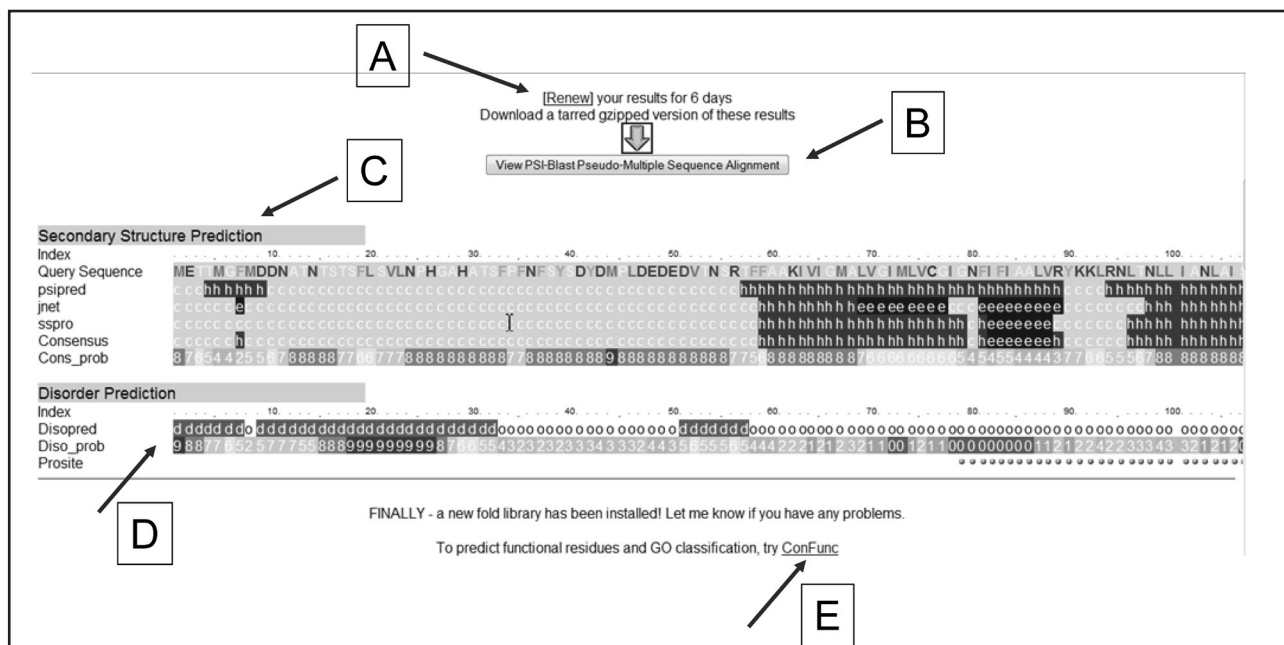


## 9. Riconoscimento di *fold*


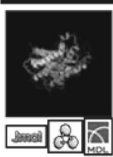

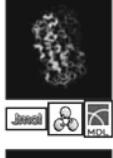

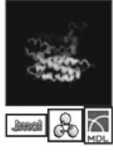
In questa esercitazione utilizzeremo alcuni sistemi di riconoscimento di *fold* tra quelli descritti nel testo.

- Collegatevi al sito [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov) ed estraete la sequenza proteica con codice gi:33112428, corrispondente al codice UNIPROT:PRK1\_HUMAN. Visualizzate la sequenza in formato FASTA. Lo scopo è verificare se la sequenza sia compatibile con qualche struttura proteica nota.
- Collegatevi al sito di PHYRE [www.sbg.bio.ic.ac.uk/phyre/index.cgi](http://www.sbg.bio.ic.ac.uk/phyre/index.cgi), inserite i dati richiesti (indirizzo di posta elettronica, ecc.) e copiate la sequenza in formato FASTA (senza la linea iniziale di commento) nell'apposito spazio. A seconda del carico di lavoro del server, l'elaborazione può durare pochi minuti o alcune ore. In ogni caso, al termine dell'analisi il sistema invia un messaggio di posta elettronica all'indirizzo indicato con un link alla pagina web che mostra i risultati (la permanenza dei risultati sul server è temporanea ma può essere estesa a 6). Porzioni della pagina dei risultati sono mostrate nelle **Figure 1 e 2**.



**Figura 1** Prima parte della pagina dei risultati prodotta dall'elaborazione di PHYRE. Il link (A) consente di confermare la permanenza dei risultati sul server per ulteriori 6 giorni. Il pulsante (B) mostra l'allineamento multiplo calcolato da PSI-BLAST e utilizzato per la ricerca nelle banche dati. Nella sezione (C) è riportata la previsione della struttura secondaria ottenuta con tre metodi, Psi-Pred, SSpro e JNet insieme a quella consenso. Anche questa previsione è utilizzata insieme al profilo nella ricerca in banca dati. I simboli indicano elica  $\alpha$  (h), foglietto  $\beta$  (e) e struttura aperiodica (c). Nella sezione (D) viene mostrata la previsione delle regioni conformazionalmente disordinate della proteina. Il collegamento (E) attiva la pagina relativa al server ConFunc ([www.sbg.bio.ic.ac.uk/~confunc/](http://www.sbg.bio.ic.ac.uk/~confunc/)), messo a punto dallo stesso gruppo di PHYRE, che consente di prevedere i residui funzionalmente importanti della proteina analizzata.

