

10. Previsione della struttura tridimensionale di una proteina

Lo scopo di questa esercitazione è di comprendere la logica alla base della modellizzazione per omologia attraverso la previsione della struttura tridimensionale di una proteasi a serina presente nel veleno dell'organismo *Deinagkistrodon acutus*, un serpente appartenente alla famiglia delle vipere. L'esercitazione, svolta sul sistema operativa Windows, prevede i seguenti passaggi:

1. Recupero di una sequenza proteica da NCBI.
2. Utilizzo dei programmi **MolIde** e **SCWRL** per la previsione della struttura.
3. Utilizzo del programma **DeepView** per la minimizzazione energetica del modello.
4. Utilizzo di **ProSA-Web** per valutare la bontà del modello.

Gli ingredienti per iniziare

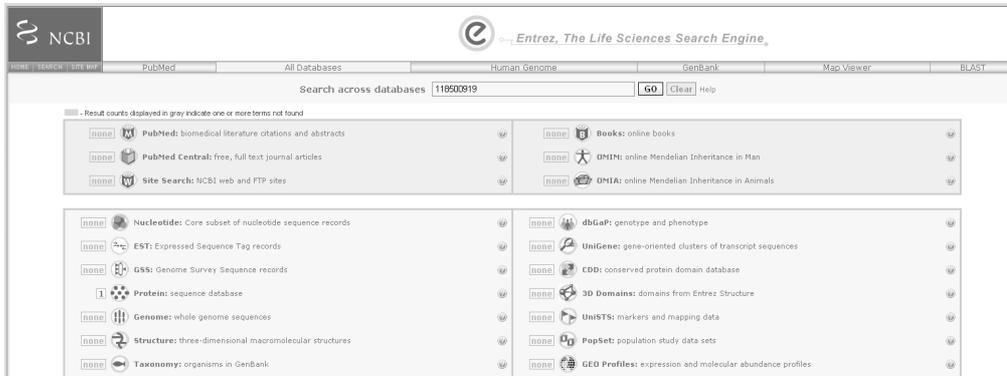
MolIde: programma di grafica e modellistica molecolare, sviluppato appositamente per la modellizzazione per omologia, scaricabile dal sito <http://dunbrack.fccc.edu/molide>. Si consiglia di installare MolIde nella directory di default (C:\FCCC). Al primo avvio del programma, installate i database **nr** e **pdbaa**, necessari per la ricerca di sequenze (questa operazione potrebbe richiedere molto tempo, dal momento che le dimensioni di nr sono di diversi gigabyte; fate riferimento al manuale di installazione per maggiori informazioni). In questa esercitazione utilizzeremo la versione 1.6 del programma.

SCWRL: programma che permette di ricostruire le catene laterali di una struttura, utilizzando librerie di rotameri. Scaricabile dal sito <http://dunbrack.fccc.edu>, il programma è accessibile tramite MolIde. Accertatevi che MolIde sia correttamente configurato per utilizzare la versione di SCWRL scaricata (nel nostro caso la versione 4.0; in MolIde cliccate su *Tools* -> *Options* -> *Scwrl*).

DeepView: programma di grafica e modellistica molecolare scaricabile dall'*Expert Protein Analysis System (ExPASy) Molecular Biology Server* di Ginevra (<http://spdbv.vital-it.ch>).

ProSA-Web: programma che fa uso di potenziali di campo medio per analizzare il profilo energetico di una struttura proteica. In questa esercitazione utilizzeremo il server disponibile sul sito <https://prosa.services.came.sbg.ac.at/prosa.php>.

1. Collegatevi al sito di **Entrez** (<http://www.ncbi.nlm.nih.gov/sites/gquery>) e cercate la proteina con codice GI 118500919 (*venom thrombin-like enzyme [Deinagkistrodon acutus]*):



Selezionate il risultato ottenuto nel campo *Protein*. Nella nuova finestra che compare, salvate in formato FASTA la proteina ottenuta, selezionando la voce *Download* in alto a destra. Salvate il file come *VSP.seq* (*Venom Serine Protease*). Aprite il file e modificate l'intestazione, lasciando solo:

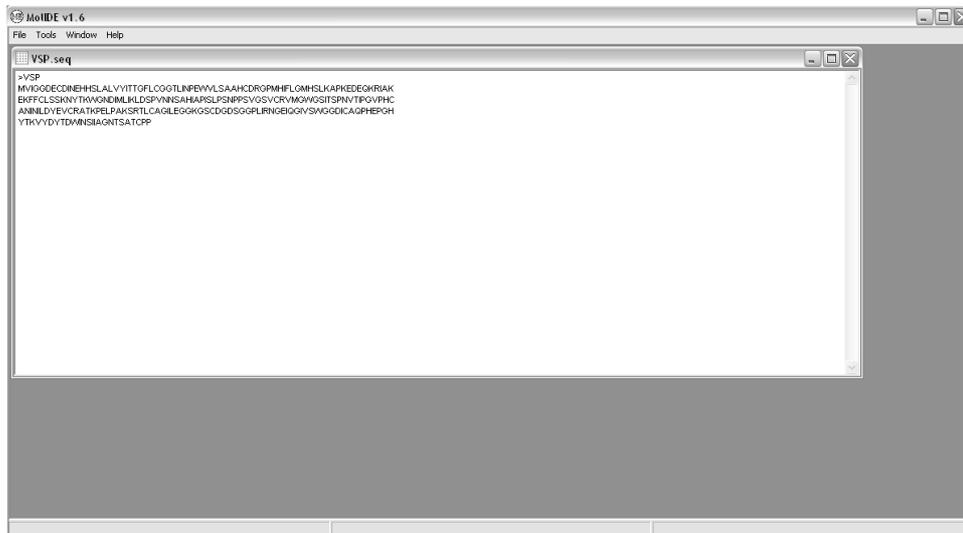
```
>VSP
MVIIGDECDINEHHSALALVYITTTGFLCGGTLINPEWVLSAAHCDRGPMHIFLGMHSLKAPKEDEQKRIAK
EKFFCLSSKNYTKWGNDIMLIKLDSPVNNSAHIAPISLPSNPPSVGSVCRVMGWSITSPNVTIPGVPHC
ANINILDYEVCRATKPELPAKSRTLCAGILEGGKGS CDGDSGGPLIRNGEIQGIVSWG DICAQPHEPGH
YTKVYDYTDWINSIIAGNTSATCPP
```

Salvate nuovamente il file.

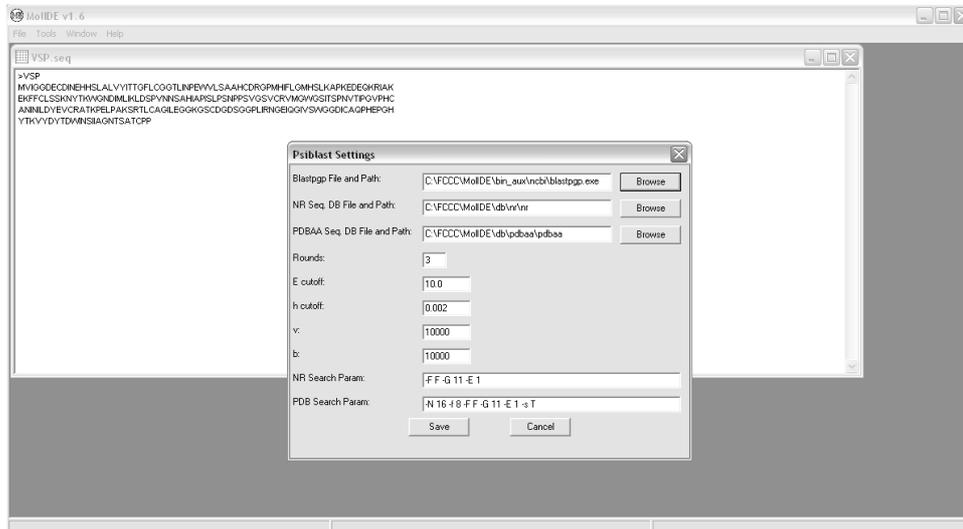
2. Lanciate **MolIde**:



3. Selezionate *File -> Open -> Sequence* dal menù di MolIde. Selezionate il file VSP.seq. Vi apparirà una pagina simile a questa:



Anzitutto effettueremo una ricerca di sequenze omologhe alla proteina VSP. In una finestra a parte appariranno le sequenze simili alla nostra query, la cui struttura è stata risolta sperimentalmente. MolIde utilizza **PSI-Blast** per effettuare la ricerca. I parametri di ricerca sono accessibili dal menù *Tools -> Options -> PSI-Blast Settings*. Notate che verranno effettuate tre iterazioni di PSI-Blast. Controllate la correttezza dei parametri inseriti (fate riferimento al manuale di MolIde per maggiori dettagli):



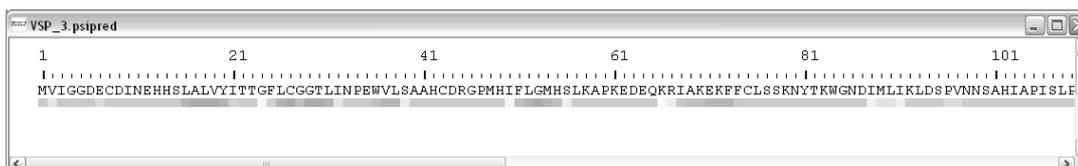
4. Chiudete la finestra dei parametri di PSI-Blast. Selezionate *Tools -> Psiblast*. Nella finestra che compare, cliccate sul pulsante *Run Psiblast*. Il programma impiegherà diversi minuti per effettuare l'operazione di ricerca. Al termine della ricerca comparirà la scritta *Done*. A questo punto MolIde avrà generato diversi file di output, uno per ogni iterazione di PSI-Blast effettuata. Nel nostro caso questi file si chiameranno *VSP_X_pdbout*, dove "X" si riferisce all'iterazione effettuata.

5. Apriamo i risultati della terza iterazione di PSI-Blast sul *database* di proteine a struttura nota (pdbaa), selezionando *File -> Open... -> PDB Hits Alignment* e scegliendo "VSP_3_pdbout". Comparirà una finestra simile a quella mostrata di seguito:

No.	Template	Method	Resolution	E-value(s)	Identities(%)	Positives(%)	Gaps(%)	Start Residue	End Residue	Alignment Length	Hit Length	Backbone only?	Description
0	1C07E	XRAY	1.90	9.00e-034	41	56	1	2	228	227	245	no	TRYPsin II <UNP TRV2_RAT>
1	1OP0A	XRAY	2.00	2.00e-033	64	74	0	2	235	234	234	no	Venom serine protease <UNP VSP2_AKGAC>
2	1OP2A	XRAY	2.10	2.00e-033	64	74	0	2	235	234	234	no	Venom serine protease <UNP VSP2_AKGAC>
3	1TR9A	XRAY	2.20	4.00e-033	39	56	1	2	229	228	224	no	TRYPsin <UNP TRV1_HUMAN>
4	1TR9B	XRAY	2.20	4.00e-033	39	56	1	2	229	228	224	no	TRYPsin <UNP TRV1_HUMAN>
5	1H4WA	XRAY	1.70	5.00e-033	40	56	1	2	229	228	224	no	TRYPsin IVA <UNP TRV4_HUMAN>
6	1BQYA	XRAY	2.50	6.00e-033	62	71	0	2	235	234	234	no	PLASMINOGEN ACTIVATOR <PIR A57290>
7	1BQYB	XRAY	2.50	6.00e-033	62	71	0	2	235	234	234	no	PLASMINOGEN ACTIVATOR <PIR A57290>
8	2R9PA	XRAY	1.40	9.00e-033	39	56	1	2	229	228	224	no	Trypsin-3 <UNP TRV3_HUMAN>
9	2R9PB	XRAY	1.40	9.00e-033	39	56	1	2	229	228	224	no	Trypsin-3 <UNP TRV3_HUMAN>
10	2R9PC	XRAY	1.40	9.00e-033	39	56	1	2	229	228	224	no	Trypsin-3 <UNP TRV3_HUMAN>
11	2R9PD	XRAY	1.40	9.00e-033	39	56	1	2	229	228	224	no	Trypsin-3 <UNP TRV3_HUMAN>
12	2RA3A	XRAY	1.46	1.00e-032	39	57	1	2	229	228	224	no	Trypsin-1 <UNP TRV1_HUMAN>
13	2RA3B	XRAY	1.46	1.00e-032	39	57	1	2	229	228	224	no	Trypsin-1 <UNP TRV1_HUMAN>
14	1F7ZA	XRAY	1.55	2.00e-032	40	55	1	2	228	227	233	no	TRYPsin II, ANIONIC <UNP TRV2_RAT>
15	3TGJE	XRAY	2.20	2.00e-032	40	55	1	2	228	227	233	no	TRYPsin <UNP TRV2_RAT>
16	1FY1A	XRAY	2.15	5.00e-032	34	52	1	2	229	228	228	no	COAGULANT FACTOR YA-TRYPsin CHIMERA <UNP TR...
17	2BYXK	XRAY	1.30	5.00e-032	40	55	1	2	228	227	243	no	CATIONIC TRYPsin <UNP TRV1_BOVIN>
18	2BYXK	XRAY	1.30	5.00e-032	40	55	1	2	228	227	243	no	CATIONIC TRYPsin <UNP TRV1_BOVIN>
19	2BYXK	XRAY	1.30	5.00e-032	40	55	1	2	228	227	243	no	CATIONIC TRYPsin <UNP TRV1_BOVIN>
20	2BYXK	XRAY	1.30	5.00e-032	40	55	1	2	228	227	243	no	CATIONIC TRYPsin <UNP TRV1_BOVIN>
21	2BYXK	XRAY	1.30	5.00e-032	40	55	1	2	228	227	243	no	CATIONIC TRYPsin <UNP TRV1_BOVIN>
22	2BYXK	XRAY	1.30	5.00e-032	40	55	1	2	228	227	243	no	CATIONIC TRYPsin <UNP TRV1_BOVIN>
23	1E2ZC	XRAY	2.60	5.00e-032	40	55	1	2	228	227	243	no	TRYPsin <UNP TRV1_BOVIN>
24	1F0TA	XRAY	1.80	5.00e-032	40	55	1	2	228	227	243	no	TRYPsin <UNP TRV1_BOVIN>
25	1PQLA	XRAY	1.90	5.00e-032	40	55	1	2	228	227	243	no	TRYPsin <UNP TRV1_BOVIN>
26	1LQEA	XRAY	2.20	5.00e-032	40	55	1	2	228	227	243	no	TRYPsin <UNP TRV1_BOVIN>
27	1UTNA	XRAY	1.15	5.00e-032	40	55	1	2	228	227	243	no	TRYPsin <UNP TRV1_BOVIN>
28	1UTOA	XRAY	1.15	5.00e-032	40	55	1	2	228	227	243	no	TRYPsin <UNP TRV1_BOVIN>
29	1UTPA	XRAY	1.30	5.00e-032	40	55	1	2	228	227	243	no	TRYPsin <UNP TRV1_BOVIN>
30	1UTQA	XRAY	1.15	5.00e-032	40	55	1	2	228	227	243	no	TRYPsin <UNP TRV1_BOVIN>
31	1OP4B	XRAY	2.30	6.00e-032	40	55	1	2	228	227	243	no	Trypsinogen, cationic precursor <UNP TRV1_BOVIN>
32	1AMEA	XRAY	2.20	1.00e-031	40	55	1	2	228	227	223	no	ANIONIC TRYPsin <UNP TRV2_RAT>
33	3FPFE	XRAY	1.49	1.00e-031	40	55	1	2	228	227	223	no	Anionic trypsin-2 <UNP TRV2_RAT>
34	3TGJE	XRAY	1.80	1.00e-031	40	55	1	2	228	227	223	no	TRYPsin <UNP TRV2_RAT>
35	1AMHA	XRAY	2.50	1.00e-031	40	55	1	2	228	227	223	no	ANIONIC TRYPsin <UNP TRV2_RAT>
36	1AMHB	XRAY	2.50	1.00e-031	40	55	1	2	228	227	223	no	ANIONIC TRYPsin <UNP TRV2_RAT>
37	3FP7E	XRAY	1.46	1.00e-031	40	55	1	2	228	227	223	no	Anionic trypsin-2 <UNP TRV2_RAT>
38	3FP8E	XRAY	1.46	1.00e-031	40	55	1	2	228	227	223	no	Anionic trypsin-2 <UNP TRV2_RAT>
39	1K93C	XRAY	2.30	1.00e-031	40	55	1	2	228	227	223	no	TRYPsin II ANIONIC <UNP TRV2_RAT>

