

Esame 12/02/2013

Andrea Passerini
passerini@disi.unitn.it

Informatica

Programma python

Scrivere una funzione `computeMinDistances(data)` che:

- prenda in ingresso un nome di file `data` che contiene un elenco di polisomi, con le coordinate spaziali dei loro ribosomi
- stampi per ciascun polisoma la distanza di ogni suo ribosoma dal ribosoma più vicino

File dei polisomi (polisomi.txt)

----- HMW13_test15_cluster_1_1.sel_coord_ -----

X (a.u.) Y (a.u.) Z (a.u.)

82 82 202

92 81 220

106 78 194

117 84 220

108 95 230

118 100 178

89 98 232

70 94 187

72 106 172

84 112 185

104 121 209

100 130 152

84 129 163

image X230113_2_0005_2_000005.xmp

13 ribosomes

=====

61 76 182

58 89 199

71 112 221

...

Esempio esecuzione

```
>>> import utility
>>> utility.computeMinDistances('polisomi.txt')
X230113_2_0005_2_000005.xmp
17.378147197
17.378147197
18.6815416923
18.6815416923
19.3132079158
19.3390796058
19.4422220952
19.4422220952
20.6155281281
20.6155281281
25.6124969497
29.7321374946
32.511536414
X230113_2_0003_2_000009.xmp
15.4272486205
15.4272486205
18.4119526395
...
```

Programma python: suggerimento

Si possono implementare 4 funzioni separate:

- 1 una che legga il file e restituisca un dizionario con nome di polisoma come chiave, e come valore la lista dei suoi ribosomi
- 2 una che data una lista di ribosomi, calcoli per ogni ribosoma la distanza minima da un altro ribosoma, e restituisca la lista di tali distanze
- 3 una che stampi una lista di distanze, ordinandole
- 4 una che realizzi il programma richiesto usando le funzioni di cui sopra

Nota

- La distanza tra due punti si calcola come distanza Euclidea:

$$d((x, y, z), (x', y', z')) = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}$$

- Il modulo `math` contiene una funzione `math.sqrt` che calcola la radica quadrata di un numero

Esercizio Shell 1

Ciascuno dei file `fasta` nella directory `fasta` descrive una catena di una proteina. La quinta lettera dell'ID dei `fasta` identifica la catena (A, B, *etc.*). Trovare la catena C più lunga.

Soluzione

`1m71c.`

Esercizio Shell 2

Contare tra i file fasta quante catene includono un **SUMO-binding motif**. La sequenza è composta da:

- Tre residui, ciascuno dei quali può essere una isoleucina (I), una leucina (L), oppure una valina (V).
- Tra il primo ed il secondo residuo, **oppure** tra il secondo ed il terzo, può essere presente (ma non è necessario!) un solo residuo aggiuntivo di tipo qualunque.
- Da uno a dieci residui, ciascuno dei quali può essere una serina (S), una treonina (T), un acido aspartico (D), oppure un acido glutammico (E)

Esempi: MDSSA**VYVLSS**MARQ, AISVLLAQAVF**LLLTS**QR

Soluzione

41