

Introducción a la Bioinformática

TP1

Enrique Alonso

- 1) DESAFÍO I: ¿Podrías buscar un ejemplo de macromoléculas que almacenen información sobre la 'identidad' de un organismo dado?

Los Ácidos Nucleicos son macromoléculas. El ADN es una de ellas y guarda información sobre la identidad del organismo.

- 2) DESAFÍO II: Proponé una forma de expresar la información contenida en la estructura primaria de las proteínas usando tipos de datos de los lenguajes de programación que conocés.

Dado que la estructura primaria de las proteínas está dada por las cadenas peptídicas y aminoácidos que la componen podríamos pensar en una lista o arreglo (array) donde cada posición sea un aminoácido.

- 3) DESAFÍO III: ¿ En qué tipo de datos podrías expresar la información de la estructura terciaria proteica?

Un array de ternas donde cada posición del array tenga información tridimensional de un punto. La unión de esos puntos daría la estructura terciaria. Solo faltaría darle un "espesor o diámetro" a esa consecución de puntos.

Por ejemplo, dado una terna (x,y,z) : [(0,0,0),(100,0,0),(100,20,20),(100,20,40)]

- 4) DESAFÍO IV: Rosalind Franklin es una científica muy relevante, que tuvo menos reconocimiento del merecido. ¿Cuáles fueron sus contribuciones en este campo? ¿Qué nos cuenta su historia acerca del mundo de la ciencia?

Rosalind Franklin fue parte del equipo que descubrió la estructura del ADN utilizando la técnica de difracción de Rayos X. A juzgar por estos hechos el mundo de la ciencia es bastante injusto. El hecho de que Watson y Crick hayan conseguido el Nobel basados en los estudios de Rosalind y apenas la mencionaran por arriba en una revista científica da muestra de ello.

- 5) DESAFÍO V: Escribí un script en Python que prediga la estructura secundaria que adoptará cada residuo (aminoácido) de la secuencia proteica dada, especificandola como H (si es una hélice), B (si es una hoja beta plegada) y L (si es un bucle o loop).

```
import sys

def getAminoPreference (aminoAcid):
    #esta función toma la tabla y retorna la preferencia de plegado de un aminoácido
    preferences = dict(      glu='H', alu='H', leu='H', met='H', gln='H', lys='H',
arg='H', his='H', val='B', lle='B', tyr='B', cys='B', trp='B', phe='B', thr='B', gly='L',
asn='L', pro='L', ser='L', asp='L' )

    return preferences.get(aminoAcid)

def getSecondaryStructure(proteticChain):

    #Dada una lista de aminoácidos retorna una lista de preferencias de plegado

    result = map(getAminoPreference, proteticChain)

    return result

def main():
    #toma como parámetro una lista expresada entre comillas y separando los elementos con
    una coma.

    #Ejemplo "glu,alu,pro"

    inboundList = sys.argv[1]

    proteticChain = inboundList.split(",")

    print (list(getSecondaryStructure(proteticChain)))

if __name__ == '__main__':
    main()
```

- 6) PARA PENSAR: ¿Cuántas proteínas puede sintetizar un organismo? ¿De qué depende la cantidad y la característica de las proteínas que puede sintetizar un organismo?

Para sintetizar proteínas se requiere de los aminoácidos (algunos sintetizados por el organismo y otros obtenidos por ingesta). Por lo tanto creo que las proteínas que un organismo puede sintetizar son todas aquellas combinaciones de aminoácidos posibles.

- 7) DESAFÍO VI: ¿Qué hace distintos a dos individuos de una especie? Propone una forma de corroborar tu respuesta realizando un diagrama de un posible método computacional para dicho fin.

Entiendo que la cadena ADN es muy similar (aunque no exactamente igual) entre dos individuos de una misma especie.

Entonces una forma de corroborarlo sería comparar ambos ADN y ver que porcentaje de similitud tienen.

SUJETO 1		SUJETO 2
A	COINCIDE	A
A		T
G	COINCIDE	G
A		C
C	COINCIDE	C
T	COINCIDE	T
T	COINCIDE	T

La coincidencia entre estos dos sujetos, por ejemplo seria del 71%.

Para el ejemplo se usaron las letras que representan Acidos nucleicos.

Tabla auxiliar usada para el script del punto 5

Amino acid	Preference		
	α -helix	β -strand	Reverse turn
Glu	1.59	0.52	1.01
Ala	1.41	0.72	0.82
Leu	1.34	1.22	0.57
Met	1.30	1.14	0.52
Gln	1.27	0.98	0.84
Lys	1.23	0.69	1.07
Arg	1.21	0.84	0.90
His	1.05	0.80	0.81
Val	0.90	1.87	0.41
Ile	1.09	1.67	0.47
Tyr	0.74	1.45	0.76
Cys	0.66	1.40	0.54
Trp	1.02	1.35	0.65
Phe	1.16	1.33	0.59
Thr	0.76	1.17	0.90
Gly	0.43	0.58	1.77
Asn	0.76	0.48	1.34
Pro	0.34	0.31	1.32
Ser	0.57	0.96	1.22
Asp	0.99	0.39	1.24