

# Allineamento di sequenze proteiche

Una proteina è composta da 20 diversi amino acidi uniti da legami peptidici. Si definisce:

**struttura primaria:** la sequenza dei residui

**struttura secondaria:** la parziale strutturazione 3-D di residui adiacenti

**struttura terziaria:** l'organizzazione strutturale globale della proteina

**struttura quaternaria:** l'associazione di più strutture terziarie

Proteine che svolgono funzioni simili o anche identiche possono differire molto nella struttura primaria, molto meno nella secondaria e nella terziaria.

Anche nella sequenza primaria sono però presenti delle conservazioni, che si rendono evidenti mediante l'allineamento.

Per gli allineamenti tra proteine ci interessa principalmente la struttura primaria, quindi la sequenza dei residui che la compongono.

**Allineare due o più proteine significa trovare il modo migliore di sovrapporle, alla ricerca di pattern e motivi comuni.**

Trovare un motivo comune tra due proteine significa individuare il motivo evolutivo per cui esse sono da considerarsi simili e da questo prendere spunto per capire i meccanismi d'azione o le caratteristiche strutturali comuni.

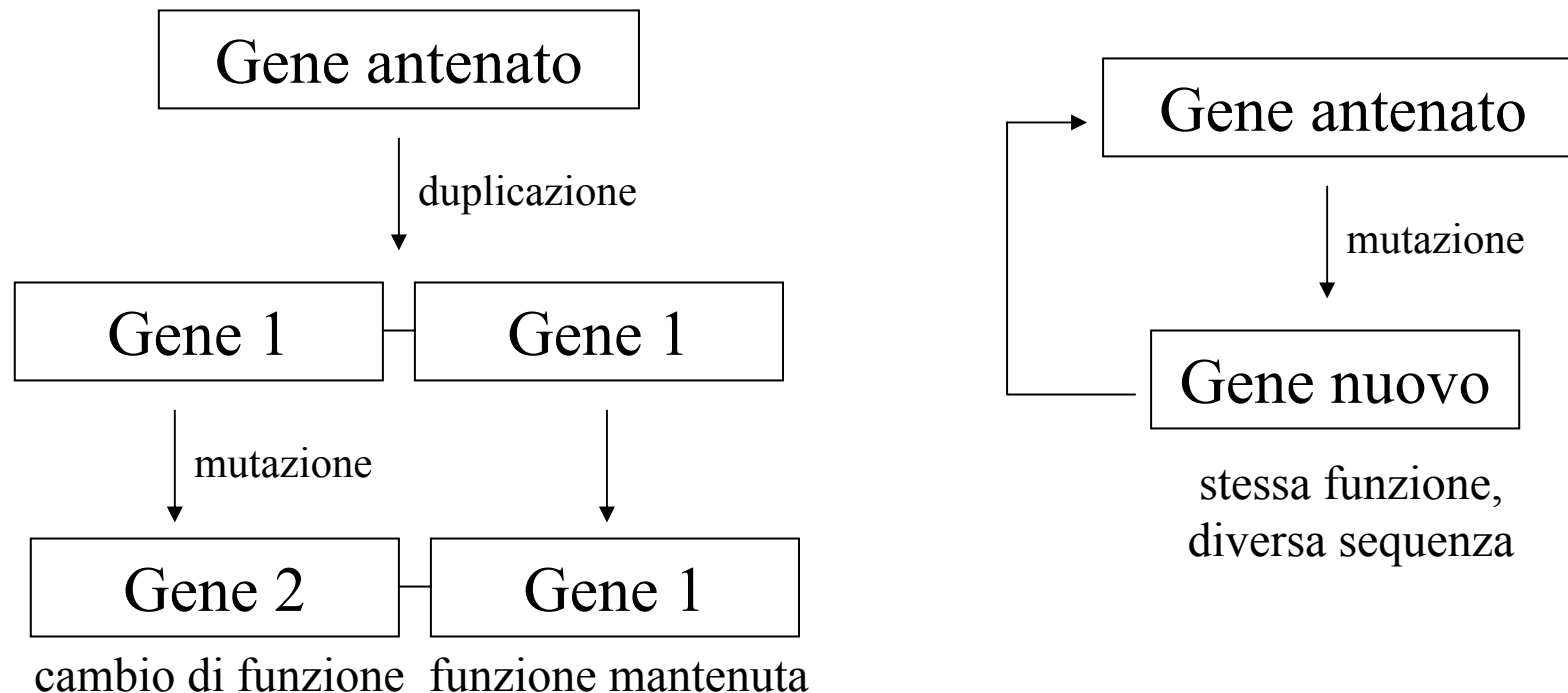
**Da un punto di vista informatico una proteina è una stringa più o meno lunga composta da un alfabeto di venti lettere.**

