Esame 08/06/2012

Andrea Passerini
passerini@disi.unitn.it

Informatica

Programma python

Scrivere una funzione `bindingResiduesHistogram(bindingfile)` che:

- prenda in ingresso un file `bindingfile` di proteine ciascuna rappresentata da :
  - nome (stile fasta)
  - sequenza di residui
  - sequenza di etichette di legame con RNA dei residui ('+' = lega , '-' non lega)
- legga le informazioni sulle proteine dal file
- estragga tutti i residui che legano RNA e ne stampi l'istrogramma

Esempio dati

```
$ cat rbp_binding
>3BSU:A
HMFYAVRRGRKTGVFLTWNECRAQVDRFPAARFKKFATEDEAWAFVRK
--------------------+--++++++----------------
.....
>1DI2:A
MPVGSLQELAVQKGWRLPEYTVAQESGPPHKREFTITCRVETFV...
-------------------------++--------------...
....
```

Esempio esecuzione

```
>>> import utility
>>> utility.bindingResiduesHistogram('rbp_binding')
A 278
C 38
E 217
D 217
G 427
F 161
I 159
H 196
K 661
M 109
L 238
N 270
Q 231
P 238
S 337
R 837
T 306
W 69
V 221
Y 226
```

Programma python: suggerimento

Si possono implementare 5 funzioni separate:

1. una che legga da file sequenza di residui e sequenza di etichette per ogni proteina e le restituisca (e.g. dizionario)
2. una che prenda in ingresso il dizionario creato e, processando tutte le sequenze, estragga tutti i residui che legano RNA
3. una che prenda in ingresso i residui e ne calcoli l'istrogramma
4. una che stampi l'istrogramma
5. una che realizzi il programma richiesto usando le funzioni di cui sopra

Esercizi da linea di comando

- Contare quante sequenze nella directory `fasta` includono un potenziale sito di N-glicosilazione. Questo e' individuato dalla seguente sequenza di aminoacidi:
  1. il primo deve essere una asparagina (N)
  2. il secondo non puo' essere una prolina (P)
  3. il terzo deve essere una serina (S) o una treonina (T)
  4. il quarto non puo' essere una prolina
- Risposta: `69`

Esercizi da linea di comando

- Stampare a schermo la prima parola (in ordine alfabetico), del dizionario `dictionary` che non sia un nome proprio (cioe' la cui iniziale non sia maiuscola.)
- Risposta: `aardvark`.