

6. Previsione di una struttura secondaria

In questa attività utilizzeremo alcuni programmi per prevedere la struttura secondaria di una proteina. Faremo riferimento inizialmente a un algoritmo che appartiene a una generazione precedente agli algoritmi basati sulle reti neurali.

- Collegatevi al sito <http://www.ncbi.nlm.nih.gov> e recuperare la sequenza proteica con codice **P98163**; salvatela sul vostro calcolatore in formato FASTA.
- Collegatevi al sito http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html che mette a disposizione un servizio di previsione di struttura secondaria basato sull'algoritmo GOR (dalle iniziali dei cognomi degli autori Garnier, Osguthorpe, Robson). Questo metodo utilizza un algoritmo di generazione precedente a quelli basati sulle reti neurali. Incollate la sequenza nell'apposito campo e lanciate il programma. Il risultato dovrebbe comparire in pochi secondi (**Figure 1** e **2**). La descrizione dell'interpretazione dei risultati è nella didascalia delle figure.



Figura 1 Previsione della struttura secondaria con il metodo GOR. La sequenza è riportata con il codice amminoacidico monoletterale. Sotto ogni residuo è riportata la conformazione prevista: h indica α -elica, b foglietto β e c struttura aperiodica.

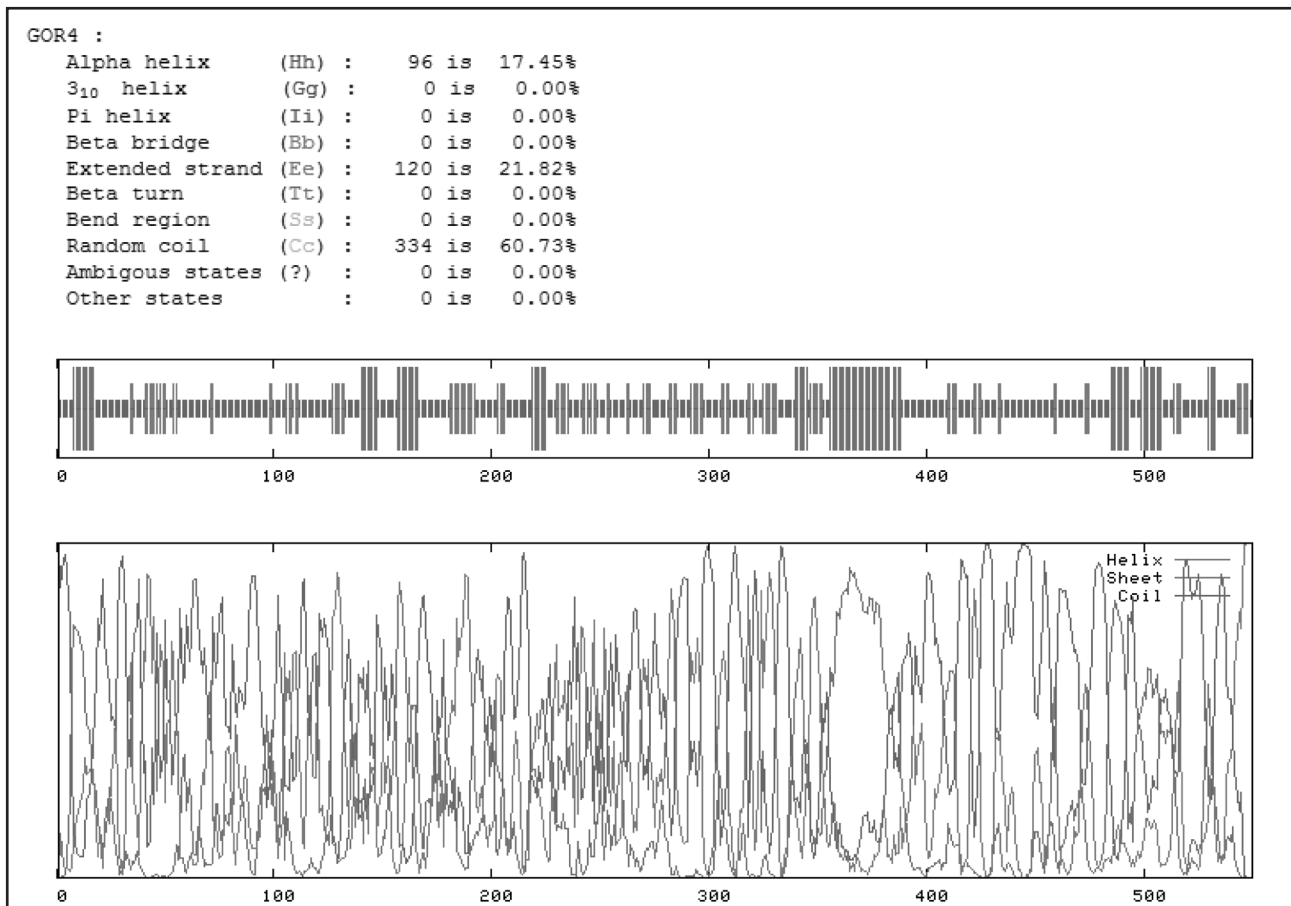


Figura 2 Statistiche sul contenuto di strutture secondarie previsto per la sequenza. Nel pannello centrale è riportata la schematizzazione della distribuzione delle diverse strutture secondarie lungo la sequenza. Nel pannello inferiore è mostrato il grafico della distribuzione delle probabilità di ciascuna delle tre conformazioni calcolate dal programma GOR. Ai residui è assegnata la struttura secondaria che ha la massima probabilità.

Nella prossima esercitazione utilizzeremo i sistemi di previsione più evoluti basati su reti neurali e Modelli di Markov Nascosti.

- Collegatevi al sito www.compbio.dundee.ac.uk/~www-jpred/advanced.html. Incollate la sequenza nel riquadro apposito. Spuntate l'opzione relativa alla scritta *Skip searching PDB before prediction* in modo da disabilitare la ricerca di sequenza omologhe nella Protein Data Bank. Così facendo si forza il programma a calcolare la previsione a partire dalla sola sequenza. Nella realtà la strategia migliore da seguire è verificare prima se esista una sequenza a struttura nota omologa da utilizzare come stampo per l'assegnazione della struttura secondaria. Questo è il modo più accurato e rigoroso di procedere. Poiché il server può impiegare diverso tempo per effettuare la previsione, si consiglia di immettere il proprio indirizzo di posta elettronica al quale arriverà la notifica di completamento della procedura di previsione. Una volta completata la previsione, i risultati possono essere visualizzati con modalità diverse o scaricati direttamente sul proprio computer (**Figure 3 e 4**).