Alanine	Ala	A	GC[CATG]
Cysteine	Cys	C	TG[CT]
Aspartic Acid	Asp	D	GA[CT]
Glutamic Acid	Glu	E	GA[AG]
Phenylalanine	Phe	F	TT[CT]
Glycine	Gly	G	GG[CATG]
Histidine	His	H	CA[CT]
Isoleucine	Ile	I	AT[CAT]
Lysine	Lys	K	AA[AG]
Leucine	Leu	L	CT[CATG], TT[AG]
Methionine	Met	M	ATG
Asparagine	Asn	N	AA[CT]
Proline	Pro	P	CC[CATG]
Glutamine	Gln	Q	CA[AG]
Arginine	Arg	R	CG[CATG], AG[AG]
Serine	Ser	S	TC[CTAG], AG[CT]
Threonine	Thr	T	AC[CATG]
Valine	Val	V	GT[CATG]
Tryptophan	Trp	W	TGG
Tyrosine	Tyr	Y	TA[CT]
STOP	-	-	TA[AG], TGA

Il codice genetico si definisce **DEGENERATO** e **RIDONDANTE**, dato che esistono più codoni per lo stesso amino acido

## Perchè cambiano le proteine

- 1) Perchè il cambiamento non comporta modifiche nè nella struttura (destabilizzazione) nè nella funzione.
- 2) Perchè sono presenti più proteine che svolgono la stessa funzione, quindi se ne cambia una la funzione è comunque preservata.



## Come cambiano le proteine

Una proteina presente in un organismo può cambiare in seguito a mutazioni nella sua sequenza codificante. Le mutazioni possono essere puntiformi o più estese.

**Sostituzione** - una base viene sostituita con un'altra

**Inserzione** - una o più basi vengono inserite

**Delezione** - una o più basi vengono tolte

**Inversione** - un tratto di DNA si inverte

Il codice genetico è ridondante, perciò non sempre una sostituzione porta ad un cambiamento di amino acido: ho una mutazione silente e la proteina resta identica a prima.

Negli altri casi, dal punto della mutazione, gli amino acidi cambiano, e la proteina può diventare da lì irriconoscibile.

## Allineamenti facili e difficili

```

VILMA ATGFR EACVG TPLKT
VILMA ATGFR EACLG TPLKT
VILMA ATEAR EACVG TPLKT
VILMA ATGPR EACMG TPLKT
VILMA ATGFR EACVG TPLKT
***** **_** *** * *****
  
```

Facile: anche  
dall'osservazione si intuisce  
il giusto allineamento

```

VILMA ATGFR EACVG TPLKT
--LMT QTG-R ERTGA TP---
--LMT QTGHR LMTGA LP-K-
--LMT QTGFR ERTGA TP---
--LMT QTG-R ERTGA TP---
**_  _** *   ---  -*
  
```

Difficile: servono dei metri  
di giudizio oggettivi per  
valutare la qualità  
dell'allineamento

## Valutare un allineamento

VILMA   ATGFR   EACVG   TPLKT   MTGAP   LPYND   QRTED  
VLKMA   ASGFR   LVCIG   KYWKT   MTGAP   LPWND   YTREN

Lunghezza: 35 residui

Bisogna prendere in considerazione **2 parametri**

**Identità di sequenza:**  $21/35 = 60 \%$

numero di residui identici appaiati dopo l'allineamento

**Similarità di sequenza:**  $4/35 = 11.5 \%$

numero di residui con caratteristiche chimico-fisiche simili appaiati dopo l'allineamento.

