

Databases

Esercizio 1

Traccia

Scaricare il fasta della sequenza genomica di human hemoglobin subunit beta (NM_000518.5).

1. Visitare il sito di ExPASy (
[expasy.org] (<http://expasy.org/>
)):
2. Provare il tool TRANSLATE (resources A..Z) per tradurre automaticamente una sequenza genica in una proteica.
3. Sottomettere la sequenza genomica scaricata
4. Quale frame è corretto (confrontare la sequenza predetta con quella reale NP_000509.1)?
5. Perché ci sono 6 frames?

Esercizio 2

Traccia

Cercare la sequenza nucleotidica e amminoacidica della rodopsina (rhodopsin), il pigmento visivo che innesca la visione nei vertebrati

1. Cominciamo dal database Nucleotide. Quante sequenze ci sono per la ricerca “rhodopsin”?
2. Limitare la ricerca al database RefSeq. Quanti record ci sono?
3. Limitare la ricerca ad homo sapiens (human), usando l’opzione advanced search. Quante sequenze nucleotidiche trova?
4. Visualizziamo l’entry “Homo sapiens rhodopsin (RHO), RefSeqGene on chromosome 3”. Quante bp ci sono nella sequenza?
5. Ci sono malattie genetiche associate a questa entry? Di tipo solo autosomico dominante? (OMIM)
6. Scaricare il file della sequenza nucleotidica del gene di rhodopsin

Esercizio 3

Traccia

Ricerca la proteina “Hemoglobin subunit beta” di Homo sapiens. Filtrare solo i record con RefSeq selezionare il risultato con codice RefSeq NP_000509.1 (accession).

1. Individuare
 - lunghezza,
 - peso molecolare
 - il refseq del trascritto
2. Salvare localmente la sequenza FASTA della PROTEINA
3. Salvare localmente la sequenza FASTA del TRASCRITTO
4. Ci sono SNP? Cos'è un SNP?
5. Ci sono malattie mendeliane note legate a questa proteina?
6. Ci sono strutture legate a questa proteina?
7. Quante risolte per NMR e quante mediante Cristallografia (X-Ray)?

Se vogliamo adesso scaricare la sequenza amminoacidica, della rodopsina (rhodopsin) per l'uomo su quale database dobbiamo andare e quali filtri utilizzare ?

1. Scaricare il FASTA della proteina e salvarlo in una directory locale.
2. Collegarsi ad OMIM sfruttando il link sulla destra. Quanti records si ottengono? Trovare almeno due mutazioni puntiformi associate a retinite pigmentosa.

Esercizio 4

Traccia

Nel database Uniprot si cerchi la proteina Transferrin receptor (TFR1) per l'uomo (P02786).

1. Quante isoforme ha ?
2. Ha la struttura risolta ? Se si, a partire da quale aminoacido è risolta.
3. Quale è il nome del gene che la codifica (entrare in HGNC)

Esercizio 5

Traccia

Nel database Uniprot si cerchi la proteina Transferrin receptor 2 (TFR2) per l'uomo (Q9UP52).

1. Quante isoforme ha, se ne ha più di una perché ?
2. Ha la struttura risolta ? Se sì, a partire da quale aminoacido è risolta.
3. Quale è il nome del gene che la codifica (entrare in HGNC)

Esercizio 6

Traccia

Scaricare le sequenze proteiche del recettore della transferrina (TFR1), ma che abbiano la struttura 3D risolta e formino un complesso con un qualsiasi ligando.

1. Utilizzare il database Protein.
2. Limitare la ricerca solo al database PDB (quelli con struttura risolta).
3. In ricerca avanzata cercare “TFR1” e “complex” in tutti i campi
4. Scegliere una entry specifica
5. In “Display Settings” selezionare “FASTA”
6. In “Send” selezionare “Complete Record” e “File”

Matrici di Punteggio

Traccia

Esercizio 1

Allineare con i 2 algoritmi le sequenze GAATTCAGTTA GGATCGA Per l'allineamento globale (NW) usare la seguente opzione

