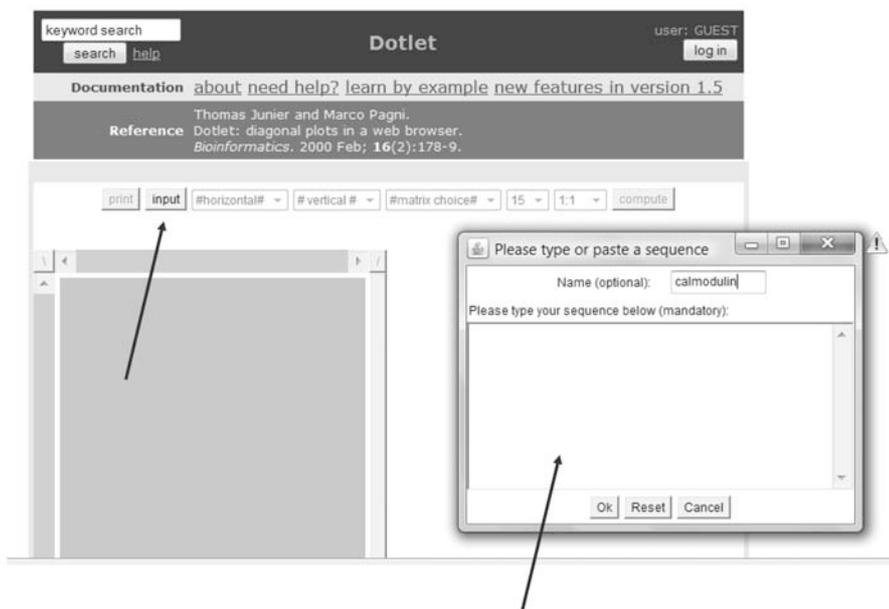


### 3. Confronto tra due sequenze

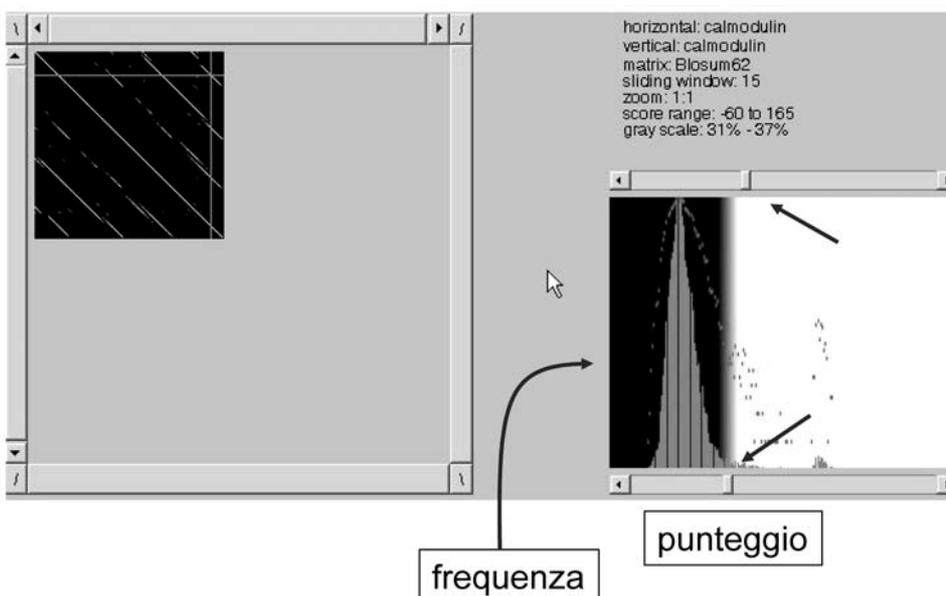
#### Esercizio 1: uso di DotLet

Il programma DotLet è accessibile dal sito <http://myhits.isb-sib.ch/cgi-bin/dotlet>, dove può essere utilizzato attraverso un'interfaccia utente semplice e intuitiva.

- Utilizzando i servizi disponibili sul sito dell'NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), trovate e memorizzate la sequenza proteica in formato FASTA della calmodulina umana con numero di accessione CAA36839. Confrontate la sequenza con se stessa per mezzo di DotLet (cliccare su "input", chiamare la sequenza "calmodulin", ed incollarla nella finestra sottostante) (**Figura 1**). Lasciare come parametri iniziali la matrice Blosum62 ed una finestra di 15 residui per il confronto, e cercare di ottenere una visualizzazione simile a quella riportata nella **Figura 2** agendo sui cursori che controllano l'estensione della scala di grigi.