I criteri per la similarità tra i residui vanno valutati bene perchè alla fine determinano la qualità dell'allineamento.

## Cosa significa allineare le proteine

Trovare il modo di convertire le sequenze inserendo il minor numero possibile di cambiamenti. Si assegnano dei criteri per valutare l'importanza dei:

**MATCH:** similarità positiva

**MISMATCH:** similarità negativa

**GAP:** similarità negativa che permette di migliorare l'allineamento globale.

Le domande a cui rispondere sono:

1. Come assegno i punteggi per ogni posizione?
2. Qual'è il migliore algoritmo che permetta di valutare tutte le possibili combinazioni?
3. L'allineamento finale è significativo?

## Le matrici di sostituzione

Confrontando gli amino acidi, è opportuno creare dei criteri di similarità che non si basino solo sull'identità assoluta, ma sulle loro caratteristiche chimico-fisiche.

Se una coppia S/T si trova appaiata nell'allineamento, lo score dovrebbe essere più alto rispetto ad una coppia S/W, lo stesso per una coppia D/E rispetto ad una coppia D/K.

Aromatici:	WYF	} caratteristiche chimico-fisiche comuni
Idrofobici:	VILMA	
Polari carichi:	KRDE	
Polari non carichi:	HCSTQNPG	

Inoltre alcuni amino acidi sembrano maggiormente conservati di altri, quindi un allineamento C/C dovrebbe essere più premiato rispetto ad un A/A

Una matrice di sostituzione è una matrice  
20 x 20 in cui per ogni posizione si  
riporta il punteggio di quella specifica  
coppia di amino acidi.

Il punteggio deriva da osservazioni di  
varia natura

Un meccanismo di punteggi che voglia tenere in  
considerazione solo le identità e non le similitudini, potrà  
sempre utilizzare una matrice con valori alti nella diagonale  
principale e zero nelle altre posizioni.

# Matrice identità

C		10																			
S		0	10																		
T		0	0	10																	
P		0	0	0	10																
A		0	0	0	0	10															
G		0	0	0	0	1	10														
N		0	0	0	0	0	0	10													
D		0	0	0	0	0	0	0	10												
E		0	0	0	0	0	0	0	0	10											
Q		0	0	0	0	0	0	0	0	0	10										
H		0	0	0	0	0	0	0	0	0	0	10									
R		0	0	0	0	0	0	0	0	0	0	0	10								
K		0	0	0	0	0	0	0	0	0	0	0	0	10							
M		0	0	0	0	0	0	0	0	0	0	0	0	0	10						
I		0	0	0	0	0	0	0	0	0	0	0	0	0	0	10					
L		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10				
V		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10			
Y		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10		
F		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	
W		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10

---

C S T P A G N D E Q H R K M I L V Y F W

## Matrici PAM

1978: Margareth Dayhoff analizzò 71 famiglie di proteine da un punto di vista evolutivo.

**Assunto:** date delle sequenze correlate, è possibile calcolare la probabilità con cui ogni aminoacido subisce una una mutazione e metterlo in relazione con la distanza evolutiva. Fatto questo, è possibile ricavare tutte le probabilità di sostituzione data una distanza evolutiva.

Si può ricavare così una tabella di Percent Accepted Mutations (PAM) per ogni distanza evolutiva a partire dalla PAM 1 (cioè una tabella che prenda in considerazione una frequenza di sostituzione di 1 amino acido ogni 100) mediante estrapolazione, serve solo un indice di mutazione.

La PAM 1, proprio perchè 1 è un passo piccolo, sarà simile alla matrice identità. E' possibile però da questa estrapolare tutte le altre matrici.