1. Quale dei 2 algoritmi restituisce l'allineamento con il punteggio maggiore? Perché?

Esercizio 2

Traccia

Allineare con i 2 algoritmi le sequenze GAATTCAGTTA GGATCGA

1. Settando la penalità per apertura (e chiusura) dei gap a 1 con i due algoritmi (NW e WS) cosa cambia? Perché?

Esercizio 3

Traccia

Utilizzando l'algoritmo NW disponibile su: [

https://www.ebi.ac.uk/Tools/psa/emboss_needle/] (https://www.ebi.ac.uk/Tools/psa/emboss_needle/)

) Allineare la sequenza di calmodulina umana (CALM1) con quella di:

1. Bos taurus (bovina)
2. Arabidopsis thaliana (pianta) (ottenere le sequenze da opportuni database...).

Mantenere i settaggi di default per i gaps, e utilizzare la matrice di score BLOSUM 62)

3. Qual è l'identità di sequenza? Quali residui differiscono pur restando simili per proprietà chimico-fisiche?

Esercizio 4

Traccia

Utilizzando l'algoritmo WS per allineamenti locali disponibile su:

[https://www.ebi.ac.uk/Tools/psa/emboss_water/] (https://www.ebi.ac.uk/Tools/psa/emboss_water/)

Allineare la sequenza delle due proteine con codice Uniprot P46065 e P21457

1. Di quali proteine si tratta? Cosa hanno in comune?
2. Qual è l'identità di sequenza? E la similitudine? Si può trattare di proteine omologhe? Perché?
3. Identificare una zona in cui l'identità è estesa a 8 residui. Che struttura secondaria ha la seconda proteina in quella zona?
4. Se si allinea la prima proteina con P51177 quali sono i punteggi di allineamento? Allineare LOCALMENTE la prima lunga regione senza gaps. Qual è? E quali sono i nuovi punteggi? Di quali zone di SII si tratta?

Blast

Esercizio 1

Traccia

Eseguire una ricerca tramite blastp su NCBI usando la seguente sequenza di 12 aminoacidi: PNLHGLFGRKTG

1. Metterla in formato FASTA. I parametri di ricerca saranno automaticamente adattati per sequenze corte.
2. Resettare l'interfaccia
3. Attivare l'opzione "Show results in a new window" per poter confrontare i parametri di default con quelli modificati automaticamente.
4. Osservare la sezione "search summary":
5. Qual è il valore di cut-off dell'E-value?
6. Come è cambiata la "word size"?
7. Qual è la matrice di punteggio?
8. Come sono variati i parametri rispetto al default e perché?

Esercizio 2

Traccia

PSI-BLAST - proteina sconosciuta

1. Un campione biologico ha rivelato la presenza della sequenza proteica di origine sconosciuta riportata in:

`http://goo.gl/siebf5`
2. Si ritiene che debba appartenere alla specie Danio Rerio (zebrafish).
3. Utilizzare PSI-BLAST con i seguenti parametri: RefSeq come database, escludendo i modelli dagli output, limitandosi all'organismo Danio rerio, PAM30 come matrice di score.
4. Di che tipo di proteina si tratta? (Guardare se ci sono domini conservati!) Quanti hits ci sono alla prima iterazione? Qual è l'hit con score più basso ed E-value più alto? Segnarsi il codice RefSeq. Quante hits hanno score >200

5. Alla seconda iterazione, qual è l'hit con score minore? Che E-value ha? E che score ha la proteina con peggior score alla iterazione precedente? Perché?
6. Quante nuove hit compaiono alla terza iterazione?
7. A quale iterazione non vengono più aggiunte hits?

Esercizio 3

Traccia

Entrare in BLASTX di NCBI e copiare la sequenza di “dinosaur” "Lost World" come input. <ftp://ftp.ncbi.nlm.nih.gov/pub/FieldGuide/lostworld.txt>
 Resetare la pagina prima di impostare i parametri Assicurarsi di includere l'intera sequenza. Cercare sul database “nr”. Escludere i modelli (XM/XP).

1. A quale proteina appartiene probabilmente questa sequenza nucleotidica?
2. Nella pagina dei risultati, guardare i risultati degli allineamenti.
3. La pagina risultante mostrerà la sequenza query scritta come proteina (utilizzando le 20 lettere corrispondenti agli amminoacidi). Il Dr. Mark Boguski che ha creato la sequenza ha lasciato un messaggio nascosto nella sequenza query in posizioni corrispondenti ai 4 gap della sequenza allineata. Qual è il suo messaggio?

