

Esame 25/07/2013

Andrea Passerini
passerini@disi.unitn.it

Informatica

Programma python

Scrivere una funzione `printMetalBindingSites(filename)` che:

- prenda in ingresso un nome di file con un elenco di proteine con i rispettivi metal binding sites
- stampi per ogni proteina il nome e l'elenco delle posizioni di ciascun residuo del metal binding site

File mbs.txt

```
chainId: 1g25A                # nome proteina
chPositions: 5 8 25 27 30 33 45 48 # elenco posizioni C e H
numSites: 2                    # numero siti
chInSite1: 0 1 4 5            # C e H primo sito
chInSite2: 2 3 6 7            # C e H secondo sito
chainId: 1ocyA
chPositions: 32 33 42 78 99 115 117
numSites: 1
chInSite1: 5 6
chainId: 1cw0A
chPositions: 3 57 62 64 67 68 69 70 71 115 137 151
numSites: 1
chInSite1: 3 6 8 9
chainId: 1fb1A
chPositions: 71 86 88 89 98 155 157
numSites: 1
chInSite1: 1 3 6
...
```

Esempio esecuzione

```
>>> import utility
>>> utility.printMetalBindingSites('mbs.txt')
1fb1A
86  89  157

1b0jA
357  420  423

1d4uA
7   10  28  31

1e29A
36  39  40  91

1b9rA
38  44  47  85

...
```

Programma python: suggerimento

Si possono implementare 4 funzioni separate:

- 1 una che legga il file e restituisca un dizionario con il nome di proteina come chiave e come valori l'elenco delle posizioni e la lista dei siti
- 2 una che scorra il dizionario e per ogni proteina ne stampi i siti chiamando una funzione ausiliaria
- 3 una funzione ausiliaria che prenda nome di proteina, posizioni ed elenco di siti, stampi il nome della proteina e per ogni sito:
 - 1 estragga ciascun elemento
 - 2 recuperi dalla lista di posizioni la posizione corrispondente all'elemento
 - 3 stampi la posizione
- 4 una che realizzi il programma richiesto usando le funzioni di cui sopra

Esercizio Shell 1

Dato il file `mbs.txt`, controllare se ci sono proteine che compaiono più di una volta.

Soluzione

No.

Esercizio Shell 2

Date le sequenze nella directory `fasta`, contare quante proteine includono almeno uno dei seguenti due motivi (sequenze di aminoacidi):

- Motivo SH3/2: una prolina, due aminoacidi qualunque, una prolina, un aminoacido qualunque, ed infine una lisina (K) oppure una arginina (R).
- Motivo SH3/5: una prolina (P), due aminoacidi qualunque, un acido aspartico (D), ed infine una tirosina (Y).

Soluzione

5