## Significato della matrici PAM

1 - Al crescere dell'indice, ci si allontana dalla matrice identità e i singoli valori di probabilità sono più “permissivi”

2 - Ogni indice dice il numero di passi evolutivi che si prevede per le proteine in analisi. Quindi PAM 100 non significa che il 100% degli aminoacidi vengono sostituiti, ma che si prevedono 100 passi evolutivi, ognuno con le sue probabilità.

3 - E' possibile che un residuo che ha una bassa probabilità di cambiare resti invariato per molti passi evolutivi.

4 - E' possibile che un aminoacido che ha un'alta probabilità di cambiare si presenti uguale dopo un certo numero di passi evolutivi.

5 - Se si confrontano proteine vicine, si devono usare PAM a basso indice, altrimenti ad alto indice.

## PAM scoring matrix

Derivano dalle PAM di probabilità e sono quelle che si usano davvero nei programmi di allineamento. Non riportano probabilità, ma punteggi (scores), calcolati in base a

$$S(a,b) = \text{int} \left( 10 \times \log \frac{M(a,b)}{C(a,b)} \right)$$

**M(a,b)**, la probabilità ricavata dalla matrice PAM di probabilità.

**C(a,b)** la probabilità di appaiamento casuale, cioè considerando la frequenza media di ogni aminoacido come evento indipendente.

I valori così ottenuti si definiscono **log odds** e permettono agli scores di essere sommati anziché moltiplicati nel calcolo globale dei punteggi di allineamento.

=> score 0 = p(osservato) = p(appaiamento casuale)

=> score < 0 = p(appaiamento casuale) > p(osservato)

=> score > 0 = p(osservato) > p(appaiamento casuale)



## Matrici BLOSUM

1992: Henikoff e Henikoff utilizzarono la banca dati definita BLOCKS (contenente allineamenti multipli di sequenze proteiche senza gap ognuno caratterizzato da varie lunghezze delle proteine e da diversi valori di conservazione minimi) come fonte di osservazione delle probabilità di sostituzione.

$$S(a,b) = \text{int} \left( k \times \log \frac{B(a,b)}{C(a,b)} \right)$$

**B(a,b)**, la probabilità ricavata dall'osservazione dei vari BLOCKS

**C(a,b)** la probabilità di appaiamento casuale, cioè considerando la frequenza media di ogni aminoacido come evento indipendente.

**k**  $3/\log(2)$  oppure  $2/\log(2)$



## Differenze PAM -BLOSUM

- 1 - Mentre le PAM fanno derivare le distanze evolutive dall'assunzione che le mutazioni siano successive, indipendenti e che quindi si possano sommare, le BLOSUM non fanno assunzioni ma si basano sull'osservazione di allineamenti esatti e reali.
- 2 - Mentre l'indice PAM crescente indica scores per proteine più distanti fra loro, esprimendo una distanza evolutiva, l'indice BLOSUM crescente indica proteine più simili tra loro, esprimendo il valore minimo di conservazione del BLOCK.
- 3 - Le PAM tendono a premiare sostituzioni aminoacidiche derivanti da mutazioni di singola base più che motivi strutturali degli aminoacidi, come fanno invece le BLOSUM

## Algoritmi ESAUSTIVI per gli allineamenti a coppie

Far scorrere una sequenza sull'altra, utilizzando un matrice identità a fattore 1 per calcolare gli score.

AAKKQV ->  
AAKWQ

score: 0

AAKKQV  
AAKWQ

score: 0

AAKKQV  
AAKWQ

score: 0

AAKKQV  
AAKWQ

score: 0

**AA**KKQV  
**AA**KWQ

score: 1

**AAKKQV**  
**AAKWQ**

score: 4

**AA**KKQV  
**AA**KWQ

score: 1

AAKKQV  
AAKWQ

score: 0

AAKKQV  
AAKWQ

score: 0

AAKKQV  
AAKWQ

score: 0

Combinazioni possibili:

$$6 \times 5 = 30$$

L'efficienza di un algoritmo in termini di tempo è proporzionale al numero di operazioni necessarie al suo completamento

⇒ Un algoritmo di questo tipo è perfetto da un punto di vista teorico, ma per due proteine di 100 residui ciascuna, richiede 10.000 operazioni.

