

Esame 20/06/2011

Andrea Passerini
passerini@disi.unitn.it

Informatica

Programma python

Date:

- Una directory (`Fasta`) contenenti sequenze proteiche in formato FASTA
- Una directory (`Sites`) contenenti liste di posizioni (limitatamente a C ed H) corrispondenti a siti di legame metallico, ad esempio (`Sites/1a71A`):

2 4 14

5 6 7 9

Programma python

Scrivere una funzione

`printResidues4Site(fastafile, sitefile, candidates)`
che:

- prenda in ingresso i nomi del file fasta, del corrispondente file con i siti, e dei residui candidati (ossia "CH")
- legga la sequenza proteica e la lista di siti
- stampi per ogni sito i residui che lo compongono
- E.g. (`utility.py`):

```
>>> import utility
>>> printRes4Site('Fasta/1a71A', 'Sites/1a71A', 'CH')
C H C
C C C C
```

Programma python: suggerimento

Si possono implementare 5 funzioni separate:

- 1 una che legga da un file FASTA in ingresso una sequenza proteica e la restituisca
- 2 una che legga da un file di siti in ingresso una lista di liste, ciascuna contenente le posizioni di un sito, e la restituisca
- 3 una funzione che prenda in ingresso una sequenza ed una stringa di possibili residui ("CH") e restituisca una sequenza con solo residui di quel tipo
- 4 una funzione che prenda in ingresso una sequenza proteica con solo residui "CH" ed una lista di siti, e stampi per ogni sito i residui che lo compongono
- 5 una funzione che utilizzi le quattro funzioni precedenti per produrre il risultato voluto

Esercizi da linea di comando

- Selezionare tra i file nella directory `Fasta` tutte le sequenze contenenti:
 - una Cisteina (C) ed una Istidina (H), separate da due o tre residui che non siano ne' Cisteine ne' Istidine
 - oppure tre Cisteine di seguito
- e.g. ...QRCPS**CGTH**YKLVPH...
- e.g. ...NDHVVG**CCC**YPGNKP

Esercizi da linea di comando

- Calcolare a partire dai file nella directory `Sites` il numero di volte in cui il primo residuo (tra C ed H), ossia quello di posizione 0, compare in un sito in una qualche proteina
- Output:

92