

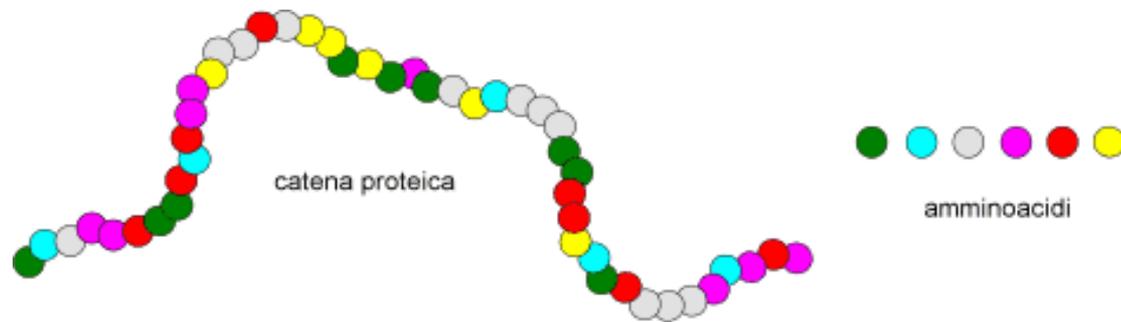


**Corso di
Proprietà di Biopolimeri**

**DETERMINAZIONE DELLA
STRUTTURA PRIMARIA
DELLE PROTEINE**

**Prof. R.Urbani
a.a. 2023-2024**

DETERMINAZIONE DELLA STRUTTURA PRIMARIA DELLE PROTEINE



Perché è importante determinare la struttura primaria delle proteine?

1. La conoscenza della sequenza amminoacidica delle proteine è essenziale per comprendere i loro meccanismi di azione
2. Il confronto fra le sequenze di proteine analoghe ha permesso di comprendere il funzionamento delle proteine e ha indicato collegamenti evoluzionistici fra le proteine e gli organismi che le producono
3. L'analisi della sequenza amminoacidica ha portato ad importanti applicazioni cliniche poiché molte malattie sono causate da mutazioni legate ad un cambiamento amminoacidico nella sequenza proteica

Quando si vuole determinare la struttura primaria completa di una proteina solitamente viene seguito il seguente protocollo:

- ✓ scissione della proteina in peptidi, mediante proteolisi chimiche o enzimatiche;
- ✓ separazione dei peptidi ottenuti;
- ✓ analisi degli amminoacidi di ciascun peptide;
- ✓ analisi sequenziale di ciascun peptide

1. Preliminari

a.
Denaturazione
delle proteine

b.
Riduzione e alchilazione
dei ponti disolfuro

c.
Determinazione
delle subunità

2. Analisi dei segmenti

a.
Frammentare le subunità
in peptidi

b.
Degradazione di
Edman

3. Ricostruzione della sequenza

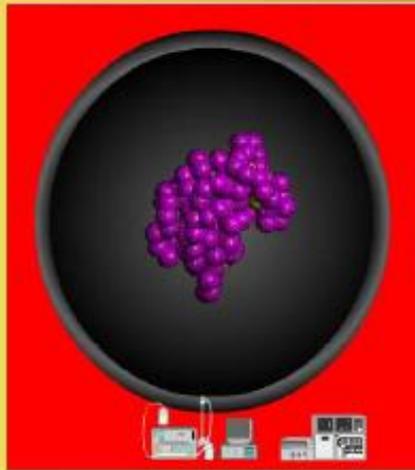
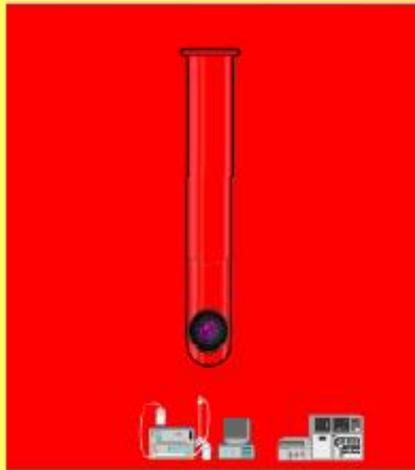
a.
Allineamento
di sequenze

b.
Assegnazione
disolfuri

1. Preliminari

a. Denaturazione delle proteine

Le proteine vengono denaturate in condizioni acide o basiche, a bassa concentrazione salina, a elevate temperature oppure usando agenti denaturanti come l'UREA e lo ione guanidinio o detergenti come l' SDS.

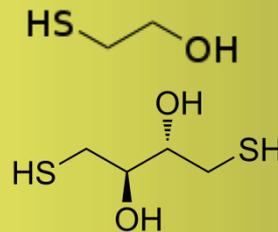


b.
**Riduzione e alchilazione
 dei ponti disolfuro**

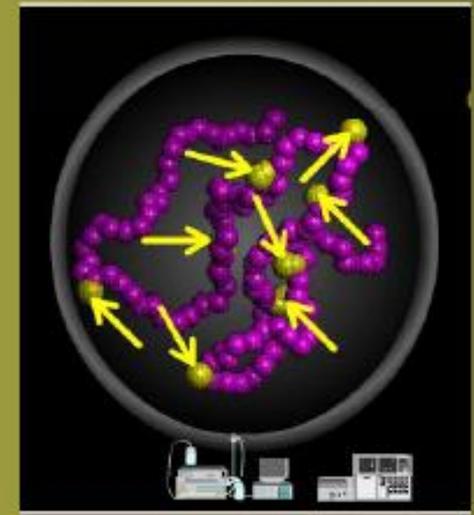
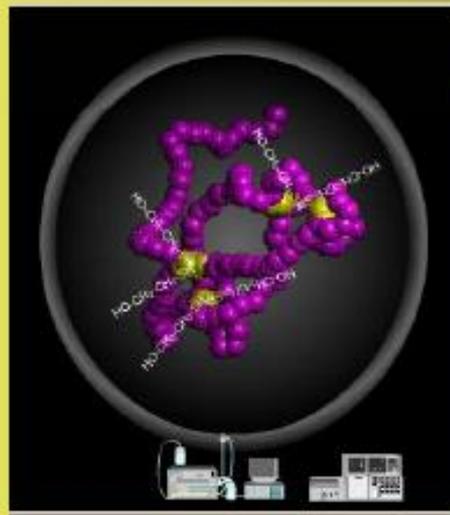
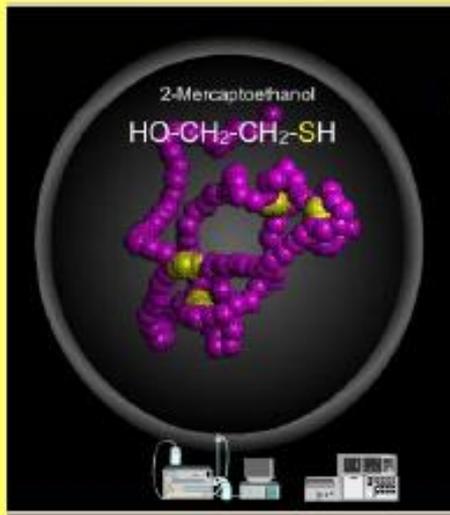
Serve per prevenire la conformazione nativa della proteina che può ostruire l'azione degli agenti proteolitici usati nella determinazione della struttura primaria

Agente riducente:

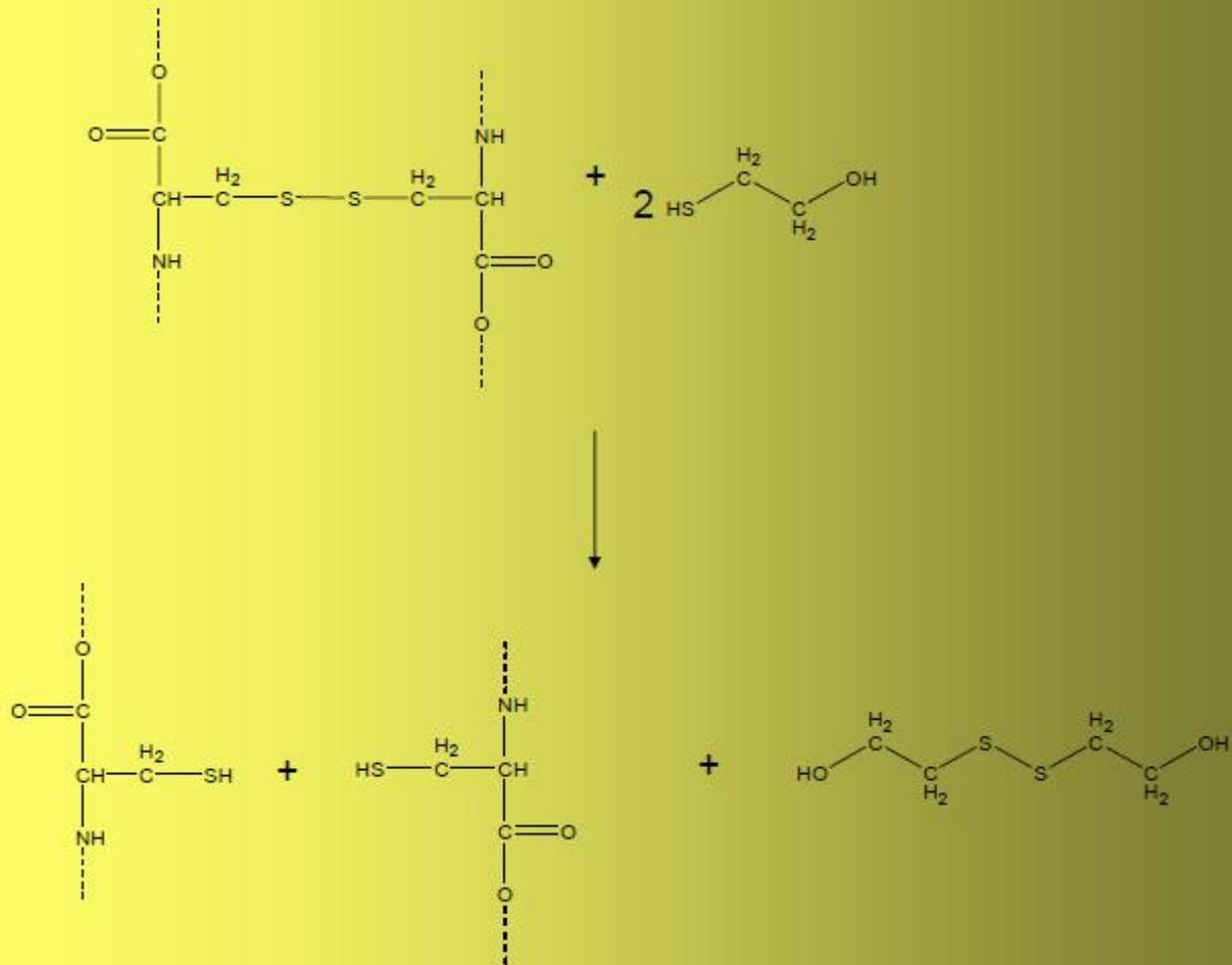
- 2-mercaptoetanolo
- ditiotreitolo
- ditioeritritolo



(2*S*,3*S*)-1,4-Bis-sulfanilbutano-2,3-diolo
 (2*S*,3*R*)-1,4-Bis-sulfanilbutano-2,3-diolo



Reazione di riduzione dei ponti disolfuro con 2-mercaptoetanol



Idrolisi di una proteina in fase liquida

La proteina viene ripresa in **HCl 6N** all'interno di una fiala in cui viene fatto il vuoto in atmosfera di azoto.

La fiala viene quindi messa in stufa a **110°C** per un tempo variabile compreso tra **24-72 ore**.

alcuni amminoacidi come **serina, treonina e tirosina** vengono parzialmente distrutti

la cisteina, la cistina, il triptofano, l'asparagina e la glutammina vengono distrutti completamente

→ **Protezione con con acido iodoacetico → S-carbossimetilcisteina**



L'S-carbossimetilcisteina a differenza della cisteina e della cistina è resistente alle condizioni di idrolisi e può essere facilmente dosata.

Triptofano: si può condurre l'idrolisi non in HCl, ma in acido p-toluen sulfonico 3M

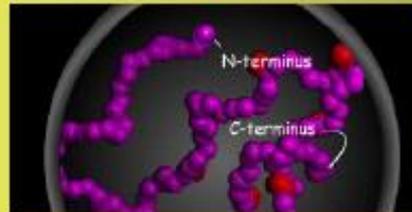
Asparagina e Glutamina: in ambiente acido vengono **deaminate** e vengono quindi dosate insieme ai corrispondenti acidi (acido aspartico e acido glutammico).

Alcuni legami peptidici sono particolarmente resistenti all'idrolisi: è questo il caso di legami in cui sono interessati **residui idrofobici (Ile-Val, Val-Val** etc).

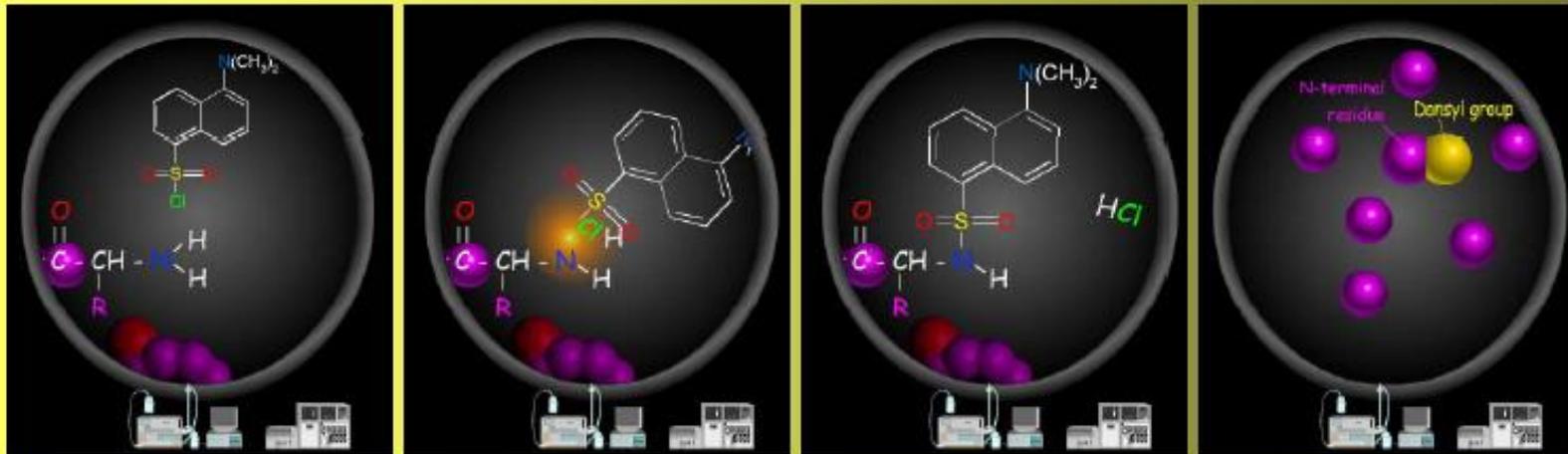
Questi legami non vengono completamente idrolizzati nelle 24 h, quindi per un dosaggio accurato degli amminoacidi idrofobici si può prolungare l'idrolisi **per 48 o 72 h.**

C.
Determinazione
delle subunità

L'identificazione dei residui N- e C- terminali permette di stabilire il numero di subunità (catene polipeptidiche) da cui è costituita una proteina.



Identificazione del residuo N-terminale:
- dansil cloruro



Se una proteina risulta costituita da più subunità, queste devono essere separate usando i metodi di purificazione quali IEC, GFC, ecc. e caratterizzate mediante SDS-PAGE o spettrometria di massa per stabilirne il peso molecolare

Identificazione del residuo C-terminale:

- Non esistono procedure paragonabili a quelle della determinazione del residuo N-terminale.
- Esistono enzimi che tagliano al C-terminale, come le **carbossipeptidasi**.
- Le carbossipeptidasi tagliano al C-terminale con una velocità che dipende dal tipo di amminoacido
- Esistono metodi chimici per identificare i residui C-terminali, quali l'**idrazinolisi**, mediante l'uso di idrazina. Tale processo è però soggetto a molte reazioni secondarie.

Separazione e quantificazione degli amminoacidi presenti nell'idrolizzato

Utilizzo di apparecchi completamente automatizzati che eseguono sia la separazione cromatografica degli amminoacidi sia la loro analisi quantitativa (sequenziatori).

Gli amminoacidi liberi presenti nell'idrolizzato devono essere derivatizzati con reagenti opportuni che ne permettano il riconoscimento.

La derivatizzazione può essere sia **pre-** che **post-colonna** cromatografica.

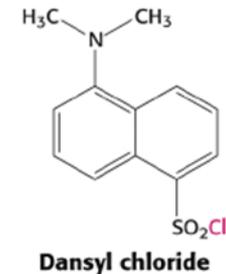
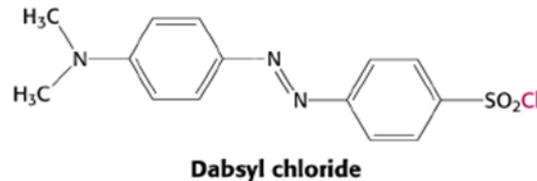
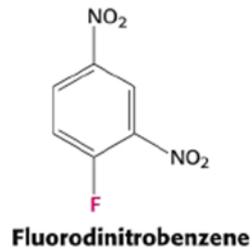
Derivatizzazione pre-colonna :

dopo aver idrolizzato la proteina, gli aminoacidi vengono derivatizzati con un reagente opportuno e poi separati in HPLC con una colonna in fase inversa.

