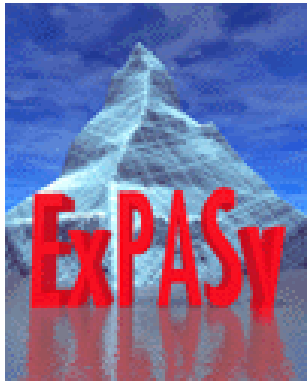


Analisi della  
struttura  
primaria delle  
proteine

La maggior parte degli strumenti per l'analisi della struttura primaria si trovano on-line sul sito

[www.expasy.org](http://www.expasy.org)



## ExPASy Molecular Biology Server

The ExPASy (Expert Protein Analysis System)

### Tools and software packages

- **Proteomics and sequence analysis tools**
  - [Proteomics](#) [[PeptIdent](#), [PeptideMass](#), ...]
  - [DNA -> Protein](#) [[Translate](#)]
  - [Similarity searches](#) [[BLAST](#)]
  - [Pattern and profile searches](#) [[ScanProsite](#)]
  - [Post-translational modification and topology prediction](#)
  - [Primary structure analysis](#) [[ProtParam](#), [pI/MW](#), [ProtScale](#)]
  - [Secondary and tertiary structure prediction](#) [[SWISS-MODEL](#), [Swiss-PdbViewer](#)]
  - [Alignment](#) [[T-COFFEE](#), [SIM](#)]
  - [Biological text analysis](#)

## DNA -> Protein

**Translate:** traduzione in silico di sequenze nucleotidiche, con possibilità di scegliere i frames e le diverse visualizzazioni (1 lettera, 3 lettere, carattere diverso per lo stop, degenerazioni)

**Backtranslation:** data una sequenza aminoacidica, cerca di “indovinare” la sequenza nucleotidica chiedendo l’organismo di appartenenza, la tavola di uso dei codoni e altre informazioni utili.

**Graphical Codon Usage Analyser:** analizza la strategia di utilizzo dei codoni per ottenere considerazioni sulla espressività della sequenza codificante.

**CAI calculator:** analizza una sequenza codificante per capire, in base a come utilizza i codoni, quanto il gene sarà espresso (non è linkato in expasy).

## Strategia dell'analisi dell'uso dei codoni

Anche se il codice genetico è degenerato, non tutti i codoni sinonimi vengono utilizzati allo stesso modo.

L'abbondanza relativa dei tRNA che portano un dato codone è altamente specie-specifica, quindi questo influenza l'utilizzo dei codoni in quella specie.

E' stato dimostrato che esiste una relazione diretta tra l'espressione di una proteina e l'utilizzo dei codoni della sua sequenza codificante: un gene altamente espresso tenderà ad utilizzare i codoni che hanno una maggior rappresentanza di anticodoni, ossia una maggior abbondanza di tRNA.

**=> Se conosco i geni altamente espressi di un organismo, posso analizzare le loro sequenze codificanti alla ricerca dei codoni migliori, e utilizzare questa tabella per predire l'espressività di una qualunque sequenza codificante**

## Protein identification and characterization

**MultiIdent tool:** permette di individuare una proteina non dalla sequenza ma da dati sperimentali come:

- una serie di pesi molecolari originati dalla frammentazione con proteasi
- un punto isoelettrico stimato e un peso molecolare stimato
- la composizione percentuale dei suoi amino acidi

**PeptideMass:** permette di prevedere il peso dei frammenti che si generano da una digestione con proteasi per una più facile identificazione in spettrometria di massa

**PeptideCutter:** analizza tutti i possibili siti di taglio di proteasi o di trattamenti chimici per decidere come lavorare su una proteina

## Similarity searches

**BLAST:** è una serie di link alternativi per tutti i possibili Blast residenti su server diversi da quello della NCBI.

**MPsrch:** ricerca in banche dati modificate usando l'algoritmo di Smith & Waterman

**FASTA3:** ricerca in banca dati usando l'algoritmo FASTA e tutte le sue varianti.

**PropSearch:** ricerca in banca dati per proteine non usando la sequenza, ma una serie di dati che il programma ricava (tipo composizione, struttura se presente, idrofobicità ecc.). Utilizza una rete neurale.

**SAMBA:** usa l'algoritmo di Smith & Waterman, ma fa tutto via hardware, con un array di 128 processori e una architettura dedicata.

## Pattern and profile searches

**InterProScan:** permette la ricerca incrociata in tutte le banche dati di profili della InterPro

**Pfam HMM search:** analizza la banca Pfam utilizzando gli Hidden Markov Models.

**FPAT:** cerca nelle banche dati utilizzando le espressioni regolari (in formato ProSite) fornite dall'utente.

**TEIRESIAS:** data una serie di proteine, cerca di generare il maggior numero di pattern proteici possibili condivisi dalla maggior parte di essi. I pattern ottenuti possono poi essere utilizzati per la ricerca in banca dati.

## Post-translational modification prediction

**ChloroP, LipoP, SignalP:** algoritmi per la ricerca di sequenze segnale data la sequenza primaria, disponibili nel centro [CBS](#) insieme a molti altri strumenti di predizione.

**MITOPROT, Plasmit, Predotar:** algoritmi per la ricerca di sequenze per l'indirizzamento mitocondriale

**NetAcet, NetOGlyc, NetNGlyc:** algoritmi per la ricerca di sequenze di modificazione post traduzionale.

## Topology prediction

**Psort, TargetP:** algoritmi per la ricerca di caratteristiche che identifichino il compartimento cellulare in cui una proteina dovrebbe trovarsi (es. citoplasma, membrana, nucleo, organelli).

## Primary structure analysis

**ProtParam:** data la sequenza, analizza moltissime caratteristiche chimico-fisiche intrinseche e produce un'analisi statistica delle proprietà della proteina

**ProtScale:** data la sequenza, costruisce dei grafici di scorrimento utilizzando valori specifici per ogni aminoacido corrispondenti a vari aspetti chimico-fisici e non.

**REP:** identifica le regioni ripetute in una proteina, per vedere se può essersi originata es. da una duplicazione genica. Oppure si può cercare in un allineamento multiplo con delle sequenze modello

**RandSeq:** produce una sequenza casuale di cui si può controllare lunghezza e composizione.

# Algoritmi a scorrimento di finestre

ProtScale è un esempio di questi algoritmi.

Ad ogni amino acido è possibile associare un valore numerico che descriva una qualunque caratteristica.

In una proteina l'amino acido in se' non ha una grande importanza, ma conta la sequenza, cioè come i residui si susseguono.

Se un algoritmo scorre lungo una proteina, non ha senso nella maggior parte dei casi guardare il valore del singolo residuo, ma è bene considerare una finestra di dimensioni fissate e per ogni punto (per ogni residuo) associare non tanto il suo valore quanto il valore ottenuto da tutta la finestra (per esempio, sommando i vari punteggi).

Inoltre, è opportuno pesare diversamente i residui che si trovano a distanze diverse rispetto al centro della finestra.

- Valori di idrofobicità di Kyte & Doolittle.
- Finestra di 9 residui
- Peso finestra: lineare

Ala: 1.800  
 Arg: -4.500  
 Asn: -3.500  
 Asp: -3.500  
 Cys: 2.500  
 Gln: -3.500  
 Glu: -3.500  
 Gly: -0.400  
 His: -3.200  
 Ile: 4.500  
 Leu: 3.800  
 Lys: -3.900  
 Met: 1.900  
 Phe: 2.800  
 Pro: -1.600  
 Ser: -0.800  
 Thr: -0.700  
 Trp: -0.900  
 Tyr: -1.300  
 Val: 4.200

MWIDIGDAFLALHNADHKTISHGDLLCPIMLVTK

