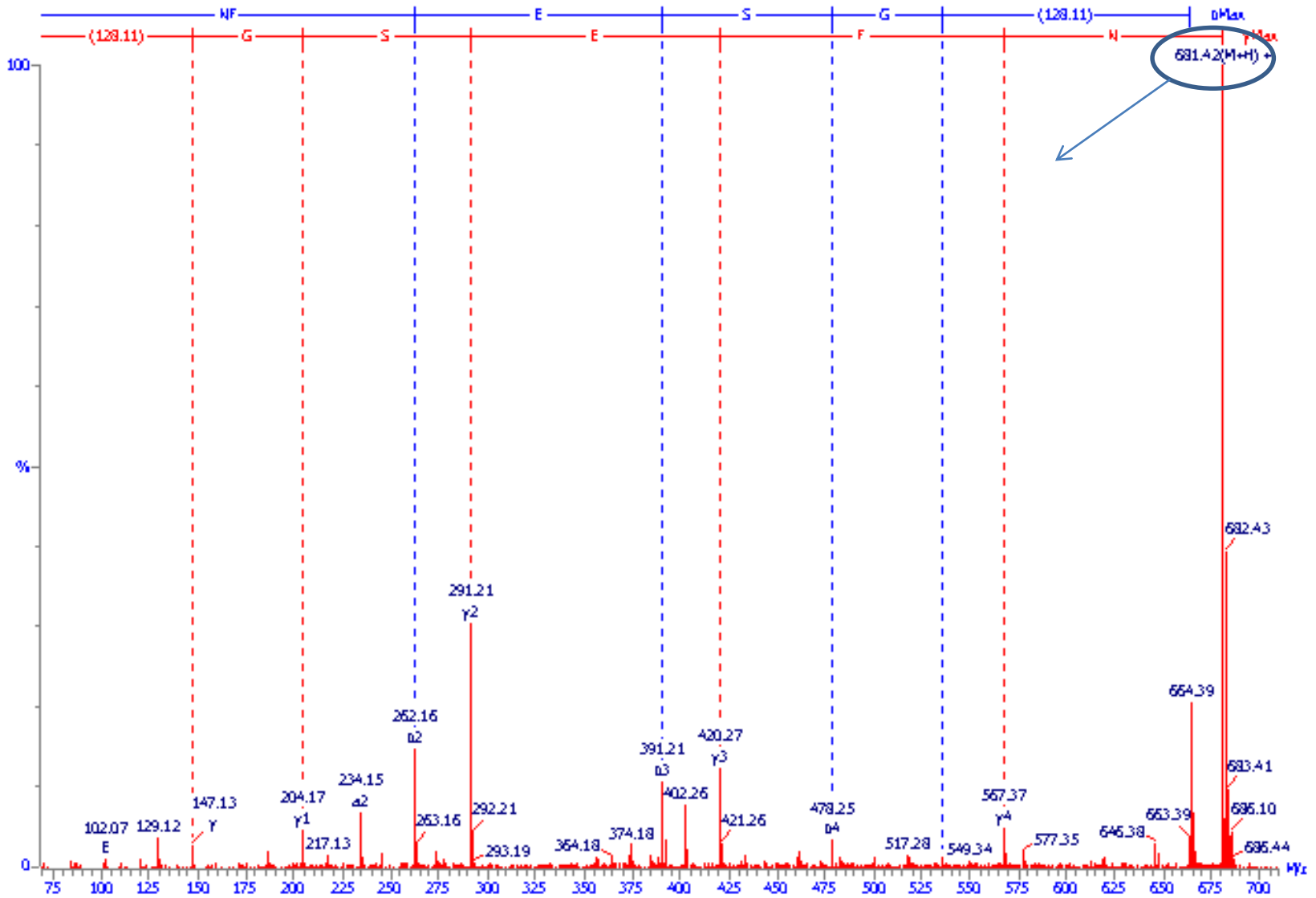
Nell'esempio il peptide frammenta soprattutto sul legame **CO – NH** e dà luogo a ioni di tipo **b** ed **y**.

1:1000; D of 681.3

snag09 AccMass 3 (Top,2, Ht,5000.0,0.00)

# NFESGK

1: TOF MSMS 681.30ES+



*Sequenziamento di un peptide mediante spettrometria di massa tandem.*

La serie “b” degli ioni è stata marcata con **linee blu** mentre quella “y” con **linee rosse**.

Può essere calcolata **la differenza di massa tra membri adiacenti della serie**;  
p. es.

$b_3 - b_2 = 391.21 - 262.16 = 129.05$  Da glutammina (E)

$y_4 - y_3 = 567.37 - 420.27 = 147.10$  Da fenilalanina (F).

In questo modo utilizzando la serie “b” o la serie “y” può essere ottenuta la sequenza che corrisponde a NFESGK

**Lo ione immonio a m/z 102** conferma solo la presenza della glutammina (E) nel peptide.

A volte può essere ottenuta l'intera sequenza della proteina; alcune proteine generano dati sufficienti per coprire il 70 - 80% dell'intera sequenza che è ragionevolmente sufficiente ad identificare la proteina.

Spesso una sequenza “marcatore” di 4 – 5 amminoacidi ottenuta da un singolo peptide di origine proteolitica è sufficiente ad identificare una proteina attraverso una banca dati.

Generalmente lo spettro MS dà informazioni sulle masse molecolari dei componenti del digerito proteico che vengono poi utilizzate per cercare la struttura primaria in una banca dati sulla base di alcuni dei picchi di massa ottenuti (protein database searching).

Se la ricerca nella banca dati non è sufficiente sia perchè la proteina non è stata catalogata, sia perchè non è già stata caratterizzata o perchè i dati non sono così accurati per discriminare tra diverse soluzioni proposte dalla banca dati, allora servono nuovi dati.

Questo può essere ottenuto purificando il campione ed utilizzando tecniche MS – MS di singoli peptidi provenienti dal digerito per tornare poi alla ricerca nella banca dati.

I dati di  $m/z$  di peptidi vengono utilizzati nella ricerca della struttura con metodi bioinformatici nei data base, p. es., **the Protein Identification Resource (PIR)**

- M. Mann, P. Højrup, P. Roepstorff, *Use of MS molecular weight information to identify proteins in sequence databases*, Biol. Mass Spectrom., 22 (1993) 338
- O.N. Jensen, A.V. Podtelejnikov, M. Mann, *Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching*, Anal. Chem., 69 (1997) 4741..



## Altri metodi bioinformatici

- MOWSE

D.D.J. Pappin, P. Højrup, A.J. Bleasby, *Rapid identification of proteins by PMF*, Curr. Biol., 3 (1993) 327

- ProFound

W. Zhang, B.T. Chait, *ProFound -an expert system for protein identification using MS PMF information*, Anal. Chem., 72 (2000) 2482

- MASCOT

D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, *Probability-based protein identification by searching sequence databases using MS data*, Electrophoresis, 20 (1999) 3551