Confrontate i risultati ottenuti con questo metodo e con il precedente e notare, se ve ne sono, le differenze.

Results

After much trouble and strife, Bob the scheduling penguin has retrieved your results! Rejoice. For your pleasure the following viewing options are available. You may bookmark this page for future reference although data is not kept on the server for more than two days.

Jalview	Full HTML	Simple HTML	Postscript	PDF	Raw Data	Download
<input type="button" value="Start Jalview"/>	<input type="button" value="View All"/>	View Simple	View PS	View Pdf	see data	here

This Jpred prediction was made with following.

Jnet version: 2.2
UniRef90 release: 15.4



Figura 3 Schermata di Jpred da cui è possibile visualizzare i risultati con diverse modalità. Per esempio, si possono visualizzare come file PDF, postscript, come allineamento in formato Jalview, oppure è possibile scaricare tutti i dati in formato compresso.

```

UniRef90 UPI000180BA03 : ----CFCKRVFQICDDELCEVEQPEQCFTCPDRMTM-----YKFSQLCDNKNDCLDGSDEMCKDDFVCGNGNM-----VTEALRCDYIDDCGDTTDECNCA
UniRef90 UPI00017C3284 : ----CHTGGKLNYSLVCDGYDDCQNCDCNPEEHRGCGD--RCIAIEWVCDGDHDCVDKSDEVNCSCHSQGLVECRNGQCIPTSFQCDGDEDCCKDGSDEFSCSD
UniRef90 UPI00015B62C5 : ----CDGYKNCQDNSDEVDCTEERCKNM---GKLCCKNRDV-CFPESAICNGRNDCCGNSDEENCT-----EKRCNELRKFCKSVNMCIPHHWTCDG
UniRef90 UPI0000ECD773 : ----CSCPGLTIAADGRMCDECVLGRHACHTDQ-DCENGSYHCV---LRCDGLS-CQDINECQESNSCHQHCINTVGSFHCDCDPGYQCRPDQLCKNTRGGYKC
UniRef90 Q7Z4F1 : ----CLQEEFQQLN--HRCVSAVQRCDGVDACGGDGSDEAGCSSDPFPTPRPVPSLPCNVLTLEDYGVFSSPGYTHLASVS
UniRef90 UPI00005EAEC0 : ----PNCSEPECFGN---CNNR-GQCI-----DGQVCDEGFTGEDCGQLACPSDCNDQGGKCVHGACVCFEGYTGEDCSEELCPVP
UniRef90 UPI00005A28C9 : ----CQCYSGFTLAEDGKICA---AVDYCASENHGC---EHECVDSYFCRGTFLNPDKKTCAKIDYCASPNQGCQHECVNTDDSYSCRCLKGFTLNPDKKTCTCR
UniRef90 B6MRG0 : ----TCPAHLV-----CEDITPGYRCACQSGHHM--DRYGRCV-----DTDECLVDNNGGCQRLCVNTAGSYHCE---CGGDVYLSEDMACITDINPCERQNK
UniRef90 O08746 : ----CKCKQGYILSTDQKTCCEQLCVNCQCYSGYTLAEDGMDYCASENHGC--EHECVNAESSYLRCHEGFALNSDKKTKSKIDYCASNNHGQHECVNAQTS
UniRef90 Q80YX1 : ----TGEDCSELICPNDCFRINGTCYCLTCPNDCCQGRGQCEEGQVCVNEGFAGADCSEKRCPADCHHRGRCLNGQCECDDGFTGADCG

OrigSeq : 1-----11-----21-----31-----41-----51-----61-----71-----81-----91-----101--
          : MVPKADSGAFLLLFLVLTVTEPLRPELRCNPGQFACRSGTIQCIPLPWQCDGWATCEDESDEANCPVTEVGRPHHGKEAVDPRQGRARGGDP SHFHAVNVAQP

Jnet : -----EEE-----EE-----EEE-----EE-----
jhmm : -----EE-----EE-----
jpsmm : -----EEEE-----EEE-----EE-----EEE-----EEEE-----

Lupas 14 : -----
Lupas 21 : -----
Lupas 28 : -----

Jnet_25 : -B--B-B---BBBBBBB-BBB---B---B---BB-B-B---BB---B-B-B---B---BBBB-B-B-BB-B---B---BBBB---BBB-BBB-----BBB-
Jnet_5 : -----B-BB-----B-----B-----B-----B-----B-----B-----
Jnet_0 : -----
Jnet Rel : 998887888610000254677777765467762450477506503662004677777777510004677777777730000267777777777777776
  