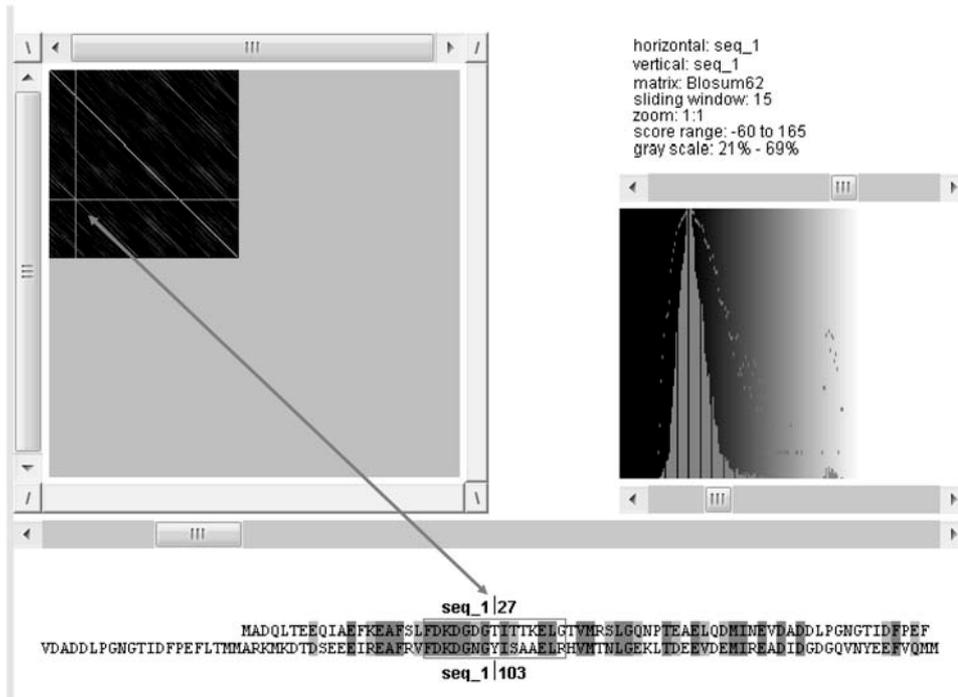


**Figura 1** Maschera di inserimento della sequenza in Dotlet. Cliccando sul pulsante input si apre una finestra nella quale può essere incollata la sequenza. Per inserire un'altra sequenza si ripete la procedura.



**Figura 2** Finestra per la variazione della scala di grigi. Il grafico riporta la distribuzione dei punteggi ottenuti da tutte le coppie di finestre di sequenza confrontate. Si noti che la maggior parte dei punteggi ricade nella distribuzione a sinistra a basso punteggio, mentre una piccola popolazione a punteggio elevato si trova a destra. Spostando i cursori si variano i punteggi limite al di sotto dei quali la cella assume il colore nero e al di sopra il colore bianco. Tra i due limiti le celle assumono un tono di grigio proporzionale al punteggio che contengono.

- Individuate le sequenze ripetute che legano il  $\text{Ca}^{++}$  «**D-x-D-G-(D/N)-G**». Quanti di questi motivi di sequenza sono presenti? *Suggerimento*: cliccando sulla matrice si attiva un reticolo che può essere spostato sulla superficie della matrice stessa con il puntatore del mouse. In basso viene riportato l'allineamento tra i due segmenti della proteina corrispondenti alla posizione del centro del reticolo sulla matrice (**Figura 3**).



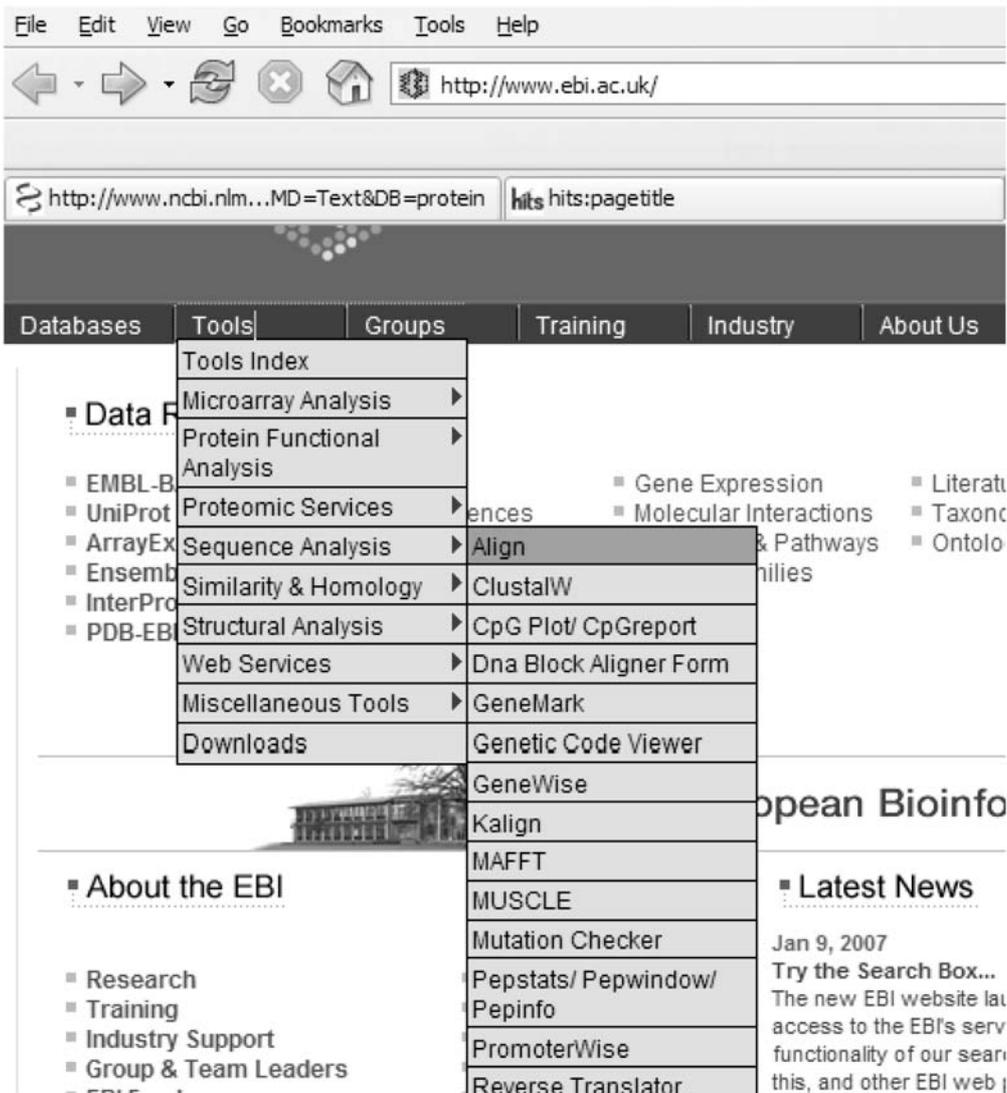
**Figura 3** Sul pannello della matrice a punti può essere spostato il reticolo il cui centro indica la cella corrispondente alle due finestre. Le sequenze contenute nelle finestre vengono evidenziate nel pannello inferiore.

- Utilizzando l'applet Protein Workshop con la struttura **1IQ5**, evidenziate sulla struttura le sequenze trovate. A quale elemento di struttura secondaria corrisponde questo motivo? *Suggerimento*: collegatevi alla banca dati PDB e cercate la struttura attraverso il codice pdb; nella parte destra della schermata c'è il pulsante per l'esecuzione di Protein Workshop.
- Utilizzando la ricerca nella banca dati proteica dell'NCBI, trovate e memorizzate le sequenze proteiche in formato FASTA delle subunità alfa 2 di *Rattus norvegicus* e alfa 4 di *chicken* del recettore neuronale dell'acetilcolina (*Neuronal acetylcholine receptor protein, alpha-2 chain precursor P12389*; *Neuronal acetylcholine receptor protein, alpha-4 chain precursor P09482*). Attraverso un confronto tra le due sequenze con DotLet, si noti la distribuzione della zona di somiglianza. Agite sulla scala di grigi per vedere la variazione del risultato. Ripetete il calcolo variando la lunghezza della finestra e la matrice di punteggio utilizzata. *Suggerimento*: nella finestra di zoom, impostate un rapporto dimensionale 1:2.
- Recuperate le sequenze in formato FASTA dal sito NCBI delle due proteine umane con codice STAM2\_HUMAN e HGS\_HUMAN. Questi due codici, come si è visto nel Capitolo 2, sono propri della banca dati SwissProt. Confrontate le due sequenze con DotLet. Correlate il risultato trovato con quanto noto sulla struttura e funzione di queste proteine. *Suggerimento*: esaminate le annotazioni delle proteine nella banca dati.

## Esercizio 2: allineamento di due sequenze con algoritmi dinamici

In questa seconda attività cercheremo di comprendere come il risultato ottenuto applicando algoritmi dinamici per l'allineamento di sequenze sia influenzato dalla variazione dei parametri della funzione di penalizzazione delle indel e dalla matrice di punteggio utilizzata.

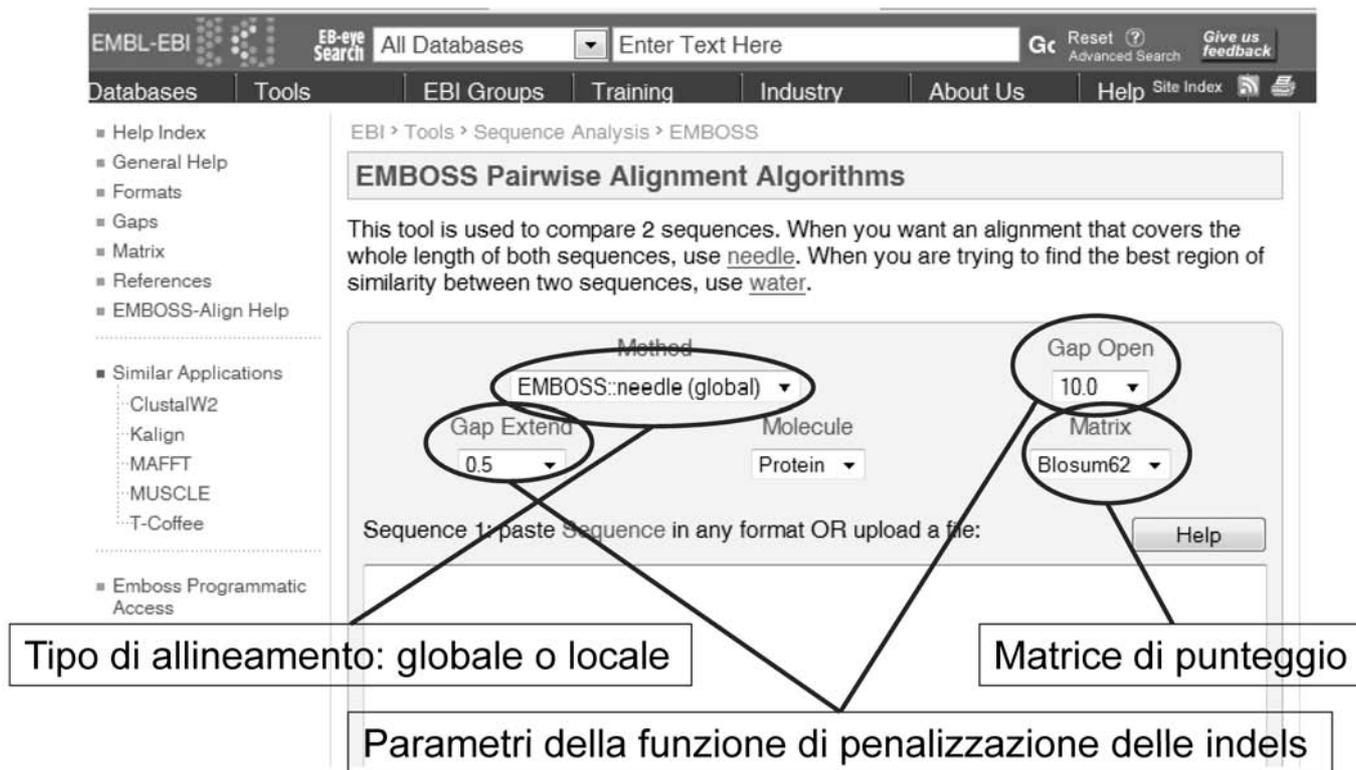
- Collegatevi al sito <http://www.ebi.ac.uk> e selezionate i servizi di allineamento di sequenze (**Figura 4**): comparirà la pagina di avvio del programma di allineamento di sequenze.



**Figura 4** Menu di accesso ai programmi di allineamento di sequenze

- Inserite negli appositi campi le due sequenze trovate nell'Esercizio 1 (*Neuronal acetylcholine receptor protein, alpha-2 chain precursor*; *Neuronal acetylcholine receptor protein, alpha-4 chain precursor*) nel formato FASTA *senza* la linea di intestazione individuata dal simbolo ">" (**Figura 5**).
- Effettuate un primo allineamento globale *needle* utilizzando i parametri standard. Qual è il risultato e che differenze ci sono rispetto ai risultati del dot-plot calcolato nel punto precedente? Le indel sono posizionate in modo simile?
- Ripetete l'allineamento aumentando la penalizzazione per l'apertura delle indel (*open gap penalty*) a 50 e per l'allungamento (*gap extension penalty*) a 10: cosa succede e perché? *Suggerimento*: controllare la posizione delle indel rispetto all'allineamento precedente. La

variazione dei parametri della funzione di penalizzazione ha modificato l'allineamento risultante. In particolare, l'aumento della penalizzazione per l'apertura e l'allungamento limita il numero e la lunghezza delle indel introdotte nell'allineamento.



**Figura 5** Modulo di avvio del programma di allineamento di sequenze. Sono evidenziati i campi i cui valori possono essere modificati dall'utente

- Ripetete l'allineamento diminuendo la penalizzazione per l'apertura a 1,0 e per l'allungamento a 0,1: cosa succede e perché?
- Ripetete l'allineamento utilizzando i parametri iniziali (apertura 10,0 e allungamento 0,5) ma diverse matrici di punteggio. Notate qualche differenza?
- Ripetete gli esperimenti utilizzando il programma di allineamento locale *water* e confrontate il risultato con quello ottenuto con *needle*.
- Ripetete l'intera procedura con le sequenze STAM2\_HUMAN e HGS\_HUMAN. In particolare aumentate la *gap open penalty* fino a 15 e calcolate l'allineamento con *needle* e *water*. Gli allineamenti risultanti sono molto diversi. L'allineamento locale isola il dominio di maggior somiglianza (**Figura 6**). Quando si diminuisce il valore della penalizzazione a 10 (ciò che viene suggerito inizialmente) allora i due tipi di allineamento risultanti diventano molto simili.

```

#####
# Program: water
# Rundate: Tue Nov 24 11:58:58 2009
# Align_format: srspair
# Report_file: /ebi/extserv/old-work/water-20091124-1158585203.output
#####

#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 15.0 ←
# Extend_penalty: 0.5
#
# Length: 136
# Identity: 47/136 (34.6%)
# Similarity: 78/136 (57.4%)
# Gaps: 0/136 ( 0.0%)
# Score: 237.0
#
#
#=====

EMBOSS_001      9 FEQDVEKATNEYNTTEDWSLIMDICKVGGSTPNGAKDCLKAIMKRVNHKV      58
                  ||:::||||:.....||..|:||||:.....||..:|. |:|.|.
EMBOSS_001      8 FERLLDKATSQLLLETDWESILQICDLIRQGDTQAKYAVNSIKKKVNDKN      57

EMBOSS_001     59 PHVALQALTLLGACVANCGKIFHLEVCSRDFATEVRAVIKKAHPKVCEK      108
                  |||||.||:..:|.|||:..|.||:.....|:~::~|:~:~:|.~|.
EMBOSS_001     58 PHVALYALEVMESVVKNCGQTVHDEVANKQIMEELKDLLKRQVEVNVVRNK      107

EMBOSS_001     109 LKSLMVEWSEEFQKDPQFSLISATIKSMKEEGITFP      144
                   :~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|
EMBOSS_001     108 ILYLIQAWAHAFRNEPKYKVVQDTYQIMKVEGHVFP      143

```

**Figura 6** Allineamento tra STAM2\_HUMAN e HGS\_HUMAN con il programma *water*, matrice BLOSUM62, *gap open* = 15,0 e *gap extension* = 0,5.