I reagenti che possono essere utilizzati sono:

- dansil cloruro;
- dabsil cloruro;
- OPA;
- 9-fluorenilmetilcloroformiato (FMOC-Cl);
- fenilisotiocianato (PITC).

I singoli aminoacidi possono essere individuati in HPLC dopo derivatizzazione con composti fluorescenti



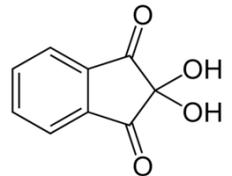
Derivatizzazione post-colonna :

gli aminoacidi presenti nell'idrolizzato vengono separati su una **colonna a scambio ionico**.
Resina utilizzata: scambiatori cationici forti costituiti da polistireni solfonati.

Gli aminoacidi vengono **solubilizzati in un tampone a pH=2** e la colonna cromatografica viene equilibrata con un tampone allo stesso pH, in modo che tutti gli aminoacidi si trovino ad un valore di **pH inferiore al loro punto isoelettrico**.

In queste condizioni tutti gli aminoacidi si legano alla resina. L'eluizione viene condotta in **gradiente** aumentando sia la forza ionica che il pH. In questo modo vengono eluiti prima gli aminoacidi acidi, quindi gli aminoacidi neutri ed infine quelli basici.

Gli aminoacidi eluiti dalla colonna vengono miscelati a 100°C con la **ninidrina** che reagisce sia con gli **aminoacidi primari che quelli secondari** formando due diversi tipi di cromofori, che possono essere rilevati rispettivamente a **570 e 440 nm**.



Metodo alternativo: **OPA (orto-ftalaldeide)** in presenza di β -mercaptoetanolo, pH=9
Si formano **derivati fluorescenti**.

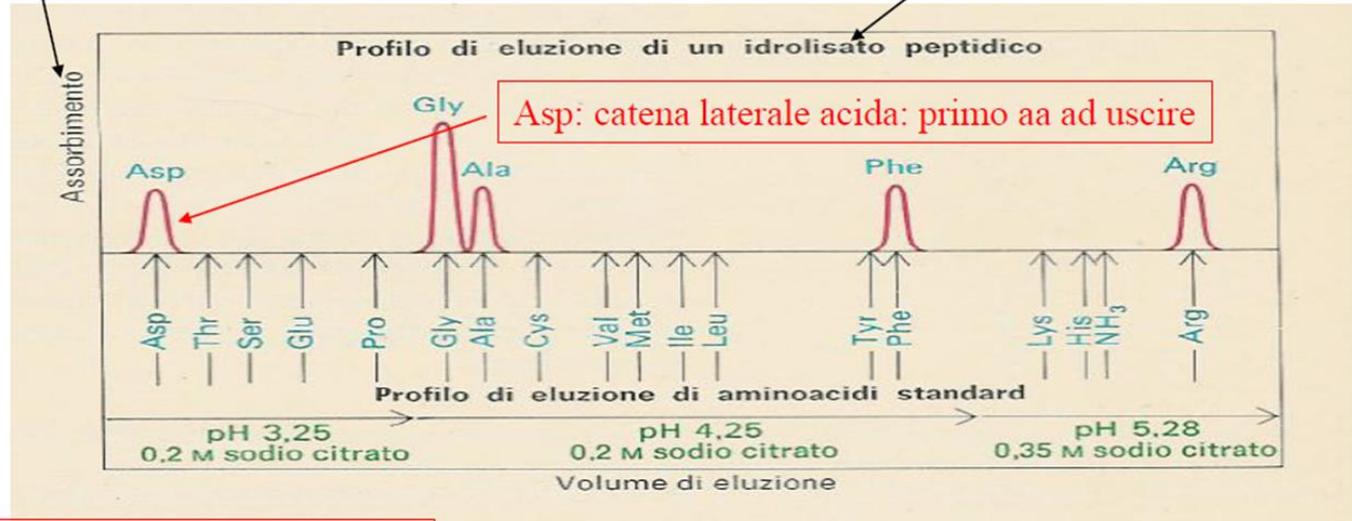
I vantaggi di questo reattivo rispetto alla ninidrina sono rappresentati dall'elevata sensibilità (10-100 pmoli) e dal fatto che il prodotto fluorescente si forma a **temperatura ambiente** in breve tempo (2 min).

αconcentrazione aa

Ala-Gly-Asp-Phe-Arg-Gly

scaldando a 110°C
in HCl 6N per 24 ore

i. composizione in a.a.



si può individuare fino a 1 µg di un aa (=quantità presente in un impronta digitale)

- ❖ cromatografia a scambio ionico
- ❖ eluzione in funzione del pH

Strategia per il sequenziamento

Non è possibile sequenziare un'intera proteina.

La strategia seguita è quella di digerire la proteina con enzimi proteolitici per generare **oligomeri**.

Il problema è quello di mettere poi **nell'ordine** giusto gli oligomeri ottenuti.

Si utilizzano quindi **due o tre diverse proteasi** che generano oligomeri la cui sequenza in parte si sovrappone.

Gli oligomeri ottenuti vengono poi separati tramite HPLC e sequenziati. Le zone di sovrapposizione indicheranno la corretta sequenza degli oligomeri nella struttura primaria della proteina.

La proteina da sequenziare deve essere scissa in frammenti sufficientemente brevi da poter essere sequenziati individualmente, poi, partendo dalla sequenza dei tratti sovrapposti, si ricostruisce la struttura primaria della proteina intatta

Protocollo per determinare la struttura primaria completa di una proteina (composizione e sequenza):

- 1) **analisi sequenziale** della porzione N-terminale della proteina carbossimetilata (degradazione di Edman);
- 2) **scissione della proteina in peptidi**, mediante proteolisi chimiche o enzimatiche;
- 3) **separazione dei peptidi** ottenuti;
- 4) **analisi degli amminoacidi** di ciascun peptide;
- 5) **analisi sequenziale** di ciascun peptide;
- 6) **allineamento dei peptidi**.

1) analisi sequenziale della porzione N-terminale della proteina carbossimetilata

La degradazione di Edman

E' un metodo molto efficace basato sulla derivatizzazione con PITC (fenilisotiocianato: Ph-N=C=S).

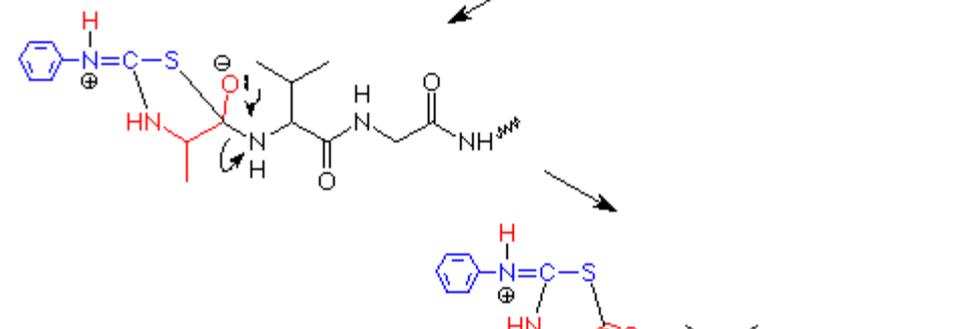
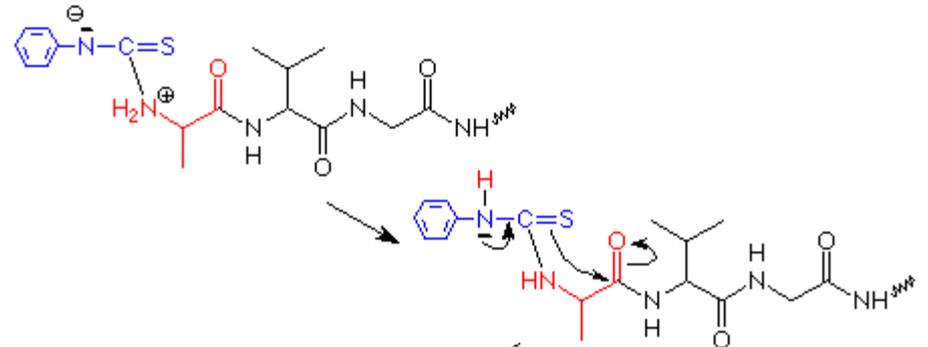
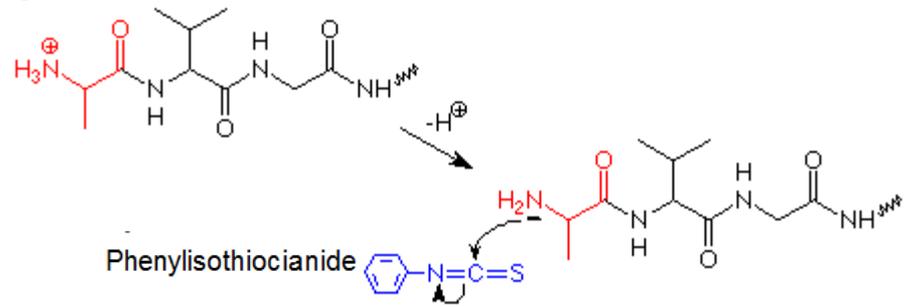
Questo permette la rimozione sequenziale e l'identificazione dei residui di amminoacidi a partire dall'N-terminale.

Gli errori che si accumulano durante la ripetizione dei vari stadi e le reazioni indesiderate limitano il numero di cicli possibili.

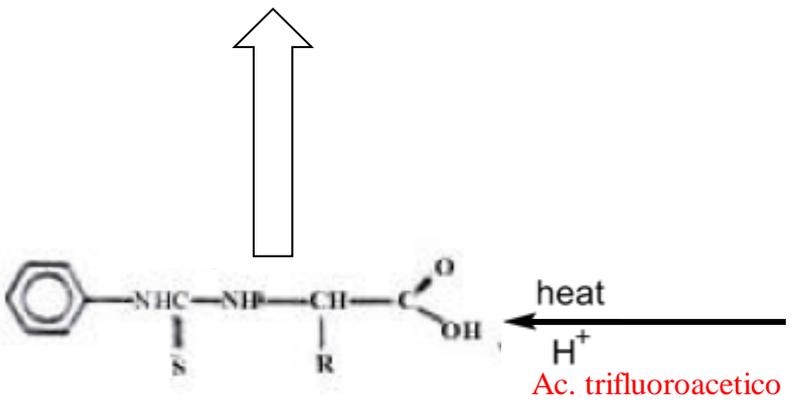
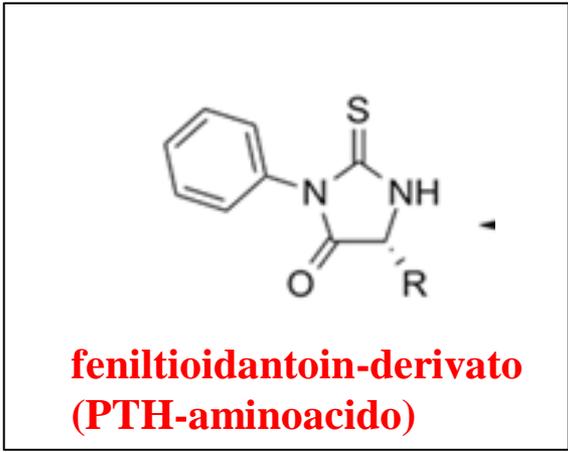
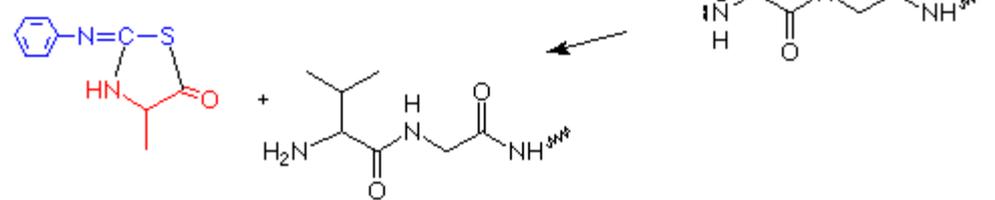
Ma il metodo è stato automatizzato ed oggi un sequenziatore in fase gassosa può raggiungere una resa del 98% riuscendo a sequenziare fino a un 70-mero con solo 5-10 picomoli di proteina.

Prima di sottoporre una proteina alla degradazione di Edman, solitamente, questa viene ridotta e carbossimetilata. Ciò permette di avere una proteina denaturata, in cui il residuo N-terminale è ben esposto, e di trasformare i residui di cisteina (che durante la degradazione di Edman vengono distrutti) in composti stabili e rilevabili come PTH derivati.

Reazione di degradazione di Edman

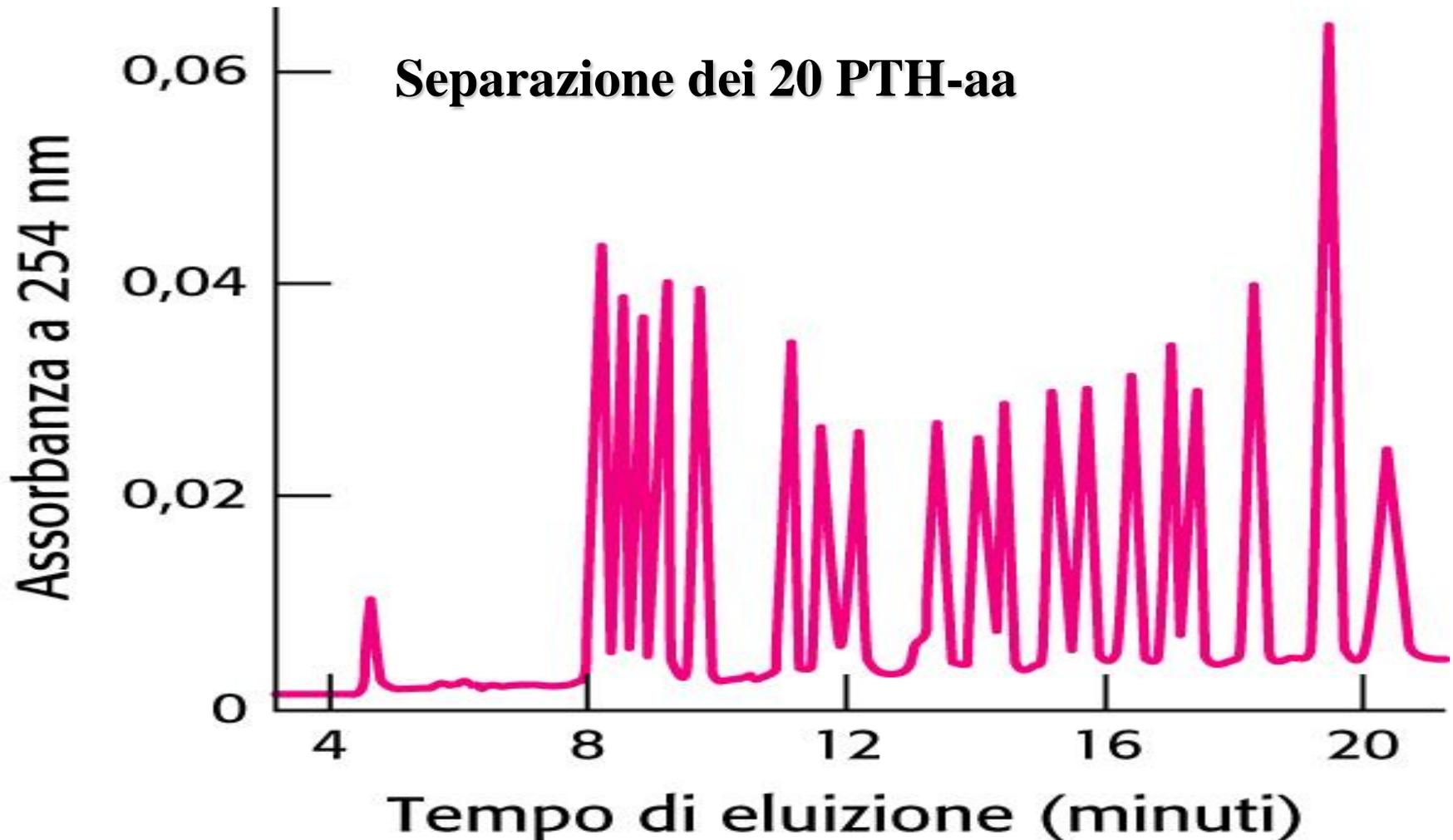


Tiazolinone (instabile)

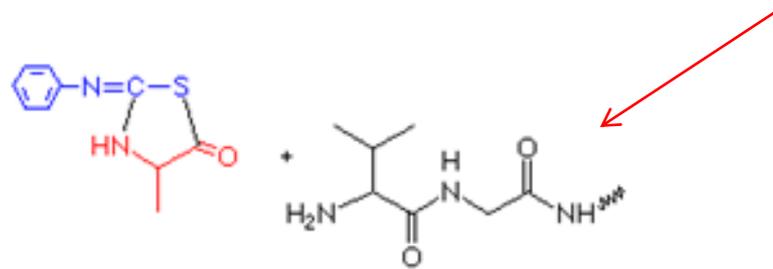


Identificazione del PTH-aa N-terminale mediante cromatografia

Il tipo di amminoacido è identificato con una cromatografia per mezzo di amminoacidi standard di riferimento