```

Figura 4 Parte del risultato ottenuto con Jalview. L'interpretazione dei simboli è riportata nel capitolo che tratta la previsione della struttura secondaria delle proteine (Capitolo 9).

- Attraverso il programma **TMHMM**, accessibile dalla pagina www.cbs.dtu.dk/services/TMHMM-2.0, verificheremo se la sequenza contenga potenziali eliche transmembrana, come potremmo attenderci da un recettore. Incollate la sequenza in formato FASTA nell'apposito riquadro ed eseguite il programma. In genere i risultati vengono presentati in pochi secondi. Nella **Figura 5** è riprodotta la schermata che li riporta. Il programma prevede la presenza di un'elica transmembrana intorno alla posizione di sequenza 350. C'è anche una certa probabilità che sia presente una breve elica transmembrana all'estremità N-terminale della proteina.

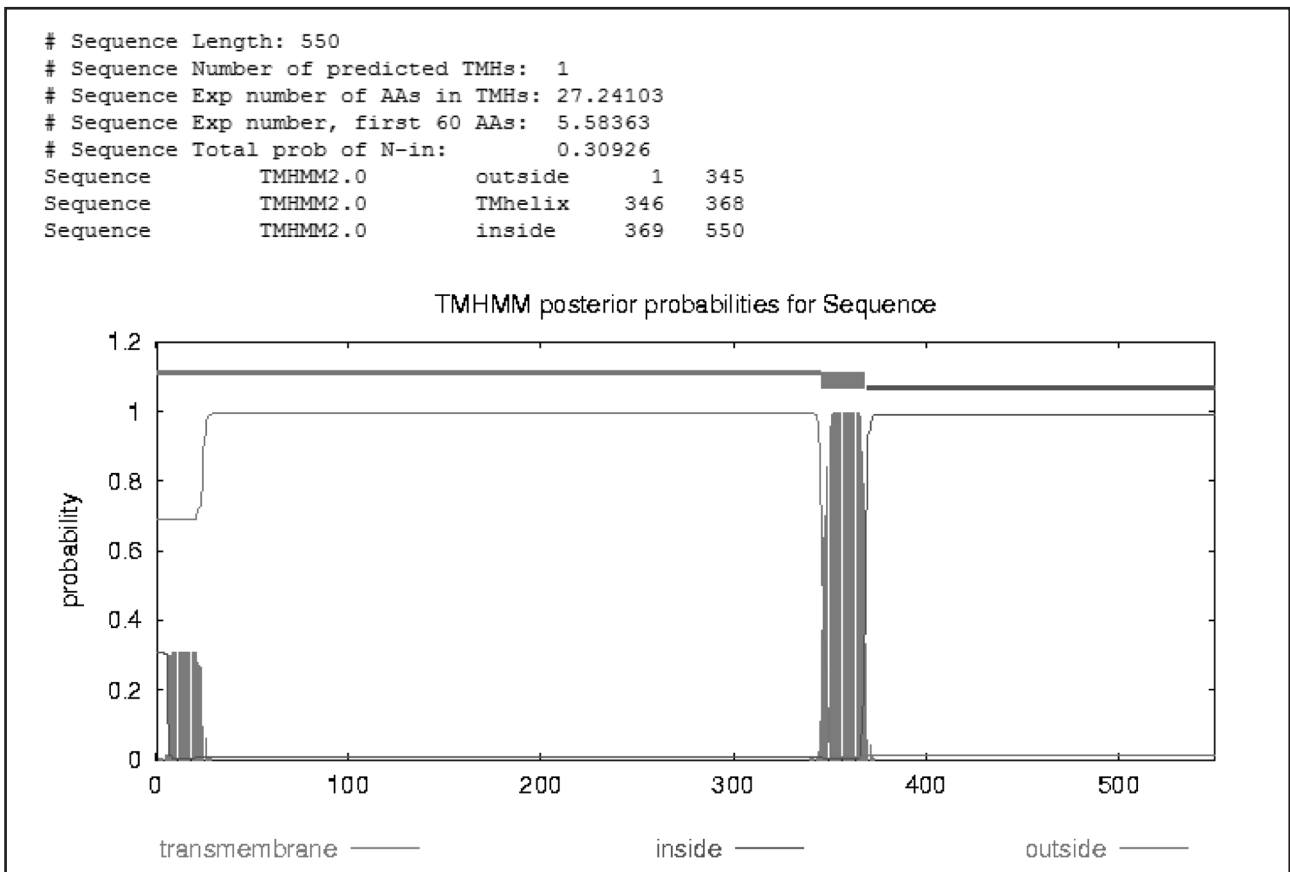


Figura 5 Risultato della previsione con TMHMM. In base al grafico la proteina contiene un'elica che attraversa la membrana in posizione centrale intorno al residuo 350. La metà N-terminale è esposta all'esterno della membrana, mentre la parte C-terminale è intracitosolica. Una zona a potenziale elica transmembrana è riconosciuta all'estremità N-terminale.

- Utilizzeremo ora il server **SignalP** disponibile su www.cbs.dtu.dk/services/SignalP per la previsione di peptidi segnale lungo una sequenza. La schermata iniziale è del tutto simile a quella mostrata da TMHMM. Procedete quindi all'immissione della sequenza nell'apposito spazio lasciando invariati tutti i parametri. Anche in questo caso i risultati dovrebbero comparire in pochi secondi. La **Figura 6** mostra il risultato, dal quale si evince che effettivamente è prevista la presenza di un peptide segnale sul lato N-terminale con probabile sito di scissione in posizione 21-22.

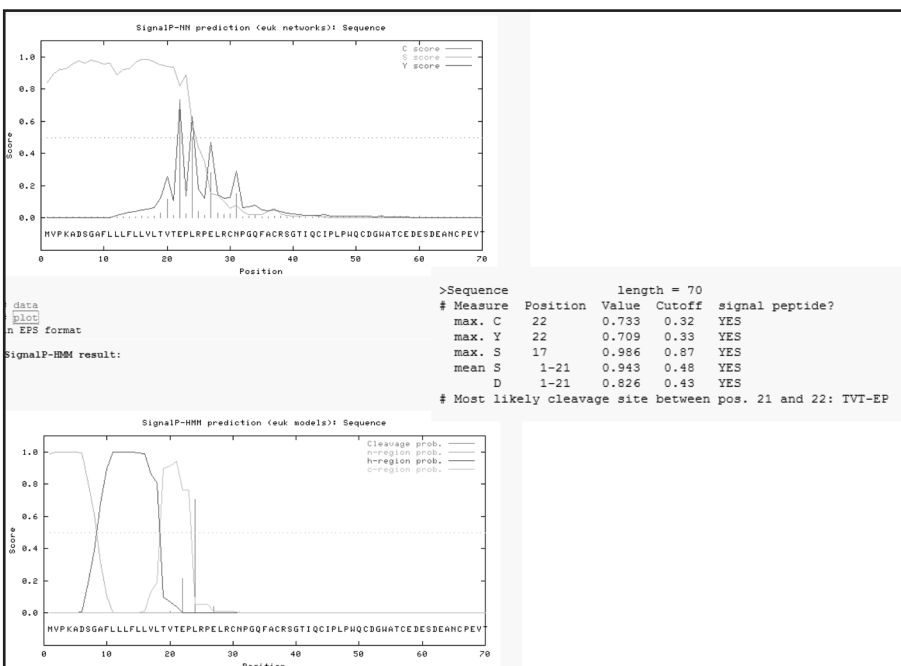


Figura 6 Risultato della previsione con SignalP. Nel pannello superiore è mostrato il grafico ottenuto con l'applicazione delle reti neurali. Sulla destra è riportato il dettaglio della previsione con il sito più probabile. Nel pannello inferiore è mostrata la previsione con gli HMM. Per maggiori dettagli sulla previsione della struttura secondaria si rimanda al Capitolo 9.