Un algoritmo così è chiamato ESAUSTIVO e si indica con  $O(nm)$  indicando con  $O$  l'ORDINE dell'algoritmo,  $n$  ed  $m$  le lunghezze delle sequenze. Se  $n=m$  ho  $O(n^2)$  che con i calcolatori di oggi si può attuare. Se ho 3 sequenze ho  $O(n^3)$ , molto più lungo, ma attuabile.

Generalizzando ho  $O(n^k)$ . E' chiaro che un algoritmo così va bene per  $n$  piccoli

Se ho molte sequenze devo rinunciare alla certezza di trovare l'allineamento perfetto con metodi esaustivi e utilizzare un approccio EURISTICO, basato su assunzioni non certe ma che velocizzano gli algoritmi.

Se devo cercare delle similarità in banca dati l'approccio esaustivo è impensabile, dato che la mia proteina query deve essere confrontata con migliaia di proteine.

## Il problema dei gap

Nell'allineare le proteine (ma anche gli acidi nucleici) bisogna tenere in considerazione il fatto che il migliore allineamento si può ottenere con l'inserimento di uno o più gap, che rispondano a fenomeni evolutivi di inserzione o delezione.

```
PLMTRWDQEQESDFGHKLP IYTREWCTRG
| | | | | | | | | | | | | | | |
CHKIPLMTRWPQQESDFGHKLPVIYTREW
```

**Miglior allineamento senza gap: score 13**

```
IPLMTRWDQEQESDFGHKLP - IYTREWCTRG
| | | | | | | | | | | | | | | |
CHKIPLMTRWPQ - QESDFGHKLPVIYTREW
```

**Miglior allineamento con gap: score 24**

## Calcolo degli scores

Date due sequenze A e B con elementi  $(a_1, a_2, a_3, a_L)$  e  $(b_1, b_2, b_3, b_L)$

Score senza gap: 
$$\text{Score} = \sum_{i=1}^L S(a_i, b_i)$$

S: punteggio assegnato dalla matrice utilizzata a quella coppia.  
L: lunghezza della proteina

Score con gap: 
$$\text{Score} = \sum_{i=1}^L S(a_i, b_i) - \sum_{j=1}^G \left( \gamma + \delta [\text{len}(j) - 1] \right)$$

G: numero di gap inseriti

$\gamma$  : penalty per l'apertura del gap

$\delta$  : penalty per l'allungamento del gap

Si può assegnare una penalità all'inserimento dei gap e alla loro estensione cosicché l'algoritmo prenda in considerazione solo inserimenti di gap veramente vantaggiosi.

es.  $\gamma = \delta = 5$

```

IPLMTRWDQEQESDFGHKLP IYTREWCTRG
| | | | | | | | | | | | | | | |
CHKIPLMTRWPQQESDFGHKLPVIYTREW

```

**Miglior allineamento senza gap: score 14**

```

IPLMTRWDQEQESDFGHKLP - IYTREWCTRG
| | | | | | | | | | | | | | | |
CHKIPLMTRWPQ - QESDFGHKLPVIYTREW

```

**Miglior allineamento con gap: score  $24 - (2 \times 5) = 14$**

L'algoritmo per analizzare tutte le possibilità non può essere basato sullo scorrimento perchè richiederebbe una quantità di operazioni e quindi di tempo non adeguata.



