

Fecha de entrega: Lunes 21 de Enero

Instrucciones: Resuelva el problema propuesto usando Python. Envíe todos los archivos necesarios para reproducir sus resultados (archivos de datos, códigos .py, notebooks .ipynb, etc.) por email a evohringer@udec.cl.

En esta tarea usted deberá procesar la estructura cristalina de la proteína hemoglobina obtenida a una resolución de 2.1 Å mediante el método de difracción de rayos X. El archivo que contiene la estructura de la hemoglobina lo puede encontrar en <http://www.rcsb.org/pdb/ngl/ngl.do?pdbid=1GZX&bionumber=1> en donde debe descargar el archivo denominado PDB.

El formato del archivo es pdb "Protein Data Bank" que es el formato común para estructuras de proteínas (Mayor información puede obtener bajo <http://www.wwpdb.org/documentation/file-format-content/format33/v3.3.html>).

El archivo contiene información sobre la técnica de cristalización, los detalles de la metodología de Rayos X empleada etc. los cuales son especificadas como *REMARK*, *SEQRES* etc:

```
REMARK      2
REMARK      2 RESOLUTION.      2.1  ANGSTROMS.
REMARK      3
REMARK      3 REFINEMENT.
REMARK      3   PROGRAM      : PROLSQ
REMARK      3   AUTHORS      : KONNERT,HENDRICKSON
REMARK      3
REMARK      3 DATA USED IN REFINEMENT.
REMARK      3   RESOLUTION RANGE HIGH (ANGSTROMS) : 2.1
REMARK      3   RESOLUTION RANGE LOW  (ANGSTROMS) : 10
REMARK      3   DATA CUTOFF          (SIGMA(F)) : 0.0
REMARK      3   COMPLETENESS FOR RANGE          (%) : 95
REMARK      3   NUMBER OF REFLECTIONS              : 24327
REMARK      3
REMARK      3 FIT TO DATA USED IN REFINEMENT.
REMARK      3   CROSS-VALIDATION METHOD              : FREE R
REMARK      3   FREE R VALUE TEST SET SELECTION      : RANDOM
```

Después de esta sección se agregan líneas que contienen la información de cada átomo en la estructura de la proteína que comienzan con *ATOM* o *HETATOM* (cuando pertenecen a la cadena peptídica):

```
ATOM      1  N   VAL A   1      18.432 -2.931   3.579  1.00 37.68      N
ATOM      2  CA  VAL A   1      19.662 -2.549   2.806  1.00 35.41      C
ATOM      3  C   VAL A   1      19.282 -1.939   1.441  1.00 34.04      C
ATOM      4  O   VAL A   1      18.421 -2.497   0.695  1.00 33.95      O
ATOM      5  CB  VAL A   1      20.659 -3.754   2.825  1.00 35.59      C
ATOM      6  CG1 VAL A   1      20.109 -4.992   2.222  1.00 37.84      C
ATOM      7  CG2 VAL A   1      21.982 -3.272   2.245  1.00 36.73      C
ATOM      8  N   LEU A   2      19.905 -0.786   1.169  1.00 29.21      N
```

ATOM	9	CA	LEU	A	2	19.749	-0.064	-0.067	1.00	27.27	C
ATOM	10	C	LEU	A	2	20.513	-0.749	-1.213	1.00	27.19	C
ATOM	11	O	LEU	A	2	21.748	-0.901	-1.212	1.00	27.58	O
ATOM	12	CB	LEU	A	2	20.204	1.339	0.210	1.00	25.79	C
ATOM	13	CG	LEU	A	2	19.275	2.508	0.284	1.00	30.66	C

Estas líneas contienen el índice del átomo, el tipo de átomo, el amino ácido al cual pertenece y su número dentro de la secuencia, las coordenadas x, y, z de este átomo en Angstrom y al final el elemento.

Escriba un programa que procese este archivo y entregue la siguiente información:

- (a) Una lista con los nombres de amino ácidos que componen la cadena peptídica (secuencia) y un histograma con la frecuencia de ocurrencia de un amino ácido en esa lista.
- (b) Las distancias entre cada uno de los cuatro átomos de hierro en Å y ordene estas en una matriz de distancia (4x4) usando numpy.
- (c) La distancia entre los átomos de tipo **CA** del mismo amino ácido presente en las cadenas A, B, C y D e identifique la distancia máxima, la mínima y el promedio.