MSA

Esercizio 1

Traccia

Nel sito Homologene scaricare le sequenze fasta che ci sono nell'entry relativa alla proteina NP_000940.1 ed allinearle con muscle EBI :

[<https://www.ebi.ac.uk/Tools/msa/muscle/>] (<https://www.ebi.ac.uk/Tools/msa/muscle/>)

(attenzione! Selezionare come output il formato Clustal!)

1. Quante sequenze si stanno allineando?
2. Cosa permette di dire che le sequenze sono in formato FASTA?
3. Quali due delle sequenze non conservano la stringa “ICLI”?

4. Quante e quali inserzioni di un singolo aminoacido sono avvenute e in quali sequenze?
5. Aprire l'allineamento in Jalview dopo averlo esportato in formato FASTA da MUSCLE. Selezionare la regione che si estende da "GQSPPE..." a "...VRDVQ" della sequenza NP_990185.1, tramite il tab Web Service lanciare JPRED. Qual è l'elemento di struttura secondaria più ricorrente, secondo la predizione di JPRED? Quante alfa eliche sono predette? Suggerimento: Usare HTML format per l'output
6. (se non fosse disponibile dal tab, collegarsi a:

<http://www.compbio.dundee.ac.uk/jpred/>

). ATTENZIONE: JPRED può essere lento!!!

Esercizio 2

Traccia

Cerchiamo l'entry 1EBM nel database PDB

1. Quali macromolecole contiene la struttura?
2. Quante catene? Cosa rappresenta la catena A? E' mutata?
3. È una proteina intera? Mancano residui? Perché?
4. Cliccare sul tab Sequence. Che informazioni troviamo?

Scarichiamo il file PDB e visualizziamolo con un editor di testo (attenti a dove lo salvate!)

1. Chi sono gli autori del lavoro strutturale?
2. Si tratta di cristallografia a raggi X o di NMR?
3. Qual è la risoluzione della struttura?
4. Cosa si trova al REMARK 200? Hanno usato luce di sincrotrone per risolvere la struttura?
5. Cosa si trova al REMARK 470? Spiegate i residui mancanti
6. Cosa si trova nel campo SEQRES?
7. Quante α -eliche e β -sheets ci sono?
8. Trovare le coordinate del carbonio alfa di Asp174

SystemsBiology

Esercizio 1

Traccia

In seguito ad una variazione locale di pH la proteina P_1 subisce un cambiamento conformazionale che la rende attiva. La forma attiva della proteina (P_{1a}) è in grado di legare la proteina P_2 in modo reversibile con le costanti cinetiche $k_{as} = 1.4 \times 10^5 M^{-1} s^{-1}$ e $k_{dis} = 5 \times 10^{-2} s^{-1}$.

Sapendo che le concentrazioni iniziali della proteina P_1 e della proteina P_2 sono rispettivamente $35 \mu M$ e $24 \mu M$ e che la costante cinetica di variazione conformazionale della proteina P_1 è $k_{conf} = 7 \times 10^3 M^{-1} s^{-1}$ simulare il sistema dinamicamente e rispondere ai seguenti quesiti:

1. Sono sufficienti 0.2 secondi (200 ms) per stabilire l'equilibrio?
2. Assumendo che a $t=20s$ il sistema sia all'equilibrio, determinare empiricamente la costante di equilibrio e confrontarla con la costante di equilibrio teorica.
3. Supponendo ora che la proteina P_{1a} sia soggetta a degradazione con una costante cinetica $k_{deg} = 1.2 \times 10^{-1} M^{-1} s^{-1}$, per quanto tempo nel sistema si può rilevare la presenza di P_3 ?