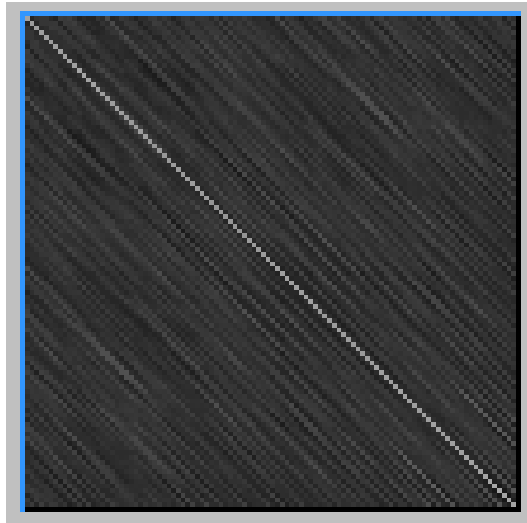
1970: Gibbs, McInter e Fish creano la visualizzazione  
DOT MATRIX

Le due proteine vengono scritte ai margini di una matrice di allineamento  $m \times n$ . Viene annerita una casella della matrice (dot) se gli amino acidi corrispondenti sono identici oppure se si imposta una matrice di identità che calcoli uno score.

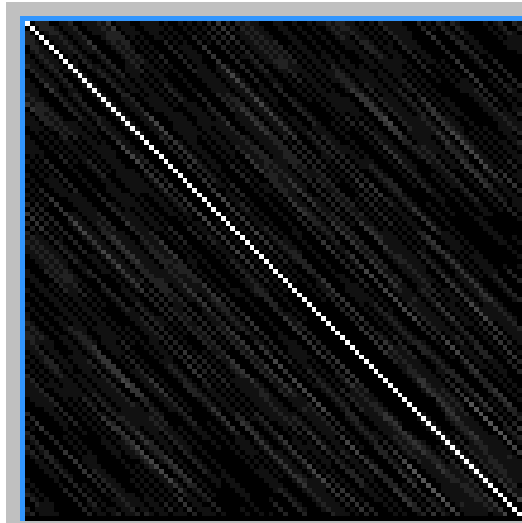
Il miglior programma per la visualizzazione di matrici di allineamento oggi è DOTLET



Le dot matrix sono ancora oggi usate per visualizzare gli allineamenti, ma non danno un valore numerico della qualità dell'allineamento



Allineamento con Blosum 62 di una proteina con se stessa. La diagonale principale indica l'identità, ma lo sfondo è chiaro.



Allineamento con la matrice identità di una proteina con se stessa. La diagonale è ora più marcata e lo sfondo meno confuso



Allineamento tra due proteine diverse, ma con zone in comune

## Alineamenti globali e locali

### Alineamento globale

```
LTGARDWEDIPLWTDWDIEQESDFKTRAFGTANCHK
||.  |  |  |  .  |  .  |  ||  ||  ||  = 14+3
TGIPLWTDWDLQESDNSCNTDHYTREWGTMNAHKA
```

### Alineamento locale

```
LTGARDWEDIPLWTDWDIEQESDFKTRAFGTANCHK
|||||||.||||| = 13+1
TGIPLWTDWDLQESDNSCNTDHYTREWGTMNAHKA
```

Qual'è il migliore tra i due?

Il computer direbbe il primo (score più alto),  
ma biologicamente è vero?

Nonostante lo score più alto, un allineamento di tipo locale nel caso precedente sembra aver trovato un **DOMINIO proteico** che nell'altro caso non è evidente.

Una banca dati è costituita da tantissime protene molto diverse e divergenti tra loro, ma sono accomunate da un numero FINITO di domini strutturali, funzionali, di classe, ecc. che ne permettono una classificazione e un riconoscimento

**⇒ in una banca dati è meglio cercare similarità locali piuttosto che globali.**

⇒ proteine molto simili saranno accomunate da regioni locali così estese che l'allineamento globale darà risultati migliori, perché minimizzerà i gaps rendendo le similarità più leggibili.

## Algoritmi dinamici di allineamento

Sono in grado di produrre il migliore allineamento senza ricalcolare lo score per ogni posizione dello scorrimento, risparmiando così molto tempo.

Il dinamismo in un algoritmo consiste nel fatto che ogni operazione determina l'operazione successiva, scartando una serie di altre operazioni inutili.

1970: **Needleman & Wunsch**: primo algoritmo per similarità globali

1981: **Smith & Waterman**: implementazione del primo per trovare anche similarità locali

## Tre fasi sono alla base degli algoritmi dinamici

date 2 proteine di lunghezza  $n$  e  $m$ , una matrice di sostituzione e un gap penalty

- 1 Produzione di una matrice  $n \times m$  con tutti i residui delle due. In ogni casella si posiziona il punteggio attribuito da una matrice di sostituzione per la coppia di residui corrispondente.
- 2 Per ogni casella trovare il massimo punteggio che si ottiene dai percorsi
  - a) diagonale: lo score della casella precedente in diagonale + il punteggio della casella
  - b) verticale: lo score della casella precedente in verticale – il penalty
  - c) orizzontale lo score della casella precedente in orizzontale – il penalty
- 3 Cercare lo score più alto di tutti tra tutte le caselle procedere verso gli scores più alti a ritroso verso l'inizio della diagonale, scrivendo i residui corrispondenti.

Fase 1: impostare la matrice e riempire ogni casella con i punteggi derivati da un'amatrice di similarità, es. PAM240

	T	F	D	E	R	I	L	G	V	Q
Q	3	1	4	6	5	1	2	2	2	9
T	8	2	3	3	3	3	3	2	4	3
F	2	10	1	1	1	4	4	1	3	1
W	2	5	0	1	1	2	2	2	1	2
E	3	1	6	9	4	1	1	2	2	6
C	3	2	1	1	1	3	3	2	3	1
I	3	4	1	1	1	8	6	1	7	1
K	3	1	3	5	6	1	2	3	2	5
G	2	1	3	2	2	1	0	10	1	2
D	3	1	10	6	3	1	1	3	1	4

## Fase 2

$$\begin{array}{l} \swarrow 3 + 2 = 5 \\ \rightarrow 8 - 5 = 3 \\ \downarrow 1 - 5 = -4 \end{array}$$

$$\begin{array}{l} \swarrow 8 + 4 = 12 \\ \rightarrow 13 - 5 = 8 \\ \downarrow 4 - 5 = -1 \end{array}$$

$$\begin{array}{l} \swarrow 5 + 1 = 6 \\ \rightarrow 18 - 5 = 13 \\ \downarrow 4 - 5 = -1 \end{array}$$

$$\begin{array}{l} \swarrow 8 + 1 = 9 \\ \rightarrow 5 - 5 = 0 \\ \downarrow 19 - 5 = 14 \end{array}$$

$$\begin{array}{l} \swarrow 14 + 2 = 16 \\ \rightarrow 12 - 5 = 7 \\ \downarrow 23 - 5 = 18 \end{array}$$

	T	F	D	E	R	I	L	G	V	Q
Q	3	1	4	6	5	1	2	2	2	9
T	3	1	4	6	5	1	2	2	2	9
F	8	2	3	3	3	3	3	2	4	3
W	8	5	4	7	9	8	4	4	6	5
F	2	10	1	1	1	4	4	1	3	1
E	3	18	13	8	8	13	12	7	7	7
W	2	5	0	1	1	2	2	2	1	2
E	2	13	18	14	9	10	15	14	9	9
E	3	1	6	9	4	1	1	2	2	6
C	3	2	1	1	1	3	3	2	3	1
I	3	5	14	22	28	25	20	15	20	17
I	3	4	1	1	1	8	6	1	7	1
K	3	7	9	17	23	36	31	26	22	21
K	3	1	3	5	6	1	2	3	2	5
G	3	4	10	14	23	31	38	34	29	27
G	2	1	3	2	2	1	0	10	1	2
D	2	4	7	12	18	26	33	48	43	38
D	3	1	10	6	3	1	1	3	1	4
	3	3	14	13	15	21	28	43	49	47

	T	F	D	E	R	I	L	G	V	Q
Q	3	1	4	6	5	1	2	2	2	9
T	8	2	3	3	3	3	3	2	4	3
F	2	10	1	1	1	4	4	1	3	1
W	2	5	6	1	1	2	2	2	1	2
E	3	1	6	9	4	1	1	2	2	6
C	3	2	1	1	1	3	3	2	3	1
I	3	4	1	1	1	1	6	1	7	1
K	3	1	3	5	6	1	2	3	2	5
G	2	1	3	2	2	1	0	10	1	2
D	3	1	10	6	3	1	1	3	1	4

Fase 3: trovare il valore più alto e procedere a ritroso lungo la diagonale cercando lo score maggiore, considerando il salto orizzontale o verticale (gap)

Il valore più alto

Allineamento risultante: **TFDERILGVQ**  
**QTFWECIKGD**

## **Assunzioni implicite nella programmazione dinamica**

1- Se una sequenza è scritta da sinistra a destra e l'altra è scritta dall'alto in basso, qualsiasi percorso valido deve procedere dall'angolo in alto a sinistra all'angolo in basso a destra. Altre strade non sono permesse. Nonostante ogni casella sia circondata da 8 caselle, solo 3 sono valide come percorso di calcolo.

2 - Se il percorso procede in verticale o in orizzontale, solo una casella deve essere conteggiata. Quindi un amino acido non può apparire due volte.

3 - Ogni casella ha uno score che dipende dal percorso a monte ed è indipendente dal percorso a valle. Il miglior percorso a monte di un residuo è sempre determinabile.

## Algoritmo di Needleman & Wunsch per similarità globali

- richiede matrici di sostituzione in cui non sono previsti valori negativi
- lungo le diagonali i valori tendono a crescere, posizionando il valore massimo sempre in nell'ultima riga o nell'ultima colonna, quindi in corrispondenza con l'ultimo residuo di una delle due sequenze.
- richiede un numero di operazioni pari al prodotto delle lunghezze delle sequenze da allineare.
- prende in considerazione in modo esaustivo tutte le possibilità di inserimento di gap.

## Algoritmo di Smith e Waterman per similarità locali

- utilizza matrici di sostituzione con valori anche negativi
- utilizza gli stessi criteri per l'assegnazione degli scores, solo che i percorsi possibili non sono solo le tre caselle a monte, ma c'è anche una nuova origine con score zero, se i valori precedenti sono negativi.
- nella terza fase segue criteri diversi:
  - a) si cercano i valori massimi non solo nelle ultime riga e colonna, perchè gli allineamenti locali possono anche cominciare dentro la sequenza
  - b) si cercano tutti i punti di inizio che stiano sopra ad un certo score soglia, ottenendo più risultati

Soglia di  
rilevazione: 40

miglior allineamento:

VQTYWAECLA  
QTFW-ECIKGD

		V	Q	T	Y	W	A	E	C	L	A
	0	0	0	0	0	0	0	0	0	0	0
Q	0	-2	4	-1	-4	-5	0	3	-6	-2	0
T	0	0	-1	3	-3	-5	1	0	-2	-2	1
F	0	-1	-5	-3	7	0	-4	-6	-5	2	-4
W	0	-6	5	-5	0	7	-6	-7	-8	-2	-6
E	0	-2	3	0	-4	-7	-1	5	-6	-3	-1
C	0	-2	-6	-2	0	-8	-2	-6	-12	-6	-2
I	0	4	-2	0	-1	-5	-1	-2	-2	-2	-1
K	0	-3	1	0	-5	-4	-1	0	-6	-3	-1
G	0	-1	-1	0	-5	-7	1	0	-4	-4	1
D	0	-2	2	0	-4	-7	0	4	-5	-4	0

$\swarrow$  0 - 5 = -5  
 $\rightarrow$  -6 - 5 = -11  
 $\downarrow$  -5 - 5 = -10  
 $\Rightarrow$  0