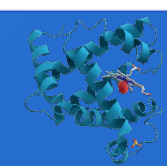


**13 - 27.04.2024**

## **Curs 4 – Metode de elucidare a structurii proteinelor**

# Elucidarea structurii primare a proteinelor



Multă vreme s-a considerat că stabilirea secvenței de aminoacizi a unei peptide ar fi imposibilă. În 1953 Frank Sanger reușește să secvențeze 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.

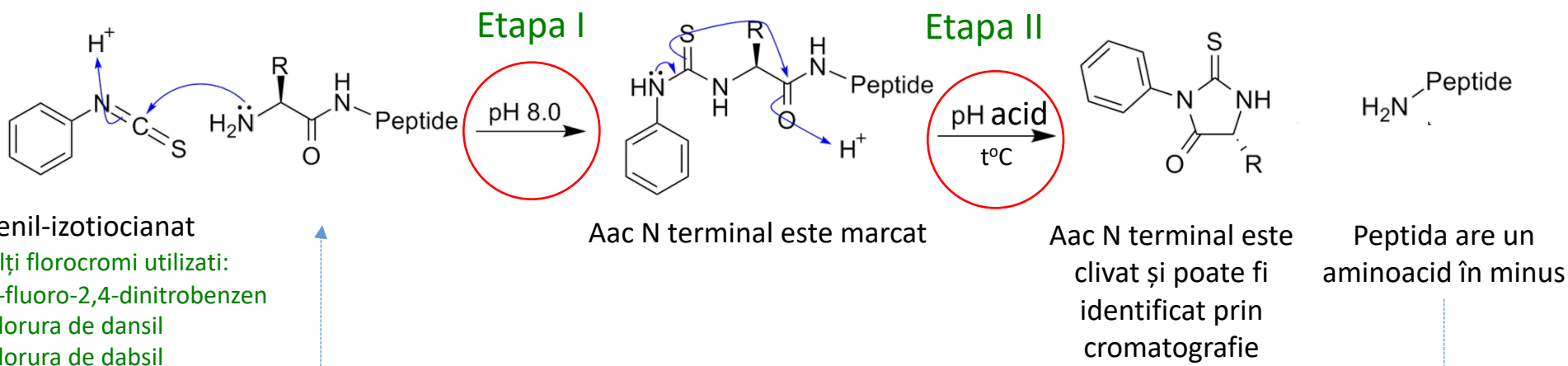
Corespunzător, **metodele de elucidare** a structurii primare a proteinelor se clasifică în **două categorii distincte**:

**1. Metode experimentale** ce implică izolarea/purificarea proteinei de interes – degradarea Edman și spectrometria de masă;

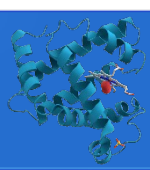
**2. Metode in-silico**, bazate pe cunoașterea prealabilă a secvenței de nucleotide și realizarea cu ajutorul calculatorului a procesului de transcriere/traducere a informației genetice.

## 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.



# Elucidarea structurii primare a proteinelor



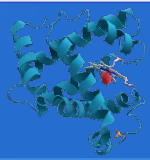
Prin repetarea reacției Edman pe peptida scurtată în prima reacție se poate stabili aminoacidul următor. Reacția Edman poate fi astfel folosită în mod secvențial, pentru a identifica unul câte unul fiecare aminoacid dintr-o secvență peptidică. **Randamentul reacției nu este însă 100%, astfel încât prin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

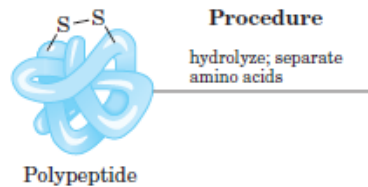
- 1. Stabilirea aminoacizilor** ce intră în alcătuirea proteinei precum și a raportului molar dintre aceștia – legăturile peptidice se hidrolizează cu HCL și aminoacizi rezultați sunt identificați prin HPLC sau cromatografie în strat subțire;
- 2. Stabilirea numărului de peptide** din structura proteinei – 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;**
- 4. Secvențierea folosind reacția Edman cu fenil-izotiocianat** a tuturor peptidelor obținute;
- 5. Asamblarea secvenței peptidelor rezultate** pentru a obține secvența proteinei inițiale.

TABLE 3-7	The Specificity of Some Common Methods for Fragmenting Polypeptide Chains
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)

# Elucidarea structurii primare a proteinelor



## Stabilirea aminoacizilor



### Procedure

hydrolyze; separate amino acids

### 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.

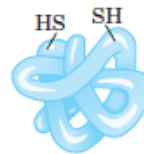
## Stabilirea numărului de peptide

## Reducerea legăturilor S-S

react with FDNB; hydrolyze; separate amino acids  
reduce disulfide bonds (if present)

2,4-Dinitrophenylglutamate detected

E (Glu) is amino-terminal residue.



## Clivarea cu agenți proteolitici

cleave with trypsin; separate fragments; sequence by Edman degradation

- (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

cleave with cyanogen bromide; separate fragments; sequence by Edman degradation

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

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

establish sequence

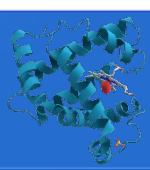
Amino terminus



Carboxyl terminus

## Asamblarea secvenței peptidelor rezultate

# Elucidarea structurii primare a proteinelor



## B. Spectrometria de masă.

Spectrometria de masă este o tehnică analitică ce se bazează pe ionizarea probei de interes și separarea ionilor funcție de raportul masă/sarcină ( $m/z$ ). Instrumentul folosit poartă numele general de **spectrometru de masă** (*mass-spectrometer*, MS).

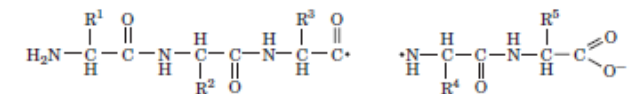
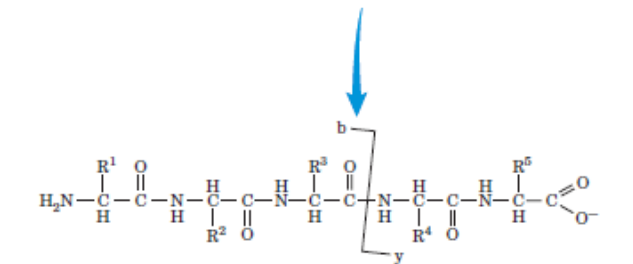
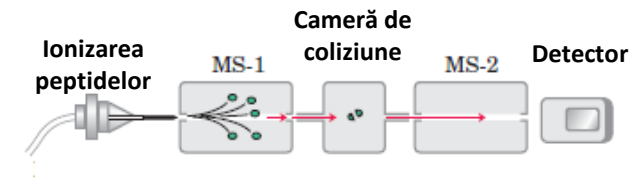
Ca și în cazul degradării Edman, proteinele de dimensiuni mari sunt inițial reduse și clivate cu agenți proteolitici ce au specificitate cunoscută.

Instrumentul utilizat pentru a secvenția proteine conține 2 detectoare de masă ce funcționează în tandem și care sunt conectate printr-o cameră de coliziune – **spectrometru de masă în tandem (MS/MS)**:

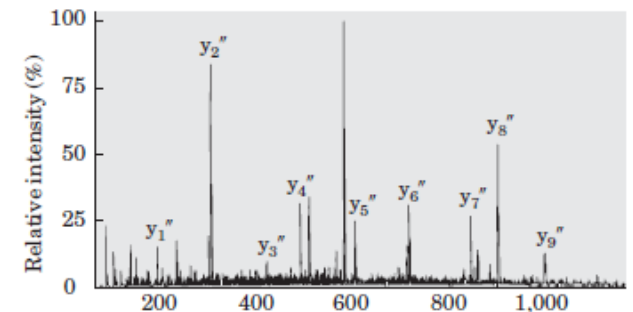
-**primul spectrometru de masă MS-1** are rolul de a separa peptidele de interes funcție de  $m/z$  și de a le livra către camera de coliziune;