Soluzione

```
*****MODEL NAME

*****MODEL NOTES

*****MODEL STATE INFORMATION
P1(0)=25e-6    %M (moli/I)
P2(0)=24e-6    %M (moli/I)

*****MODEL PARAMETERS
k_conf=7e3  \%M^-1 s^-1
k_as=1.4e5  \%M^-1 s^-1
k_dis=5e-2  \%s^-1
k_deg=1.2e-2 %aggiunto successivamente
              %all'esercizio 3

*****MODEL VARIABLES
```



```

*****MODEL REACTIONS
P1 => P1a : r1
vf=k_conf*P1
P1a + P2 <=> P3: r2
vf=k_as*P1a*P2
vr=k_dis*P3
P1a => :r3      %aggiunto successivamente
vf=k_deg*P1a    %all'esercizio 3

*****MODEL FUNCTIONS

*****MODEL EVENTS

*****MODEL MATLAB FUNCTIONS

```

Risposte

Domanda 1 Sono sufficienti per la prima reazione sono sufficienti, ma per la seconda no. Né P_3 né P_{1a} hanno raggiunto l'equilibrio. Già dopo 10 secondi invece è visibile l'equilibrio raggiunto da tutte le specie. Quindi non sono sufficienti per stabilire l'equilibrio dell'intero sistema.

Domanda 2

$$k_{eq} = \frac{P_3}{P_{1a} * P_2} = 6.8 \times 10^6$$

$$K_{as} * P_{1a} * P_2 = k_{dis} * P_3$$

Dobbiamo ora calcolare la costante di equilibrio empirica:

$$P_3(20)c.a. = 2.3 \times 10^{-5}$$

$$P_{1a}(20)c.a. = 1.13 \times 10^{-5}$$

$$P_2(20)c.a. = 1.08 \times 10^{-7}$$

$$k_{eq} = \frac{P_3(20)}{P_{1a}(20) * P_2(20)} = \frac{2.3 \times 10^{-5}}{1.13 \times 10^{-5} * 1.08 \times 10^{-7}} = 6.8 \times 10^6$$

Domanda 3 Per capirlo aggiungiamo la reazione 3, con una nuova costante k_{deg} .

Dopo circa 10^5 secondi (quindi circa 28 ore) abbiamo raggiunto lo zero (più o meno) per P_3 .

Spiegazione: dopo la prima reazione, di dissociazione di P_1 , P_{1a} tende a dissiparsi, quindi non è possibile dopo le 28 ore che si formi P_3 e quindi poi tende a sparire.

Esercizio 2

Traccia

La proteina P, è sintetizzata dai ribosomi con una costante cinetica kl pari a $5 \times 10^{-7} M^{-1} s^{-1}$. È noto che la proteina P_1 dimerizza (formando il dimero P_2) con una $K_{Dim.} = 5 nM$ e che il dimero può reversibilmente dissociare con una costante cinetica di dissociazione $k_{dim_d} = 5 \times 10^{-5} s^{-1}$.

Nella stessa cellula, un enzima E lega irreversibilmente un cofattore C con una costante cinetica $k_{cof} = 8.4 \times 10^4 M^{-1} s^{-1}$ a formare l'enzima attivo E_a . Quest'ultimo catalizza l'attivazione del dimero P_2 , trasformandolo quindi in P_{2a} , con una costante di catalisi $k_{cat} = 1.2 \times 10^{-2} s^{-1}$ ed una costante di Michaelis $K_M = 5 \mu M$. Il dimero attivato lega poi un recettore intracellulare R in modo reversibile a formare il complesso $P_{2a}R$ con costanti cinetiche di associazione e dissociazione rispettivamente $k_{Ra} = 1.3 \times 10^5 M^{-1} s^{-1}$ e $k_{Rd} = 10^{-2} s^{-1}$. Il complesso $P_{2a}R$ dissocia poi in modo irreversibile nel dimero P_2 inattivo e nel recettore attivo R , con costante cinetica $k_{Ract} = 4 \times 10^{-1} s^{-1}$. Sapendo che le concentrazioni iniziali delle specie molecolari presenti nel sistema sono: $P_1(0) = 1.5 \mu M$, $E(0) = 10.5 \mu M$, $C(0) = 5.3 \mu M$, $R(0) = 320 \mu M$, simulare dinamicamente il sistema e rispondere ai seguenti quesiti:

1. Dopo quanto tempo il recettore R è totalmente saturato da P_{2a} ? Qual è il rispettivo valore massimo di produzione di R_a , la sua forma attiva?
2. Assumendo che a $t=20s$ il sistema sia all'equilibrio, determinare empiricamente la costante di equilibrio e confrontarla con la costante di equilibrio teorica.
3. Supponendo ora che la proteina P_{1a} sia soggetta a degradazione con una costante cinetica $k_{deg} = 1.2 \times 10^{-1} M^{-1} s^{-1}$, per quanto tempo nel sistema si può rilevare la presenza di P_3 ?

Soluzione

*****MODEL NAME

Esercizio 2

*****MODEL NOTES

*****MODEL STATE INFORMATION

P1(0)=1\item5e-6 %M (moli/I)

E(0)=10.5e-6 %M (moli/I)

C(0)=5.3e-6 %M (moli/I)

R(0)=320e-6 %M (moli/I)

*****MODEL PARAMETERS

```

k_1=5e-7      %M^-1 s^-1
k_Dim=5e-9    %M^-1 s^-1
kdim_d=5e-5    %s^-1
k_cof=8.4e4    %M^-1s^-1 %alla terza domanda cambia
k_cat=1.2e-2   %s^-1
KM=5e-6        %M
k_Ra=1.3e5
k_Rd=1e-2
k_Ract=4e-1

*****MODEL VARIABLES
kdim_a=kdim_d/k_Dim
VMax=k_cat*Ea

*****MODEL REACTIONS
P1 => P1 : r1 % biosintesi della proteina P1
    vf=k_1
P1 + P1 <=> P2: r2 %dimerizzazione della proteina P1
    vf=kdim_a * P1^2
    vr=kdim_d * P2
E + C => Ea : r3 %enzima lefa cofattore irreversibilmente e diventa attiva
    vf=kcof*E*C
P2 => P2a :r4 %attivazione enzimatica di P2
    vf=VMax*P2)/(P2+KM)
P2a + R <=> P2aR :r5 %il dimero attivato
    vf=kRa * P2a * R
    vr=kRd*P2aR
P2aE => P2 + Ra :r6 %di
    vf=kRact*P2aR

*****MODEL FUNCTIONS

*****MODEL EVENTS

*****MODEL MATLAB FUNCTIONS

```

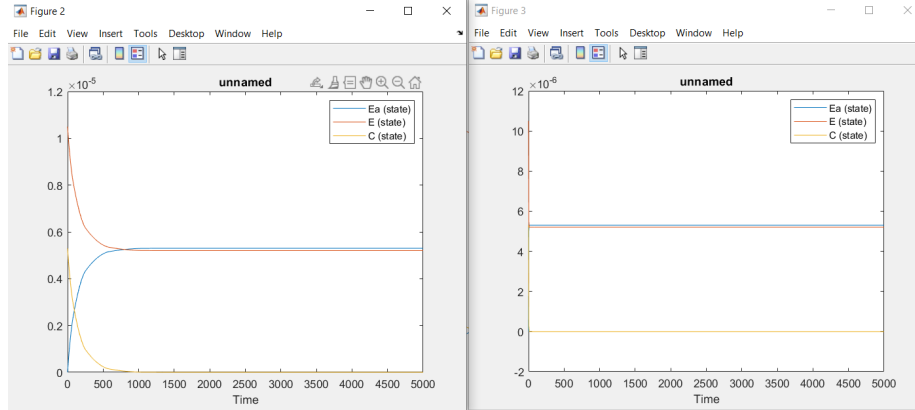
Risposte

Domanda 1 Dopo i 5500 secondi il reagente R è completamente saturato.

Domanda 2 Decresce più velocemente E.

Domanda 3 Le due figure differiscono, perchè con 10^2 su si vedono bene le curve con cui aumentano e diminuiscono i reagenti sui 5000 secondi; inizialmente con 10^4 non si nota quasi, perché troppo veloce per essere visualizzata bene.

La misura di E, C e Ea cambia (ma non abbiamo visto come).



Domanda 4 Per vederlo abbiamo runnato per 400 secondi:
La concentrazione di Ra è circa 2.1×10^{-5} , mentre quella mutata è 14.5×10^{-6} .
Calcoliamo la percentuale di Ra mutata relativamente alla quantità normalmente prodotta: dopo 400 secondi la proteina mutata è il 70% di quella non mutata.

Domanda 5 Raddoppiamo $R(0)$ per verificare la differenza con Wild Type in caso di overespressione e poi lo dimezziamo per vedere una down-regolazione:
In 2000 secondi si nota sui grafici la differenza:

