

ScanProsite è uno strumento accessibile da Expasy che permette di cercare specifici motivi proteici nel database Prosite di sequenze proteiche. I formati che si possono fornire come identificativi della sequenza sono:

- Knowledgebase UniProt
- Protein Data Bank (PDB)
- sequenza in formato FASTA
- modelli e profili Prosite

Lo strumento ScanProsite permette anche di ricercare allineamenti di motivi proteici specifici nel database di sequenze proteiche. Anche per questo genere di ricerche i formati accettati da ScanProsite sono modelli e profili Prosite o forniti dall'utente. Di default i database in cui vengono analizzate le proteine sono UniProtKB / Swiss-Prot, con in aggiunta le varianti generate da splicing alternativo. In alternativa possono essere scelti altri database da sottoporre ad analisi come TrEMBL e/o PDB.

- UniProtKB/Swiss-Prot  including splice variants  
 UniProtKB/TrEMBL  PDB

E' inoltre possibile regolare i propri limiti ricerca specificando le opzioni filtro e pattern.

**Filter(s):**

- On taxonomy:  (e.g. *Eukaryota*; *Escherichia coli*;)
  - On description:  (e.g. *protease*)

**Pattern option(s):**

- Allow at most  X sequence characters to match a conserved position in the pattern
- Match mode

Le modalità di scansione effettuabili con ScanProsite possono essere:

1. VELOCE (quick scan mode)
2. AVANZATA (Advanced Scan mode)

**Quick Scan mode**

**PROSITE access**

add wildcard "\*"


Browse:

- by documentation entry
- by ProRule description
- by taxonomic scope
- by number of positive hit

**PROSITE tools**

**Scan a sequence against PROSITE patterns and profiles - quick scan**


(Output includes graphical view and feature detection)



Enter your sequence or a UniProtKB (Swiss-Prot or TrEMBL) ID or AC [ help ]:

exclude patterns with a high probability of occurrence

- **ScanProsite** - advanced scan
- **PRATT** - allows to interactively generate conserved patterns from a series of unaligned proteins..
- **MyDomains - Image Creator** <sup>new</sup> - allows to generate custom domain figures.



È possibile inserire una o più sequenze (1 per riga fino ad un massimo di 8 sequenze) per eseguire la scansione contro i modelli e profili Prosite.

Se si desidera eseguire la scansione della sequenza contro tutti gli elementi Prosite si deve incollare la sequenza nella casella di testo. Il formato di sequenza da inserire può essere, in formato FASTA UniProtKB (Swiss-Prot o TrEMBL) o sequenza PDB. In tutti i casi i dati in ingresso saranno esaminati e confrontati con tutti gli elementi Prosite. Se si desidera eseguire la scansione di una sequenza contro tutti gli elementi Prosite tranne quelle con una alta probabilità di *occurrence*, si deve selezionare la casella sotto il pulsante di scansione:

exclude patterns with a high probability of occurrence

## Advanced Scan mode



Home ScanProsite ProRule Documents Downloads Links Fur

Sequence(s) to be scanned:
Motif(s) to scan for:

Enter:

- UniProtKB(Swiss-Prot and TrEMBL) AC and/or ID  
(e.g. P00747, ENTK\_HUMAN)
- PDB identifier(s)
- your own protein sequence(s)

Exclude motifs with a high probability of occurrence

Do not scan profiles

Enter:

- PROSITE AC and/or ID (e.g. P55088, CHEB)
- your own pattern(s)

Protein database(s):

UniProtKB/Swiss-Prot  including splice variants

UniProtKB/TrEMBL  PDB

randomize databases

excluding fragments

Filter(s):

- On taxonomy:  (e.g. Eukaryota; Escherichia coli;)
- On description:  (e.g. protease)

Pattern option(s):

- Allow at most  X sequence characters to match a conserved position in the pattern
- Match mode

Output:

Show low level score

Retrieve complete sequences

Your e-mail:

Format

Show only sequences with at least  hit(s)

