

Tema d'esame - Teoria

Rispondere in modo sintetico ai seguenti quesiti:

- 1) Descrivere brevemente le caratteristiche del Protein Data Bank ed il contenuto di un file .pdb
- 2) Spiegare la differenza tra omologia e similitudine. È possibile che due sequenze abbiano un'identità di sequenza del 57% e una similitudine del 21%? Perché?
- 3) Siano date le due sequenze di amminoacidi:

$s_1 =$

MVALMGTKAADL

$s_2 =$

MVAIMASRAGEI

Allinearle senza inserire gaps e senza costruire alcuna matrice e rispondere ai seguenti quesiti.

- a) Stimare quantitativamente il grado di identità.
- b) Quantificare la similitudine utilizzando la matrice di punteggio data.

Illustrare in ambo i casi il calcolo svolto.

Cosa si può concludere circa l'omologia?

- 4) Che tipo di allineamento di sequenza si ottiene applicando gli algoritmi di Needleman-Wunsch e Waterman-Smith? Spiegare brevemente in cosa differiscono tra loro, e qual è la sostanziale differenza rispetto ai metodi euristici.
- 5) Che differenza c'è tra allineamento globale e locale di sequenza? Illustrare come si può quantitativamente valutare la significatività di un allineamento globale.
- 6) Che cosa s'intende per "twilight zone" nell'allineamento di sequenze proteiche?

Tema d'esame - Laboratorio

Esercizio 1

Utilizzando NCBI Gene, individuare l'entry relativa al gene che codifica per la proteina 'frequenin'. Rispondere alle seguenti domande, spiegando il procedimento seguito.

- a) Qual è il nome del gene e il relativo id in *homo sapiens* e *drosophila melanogaster*?
- b) Su quali cromosomi si trova nei due organismi di cui sopra?
- c) Qual è il codice Uniprot della proteina espressa da questo gene nell'uomo?
- d) Quali sono le principali funzioni molecolari della proteina espressa?
- e) È possibile affermare che la proteina si trova nella membrana post-sinaptica? Perché?
- f) Quanti articoli PubMed sono collegati a questo gene nella specie *bos taurus* dalla pagina del database Gene?

Esercizio 2

Ricerca in BLASTP le sequenze simili a NP_000781 nell'organismo *Danio rerio* (7955), ricercando nel database di sequenze con codice RefSeq escludendo le sequenze modellate. Rispondere alle seguenti domande:

- a) A quale organismo appartiene la sequenza di input? Che proteina è?
- b) Quante sono le hits trovate e a quale superfamiglia appartengono?
- c) Quante hits hanno scope compreso tra 80 e 200? Qual è il loro codice RefSeq?
- d) Qual è la hit (nome proteina e codice RefSeq) che rappresenta l'allineamento locale che ricopre la porzione minore rispetto alla lunghezza della query? Qual è l'E-value? Qual è l'identità di sequenza?

Esercizio 3

Cercare sul database UNIPROT la proteina OUTER MEMBRANE PHOSPHOLIPASE A.

- a) Quale entries ci sono relative a *Escherichia coli* (qualunque ceppo)? Qual è la lunghezza della catena polipeptidica più frequente?
- b) Selezionare la entry relativa al ceppo K12 (PA1_ECOLI). Con quale metodo sono state risolte le strutture tridimensionali? Qual è il file PDB a maggior risoluzione che risolve solo la catena A? Cosa si intende per catena? Qual è la percentuale di sequenza risolta in questa struttura?

Aprire il relativo file PDB con Pymol e rispondere a questi ulteriori quesiti.

- c) Quante sono le α -eliche? Elencare i residui che complessivamente formano la struttura ad α -elica più lunga.
- d) Sono presenti ioni Ca^{2+} o Mg^{2+} ?
- e) Che tipo di interazione stabilizza la coppia di residui Ile41 e Lys68? Indicare i gruppi chimici coinvolti e la loro distanza.