I risultati indicano che la sequenza sonda è compatibile con la struttura dei recettori a 7 eliche transmembrana, come la rodopsina o i recettori beta adrenergici accoppiati alle proteine G. Si può visualizzare il modello costruito dal programma in corrispondenza di ciascuno stampo attraverso l'icona *Jmol*.

Fold Recognition								
View Alignments	SCOP Code	View Model	E-value	Estimated Precision	BioText	Fold/PDB descriptor	Superfamily	Family
	d1u19a (length:348) 15% i.d.		1.4e-35	100 %	n/a	Family A G protein-coupled receptor-like	Family A G protein-coupled receptor-like	Rhodopsin-like
	d1ln6a (length:309) 15% i.d.		7.4e-34	100 %	n/a	Family A G protein-coupled receptor-like	Family A G protein-coupled receptor-like	Rhodopsin-like
	c2r4sA (length:342) 20% i.d.		8.6e-33	100 %	n/a	<b>PDB header:</b> signaling protein	<b>Chain:</b> A; <b>PDB Molecule:</b> beta-2 adrenergic receptor;	<b>PDBTitle:</b> crystal structure of the human beta2 adrenoceptor

**Figura 2** Parte relativa all'identificazione dell'omologia strutturale e del riconoscimento di *fold*. Nella colonna (A) è riportato un collegamento all'allineamento tra la sequenza sonda e la sequenza della struttura della banca dati, oltre a diverse altre informazioni sulla qualità dell'allineamento, sui possibili siti funzionali e sulle zone di superficie che costituiscono avvallamenti o tasche. Nella colonna (B) è riportata la percentuale di identità tra la sequenza sonda e lo stampo strutturale trovato. Il codice dello stampo nella banca dati SCOP ha la seguente struttura: [d/c][codice PDB][identificatore di catena] [numero del dominio]. Le iniziali 'd' o 'c' indicano se lo stampo consiste in un dominio di una struttura più grande contenuto in SCOP oppure in una struttura completa. L'eventuale numero di dominio è l'indice assegnato da SCOP per identificare quel particolare dominio all'interno di una proteina multidominio. Nella colonna (C) sono riportati i collegamenti per visualizzare il modello della proteina bersaglio costruito sulla base dell'allineamento con lo stampo strutturale. L'icona *Jmol* attiva un semplice programma Java attraverso il quale si può visualizzare e manipolare il modello. Alternativamente si possono trasferire sul proprio computer le coordinate del modello in formato pdb, cliccando sull'icona con l'immagine del modello. Le colonne (D) ed (E) indicano l'*E-value* relativo al confronto e il grado di omologia con lo stampo in questione. Le caselle di (E) sono colorate in base a una scala che va dal rosso al blu (alta e bassa confidenza). Questi valori indicano la verosimiglianza dell'omologia ma non forniscono alcuna indicazione sull'accuratezza del modello tridimensionale, che dovrebbe comunque essere considerato un'ipotesi approssimativa della struttura reale. Le altre colonne indicano la classificazione strutturale dello stampo.

- Collegatevi al sito <http://bioinf4.cs.ucl.ac.uk:3000/psipred> e selezionate la voce *Fold Recognition - pGenThreader* nella sezione *Choose Prediction Method*. Il sistema mostrerà la pagina in cui è possibile immettere la sequenza sonda (**Figura 3**) Anche in questo caso si deve indicare il proprio indirizzo di posta elettronica, in modo da ricevere la notifica del termine dell'analisi con i relativi risultati (che è possibile visualizzare in diversi formati). I risultati «migliori» vengono riportati nella forma mostrata in **Figura 4**. Anche in questo caso sono state identificate strutture con architettura a 7 eliche transmembrana. Si noti come i potenziali di solvatazione tendano a dare valori positivi (che indicherebbero instabilità). Si può ipotizzare che la ragione risieda nel fatto che la proteina è di membrana, per cui espone sulla superficie di contatto con i fosfolipidi residui idrofobici. Poiché i potenziali originali sono stati calcolati su un insieme di proteine globulari solubili, la presenza di residui idrofobici sulla superficie è un evento poco frequente e quindi, secondo il formalismo della funzione inversa di Boltzmann (Capitolo 11), energeticamente sfavorevole.

- Ora esamineremo il caso in cui le ricerche producono un risultato negativo. Recuperate la sequenza Q3ZTS4\_HUMAN dalla banca dati UNIPROT (www.uniprot.org) e ripetete le due ricerche che abbiamo descritto. In questo caso nessuno dei due sistemi trova somiglianze significative con la sequenza sonda. Le ragioni possono essere sostanzialmente due: 1) La proteina è ipotetica e la sua esistenza non è certa, come si può leggere nelle annotazioni della scheda di UNIPROT. 2) Non è stata ancora risolta una struttura omologa.