Maximum of matched sequences

Se si desidera eseguire la scansione della sequenza contro tutti i motivi Prosite:

Incollare la sequenza nella casella di testo della sezione 'Sequence (s) to be scanned'. La sequenza deve essere o (solo aa, 1 lettera di identificazione, senza numeri), in formato FASTA, o in formato UniProtKB. Cliccando poi su 'avviare la Scansione' la sequenza sarà scansionata contro tutti i motivi Prosite tranne quelli con una alta probabilità di *occurrence*. Di default, gli *hit* di profili Prosite con un punteggio superiore al punteggio soglia definito saranno mostrati, ed i risultati saranno catalogati in modalità 'graphical rich view' (è possibile modificare questi parametri).

- Se si desidera eseguire un'analisi di sequenze UniProtKB (Swiss-Prot o TrEMBL) / PDB contro tutti i motivi Prosite: inserire l'identificativo della sequenza (CA P05130 ad esempio) o carta d'identità (per esempio ENTK\_HUMAN) o incollare la sequenza, nella casella di testo della sezione 'Sequence (s) to be scanned'.
- Se si desidera eseguire la scansione del UniProtKB con un motivo proteico particolare, digitare il pattern nella casella di testo della sezione 'motif (s) to scan for'. Si dovrebbe usare la sintassi del modello Prosite e digitare il modello su una sola riga (non digitare 'Invio' all'interno del modello).

E' possibile acquisire sequenze multiple nello stesso momento (massimo 8 sequenze se la scansione è contro tutti i motivi Prosite, 16 se la scansione è contro più di 1 motivo, 1000 se contro un unico motivo).

Se si desidera eseguire la scansione di varie sequenze, è necessario inserirle in formato FASTA o UniProtKB. I parametri di punteggio e di visualizzazione dei risultati possono essere modificati. Il valore di soglia ad esempio dei punteggi può essere abbassato per includere nuovi risultati e il formato di visualizzazione degli output può essere scelto tra:

- rich format view (di norma formato di visualizzazione di default), HTML con rappresentazione grafica degli hits.
- HTML output, visualizzazione senza rappresentazione grafica.
- Text output: visualizzazione solo di testo (senza nessun link HTML)
- Text FASTA output: visualizzazione del solo testo in formato FASTA

In rete è accessibile un manuale per l'uso di ScanProsite dal link <http://expasy.org/tools/scnpsit3.html>

### ESEMPIO:

Hits for all PROSITE (release 20.67) motifs on sequence tesina :

found: 1 hit in 1 sequence


TESINA (290 aa)

```
MAALRULLSCVRCGLRPPVRCFAMRPFASGANFEYIIAEKRGKONTVGLIQLNRPKALNALCDGLI  
DELNQLAKT FEEDPAUGA IULTGGDKAFAGADIKEMQNLSPQDCYSSKTLKHWDHLTQKGFPIA  
IANGYAFGGGCELAMCDI IYAGEKAQFAQPE ILIGTIPGAGGTQLRTRAVGKSLAMENULTGDR I  
SAQDAKQAGLUSKICVETLVEEAIQC AEKIASNSKIUVAM&ESUN&AFEMTLTEGSKLEKILFY  
STFATDDRKEGATAFVDRKKNFYDQ
```

ruler: 1 100 200 300 400 500 600 700 800 900 1000

hits by patterns: [1 hit (by 1 pattern) on 1 sequence]

Hits by PS00166 ENOYL\_COA\_HYDRATASE Enoyl-CoA hydratase/isomerase signature:

TESINA  (290 aa)

131 - 151: IANGYAFGGGCELAMCDI

Legend:

- disulfide bridge
- active site
- other 'ranges'
- other sites

horizontal scaling: 0.6

do not show text labels:

do not show sites in hits:

do not show ranges in hits:

redisplay

Nell'esempio mostrato la sequenza proteica analizzata è la ENOYL COA HYDRATASE. Il tool ScanProsite utilizzato è in versione advanced. La sequenza viene fornita (in formato FASTA) e il risultato è visualizzato nello screenshot sovrastante. Viene riconosciuta una banda (colore arancione) corrispondente al sito attivo della proteina e cliccandoci sopra si apre una pagina di PROSITE che riassume caratteristiche, funzioni e relazioni evolutive della proteina,

