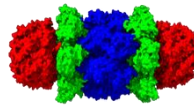


## Analiza globală a proteinelor sau proteomica



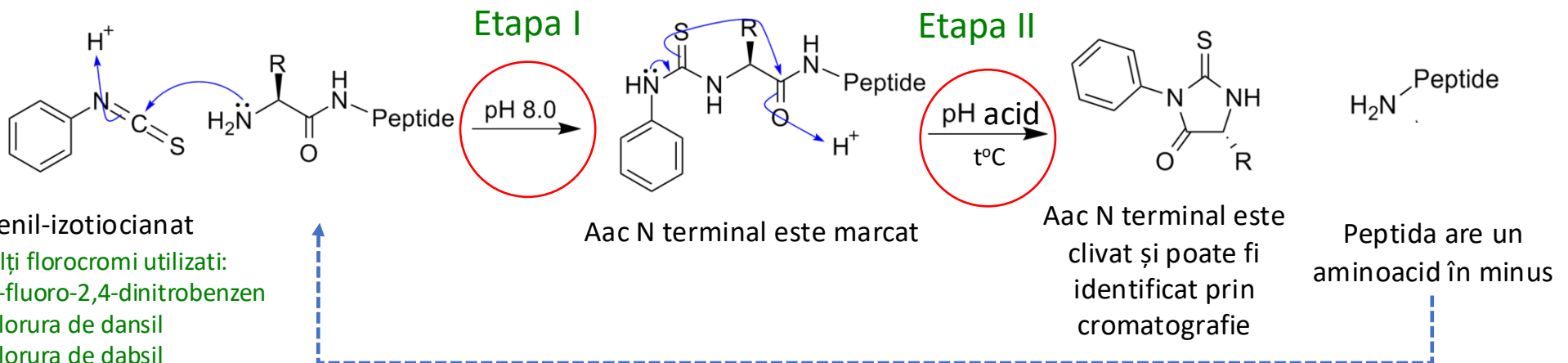
Multă vreme s-a considerat că stabilirea secvenței de aminoacizi a unei peptide ar fi imposibilă. În 1953 Frank Sanger reușește să secvențieze pentru prima dată în istorie o proteină – **insulina**. Abia după aproximativ 10 ani se stabilesc mecanismele din spatele procesului de transcriere și traducere a informației genetice și apare astfel posibilitatea de a stabili secvența de aminoacizi plecând de la o secvență de ADN.

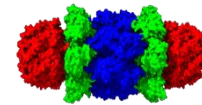
Corespunzător, **metodele de elucidare** a structurii primare a proteinelor se clasifică în **două categorii distincte**:

- 1. Metode experimentale** ce implică izolarea/purificarea proteinei de interes – degradarea Edman și spectrometria de masă;
- 2. Metode *in-silico***, bazate pe cunoașterea prealabilă a secvenței de nucleotide și realizarea cu ajutorul calculatorului a procesului de transcriere/traducere a informației genetice.

## 1. Metode experimentale de elucidare a secvenței proteinelor (peptidelor)

**A. Metoda Edman** sau **metoda degradării Edman** - are la bază reacția Edman ce constă în marcarea aminoacidului N terminal cu un cromofor specific, clivarea și identificarea lui.





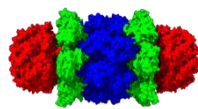
Prin repetarea reacției Edman pe peptida scurtată în prima reacție se poate stabili aminoacidul următor. Reacția Edman poate fi astfel folosită în mod secvențial, pentru a identifica unul câte unul fiecare aminoacid dintr-o secvență peptidică. **Randamentul reacției nu este însă 100%, astfel încât prin folosirea repetată a reacției Edman se pot stabili primi aproximativ 30-40 aminoacizi dintr-o proteină.**

În cazul proteinelor de dimensiuni mari, procesul de **secvențiere Edman** constă în parcurgerea **următoarelor etape:**

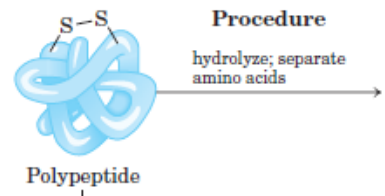
- 1. Stabilirea aminoacizilor** ce intră în alcătuirea proteinei precum și a raportului molar dintre aceștia – legăturile peptidice se hidrolizează cu HCL și aminoacizi rezultați sunt identificați prin HPLC sau cromatografie în strat subțire;
- 2. Stabilirea numărului de peptide** din structura proteinei – se realizează o reacție Edman folosind 1-fluoro-2,4-dinitrobenzen (FDNB) și se hidrolizează proteina cu HCL, numărul de aminoacizi derivați rezultați indică numărul de capete N-terminale existente;
- 3. Reducerea legăturilor S-S și clivarea cu agenți proteolitici cu specificitate cunoscută** - se generează peptide de dimensiuni convenabile **într-o manieră predictibilă;**

Reagent (biological source)*	Cleavage points <sup>†</sup>
Trypsin (bovine pancreas)	Lys, Arg (C)
<i>Submaxillaris</i> protease (mouse submaxillary gland)	Arg (C)
Chymotrypsin (bovine pancreas)	Phe, Trp, Tyr (C)
<i>Staphylococcus aureus</i> V8 protease (bacterium <i>S. aureus</i> )	Asp, Glu (C)
Asp-N-protease (bacterium <i>Pseudomonas fragi</i> )	Asp, Glu (N)
Pepsin (porcine stomach)	Leu, Phe, Trp, Tyr (N)
Endoproteinase Lys C (bacterium <i>Lysobacter enzymogenes</i> )	Lys (C)
Cyanogen bromide	Met (C)

# Degradarea Edman – privire de ansamblu



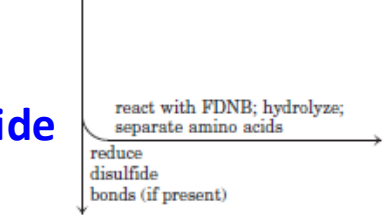
## Stabilirea aminoacizilor



Result					
A	5	H	2	R	1
C	2	I	3	S	2
D	4	K	2	T	1
E	2	L	2	V	1
F	1	M	2	Y	2
G	3	P	3		

**Conclusion**  
 Polypeptide has 38 amino acid residues. Trypsin will cleave three times (at one R (Arg) and two K (Lys)) to give four fragments. Cyanogen bromide will cleave at two M (Met) to give three fragments.

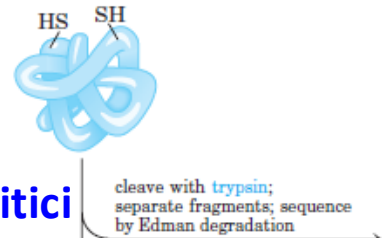
## Stabilirea numărului de peptide Reducerea legăturilor S-S



2,4-Dinitrophenylglutamate detected

E (Glu) is amino-terminal residue.

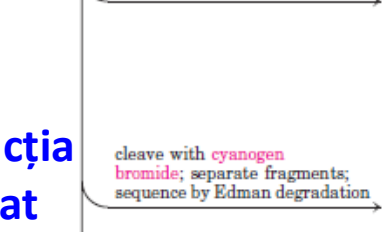
## Clivarea cu agenți proteolitici



- (T-1) GASMALIK
- (T-2) EGAAAYHDFEPIDPR
- (T-3) DCVHSD
- (T-4) YLIACGPMTK

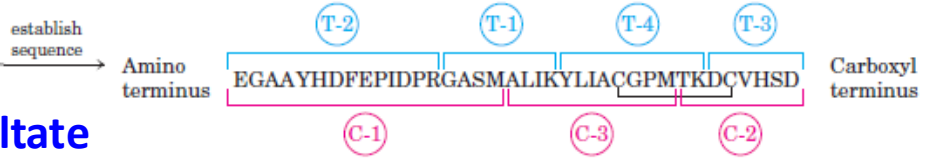
(T-2) placed at amino terminus because it begins with E (Glu).  
 (T-3) placed at carboxyl terminus because it does not end with R (Arg) or K (Lys).

## Secvențierea folosind reacția Edman cu fenil-izotiocianat

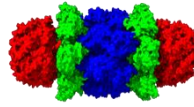


- (C-1) EGAAAYHDFEPIDPRGASM
- (C-2) TKDCVHSD
- (C-3) ALIKYLIACGPM

(C-3) overlaps with (T-1) and (T-4), allowing them to be ordered.



## Asamblarea secvenței peptidelor rezultate



Pentru a se putea stabili secvența unei proteine prin degradarea Edman, este esențial ca proteina de interes să fie izolată și purificată prin una din metodele precizate anterior. De asemenea, costul reactivilor necesari secvențierii este ridicat și din acest motiv, degradarea Edman nu s-a impus în lumea științifică. Dezvoltarea și ieftinirea tehnicilor de secvențiere ADN au permis apariția unei alternative foarte convenabile: traducerea *in-silico* (computerizată) a secvențelor de ADN în proteine.

Totuși, majoritatea modificărilor post-traducere ale proteinelor nu sunt codificate direct în secvența de ADN, pentru identificarea lor fiind necesară analiza directă a acestor molecule. Acest lucru este astăzi posibil prin intermediul unei tehnici ce își are originile în fizică și chimie: **spectrometria de masă**. Tehnica și-a găsit atât de numeroase aplicații în studiul proteinelor, încât a contribuit direct la apariția și dezvoltarea unui domeniu nou: **proteomica**.

Cuvântul **proteomică** este o preluare a termenului **proteomics** din limba engleză. La origine, termenul provine din cuvântul **proteom**, un termen compus prin unirea a 2 cuvinte:

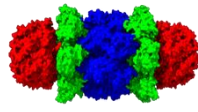
- **protein** – proteină;
- **genome** – materialul genetic al unui organism.

Termenul de **proteom** a fost introdus în 1994 de către Mark Wilkins pentru a denumi, prin analogie cu termenul genom, **întregul set de proteine care este sau poate fi exprimat de o celulă, țesut, organism**.

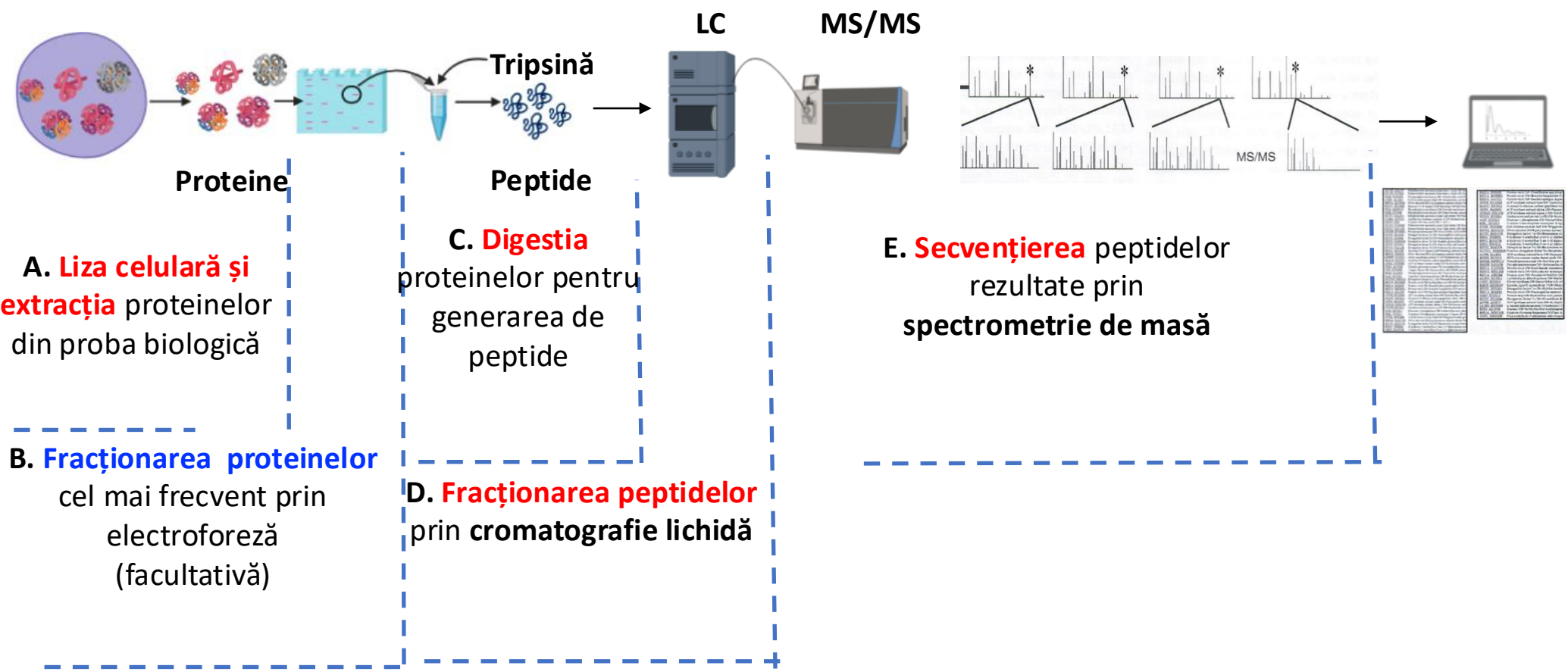
Termenul sau prescurtarea de **omics** – reprezintă un neologism al limbii engleze ce se referă la **un set de metode sau tehnici folosite pentru a studia simultan setul complet de molecule de un anumit tip (proteine, lipide, metaboliți) dintr-o probă**. Un asemenea studiu/metodă ce își propune să caracterizeze întregul set de molecule de un anumit tip mai poartă numele de **studiu/metodă extensiv(ă)** sau **studiu/metodă la scară mare (large-scale study)**, pentru a-l diferenția de studiile obișnuite din punct de vedere al:

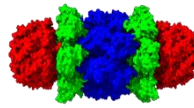
**Proteomics** sau termenul importat în limba română - **proteomică** - se referă așadar la **studiul simultan al întregului set de proteine produse la un moment dat de o celulă, țesut sau organism**.

# Ce presupune un studiu de proteomică?

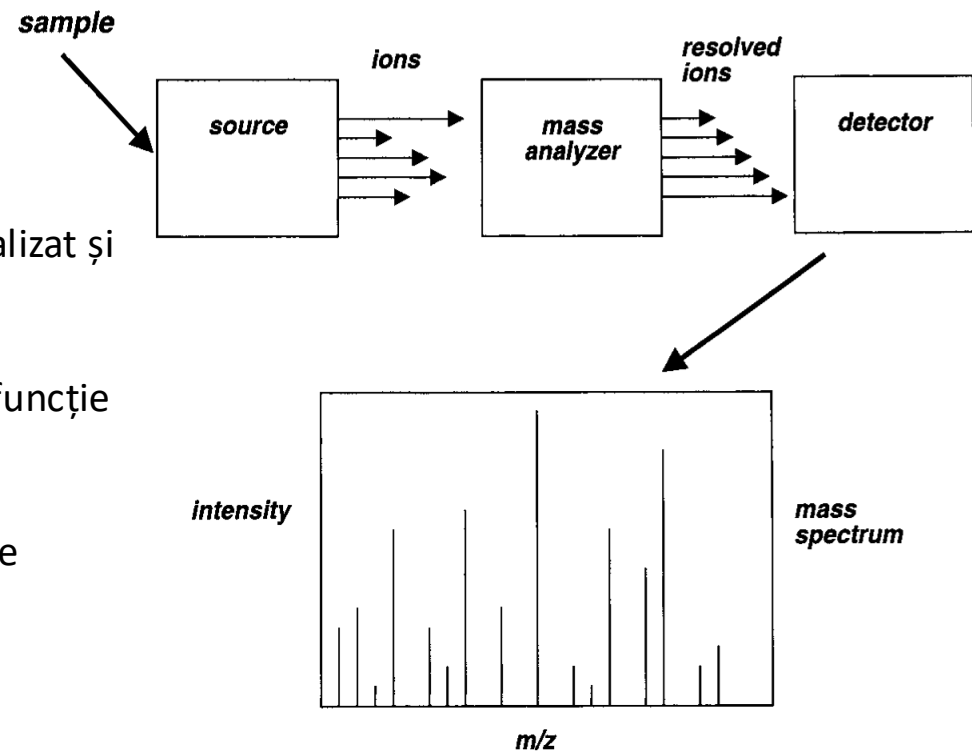


Pentru studiul simultan al întregului set de proteine produse la un moment dat de o celulă, țesut sau organism au apărut diverse metode de proteomică, dar care au în comun parcurgerea următoarelor etape majore:

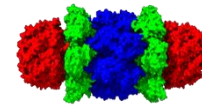




Elementul cheie în întreaga abordare îl reprezintă utilizarea spectrometriei de masă pentru a caracteriza peptidele și, apoi, proteinele dintr-o probă. **Spectrometria de masă este o tehnică analitică prin care este măsurat raportul masă-sarcină ( $m/z$ ) al moleculelor încărcate electric.** Raportul  $m/z$  poate fi mai apoi utilizat pentru a identifica cu precizia masa moleculară ( $M_w$ - molecular weight) a moleculelor neutre. Rezultatul unei analize de spectrometrie de masă este **un spectru de masă – un grafic ce are pe axa OX valoarea  $m/z$  în Da a ionilor analizați, iar pe OY intensitatea semnalului corespunzătoare ionului analizat.** Instrumentul ce înregistrează spectrul de masă poartă numele de **spectrometru de masă** și este alcătuit din 3 componente principale:



- O **sursă de ioni** – asigură trecerea probei în fază gazoasă, ionizarea moleculelor din proba de analizat și transferul ionilor către
- Un **analizor de masă** – în care ionii sunt separați funcție de  $m/z$  și sunt transferați către
- Un **detector** sensibil la prezența ionilor separați de analizor care trimite semnale către un computer ce înregistrează **spectrul de masă**.



## A. Surse de ioni utilizabile în proteomică

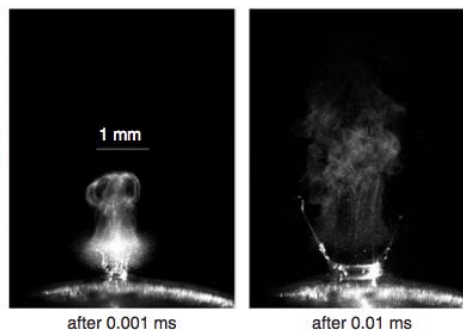
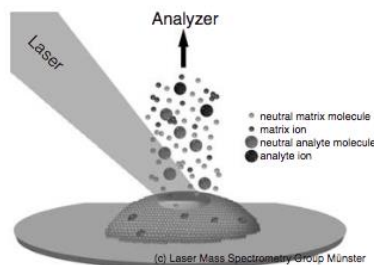
Moleculele neutre pot fi convertite în ioni prin diverse mecanisme chimice sau fizice.. Pentru a putea fi preluate de analizorul de masă, ionii generați în sursă trebuie să fie în stare gazoasă. În cazul moleculelor biologice, metodele de ionizare trebuie să fie foarte eficiente, dar să nu distrugă/fragmenteze analitul. Din aceste motive, în studiile de proteomică s-au impus preponderent două tipuri de ionizare:

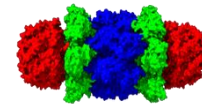
**1. MALDI - Matrix assisted LASER desorption ionization - Proba de analizat amplasată într-un matrice cu caracter acid este supusă unor pulsuri laser repetate de 50-200 ori.** Pulsurile au loc în general în vacuum și generează explozii de mici dimensiuni ale probei de analizat, peptidele și proteinele ionizând norul ce apare în urma exploziei. Deoarece mediul este acid, **ionii ce se formează sunt preponderent pozitivi.** Matricea utilizată în MALDI are rolul de a genera protoni, dar de asemenea și de a prelua majoritar energia pulsurilor laser. Moleculele din matrice ionizează și ele, însă masa lor moleculară (100-200Da) este semnificativ mai mică decât cea a peptidelor sau proteinelor (peste 1000 Da), fiind ușor de ignorat în analizorul de masă.

### Compuși frecvent utilizați ca matrice în MALDI

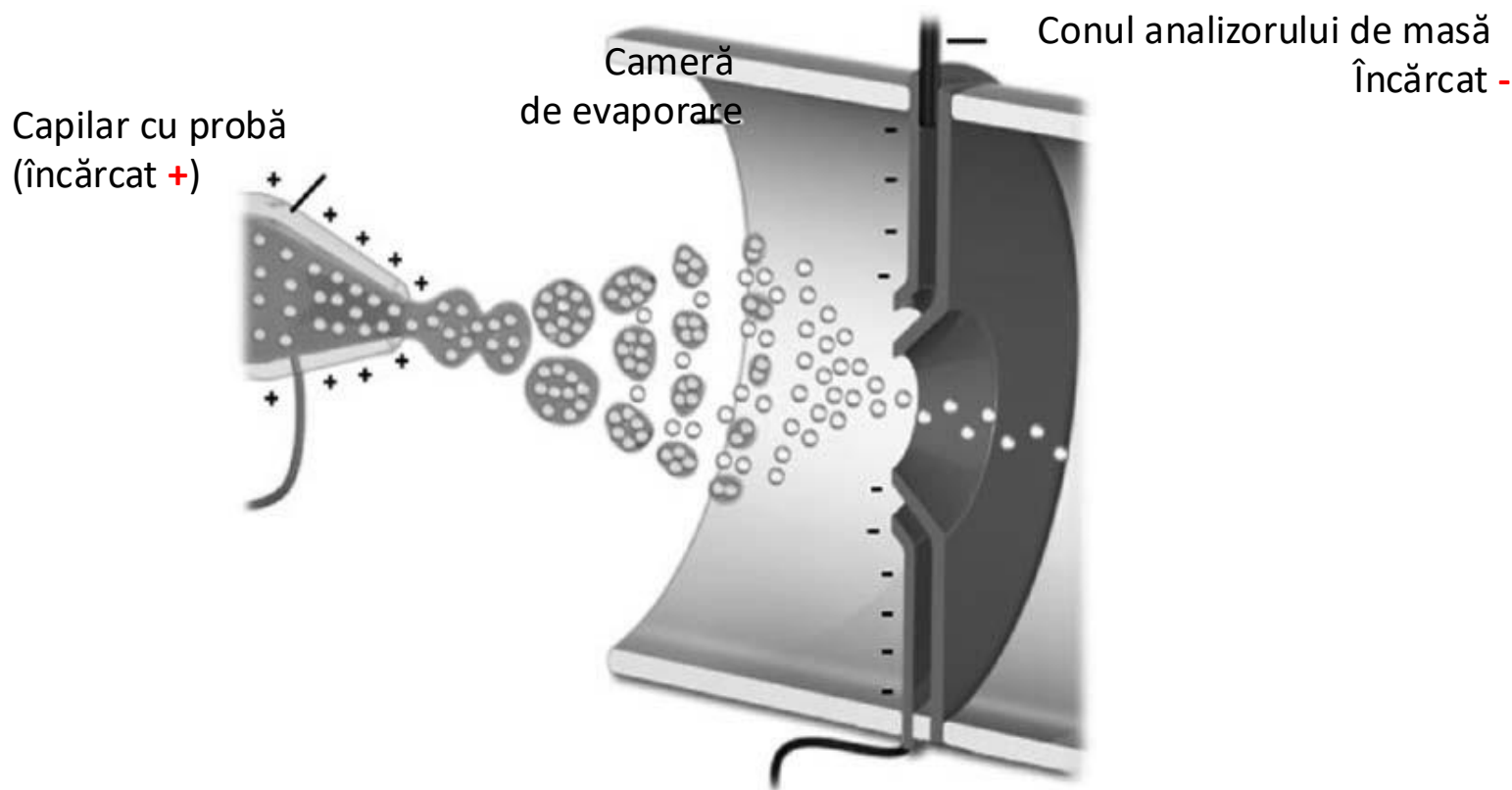
Matrix	Structure	Wavelength	Major applications
Nicotinic acid		UV 266 nm	Proteins, peptides, adduct formation
2,5-Dihydroxybenzoic acid (plus 10% 2-hydroxy-5-methoxybenzoic acid)		UV 337 nm, 353 nm	Proteins, peptides, carbohydrates, synthetic polymers
Sinapinic acid		UV 337 nm, 353 nm	Proteins, peptides
α-Cyano-4-hydroxycinnamic acid		UV 337 nm, 353 nm	Peptides, fragmentation
3-Hydroxy-picolinic acid		UV 337 nm, 353 nm	Best for nucleic acids
6-Aza-2-thiothymine		UV 337 nm, 353 nm	Proteins, peptides, non-covalent complexes; near-neutral pH
k,m,n-Di(tri)hydroxy-acetophenone		UV 337 nm, 353 nm	Protein, peptides, non-covalent complexes; near-neutral pH
Succinic acid	$\text{HOOC}-\text{CH}_2-\text{CH}_2-\text{COOH}$	IR 2.94 μm, 2.79 μm	Proteins, peptides
Glycerol	$\text{H}_2\text{C}-\text{CH}-\text{CH}_2$ $\text{OH} \quad \text{OH} \quad \text{OH}$	IR 2.94 μm, 2.79 μm	Proteins, peptides, liquid matrix

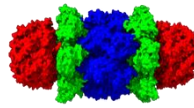
IR = infrared; UV = ultraviolet.





2. **ESI** - **electrospray ionization** – dacă în MALDI proba de analizat trebuie să fie solidă, în ESI aceasta este în soluție. Proba de analizat dizolvată într-un solvent volatil și acid este pulverizată foarte fin printr-un capilar într-o cameră de evaporare ce comunică cu conul analizorului de masă. Solventul este treptat eliminat din picăturile formate în camera de evaporare, analitul devenind astfel gaz. Prin evaporarea solventului acid se generează un exces de protoni, ceea ce face ca analitul să ionizeze. Între capilarul ce pulverizează proba și conul analizorului de masă există o diferență mare de voltaj ce direcționează ionii către conul detectorului de masă.



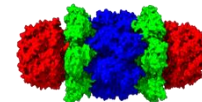


În cazul peptidelor, sarcina lor netă este variabilă, astfel:

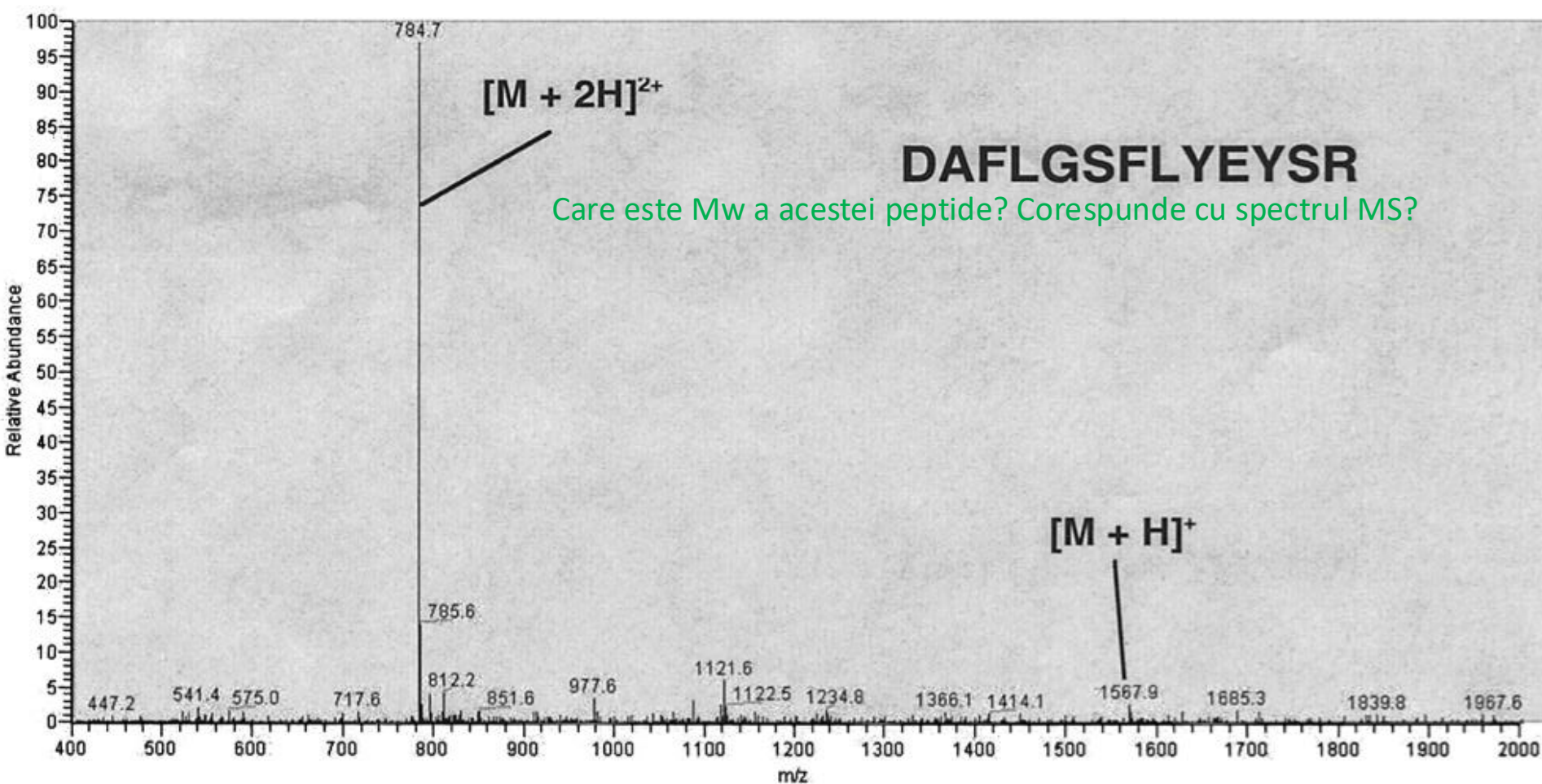
1. În soluție, **gradul de ionizare al peptidelor depinde de pH**:
  - catenele laterale ale aminoacizilor dicarboxilici sunt ne-ionizate la valori de pH mai mici de 3.0 și ionizate (încărcate negativ) la valori de pH mai mari de 7.0
  - capătul N terminal și aminoacizii di-aminici sunt ionizați la pH mai mic de 8.5
  - funcție de valoare de pH a mediului, peptidele pot fi așadar încărcate pozitiv ( $\text{pH} < 3.5$ ) sau negativ (la valori bazice de pH). De cele mai multe ori ionizarea peptidelor se face în mediu acid, ionii rezultând prin protonarea resturilor  $\text{NH}_2$  și având astfel sarcină pozitivă.
2. În **ESI peptidele nu ionizează uniform**. Orice peptidă obținută prin utilizarea tripsinei în faza de hidroliză va avea cel puțin 2 radicali ionizabili: capătul N terminal și restul de Lys/Arg unde a avut loc hidroliza la capătul C-terminal. Prin ESI, din peptida va produce așadar cel puțin 3 tipuri de molecule:
  - **peptida neionizată** – nu poate fi preluată de analizorul de masă și deci nu va apărea în spectrul de masă;
  - **peptida ionizată la un capăt** – are o sarcină pozitivă, va apărea în spectrul de masă cu  $m/z = M_w + 1A_H / 1 = M_w + 1 \text{ Da}$
  - **peptida ionizată la ambele capete** – are două sarcini pozitive, va apărea în spectrul de masă cu  $m/z = M_w + 2A_H / 2 \text{ Da}$ .

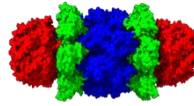
**Și proteinele native pot fi ionizate prin ESI. Proteinele** prin catenele laterale ale aminoacizilor expuși la exterior pot accepta 10-30 protoni. Similar cu o peptidă, și **o proteină va genera o populație de ioni cu  $m/z$  cuprins între  $m/10$ ..... $m/30$ .**

**Distribuția populației de ioni proveniți din aceeași moleculă dar având cu  $m/z$  diferite poartă numele de anvelopă multi-sarcină.** În cazul **peptidelor, anvelopa multi-sarcină conține 2 semnale** în cele mai multe cazuri, **3 semnale dacă în secvență există aminoacizi bazici.** Anvelopa multi-sarcină a proteinelor conține **numeroase semnale** și poate fi utilizată pentru identificarea masei moleculare a proteinei de interes printr-un proces automatizat de deconvoluție a semnalelor.

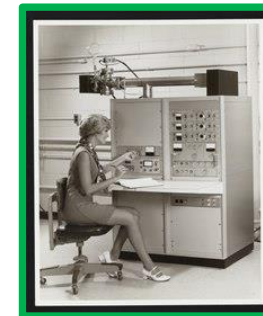
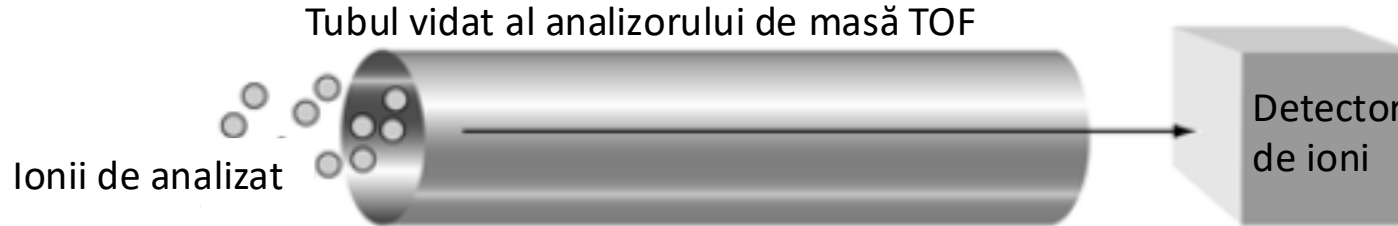


Spectrul ESI-MS al peptidei cu secvența indicată. Cele două semnale corespunzătoare celor două stări de ionizare sunt evidențiate



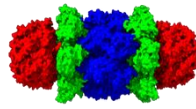


1. **Analizor de masă TOF** – time of flight – ionii generați de sursă sunt transferați într-un tub vidat și accelerați printr-un tub cu ajutorul unui câmp electric constant. Viteza de deplasare a ionilor este dependentă de sarcina acestora – cu cât sarcina este mai mare, cu atât au fost accelerați mai puternic și deci viteza de deplasare va fi mai mare. Pentru ionii cu sarcini egale, viteza de deplasare depinde doar de masa ionilor – ionii mai grei vor avea o viteză mai mică, iar cei mai ușori o viteză mai mare. **Analizorul de masă separă ionii pe baza timpului necesar ca aceștia să parcurgă în zbor o distanță prestabilită - lungimea tubului vidat al detectorului TOF.**

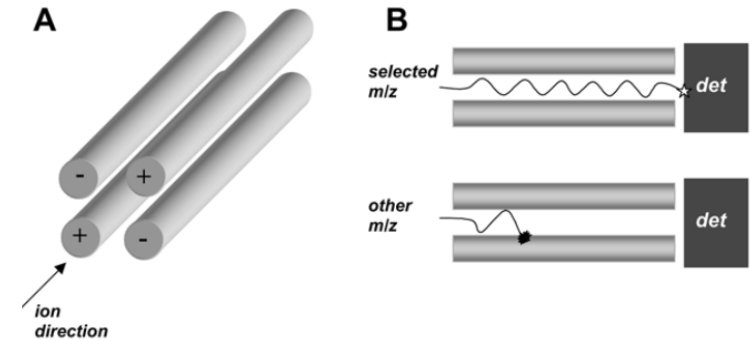


Bendix MA-2  
Time-of-Flight  
Mass  
Spectrometer,  
1960s





2. **Analizor de masă quadrupol** — conține 4 bare metalice dispuse în paralel ce alcătuiesc 4 electrozi pe care se aplică, controlat, în mod repetat, cu o anumită frecvență, curenți de voltaje variabile. În acest fel se generează un câmp magnetic ce forțează ionii să se deplaseze între cele 4 bare pe o traiectorie în spirală spre capătul opus celui în care au intrat în quadrupol. Prin controlul foarte strict al voltajului aplicat și al frecvențelor cu care



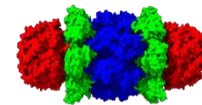
acesta se schimbă, doar ionii de un anumit  $m/z$  se vor deplasa spre ieșirea din quadrupol unde este amplasat detectorul. Prin modificarea controlată periodică a voltajului aplicat (**scanare**) și sincronizarea acestor modificări cu informațiile de la detector se poate identifica prezența de ioni cu valori  $m/z$  diferite.

În proteomică analizorul quadrupol de sine stătător este mai puțin utilizat. Cel mai frecvent 3 asemenea analizoare sunt grupate și funcționează sincronizat (**in tandem**) pentru a forma un spectrometru de masă numit **triplu-quad**.

Într-un triplu-quad, cele **3 detectoare au funcții diferite**:

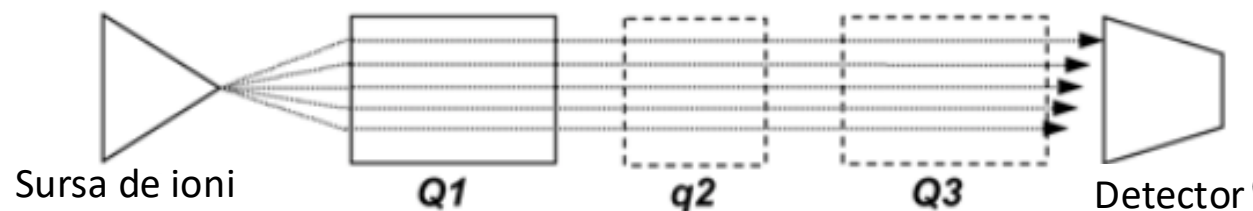
- două dintre detectoarele quadrupol notate **Q1** și **Q3** ce funcționează după principiile descrise mai sus, separând ionii funcție de  $m/z$  – **sunt realmente analizoare de masă**;
- un analizor quadrupol notat **q2** ce funcționează ca o cameră în care ionii, indiferent de  $m/z$ , sunt reținuți pentru o perioadă de timp și pot fi eliberați simultan prin modificarea tensiunii aplicate. Frecvent, acest quadrupol este numit **celulă de coliziune** deoarece aici ionii sunt bombardați cu un gaz neutru precum  $N_2$ , He sau Ar. Impactul dintre moleculele gazului și cele ale ionilor duce la fragmentarea ionilor printr-un proces numit **disociere indusă prin coliziune** – **collision-induced dissociation, CID**.

**Într-un triplu-quad, cele 3 detectoare quadrupol sunt dispuse unul după celălalt, ionii parcurgându-i în ordinea Q1, q2, Q3.**

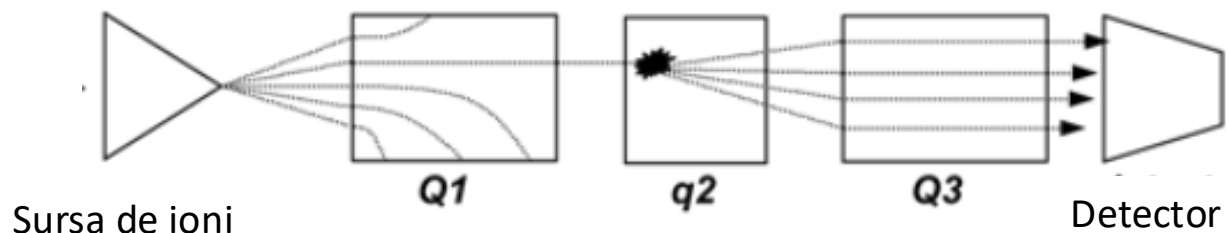


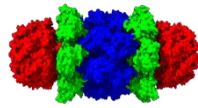
Un detector **triplu-quad** poate funcționa în **două moduri**:

1. **Scanare completă (full-scan, MS<sup>1</sup>)** – ionii ce provin de la sursă sunt separați în **Q1** ce **funcționează în acest caz ca un analizor de masă în adevăratul sens al cuvântului**. Ionii separați trec fără să interacționeze cu q2 și Q3 pentru a ajunge la detector. Acest mod este implementat în general timp de 1 s, timp în care spectrometrul de masă identifică toți ionii ce pătrund în Q1. **Aplicat în cazul peptidelor – toate peptidele ce sunt livrate spectrometrului de masă timp de 1s de către un sistem nanoHPLC prin intermediul unei interfețe ESI.**



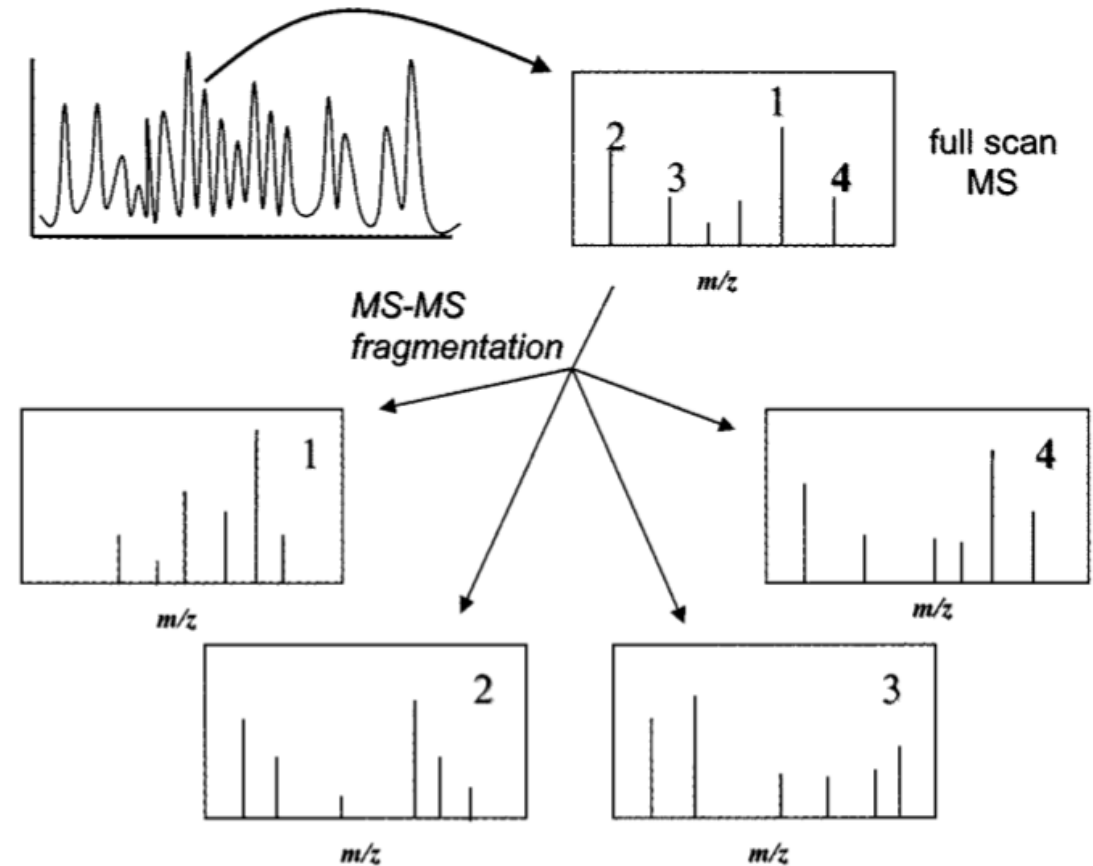
2. **Mod în tandem** sau **MS-MS**, uneori notat și **MS/MS**, sau **MS<sup>2</sup>** – toți cei 3 cuadrupoli funcționează sincronizat. În acest caz **Q1** joacă rol de **filtru de masă**, permițând doar ionilor cu un anumit  $m/z$  să treacă în **q2**. **Ionii ce ajung în camera de coliziune poartă numele de ioni precursori și sunt fragmentați, iar fragmentele rezultate sunt analizate pe bază de  $m/z$  în Q3**. În acest mod Q3 joacă rolul de analizor de masă, însă un separator ionii din proba de analizat, ci fragmentele rezultate prin disocierea acestora. Valorile  $m/z$  ale fragmentelor generate împreună cu  $m/z$  ionului precursor sunt mai apoi utilizate pentru a stabili natura/structura acestuia. **Acesta este modul de funcționare ce permite fragmentarea și secvențiere a peptidelor. În proteomică, un triplu-quad este folosit alternativ, pentru a realiza o scanare completă și identificarea peptidelor din probă, și mai apoi în modul MS/MS pentru a determina secvența acestora din fragmentele generate.**

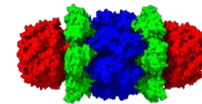




Analizoarele de masă capabile să realizeze analize **MS în tandem** pot fi deosebit de utile pentru analiza în masă a unui foarte mare de peptide deoarece pot fi programate pentru realiza în mod automat trecerea de la un mod de măsurare la altul astfel:

- A.** Instrumentul funcționează în modul scanare completă și identifică ce peptide există în proba furnizată de sursă;
- B.** Peptidele cele mai abundente sunt selectate pe rând și fragmentate prin CID; spectrele MS/MS sunt înregistrate pentru fiecare peptidă în parte
- C.** Instrumentul revine la modul scanare completă și ciclul se reia.





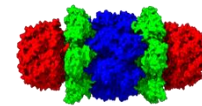
## Amprentarea masică (peptide mass fingerprinting, PMF)

Această metodă de identificare pleacă de la ideea că, având la dispoziție genomul complet al unui organism s-ar putea în principiu genera cu ajutorul calculatorului o listă cu toate proteinele din organismul respectiv. Cunoscând modul în care acționează proteazele (**tripsina hidrolizează legătura peptidică la un rest de K sau R, dacă nu este urmat de P**) tot cu ajutorul calculatorului s-ar putea genera o listă cu toate peptidele posibil a rezulta prin digestia fiecărei proteine ce alcătuiește proteomul speciei respective. Pentru peptidele teoretice generate s-ar cunoaște originea (**proteina din care provine**), lungimea, secvența și deci i s-ar putea calcula masa. În acest fel ar fi posibilă realizarea unor **liste extinse cu masele teoretice ale peptidelor obținute prin digestii virtuale, fiecărei peptide fiindu-i asociată proteina din care provine**. Unele dintre **aceste peptide**, cu precădere cele de peste 6 aminoacizi **au mase unice**, ceea ce înseamnă că **fiecărei proteine îi sunt asociate un număr variabil de valori masice unice – o amprentă masică a proteinei**, realizate pe baza maselor peptidelor.

Prin procesarea în laborator a unei probe provenite din organismul de interes, proteinele existente în probă vor fi hidrolizate, iar peptidele generate analizate cel mai frecvent print-un spectrometru de masă **MALDI-TOF (sursă de tip MALDI cuplată cu analizor de masa TOF)**. Instrumentul va identifica valorile  $m/z$  pentru fiecare peptidă din probă pe baza cărora se poate **calcula masa reală a peptidelor existente** în proba de analizat. Dacă **masa reală măsurată este unică** și se regăsește în lista **extinsă cu masele teoretice ale peptidelor obținute prin digestii virtuale, atunci este foarte probabil ca peptida reală să fie identică cu cea teoretică**, indicând așadar prezența proteinei parentale în probă. Cu cât se identifică mai multe peptide reale cu corespund ca masă cu peptidele virtuale aparținând aceleiași proteine, cu atât nivelul de încredere că proteina respectivă este în probă este mai mare.

### Peptide Mass list

.	
.	
.	
1528.7685	
1528.7989	
1529.0002	
1529.2001	
1529.2454	
1529.5006	
1529.6997	
<b>1529.7348</b>	← <b>1529.7348</b> <i>measured mass</i>
1529.9978	
1530.2332	
1530.4567	
1531.0003	
1531.0107	
1531.3004	
1531.5656	
.	
.	
.	



Prin identificare potrivirilor dintre masele peptidelor măsurate cu ajutorul unui spectrometru de masă și masele peptidelor teoretice posibil a exista în organismul respectiv se pot astfel identifica proteinele existente într-o probă necunoscută. **Succesul amprentării masice pentru identificarea proteinelor depinde de 2 factori cheie:**

- **Precizia** cu care se măsoară  $m/z$ -urile și deci cu care se calculează masele peptidelor reale. **Din acest motiv instrumentele utilizate sunt cel mai frecvent MALDI-TOF;**
- **Acuratețea** datelor genomice pe baza cărora se creează lista de mase teoretice.

Limitările principale în aplicarea metodelor de amprentare masică se referă la:

1. deși bazele de date cu secvențe sunt deosebit de mari, nu conțin decât rar genomul complet și anotat (poziția genelor este descrisă, codonii start și stop sunt identificați) al unei specii;
2. în organisme superioare există numeroase proteine omologe, foarte similare ca sevă, al căror amprentă masică este foarte similară sau chiar identică;
3. bazele de date cu secvențe conțin rar date despre modificările post-traducere ale proteinelor. În cazul unei proteine ce suferă PTM, amprenta masică teoretică este semnificativ diferită de cea reală și nu mai poate fi identificată prin această metodă.