La rimanente catena polipeptidica, più corta di un residuo, viene nuovamente sottoposta **ad un altro ciclo** di degradazione di Edman

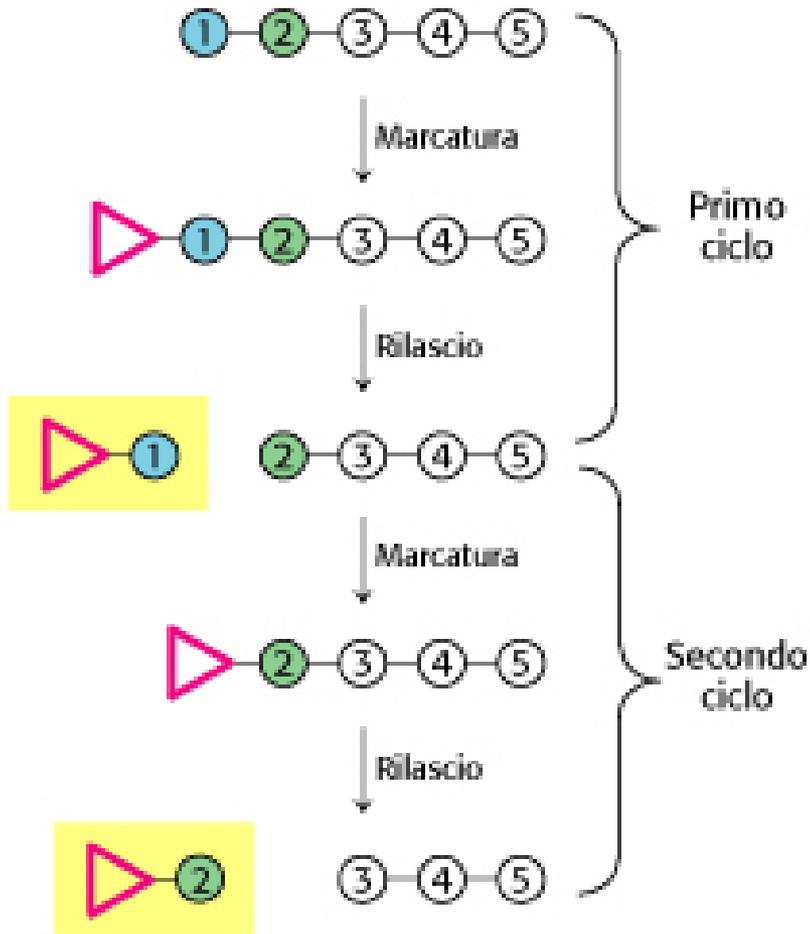


Sequenziatori automatici di proteine

più di 60 residui della
porzione N-terminale della proteina

quantità: **pmoli**

DEGRADAZIONE DI EDMAN



Dopo circa **30 cicli** diventa difficile determinare la sequenza amminoacidica.

Ciò accade perché, dopo un numero elevato di cicli, gli **effetti cumulativi di reazioni incomplete e di reazioni secondarie** del processo di degradazione di Edman rendono impossibile una identificazione certa dell'amminoacido rilasciato e quindi della sequenza amminoacidica.

Per tale motivo la proteina da sequenziare è scissa in **piccoli frammenti polipeptidici**.

Una volta determinata la sequenza di ognuno di essi, si procede con la determinazione della sequenza della proteina completa mediante il metodo della **sovrapposizione dei peptidi**.

Degradazione di Edman \Rightarrow Lunghezza massima \approx **50-70 residui**

 perdita di attendibilità

2) scissione della proteina in peptidi, mediante proteolisi chimiche o enzimatiche

dalla semplice analisi sequenziale della porzione N-terminale di una proteina, non è possibile determinarne la struttura completa, poiché solitamente le proteine hanno dimensioni maggiori di 60 amminoacidi.

Occorre quindi frammentarle in peptidi più piccoli e completamente sequenziabili.

Questa procedura prevede la scissione dei legami peptidici in punti specifici utilizzando sia **metodi chimici** che **enzimatici**.

Tagli specifici con metodi chimici o enzimatici

Reagente

Sito di taglio

Taglio chimico

Bromuro di cianogeno

Lato carbossilico di residui Met

O-iodobenzoato

Lato carbossilico di residui Trp

Idrossilammina

Legami Asp-Gly

2-Nitro-5-tiocianobenzoato

Lato aminico di residui di Cys

Taglio enzimatico

Tripsina

Lato carbossilico di residui Lys e Arg

Clostripaina

Lato carbossilico di residui Arg

Proteasi dello Stafilococco

Lato carbossilico di residui Asn e Gln

Eso- ed endo-peptidasi

METODI ENZIMATICI

Determinazione della mappa triptica

Endopeptidasi :

Tripsina

$R_{n-1} =$ (C terminale) (Arg, Lys)

$R_n \neq$ Pro

Chimotripsina

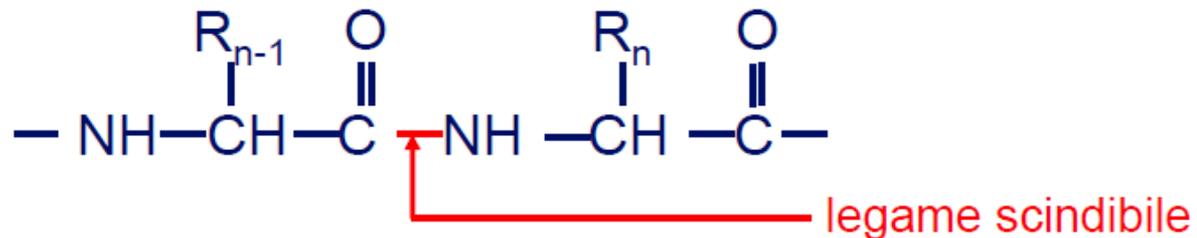
$R_{n-1} =$ grossi residui idrofobici (C terminale)
(Phe, Trp, Tyr)

$R_n \neq$ Pro

Pepsina

$R_n =$ Leu, Phe, Trp, Tyr (N terminale)

$R_{n-1} \neq$ Pro



METODI ENZIMATICI

Tabella 6.7. Specificità di alcuni metodi per la frammentazione delle catene polipeptidiche

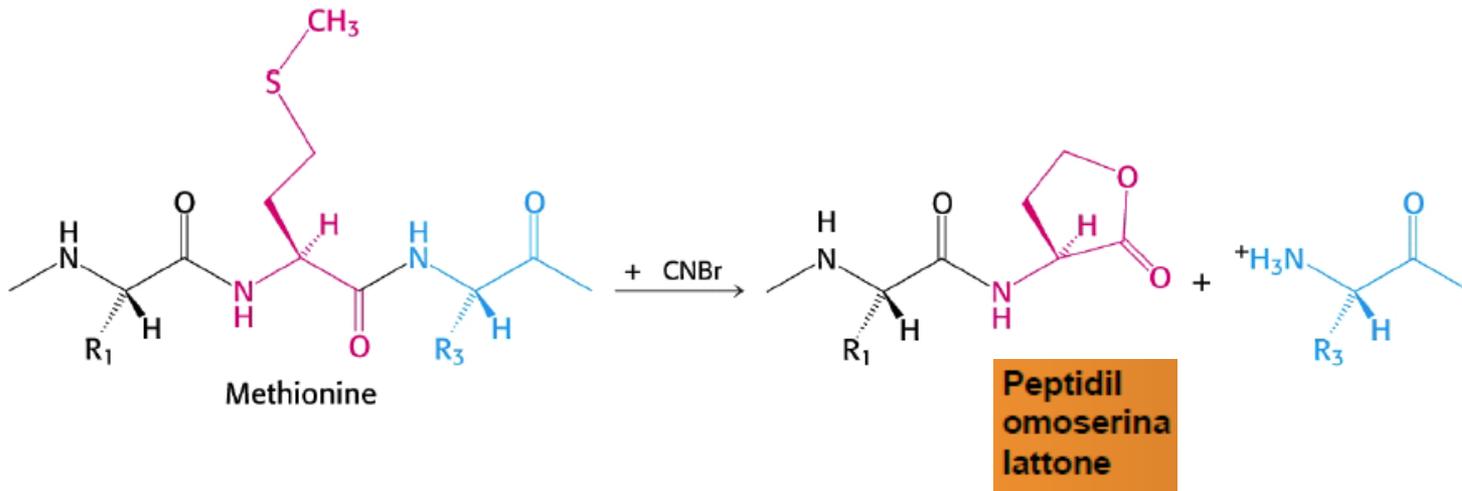
<i>Trattamento*</i>	<i>Punti di rottura**</i>
Tripsina	Lys, Arg (C)
Proteasi di <i>Submaxillarus</i>	Arg (C)
Chimotripsina	Phe, Trp, Tyr (C)
Proteasi V8 di <i>Staphylococcus aureus</i>	Asp, Glu (C)
Asp-N-Proteasi	Asp, Glu (N)
{ Pepsina	Phe, Trp, Tyr (N)
{ Bromuro di cianogeno	

* Tutti gli enzimi o i reagenti elencati nella tabella sono disponibili in commercio.