## Enoyl-CoA hydratase/isomerase signature

### Description:

Enoyl-CoA hydratase (EC 4.2.1.17) (ECH) [1] and D3,D2-enoil-CoA isomerase (EC 5.3.3.8) (ECI) [2] are two enzymes involved in fatty acid metabolism. ECH catalyzes the hydration of 2-trans-enoil-CoA into 3-hydroxyacyl-CoA and ECI shifts the 3- double bond of the intermediates of unsaturated fatty acid oxidation to the 2-trans position.

Most eukaryotic cells have two fatty-acid  $\beta$ -oxidation systems, one located in mitochondria and the other in peroxisomes. In mitochondria, ECH and ECI are separate yet structurally related monofunctional enzymes. Peroxisomes contain a trifunctional enzyme [3] consisting of an N-terminal domain that bears both ECH and ECI activity, and a C-terminal domain responsible for 3-hydroxyacyl-CoA dehydrogenase (HCDH) activity.

In *Escherichia coli* (gene *fadB*) and *Pseudomonas fragi* (gene *faoA*), ECH and ECI are also part of a multifunctional enzyme which contains both a HCDH and a 3-hydroxybutyryl-CoA epimerase domain [4].

A number of other proteins have been found to be evolutionarily related to the ECH/ECI enzymes or domains:

- 3-hydroxybutyryl-coa dehydratase (EC 4.2.1.55) (crotonase), a bacterial enzyme involved in the butyrate/butanol-producing pathway.
- Naphthoate synthase (EC 4.1.3.36) (DHNA synthase) (gene *menB*) [5], a bacterial enzyme involved in the biosynthesis of menaquinone (vitamin K2). DHNA synthase converts O-succinyl-benzoyl-CoA (OSB-CoA) to 1,4-dihydroxy-2-naphthoic acid (DHNA).
- 4-chlorobenzoate dehalogenase (EC 3.8.1.6) [6], a *Pseudomonas* enzyme which catalyzes the conversion of 4-chlorobenzoate-CoA to 4-hydroxybenzoate-CoA.
- A *Rhodobacter capsulatus* protein of unknown function (ORF257) [7].
- *Bacillus subtilis* putative polyketide biosynthesis proteins *pkSH* and *pkSL*.
- *Escherichia coli* carnitine racemase (gene *carD*) [8].
- *Escherichia coli* hypothetical protein *ygfG*.
- Yeast hypothetical protein YDR036c.

## visualizza il pattern funzionale della regione evidenziata

Consensus pattern: [LIVM]-[STAG]-x-[LIVM]-[DENQRHSTA]-G-x(3)-[AG](3)-x(4)-[LIVMST]-x-[CSTA]-[DQHP]-[LIVMFYA]

e fornisce una serie di link ad articoli riguardanti la tipizzazione della macromolecola.

### References:

- 1 *Authors* Minami-Ishii N., Taketani S., Osumi T., Hashimoto T.  
*Source* Eur. J. Biochem. 185:73-78(1989).
- 2 *Authors* Mueller-Newen G., Stoffel W.  
*Source* Biol. Chem. Hoppe-Seyler 372:613-624(1991).
- 3 *Authors* Palosaari P.M., Hiltunen J.K.  
*Title* Peroxisomal bifunctional protein from rat liver is a trifunctional enzyme possessing 2-enoil-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, and delta 3, delta 2-enoil-CoA isomerase activities.  
*Source* J. Biol. Chem. 265:2446-2449(1990).  
*PubMed ID* [2303409](#)
- 4 *Authors* Nakahigashi K., Inokuchi H.  
*Title* Nucleotide sequence of the *fadA* and *fadB* genes from *Escherichia coli*.  
*Source* Nucleic Acids Res. 18:4937-4937(1990).  
*PubMed ID* [2204034](#)