-în **camera de coliziune** peptidele ionizate sunt bombardate cu un gaz neutru (He, N sau Ar). Are loc astfel **scindarea indusă prin coliziune a legăturilor peptidice**. Această scindare este una predictibilă și reproductibilă și duce la formarea a două serii de ioni (**b și y**) ce sunt transferați către al doilea detector de masă;

-**al doilea spectrometru de masă MS-2** separă cele 2 serii de ioni și pe baza diferențelor de  $m/z$  se stabilește secvența peptidei din care provin. Procesul de scindare și separare a ionilor formați este apoi reluat pentru o altă peptidă ce a fost 'păstrată' în MS1.

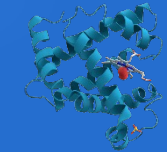


(a) Scindarea legăturii peptidice indusă prin coliziune



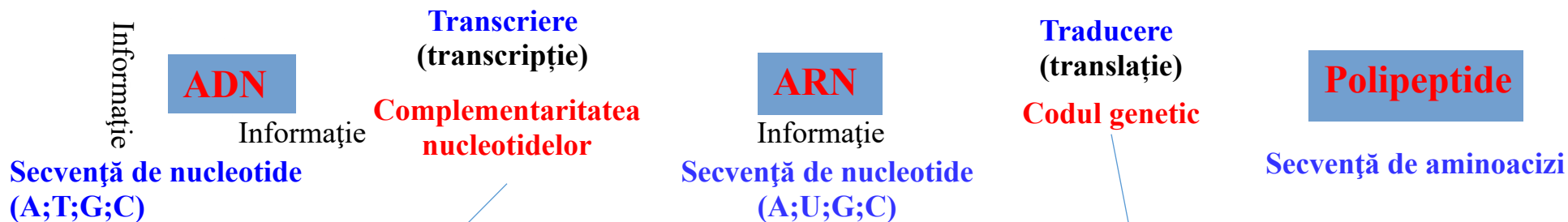
(b) Spectrul MS al unei serii de ioni y

# Elucidarea structurii primare a proteinelor



## 2. Metode *in-silico* stabilire a secvenței proteinelor (peptidelor)

Dogma centrală a biologie moleculare postulează că **fiecare tripletă de nucleotide (codon) din ADN codifică câte un aminoacid** (exceptând codonii STOP), corespondența codon-aac fiind dată de **codul genetic**. În principiu, dacă se cunoaște secvența unei gene, stabilirea secvenței de aminoacizi codificată este o chestiune simplă – se găsește corespondența aminoacizilor pe baza codonilor din molecula de ARNm generată.



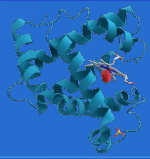
Nucleic Acid	Nucleobases	Base complement
DNA	adenine(A), thymine(T), guanine(G), cytosine(C)	A=T, G≡C
RNA	adenine(A), uracil(U), guanine(G), cytosine(C)	A=U, G≡C

1st base	2nd base						3rd base				
	U		C		A			G			
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U		
	UUC	(Leu/L) Leucine	UCC		UAC	(His/H) Histidine	UGC	(Arg/R) Arginine	C		
	UUA		UCA		UAA <sup>[B]</sup>		Stop (Ochre)		UGA <sup>[B]</sup>	Stop (Opal)	A
	UUG		UCG		UAG <sup>[B]</sup>		Stop (Amber)		UGG	(Trp/W) Tryptophan	G
C	CUU		(Ile/I) Isoleucine	CCU	(Pro/P) Proline		CAU		(Asn/N) Asparagine	CGU	(Ser/S) Serine
	CUC	CCC		CAC		CGC	C				
	CUA	CCA		CAA		CGA	A				
	CUG	CCG		CAG		CGG	G				
A	AUU	(Val/V) Valine	ACU	(Thr/T) Threonine	AAU	(Lys/K) Lysine	AGU	(Gly/G) Glycine	U		
	AUC		ACC		AAC		AGC		C		
	AUA		ACA		AAA		AGA		A		
	AUG <sup>[A]</sup>		ACG		AAG		AGG		G		
G	GUU	(Ala/A) Alanine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic acid	GGU	(Gly/G) Glycine	U		
	GUC		GCC		GAC		GGC		C		
	GUA		GCA		GAA		GGA		A		
	GUG		GCG		GAG		GGG		G		

Proprietățile codului genetic?



# Elucidarea structurii primare a proteinelor



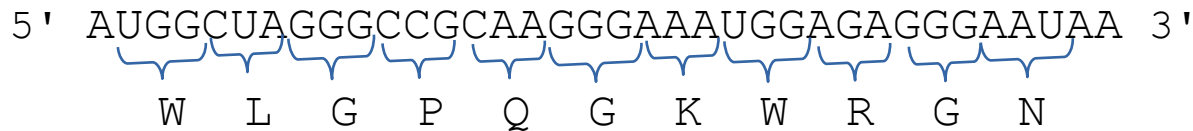
## 2. Care este prima nucleotidă a primului codon – cadrul de lectură – de unde începe traducerea informației genetice?

Pe molecula de ARNm (și corespunzător catena sens) există 3 posibilități de a citi informația genetică – 3 cadre de lectură diferite:

### Incepând cu nucleotida 1:



### Incepând cu nucleotida 2:

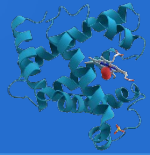


### Incepând cu nucleotida 3:



**Pe o molecula de ADN dublu catenar există 6 posibilități de traducere a informației genetice: 3 cadre de lectură pe o catenă și 3 cadre de lectură pe cealaltă catenă**





- SMS**
- gataatgagagagatattatata  
gaaacagagatataaccagg  
cagaccgagagagagagg  
cacacacacagacgagatata
- Format Conversion**
- Combine FASTA
  - EMBL to FASTA
  - EMBL Feature Extractor
  - EMBL Trans Extractor
  - Filter DNA
  - Filter Protein
  - GenBank to FASTA
  - GenBank Feature Extractor
  - GenBank Trans Extractor
  - One to Three
  - Range Extractor DNA
  - Range Extractor Protein
  - Reverse Complement
  - Split Codons
  - Split FASTA
  - Three to One
  - Window Extractor DNA
  - Window Extractor Protein
- Sequence Analysis**
- Codon Plot
  - Codon Usage
  - CpG Islands
  - DNA Molecular Weight
  - DNA Pattern Find
  - DNA Stats
  - Fuzzy Search DNA
  - Fuzzy Search Protein
  - Ident and Sim
  - Multi Rev Trans
  - Mutate for Digest
  - ORF Finder
  - Pairwise Align Codons
  - Pairwise Align DNA
  - Pairwise Align Protein
  - PCR Primer Stats
  - PCR Products
  - Protein GRAVY
  - Protein Isoelectric Point
  - Protein Molecular Weight
  - Protein Pattern Find
  - Protein Stats
  - Restriction Digest
  - Restriction Summary
  - Reverse Translate
  - Translate
- Sequence Figures**
- Color Align Conservation
  - Color Align Properties
  - Group DNA
  - Group Protein
  - Primer Map
  - Restriction Map
  - Translation Map
- Random Sequences**
- Mutate DNA
  - Mutate Protein
  - Random Coding DNA
  - Random DNA Sequence
  - Random DNA Regions
  - Random Protein Sequence
  - Random Protein Regions
  - Sample DNA
  - Sample Protein
  - Shuffle DNA
  - Shuffle Protein

## Sequence Manipulation Suite: Version 2

- The Sequence Manipulation Suite is a collection of JavaScript programs for generating, formatting, and analyzing short DNA and protein sequences. It is commonly used by molecular biologists, for teaching, and for program and algorithm testing.
- See the [about the Sequence Manipulation Suite](#) page for more information about individual Sequence Manipulation Suite programs.
- You can easily [mirror the Sequence Manipulation Suite](#) on your own web site, or you can use it [off-line](#).
- This version of the Sequence Manipulation Suite represents a complete re-write of the previous version. The new version is much faster and has many new features. The [previous version](#) of the Sequence Manipulation Suite can still be accessed.
- Send questions and comments to [stothard@ualberta.ca](mailto:stothard@ualberta.ca).

[new window](#) | [home](#) | [citation](#)

Mon Nov 6 02:56:29 2017

Valid XHTML 1.0; Valid CSS

Format: Abstract ▾

Send to ▾

**Biotechniques.** 2000 Jun;28(6):1102, 1104.

### The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences.

[Stothard P<sup>1</sup>](#).

[Author information](#)

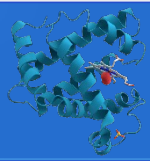
PMID: 10868275

[Indexed for MEDLINE]



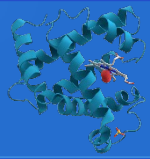
[https://mail.uaic.ro/~marius.mihasan/research/mirrored\\_sites\\_tools/sms2/index.html](https://mail.uaic.ro/~marius.mihasan/research/mirrored_sites_tools/sms2/index.html)

# Elucidarea structurii primare a proteinelor

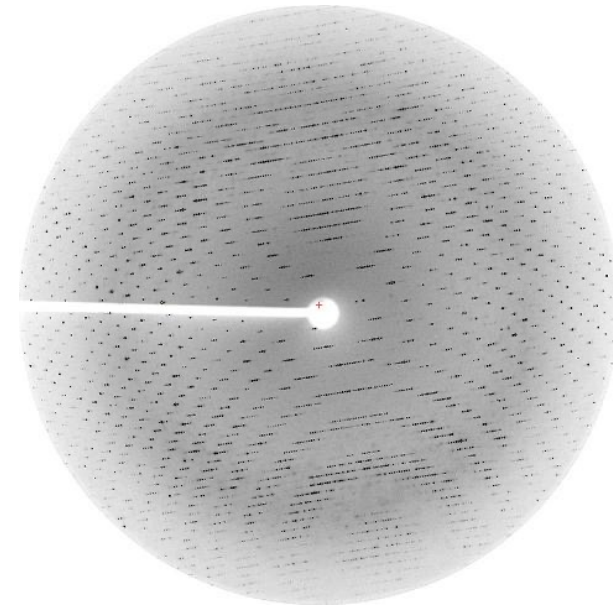
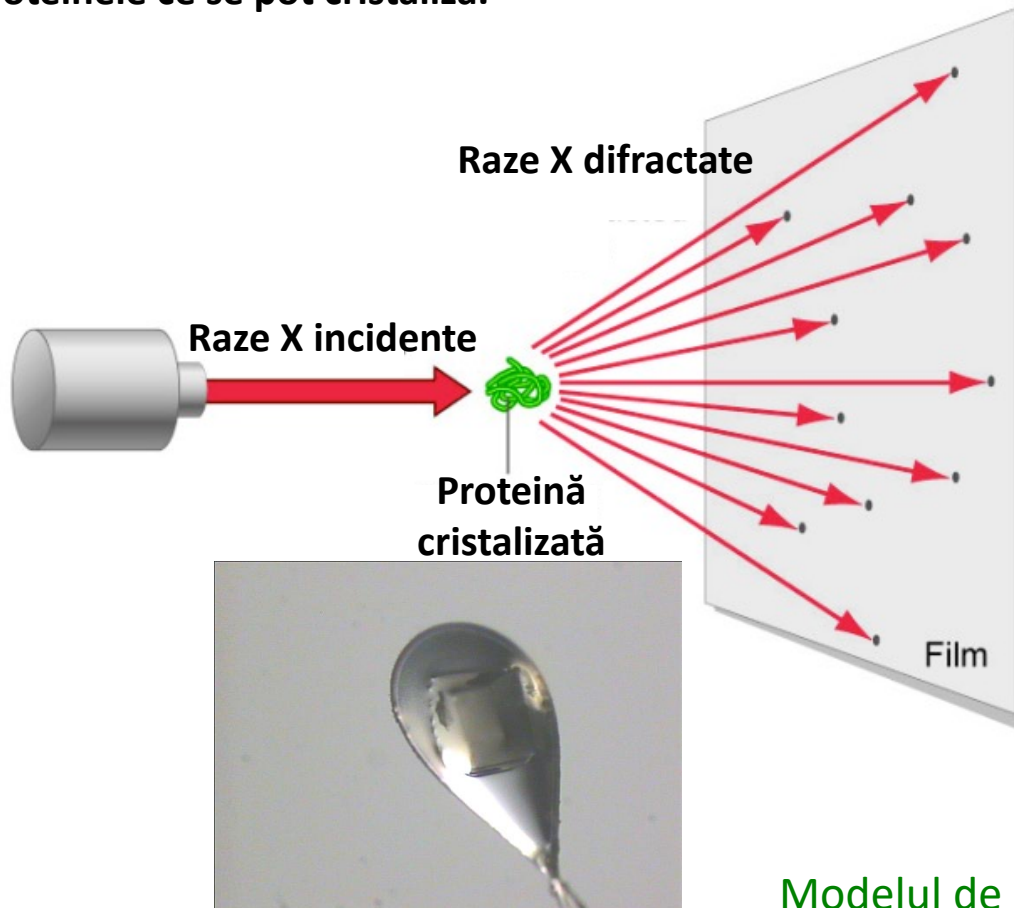


	<b>Avantaje</b>	<b>Dezavantaje</b>
<b>1. Metode experimentale de elucidare a secvenței proteinelor (peptidelor)</b>	<b>Secvențiază strict proteina exprimată, poate identifica și aminoacizi mai puțin comuni</b>	<b>Necesită purificarea prealabilă a proteinei de interes</b>
a. Degradarea Edman		<b>Extrem de lungă ca timp necesar</b>
b. Spectrometria de masă	<b>Timp de lucru scurt</b>	<b>Costul de achiziție al instrumentelor</b>
<b>2. Metode <i>in-silico</i> stabilire a secvenței proteinelor (peptidelor)</b>	<b>Extrem de rapidă</b>	<b>Necesită cunoașterea prealabilă a secvenței de ADN Nu întotdeauna secvența unei gene este complet tradusă în catena polipeptidică (exoni vs introni) Nu identifică aminoacizii modificați și cei mai puțin comuni.</b>

# Elucidarea structurii tridimensionale a proteinelor

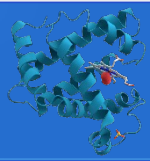


**1. Cristalografia cu raze X** – constă în proiectarea unui fascicul cu raze X pe o proteină cristalizată și înregistrarea modului de difracție a razelor X ca urmare a ciocnirii cu atomii din structura proteinei sub forma unor puncte de un film sensibil. **Modelul de difracție înregistrat este apoi folosit pentru a calcula cu ajutorul computerului poziția clară în spațiu în coordonate XZY a fiecărui atom.** Metoda se poate aplica doar pentru proteinele ce se pot cristaliza.



Modelul de difracție pentru mioglobină conține 25000 puncte

# Elucidarea structurii tridimensionale a proteinelor



**2. Rezonanța magnetică nucleară – RMN** – permite stabilirea structurii tridimensionale a unei proteine în soluție. Proteina este amplasată într-un magnet puternic și este bombardată cu radiație electromagnetică de o frecvență stabilită. Unii atomi din structura moleculei proteice ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{19}\text{F}$  și  $^{31}\text{P}$ ) au capacitatea de a absorbi radiația incidentă iar spectrul de absorbție oferă informații privind legăturile în care acești atomi sunt implicați și atomii vecini. Cu ajutorul calculatorului **se generează o familie de structuri foarte apropiate** ce reprezintă conformații diferite ale scheletului polipeptidic și radicalilor R care îndeplinesc constrângerile rezultate din analiza spectrelor de absorbție.

