

Esame 27/11/2013

Andrea Passerini
passerini@disi.unitn.it

Informatica

Programma python

Scrivere una funzione `extractSites(filename)` che:

- prenda in ingresso un nome di file, con un elenco dei residui di una proteina e, ove presente, lo ione metallico che lega il residuo
- stampi la sottosequenza compresa tra il primo e l'ultimo residuo leganti uno ione, con in maiuscolo i residui che legano, ed in minuscolo quelli che non legano

File mbs/1fp0A

1fp0A	0	M		
1fp0A	1	R		
1fp0A	2	G		
1fp0A	3	S		
1fp0A	4	H		
1fp0A	5	H		
...				
1fp0A	27	C	90	ZN
1fp0A	28	R		
1fp0A	29	V		
1fp0A	30	C	90	ZN
...				
1fp0A	67	L		
1fp0A	68	C	89	ZN
...				

Esempio esecuzione

```
>>> import utility  
>>> utility.extractSites('mbs/1fp0A')  
CrvCqkpgdlvmCnqCefcfHldChlpalqdvpggeewsCslC
```

Programma python: suggerimento

Si possono implementare 4 funzioni separate:

- 1 una che legga il file e restituisca la sequenza dei residui della proteina
- 2 una che legga il file (lo stesso) e restituisca la lista di posizioni dei leganti
- 3 una che date sequenza di residui e lista di posizioni leganti, stampi la sottosequenza dei residui tra la prima e l'ultima posizione legante, in maiuscolo se leganti e minuscolo se non leganti
- 4 una che realizzi il programma richiesto usando le funzioni di cui sopra

Shell: Esercizio 1

L'intestazione dei file fasta nella directory `fasta` contiene sia il nome della proteina, sia la catena da cui è presa la sequenza.

Esempio: `>1a02:f`, `>1tg0:a`, `>2pnv:a`.

Stampare a schermo la lista delle catene presenti in *tutti* i file fasta ed il rispettivo numero di ripetizioni. La lista deve essere ordinata per numero di ripetizioni.

Risultato

```
105 a
 41 b
 16 c
...
```

Shell: Esercizio 2

Contare quante delle sequenze nella directory *fasta* contengono *almeno uno* dei due seguenti motivi di fosforilazione:

- Una arginina (R); un amino acido qualunque; una non-prolina (P); una serina (S) oppure una treonina (T); una non-prolina (P); ed una prolina (P).
- Una arginina (R); due amino acidi qualunque; una non-prolina (P); una serina (S) oppure una treonina (T); un aminoacido tra isoleucina (I), valina (V), leucina (L), metionina (M); ed una prolina.

Risultato

20