

Facultad de Ciencias
Algoritmos Computacionales

Síntesis de Proteínas en Python

Jimena Ciros Ortiz
Laura Patricia Hernández Manríquez
María Guadalupe Martínez Arellano

Junio 5,2019



Objetivos

- ▶ Realizar un programa en el cual se lleve a cabo el proceso de síntesis de proteínas desde el proceso de duplicación del DNA hasta la traducción
- ▶ Calcular la longitud de una cadena de DNA y graficar la doble hélice con matplotlib
- ▶ Simulación de una proteína

Introducción



El proceso de síntesis de proteínas comienza con la duplicación de una cadena de DNA, seguida de la transcripción de la cadena duplicada a una cadena de RNA y finalmente de la traducción en la cual a partir de cada codón en la cadena de RNA se obtiene un aminoácido correspondiente, lo que forma una cadena de aa, que es la estructura primaria de una proteína.

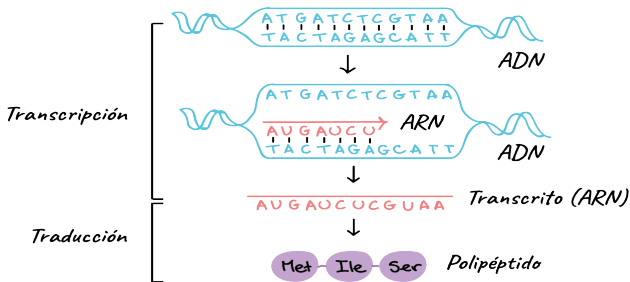


Figure: Síntesis de proteínas



Las cadenas de DNA se obtuvieron de la página de National Center for Biotechnology Information (NCBI)

- ▶ Se duplicación de la cadena de DNA
- ▶ Transcripción de la cadena de DNA a RNA
- ▶ Representación de la transcripción



- ▶ Traducción de la cadena de RNA a cadena de aminoacidos que forma la proteína
- ▶ Calculo de la longitud de una cadena de DNA
- ▶ Grafica del DNA con Matplotlib3D
- ▶ Animación de una proteína con pygame



Pygame es una biblioteca multimedia (que trabaja sobre las librerías SDL) que permiten la creación de videojuegos en dos dimensiones de una manera sencilla. Las SDL (Simple DirectMedia Layer) son un conjunto de bibliotecas que proporcionan funciones básicas para realizar operaciones de dibujo 2D, gestión de efectos de sonido, música, carga y gestión de imágenes.



Figure: Pygame

Módulos de sistema Entre los módulos de sistema que Python nos provee a través de su librería estándar, podemos destacar tres: `os`, `sys`, y `subprocess`.

El módulo `os` nos permite acceder a funcionalidades dependientes del Sistema Operativo. Sobre todo, aquellas que nos refieren información sobre el entorno del mismo y nos permiten manipular la estructura de directorios (para leer y escribir archivos).



Figure: Python



Las cadenas de nucleótidos de DNA obtenidas en la página fueron guardadas en archivos .txt, con un código se realizaba la duplicación de la cadena de DNA y se guarda en otro archivo .txt

```
TACCCAGTCTTTTTCGGGAGAGGTACAGATGAGTGCTATGTAGGGTTTGGTGACTCCTTCACCGAAAAGTCTAGTAGA
ACGAAACGGTCAAACCCCAACCCTGAAAACGGTTACATAAAGAGAAAACAGATATTAAGAGAGGTTAGAGCTGACCAAG
ATTTGTCTCCGGGTCTGTTCAAACGTGTGTACCGACCCGGTTACGGAGAGAGAAGGAGTGATATAAAGGTTTGTGTGA
AAACGAGGTTAATAAGGAGTTTGACTGGAGTTTACATTTATTAAGAAGTGAGCGGAGCACCGTCTTCGTGTTTGAACA
CAAGTTGAACACAAGCATAGGTAGTCAAACAGTGTGAACAAGGACAATTAAGTCCATTTCTTATGAGTCTCGTTCACA
GTGTTTGTACCGTTCAATAAGAACAATGTCAACAACCAAGAAGTCACAGATATTGTAGATGTAAGGTTAATTCCAGTGA
CCAGGTGTCAGTCTGTTATTGTTATTGAGATTTTCGTTCAACAAGACAAGGTGAAGACTAAAGTCACCGTAACAGAAGA
CCAAACGGGTACTACGGTGTAAGTCGTAGTACCAGACCTGGTCACAGAGGTACCATGAAGAGGAGGTATCTGTAACAGT
CTCTTACGTCATGTATAAGTGAGAGTTAGTCCTGGGGTCCCCGGTTCGTCTCTGGTGTCTGTTGGGTATGATAGGACGAC
CAGTGTAACAACCGAAAATAGAAGAATATC
```


Resultados



Con otro código hace la transcripción de la cadena duplicada a una cadena de RNA, y guarda la cadena de RNA en otro archivo .txt

```
AUGGGUCAGAAAAAGCCCUCUCCAUGUCUACUCACGAUACAUCCCAAACCACUGAGGAAGUGGCUUUUUCAGAUCAUCU
UGCUUUGCCAGUUUGGGGUUGGGACUUUUGCCAAUGUAUUUCUCUUUGUCUAUAAUUUCUCUCCAAUCUCGACUGGUUC
UAAACAGAGGCCAGACAAGUUUGACACAUGGCUGUGGCCAAUGCCUCUCUCCUACAUAAUUUCCAAACAACACU
UUUGCUCCAAUUUUCCUCAAACUGACCUCAAAUGUAAAUAAUUCUUCACUCGCCUCGUGGCAAGAAGCACAAACUUGU
GUUCAACUUGUGUUCGUAUCCAUCAGUUUGUCACACUUGUUCUGUUAAUUCAGGUAAAGGAUACUCAGAGCAAGUGU
CACAAACAUGGCAAGUUUUUUGUUACAGUUGUUGGUUCUUCAGUGUCUAUAACAUCUACAUUCCAAUUAAGGUCACU
GGUCCACAGUCAGACAAUAAACAAUAAUCUAAAAAGCAAGUUGUUCUGUUCCACUUCUGAUUUUCAGUGGCAUUGUCUUCU
GGUUUGCCCAUGAUGCCACAUCAGCAUCAUGGUCUGGACCAGUGUCUCCAUGGUACUUCUCCUCCAUAAGACAUUGUCA
GAGAAUGCAGUAUAUUACUCUCUCAAUCAGGACCCCAGGGGCCAAGCAGAGACCACAGCAACCCAUACUAUCCUGCUG
GUCACAUUUUGUGGCUUUUUAUCUUCUUAUAG
```

Resultados



Con otro código se realiza un la representación de la transcripción de la cadena de DNA a RNA.

```
A---U
C-----G
T-----A
C-----G
A-----U
A--U
AU
GC
C--G
A-----U
C-----G
G-----C
G----C
T--A
AU
GC
A--U
C-----G
C-----G
```



Con otro código se hace la traducción de la cadena de RNA, utilizando un diccionario para obtener el aminoacido correspondiente a cada codon, creamos la cadena de aminoacidos.

```
MetGlyGlnLysLysProSerProCysLeuLeuTheIleHisProLysProLeuArgLysTrpLeuPheArgSerSerC  
ysPheAlaSerLeuGlyLeuGlyLeuLeuProMetTyrPheSerLeuSerIleIleSerLeuGlnSerArgLeuValLe  
uAsnArgGlyProAspLysPheAspThrTrpLeuTrpProMetProLeuSerSerSerLeuTyrPheGlnThrThrLeu  
LeuLeuGlnLeuPheLeuLysLeuThrSerAsnValAsnAsnSerSerLeuAlaSerTrpGlnGluAlaGlnThrCysV  
alGlnLeuValPheValSerIleSerLeuSerHisLeuPheLeuLeuIleGlnValLysGluTyrSerGluGlnValSe  
rGlnThrTrpGlnValIleLeuValThrValValGlySerSerValSerIleThrSerThrPheGlnLeuArgSerLeu  
ValHisSerGlnThrIleThrIleThrLeuLysAlaSerCysSerValProLeuLeuIleSerValAlaLeuSerSerG  
lyLeuProMetMetProHisSerAlaSerTrpSerGlyProValSerProTrpTyrPheSerSerIleAspIleValAr  
gGluCysSerThrTyrSerLeuSerIleArgThrProGlyAlaLysGlnArgProGlnGlnProIleLeuSerCysTrp  
SerHisLeuLeuAlaPheIlePheLeuEND
```



Calculo de la longitud del DNA con sympy y scipy

Una molécula de ADN tiene la forma de una doble hélice. El radio de cada hélice es alrededor de 10 angstroms ($1\text{\AA}=10^{-8}$). Cada hélice se eleva unos 34 Å cada vuelta completa y se tiene que hay aprox. 2.9×10^8 vueltas completas.

Para calcular la longitud del DNA aplicaremos longitud de arco que está definida como:

$$s(t) = \int_a^t ||r'(t)|| dt$$

$$r(t) = \left(10 \cos(t), 10 \sin(t), \frac{34}{2\pi} t \right)$$

$$r'(t) = \left(-10 \sin(t), 10 \cos(t), \frac{34}{2\pi} \right)$$

$$||r'(t)|| = \sqrt{100 + \left(\frac{34}{2\pi}\right)^2}$$

$$s(t) = \int_0^{2.9 \times 10^8} \sqrt{100 + \left(\frac{34}{2\pi}\right)^2} dt = 2.07 \times 10^{10} \text{\AA} = 2 \text{ m}$$

Resultados



Grafica del DNA con matplotlib3D, el cual nos permite la visualización de datos en tercera dimensión.

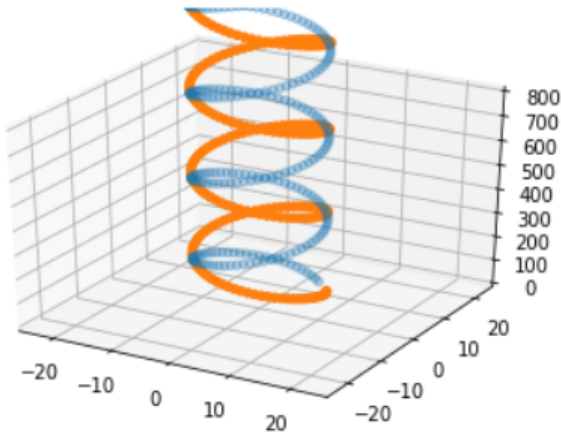


Figure: Grafica del DNA

Simulación de una proteína con pygame.

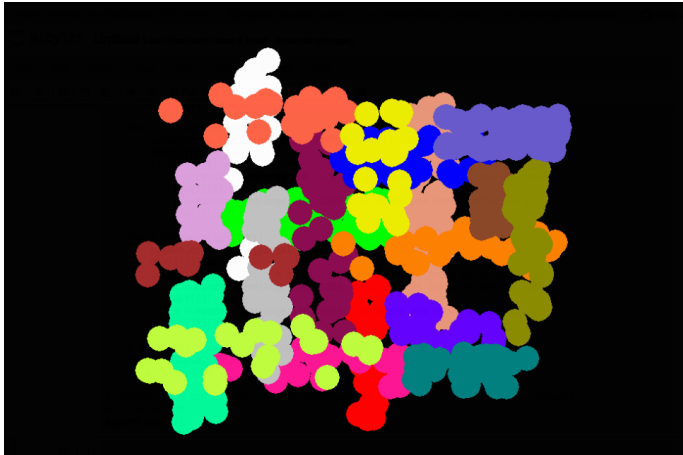


Figure: Simulación de una proteína



- [1] TRUDY MCKEE, JAMES R. MCKEE *Bioquímica: las bases moleculares de la vida* , quinta edición, McGraw-Hill Interamericana, México, DF, 2009.
- [2] <https://www.ncbi.nlm.nih.gov/>
- [3] <https://www.pythonmania.net/es/2010/03/23/tutorial-pygame-introduccion/>
- [4] <https://uniwebsidad.com/libros/python/capitulo-10/modulos-de-sistema>
- [5] <https://github.com/pattyhdz/ProyectoAlgoritmos.git>

A decorative graphic consisting of multiple overlapping, curved lines in shades of orange, peach, and light grey. The lines sweep from the left side of the frame towards the right, where they fan out and end in small, glowing white dots, resembling a stylized wave or a burst of energy.

Gracias por su atención