

# Esame 15/02/2017

Andrea Passerini  
passerini@disi.unitn.it

Informatica

## Programma python

Scrivere un programma python che:

- prenda in ingresso un file PDB di una proteina che lega eteroatomi (e.g. metalli)
- stampi per ogni eteroatomo *distinto* statistiche relative a:  
1) la distanza media con i suoi leganti; 2) i conteggi dei diversi aminoacidi che lo legano
- stampi le stesse statistiche aggregate per catena proteica

# Esemplio file

```
...  
SHEET      2 AA5 2 THR E1093 TYR E1095 -1 O THR E1093 N CYS E1089  
SHEET      1 AA6 2 PHE F1087 CYS F1090 0  
SHEET      2 AA6 2 THR F1093 TYR F1095 -1 O THR F1093 N CYS F1089  
LINK        SG CYS A1078 ZN ZN A1203 1555 1555 2.26  
LINK        SG CYS A1081 ZN ZN A1203 1555 1555 2.21  
LINK        SG CYS A1089 ZN ZN A1202 1555 1555 2.29  
LINK        SG CYS A1090 ZN ZN A1202 1555 1555 2.24  
LINK        SG CYS A1096 ZN ZN A1203 1555 1555 2.24  
LINK        SG CYS A1100 ZN ZN A1203 1555 1555 2.21  
LINK        NE2 HIS A1108 ZN ZN A1202 1555 1555 1.98  
LINK        SG CYS A1112 ZN ZN A1202 1555 1555 2.44  
LINK        SG CYS B1078 ZN ZN B1202 1555 1555 2.21  
LINK        SG CYS B1081 ZN ZN B1202 1555 1555 2.25  
LINK        SG CYS B1089 ZN ZN B1203 1555 1555 2.23  
LINK        SG CYS B1090 ZN ZN B1203 1555 1555 2.26  
LINK        SG CYS B1096 ZN ZN B1202 1555 1555 2.29  
LINK        SG CYS B1100 ZN ZN B1202 1555 1555 2.29  
LINK        NE2 HIS B1108 ZN ZN B1203 1555 1555 2.55  
LINK        SG CYS B1112 ZN ZN B1203 1555 1555 2.29  
LINK        SG CYS C1078 ZN ZN C1202 1555 1555 2.27  
LINK        SG CYS C1081 ZN ZN C1202 1555 1555 2.26  
...
```

# Esempio esecuzione

```
> python binding_sites.py
insert file name: 5mq4.pdb
```

SITE	AVG_DIST	LIGANDS
ZN_C1202	2.22	{ 'CYS' : 4 }
ZN_C1201	2.26	{ 'CYS' : 3, 'HIS' : 1 }
ZN_F1202	2.23	{ 'CYS' : 3, 'HIS' : 1 }
ZN_F1203	2.23	{ 'CYS' : 4 }
ZN_E1203	2.31	{ 'CYS' : 4 }
ZN_E1202	2.18	{ 'CYS' : 3, 'HIS' : 1 }
ZN_B1202	2.26	{ 'CYS' : 4 }
ZN_B1203	2.33	{ 'CYS' : 3, 'HIS' : 1 }
ZN_A1203	2.23	{ 'CYS' : 4 }
ZN_A1202	2.24	{ 'CYS' : 3, 'HIS' : 1 }
ZN_D1202	2.31	{ 'CYS' : 3 }
ZN_D1203	2.31	{ 'CYS' : 4 }

  

CHAIN	AVG_DIST	LIGANDS
A	2.23	{ 'CYS' : 7, 'HIS' : 1 }
C	2.24	{ 'CYS' : 7, 'HIS' : 1 }
B	2.30	{ 'CYS' : 7, 'HIS' : 1 }
E	2.25	{ 'CYS' : 7, 'HIS' : 1 }
D	2.31	{ 'CYS' : 7 }
F	2.23	{ 'CYS' : 7, 'HIS' : 1 }

## Programma python: suggerimento

Si possono implementare 5 funzioni separate:

- 1 Una che legga il file e restituisca un dizionario che mappa da eteroatomo ad un elenco dei suoi leganti (notare che servono sia il residuo legante che la distanza)
- 2 Una che dato un elenco di leganti, calcoli le statistiche di distanza media e conteggi dei residui
- 3 Una che dato un dizionario di siti, ne stampi le statistiche usando la funzione sopra
- 4 Una che modifichi il dizionario di siti aggregandoli per catena (per poter usare di nuovo la funzione di stampa)
- 5 una che realizzi il programma richiesto usando le funzioni di cui sopra

## Shell: esercizio #1

Nel file `5mq4.pdb`, la sequenza della proteina è riportata nelle righe che cominciano per `SEQRES`. Esempio:

```
SEQRES      1  A   127   SER MET  SER  LYS  ASN  ...
SEQRES      2  A   127   ILE  ARG  ARG  LEU  ARG  ...
...
SEQRES     10  A   127   GLN  SER  ALA  THR  ALA  ...
SEQRES      1  B   127   SER MET  SER  LYS  ASN  ...
...
```

Stampare a schermo i cinque aminoacidi più frequenti.

## Risultato atteso

```
60 ALA
60 LEU
72 LYS
84 GLN
96 GLU
```

## Shell: esercizio #2

Dato il file `sequences.fasta`, contare quante sequenze contengono:

- Una alanina (A), una valina (V) o una prolina (P); seguita da un aminoacido qualunque; seguito da tre residui, ciascuno una serina (S) oppure una treonina (T)
- Una serina o una treonina; seguita da due o tre aminoacidi qualunque; seguiti da una fenilalanina (F). Deve apparire alla fine della sequenza.

Quante sequenze contengono almeno uno dei due motivi?

Quante almeno due ripetizioni del primo ed una del secondo?

*Le espressioni regolari devono essere il più brevi possibile.*

## Risultato atteso

317, 15, 325, 4.

## Modalita' di esecuzione e consegna

- 1 Avviare la macchina in modalita' `ESAME`
- 2 Autenticarsi con nome utente `sci-esame` e password fornita dal docente
- 3 Il testo del compito ed i file necessari si trovano in una cartella `Testo` sul Desktop
- 4 Realizzare il programma python come file `programma.py` e scrivere gli esercizi da linea di comando in un file di testo `linea_di_comando.txt`
- 5 Creare sul Desktop una cartella con *nome\_cognome* e metterci i due file realizzati.
- 6 Eseguire il logout ma NON spegnere la macchina