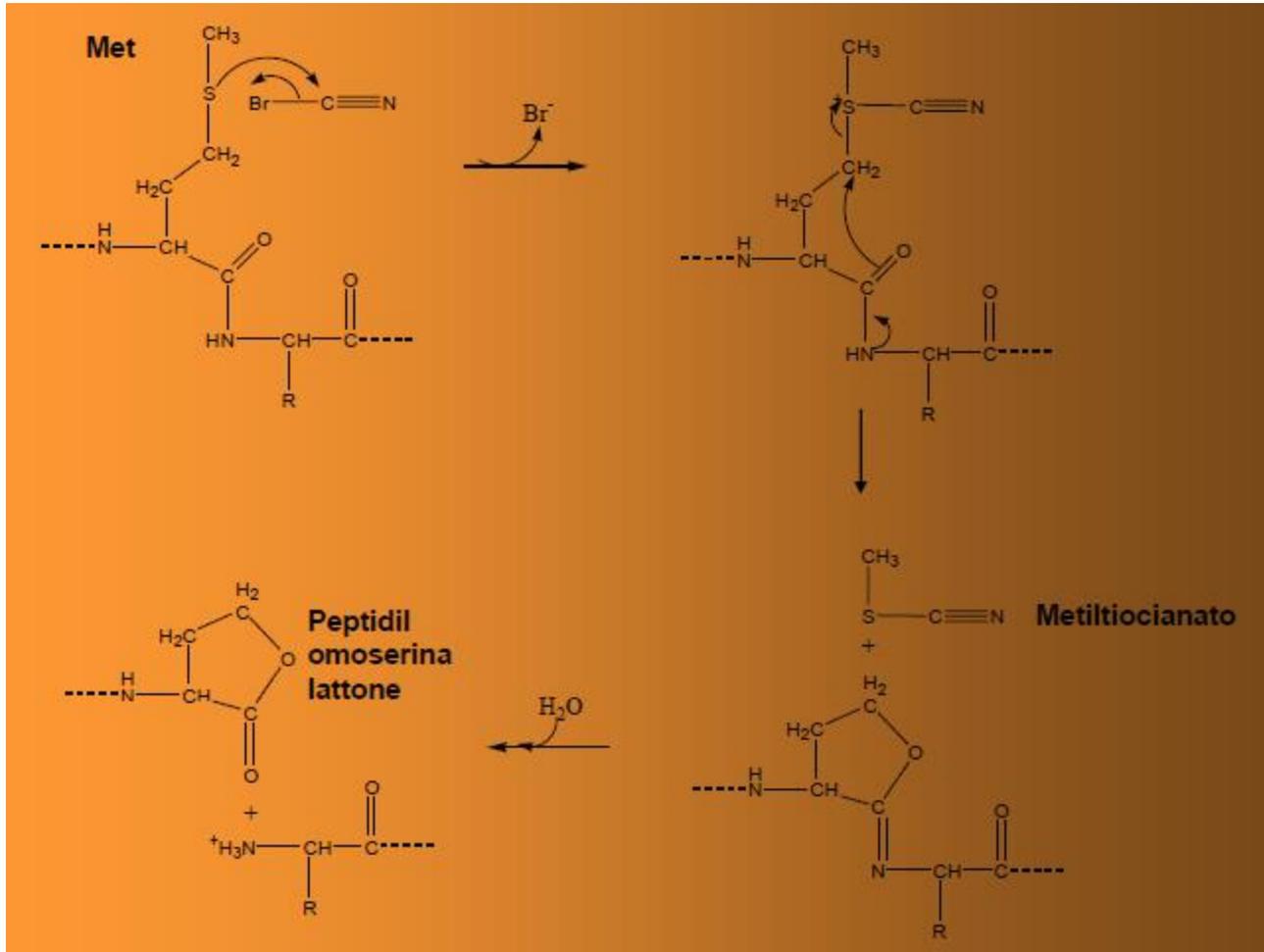
** I residui rappresentano i siti di riconoscimento primario delle proteasi; i legami peptidici possono essere rotti sia dal lato carbonilico (C) che da quello amminico (N) degli amminoacidi indicati nella tabella.

Metodo chimico

La proteolisi limitata può essere ottenuta anche mediante peptidasi chimiche (non enzimatiche), ad es. il BROMURO di CIANOGENO (CNBr) che scinde sul lato C dei residui di Met



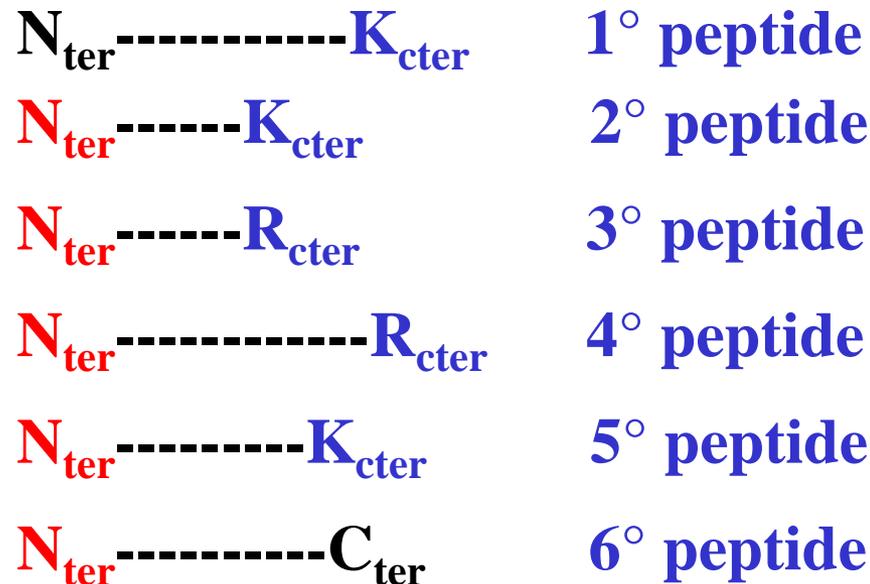
Meccanismo di idrolisi con BrCN



ESEMPIO: Proteina di 150 aa



↓
Idrolisi con **TRIPSINA**, peptidasi dotata di elevata specificità: taglia i legami peptidici dal lato C (verso il terminale carbossilico) dei residui carichi positivamente **Arg** (R) e **Lys** (K), se il residuo successivo non è Pro



Separazione dei PEPTIDI mediante cromatografia

Determinazione della sequenza N-terminale di ciascun peptide

N_{ter} ----- K_{cter}

1° peptide

N_{ter} ----- K_{cter}

2° peptide

N_{ter} ----- R_{cter}

3° peptide

N_{ter} ----- R_{cter}

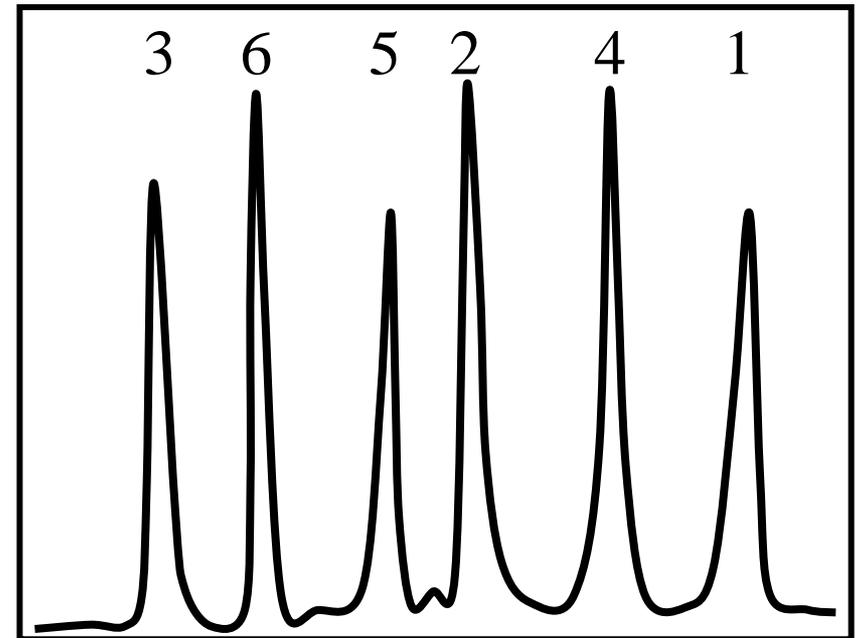
4° peptide

N_{ter} ----- K_{cter}

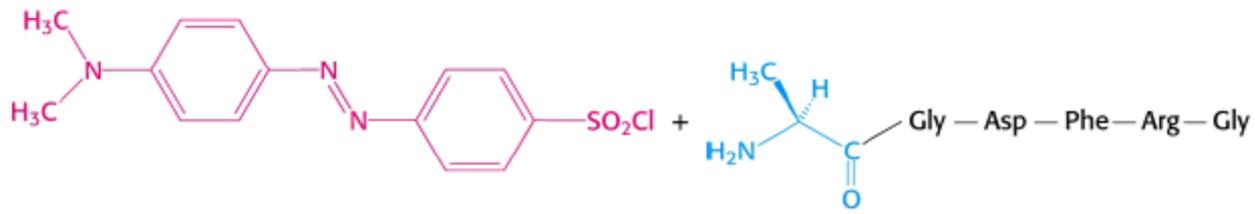
5° peptide

N_{ter} ----- C_{ter}

6° peptide

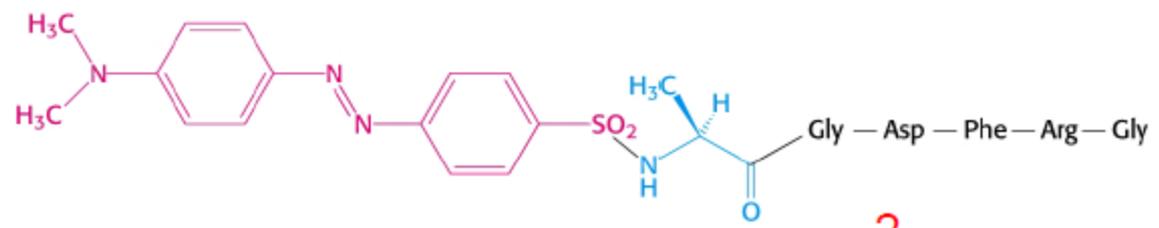


- E' necessario adoperare un metodo diverso per generare una serie di frammenti peptidici differenti
- Anche in questo caso si procederà alla determinazione della sequenza N-terminale di ciascun peptide ottenuto
- Le due serie di sequenze peptidiche sovrapposte consentono di ricostruire la sequenza di ciascun polipeptide

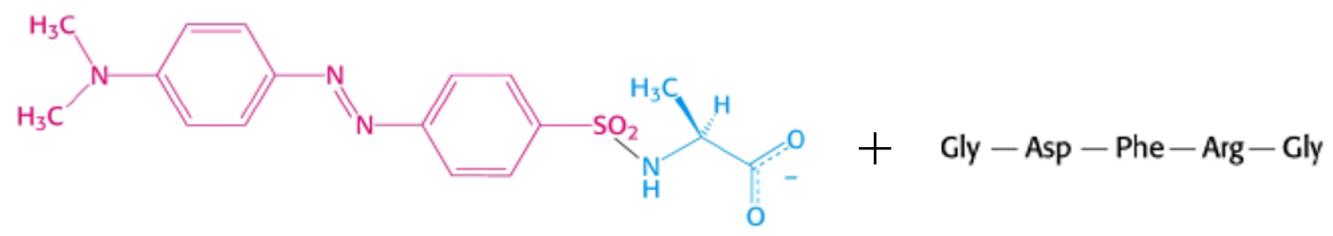


Dabsyl chloride

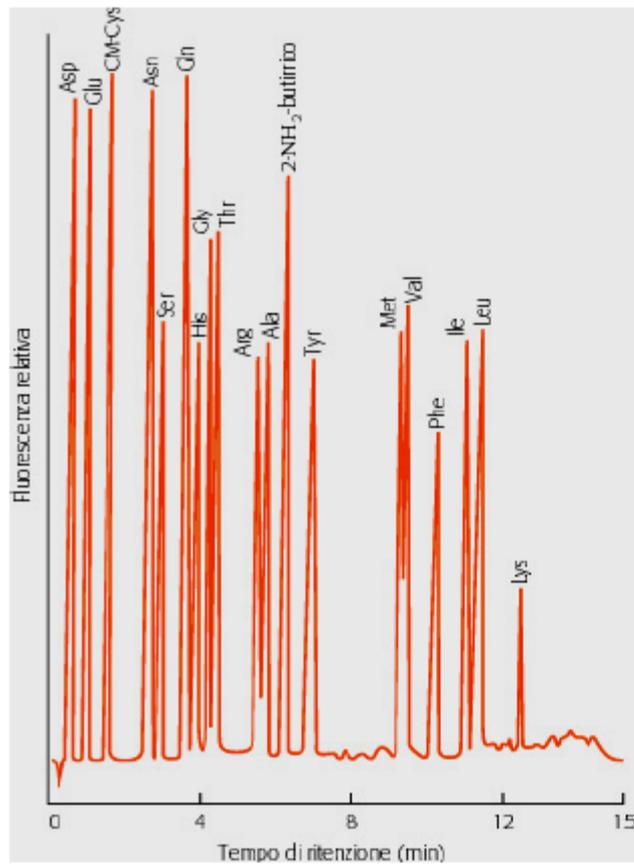
Labeling ← 1.



Hydrolysis ← 2.

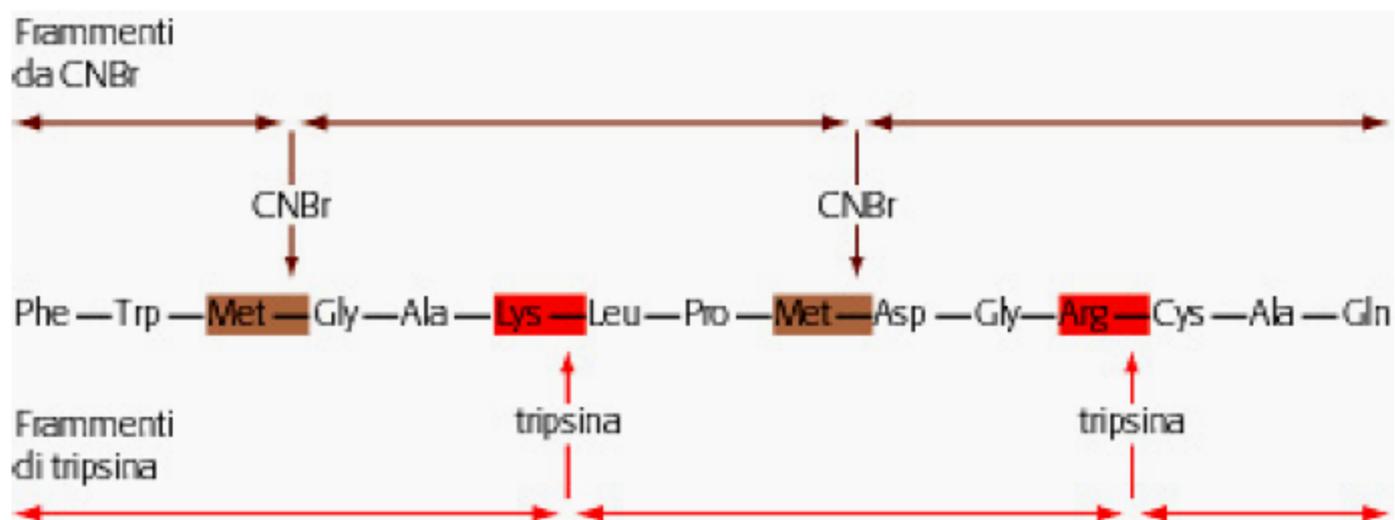


Dabsyl alanine

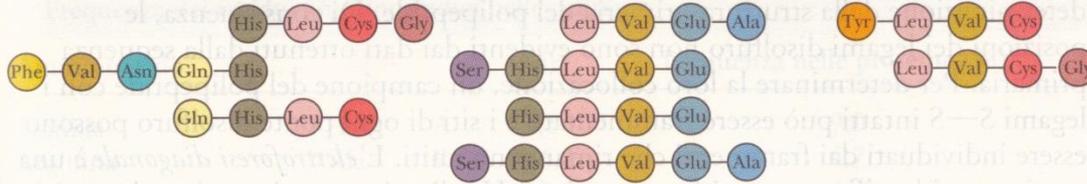


Determinazione della
composizione a.a.
mediante separazione
in HPLC dei derivati
fluorescenti

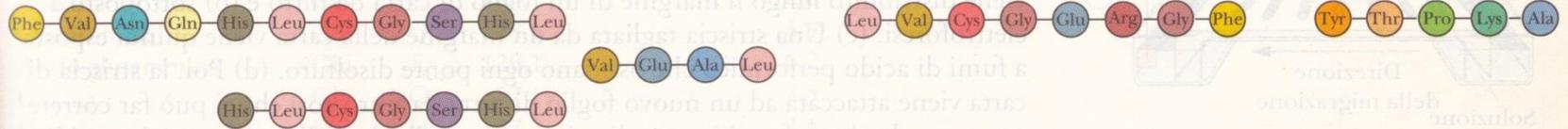
6. Ricostruzione della sequenza della proteina



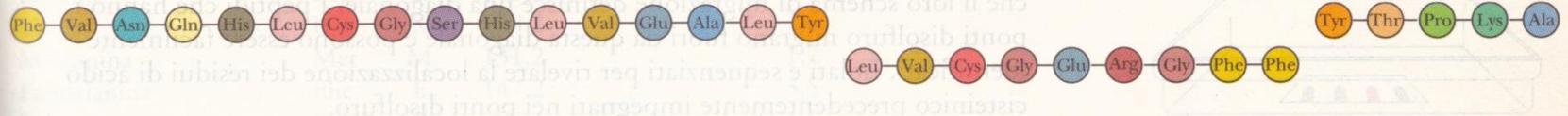
Peptidi derivanti dall'idrolisi acida parziale:



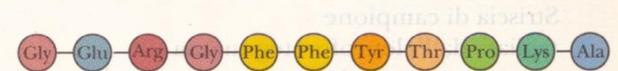
Peptidi derivanti dal trattamento con pepsina:



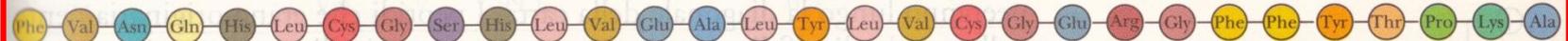
Peptidi derivanti dal trattamento con chimotripsina:



Peptidi derivanti dal trattamento con tripsina:



Sequenza dedotta:



3. Ricostruzione della sequenza

a. Allineamento di sequenze

L'allineamento di sequenza viene fatto mettendo a confronto le sequenze amminoacidiche di un set di frammenti peptidici con quelli di un secondo set i cui siti di cleavage si sovrappongono con quelli del primo set. La proteina viene ridotta prima del cleavage!

Ad es. Primo set di peptidi ottenuti con Tripsina (che taglia Lys/Arg al C terminale)
Secondo set di peptidi ottenuto con CNBr (che taglia Met al C terminale)

Trypsin digest
No. Sequence

1:K
2:SR
3:DR
4:FER
5:NLTK
6:NVACK
7:ETGSSK
8:TTQANK
9:ETAAAK
10:YPNCAYK
11:NGQTNCYQSYSTMSITDCR
12:HIIVACEGNPYVPVHFDASV
13:QHMDSSTSAASSSNYCNEMMK
14:CKPVNTFVHESLADVQAVCSQK

Cyanogen bromide digest
No. Sequence

1:M
2:KETAAAKFERQHM
3:DSSTSAASSSNYCNEM
4:SITDCRETGSSKYPNCAYKTTQ ANKHIIVACEGNPYVPVHFDAS V
5:KSRNLTKDRCKPVNTFVHESLA DVQAVCSQKNVACKNGQTNCYQ SYSTEM

Determinazione del segmento C-terminale

● Trypsin digest

● No. Sequence

1:	K
2:	SR
3:	DR
4:	FER
5:	NLT
6:	NVACK
7:	ETGSSK
8:	TTQANK
9:	ETAAAK
10:	YPNCAYK
11:	NGQTNCYQSYSTM
12:	HIIVACEGNPYVPVHFDAS
13:	QHMDSS TSAASSSNYCNE
14:	CKPVNTFVHESLADVQAVCSQ

C-terminal segment?

● Cyanogen bromide digest

● No. Sequence

1:	M
2:	KETAAAKFERQHM
3:	DSSTSAASSSNYCNE
4:	<u>SITDCRETGSSKYPNCAYKTTO</u> <u>ANKHIIVACEGNPYVPVHFDAS</u> <u>V</u>
5:	KSRNLTKDRCKPVNTFVHESLA DVQAVCSQKNVACKNGQTNCYQ SYSTEM

C-terminal segment?

Ricostruzione della sequenza

```

* Trypsin digest
No. Sequence
-----
1:K
2:SR
3:DR
4:FER
5:NLTK
6:NVACK
7:ETGSSK
8:TQANK
9:RTAAAK
10:YFNCAVK
11:NGQTNCYQSYSTMSITDCR
12:HIIVACEGNPYVPVHFDASV
13:QHMDSS TSAASSSNYCNE MMK
14:CKPVNTFVHESLADVQAVCSOK
    
```

```

* Cyanogen bromide digest
No. Sequence
-----
1:M
2:KETAAAKFERQHM
3:DSSTSAASSSNYCNE M
4: SITDCRETGSSKYPNCAVKTTQ
   ANKHIIIVACEGNPYVPVHFDAS
   V
5:KSRNLTKDRCKPVNTFVHESLA
   DVQAVCSQRNVACKRNGQTNCYQ
   SYSTM
    
```

```

* Trypsin digest
No. Sequence
-----
1:K
2:SR
3:DR
4:FER
5:NLTK
6:NVACK
7:ETGSSK
8:TQANK
9:ETAAAK
10:YFNCAVK
11:NGQTNCYQSYSTMSITDCR
12:HIIVACEGNPYVPVHFDASV
13:QHMDSS TSAASSSNYCNE MMK
14:CKPVNTFVHESLADVQAVCSOK
    
```

```

* Cyanogen bromide digest
No. Sequence
-----
1:M
2:KETAAAKFERQHM
3:DSSTSAASSSNYCNE M
4: SITDCRETGSSKYPNCAVKTTQ
   ANKHIIIVACEGNPYVPVHFDAS
   V
5:KSRNLTKDRCKPVNTFVHESLA
   DVQAVCSQRNVACKRNGQTNCYQ
   SYSTM
    
```

```

Trypsin: -----
          70      80      90      100
CNBr:  -----SITDCRETGSSKYPNCAVKTTQANK
    
```

```

Trypsin: -----NGQTNCYQSYSTMSITDCR-----
          60      70      80
CNBr:  -----SITDCRET
    
```

```

* Trypsin digest
No. Sequence
-----
1:K
2:SR
3:DR
4:FER
5:NLTK
6:NVACK
7:ETGSSK
8:TQANK
9:ETAAAK
10:YFNCAVK
11:NGQTNCYQSYSTMSITDCR
12:HIIVACEGNPYVPVHFDASV
13:QHMDSS TSAASSSNYCNE MMK
14:CKPVNTFVHESLADVQAVCSOK
    
```

```

* Cyanogen bromide digest
No. Sequence
-----
1:M
2:KETAAAKFERQHM
3:DSSTSAASSSNYCNE M
4: SITDCRETGSSKYPNCAVKTTQ
   ANKHIIIVACEGNPYVPVHFDAS
   V
5:KSRNLTKDRCKPVNTFVHESLA
   DVQAVCSQRNVACKRNGQTNCYQ
   SYSTM
    
```

```

* Trypsin digest
No. Sequence
-----
1:K
2:SR
3:DR
4:FER
5:NLTK
6:NVACK
7:ETGSSK
8:TQANK
9:ETAAAK
10:YFNCAVK
11:NGQTNCYQSYSTMSITDCR
12:HIIVACEGNPYVPVHFDASV
13:QHMDSS TSAASSSNYCNE MMK
14:CKPVNTFVHESLADVQAVCSOK
    
```

```

* Cyanogen bromide digest
No. Sequence
-----
1:M
2:KETAAAKFERQHM
3:DSSTSAASSSNYCNE M
4: SITDCRETGSSKYPNCAVKTTQ
   ANKHIIIVACEGNPYVPVHFDAS
   V
5:KSRNLTKDRCKPVNTFVHESLA
   DVQAVCSQRNVACKRNGQTNCYQ
   SYSTM
    
```

```

Trypsin: -----QHMDSS TSAASSSNYCNE MMK-----
          10      20      30
CNBr:  -----KSRNLTK
    
```

```

Trypsin: -----QHMDSS TSAASSSNYCNE MMK-----
          1      10      20
CNBr:  -----KETAAAKFERQHM-----
    
```

Sequenza finale

```
1      10      20      30
KETAAAKFERQHMSSTSAASSSNYCNEMM

31     40     50     60
KSRNLTKDRCKPVNTFVHESLADVQAVCSQ

61     70     80     90
KNVACKNGQTNCYQSYSTMSITDCRETGSS

91     100    110    120
KYPNCAYKTTQANKHIIVACEGNPYVPVHF
      From Trypsin digest
121   From CNBr digest
DASV  From Both
```

Separazione dei peptidi ottenuti: la scelta del metodo di separazione dei peptidi, ottenuti dalla scissione di una proteina, dipende dal numero e dalle dimensioni dei peptidi

Si può scegliere una **separazione cromatografica** (*fase inversa, scambio ionico, gel filtrazione, etc*)
o **elettroforetica**.

Lezione:
ELETTROFORESI

Lezione:
Gel Permeation Crom.

Lezione:
Spettr. Massa