**The PSIPRED Protein Structure Prediction Server**

The PSIPRED protein structure prediction server allows you to submit a protein sequence, perform a prediction of your choice and receive the results of the prediction via e-mail. You may select one of three prediction methods to apply to your sequence: PSIPRED - a highly accurate method for protein secondary structure prediction, MEMSAT - our widely used transmembrane topology prediction method and GenTHREADER - a sequence profile based fold recognition method. More...

**For queries regarding PSIPRED:** [psipred@cs.ucl.ac.uk](mailto:psipred@cs.ucl.ac.uk)

**Choose Prediction Method**

- Predict Secondary Structure (PSIPRED v2.6)
- Predict Transmembrane Topology (MEMSAT3)
- SVM Prediction of Transmembrane Topology (MEMSATSVMSVM)
- Fold Recognition (GenTHREADER - quick)
- Fold Recognition (pGenTHREADER - with profiles and predicted secondary structure)
- Fold Recognition (pDomTHREADER - annotates multiple domain on chains)

Help...

**Input Sequence (single letter amino acid code)**

```
NYLRTVSYVSTNALLAIAIDRYLAIVHPLRPRMKCQTATGLIALVNTVSIILAIIPSAFYTTETVLVIVK
SQEKIFCGQIWFVDQQLYYKSYFLFIFGIEFVGPVVTMLCYARISRELWFKAVPGFQTEQIRKRLACRR
KTVLVLMCILTAYVLCWAPFYGFYIVRDFPFVVFVKEKHYLTAFYIVECIAMSNMINTLCEFTVVKNDTV
KYFKKIMLLHWKASVYNGGKSSADLDLKIIGMPATEEVDCIRLK
```

Help...  
If you wish to test these services follow this link to retrieve a test fasta sequence.

**Figura 3** Pagina iniziale del server PSIPRED che offre accesso a diversi servizi di analisi di sequenza. In questa esercitazione selezioniamo la voce indicata dalla freccia corrispondente al programma pGenTHREADER.

**UCL Department Of Computer Science**  
**Bioinformatics Group**

Search Group: UCL Home >> Departments of Computer Science >> Bioinformatics Group >> psipred

**Results for mgentheader job 11101**

**Gentheader Scores**

Displaying only results rated MEDIUM confidence or better. Full results are sent by email.

Conf.	Net Score	p-value	PairE	SolvE	Aln Score	Aln Len	Str Len	Seq Len	View Alignment	SCOP Codes	CATH Codes	Structure	Functional Data
CERT	83.153	2e-07	-513.5	13.3	350.0	337	348	393	<a href="#">Start Jalview</a>	<a href="#">Search SCOP for 1u19A</a>	<a href="#">Search CATH for 1u19A</a>		<a href="#">Function Link</a>
CERT	78.523	6e-07	-506.9	10.1	327.0	329	350	393	<a href="#">Start Jalview</a>	<a href="#">Search SCOP for 2z73A</a>	<a href="#">Search CATH for 2z73A</a>		<a href="#">Function Link</a>
CERT	71.719	3e-06	-498.8	13.4	283.0	272	274	393	<a href="#">Start Jalview</a>	<a href="#">Search SCOP for 2vt4A</a>	<a href="#">Search CATH for 2vt4A</a>		<a href="#">Function Link</a>
HIGH	51.132	3e-04	-245.2	9.5	197.0	215	216	393	<a href="#">Start Jalview</a>	<a href="#">Search SCOP for 2r4rA</a>	<a href="#">Search CATH for 2r4rA</a>		<a href="#">Function Link</a>

1 2 3 4 5 6 7 8 9 10 11 12 13 14

**Figura 3** Risultato in formato HTML di pGenTHREADER. Le colonne numerate indicano:

- 1) Grado di confidenza dell'omologia. CERT = grado elevato corrispondente a un  $P\text{-value} < 10^{-4}$ ; HIGH =  $P\text{-value} < 10^{-3}$ .
- 2) Punteggio emesso dalla rete neurale (vedi Capitolo 11).
- 3)  $P\text{-value}$ , inteso come la probabilità di osservare per caso un punteggio maggiore o uguale a quello ottenuto.
- 4) Potenziale di coppia.
- 5) Potenziale di solvatazione.
- 6) Punteggio dell'allineamento sequenza sonda- sequenza stampo.
- 7) Lunghezza dell'allineamento.
- 8) Lunghezza della sequenza stampo.
- 9) Lunghezza della sequenza sonda.
- 10) Attivazione del programma java Jalview per visualizzare l'allineamento.
- 11) Link alla struttura stampo nella banca SCOP.
- 12) Link alla struttura stampo nella banca CATH.
- 13) Link alla struttura stampo nella banca PDBsum.
- 14) Link a informazioni di carattere funzionale.