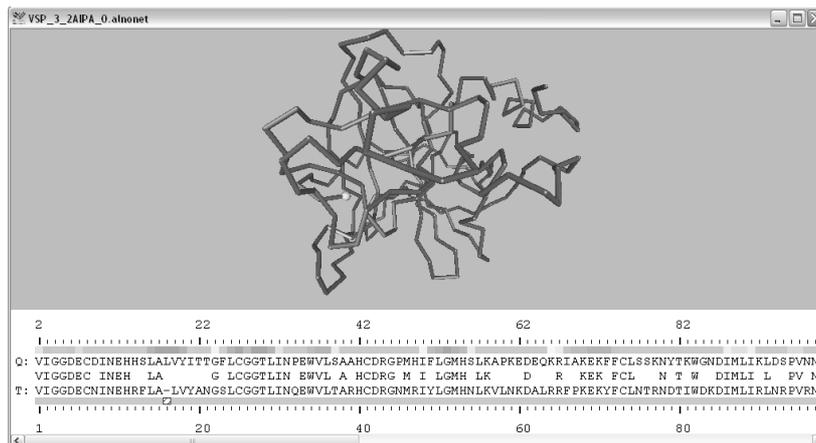
In questa schermata sono mostrati i risultati della ricerca di PSI-Blast, ordinati in base all'*E-value*. Le informazioni riportate sono: il codice PDB, la tecnica con la quale la struttura è stata risolta, la risoluzione della struttura, dati relativi all'allineamento (tra cui molto importante è la percentuale di identità di sequenza) e la descrizione della proteina. Ordinate i risultati per percentuale d'identità, cliccando con il tasto sinistro del mouse sul nome della colonna relativa. Per ora tenete da parte questa finestra.

6. Può essere utile, in un progetto di modellizzazione per omologia, prevedere la struttura secondaria della proteina da modellare. Cliccate nuovamente sulla finestra contenente la proteina VSP (VSP.seq) e selezionate *Tools -> Psipred* dal menu di MolIde. Successivamente, cliccate su *Run Psipred* nella finestra che vi comparirà. Quando PsiPred avrà terminato, selezionate il file VSP_3.psipred dal menù *File -> Open -> Sec. Struct. Pred.* Notate che, per effettuare la predizione della struttura secondaria, PsiPred ha bisogno dei risultati di PSI-Blast. Vi apparirà una finestra simile alla seguente:



La barra al di sotto della sequenza indica la predizione di struttura secondaria (in rosso le α -eliche, in verde i filamenti β , in grigio regioni a struttura secondaria non regolare).

7. Torniamo alla finestra con i risultati di PSI-BLAST e clicchiamo due volte con il tasto sinistro del mouse sul numero che appare in corrispondenza della prima colonna, adiacente al codice PDB 2AIPA (l'ultima lettera si riferisce alla catena), che utilizzeremo come stampo strutturale. La proteina cui fa riferimento questo codice, *Protein C activator* di *Agkistrodon contortrix*, condivide con VSP una percentuale di identità del 69%, e copre quasi completamente la lunghezza della sequenza da modellare. Vi apparirà una finestra simile alla seguente:



Nella parte superiore di questa finestra sarà visualizzato lo stampo strutturale, mentre nella parte inferiore è mostrato l'allineamento tra stampo e proteina da modellare. È possibile modificare manualmente l'allineamento inserendo o cancellando *gap* (utilizzate il tasto sinistro del mouse tenendo premuto contemporaneamente il tasto SHIFT o CTRL della tastiera, rispettivamente, e sovrapponetevi il cursore nel punto in cui volete aggiungere o togliere *gap*). Nel nostro esempio, il *gap* in posizione 16 va eliminato e spostato in posizione 19, in modo che la "I" di VSP in posizione 21 sia vista come un'inserzione e che la regione "LVY" (18-20) risulti allineata correttamente.

8. Quando avete terminato di modificare manualmente l'allineamento, trasferite le coordinate della catena principale dello stampo alla proteina da modellare, selezionando *Tools -> Copy Backbone*. Se tutto è andato bene, un messaggio vi avvertirà che il modello preliminare è stato costruito con successo.
9. A questo punto utilizzate SCWRL per inserire le catene laterali nel modello preliminare selezionando *Tools -> Build Side Chains (SCWRL)*. Nella finestra che compare cliccate sul pulsante *Run SCWRL*. Al termine dell'operazione, cliccate su *Done*. MolDe avrà generato un file nella stessa cartella all'interno della quale è contenuto il file VSP.seq chiamato VSP_3_2AIPA_0.pdb. Questo file contiene le coordinate del modello di VSP, costruito tramite allineamento con 2AIPA. Mancano però le inserzioni strutturali che VSP presenta rispetto a 2AIPA (posizioni 21 e 163-164 dell'allineamento). È necessario quindi ricostruire queste regioni variabili. Per farlo, utilizzeremo il programma **Loopy**, implementato in MolDe.
10. Loopy ha bisogno di conoscere i punti di ancoraggio dell'ansa da costruire. Selezionate, con il tasto destro del mouse la lettera Y in posizione 20, adiacente all'inserzione del residuo I in posizione 21. Vi apparirà un menù. Selezionate *Set Loop Left Anchor*: apparirà un cerchio azzurro intorno alla lettera selezionata. Fate lo stesso con la lettera T in posizione 22, ma questa volta selezionate *Set Loop Right Anchor*. Al termine di questa operazione cliccate con il tasto destro in corrispondenza della lettera I in posizione 21 e selezionate *Build Loop*. Nella finestra che comparirà, cliccate sul tasto *Run Loopy*. Dopo un po' MolDe vi avvertirà che la prima inserzione è stata costruita. Questa sarà sovrascritta sul file VSP_3_2AIPA_0.pdb.
11. Ripetete i passaggi precedenti per l'inserzione 163-164 di VSP e per eventuali altre inserzioni che osservate nell'allineamento.

12. Ora il modello è pronto, ma è necessario minimizzarlo energeticamente per rilassare la posizione della catena principale e delle catene laterali. Aprite il programma DeepView e, tramite la voce *File -> Open PDB File*, caricate il file *VSP_3_2AIPA_0.pdb*. Selezionate *all* dalla voce *Select* del menù. Successivamente andate in *Prefs -> Energy minimization...* e, per velocizzare il processo, selezionate solo 100 iterazioni dell'algoritmo *Steepest Descents*. Infine, selezionate *Tools -> Energy Minimization...* per minimizzare il modello. Al termine dell'operazione salvate il file selezionando *File -> Save... -> Current Layer*.

13. Collegatevi al sito di ProSA-web per effettuare la validazione del modello con ProSA. Inserite nel campo apposito il vostro file con il modello e cliccate su *Analyse*:

14. I risultati di ProSA-web indicheranno che la qualità globale della nostra struttura è confrontabile con una struttura NMR di pari lunghezza:

