



## Práctica 1.

### El dogma central de la biología molecular

Francisco Contreras Ibarra ✉, José Ethan Ortega González ✉ y Alvaro Ramírez López ✉

#### PRACTICA 1

Entrega 01/03/2024

##### Números de cuenta

Francisco Contreras Ibarra  
316083786  
José Ethan Ortega González  
316088327  
Alvaro Ramírez López  
316276355

##### Correos

Francisco Contreras Ibarra  
franciscoc.ibarra@ciencias.unam.mx  
José Ethan Ortega González  
ethan@ciencias.unam.mx  
Alvaro Ramírez López  
alvaro@ciencias.unam.mx

### 1. Parte 1: explorando el DNA

Son el equipo a cargo de un grupo de investigación multidisciplinario donde hay de todo: biólogos, matemáticos, terrólogos, etc. Resulta que este grupo se dedica a investigar las formas de vida que hay en ambientes extremos, ésta vez fue el turno de [un manantial hidrotermal en el Parque Nacional de Yellowstone, USA](#). Los biólogos encontraron un bicho interesante, lo aislaron, lo secuenciaron y luego les pasaron a ud [este archivo](#) FASTA. Los archivos con formato .fasta son comunes en bioinformática, son archivos de texto plano con secuencias de ácidos nucleicos o aminoácidos. Por ejemplo:

>Encabezado de la secuencia

```
GTGGCCTTGACGCACGAGGCGGTCTGGCAGCACGTTCTGGAGCACATCCGCCGCAACAGGTACCCCGCTTTGAGCT
TTTGGTGGTGCCCGGAGCGCCAGCCGGTCCAGGTGGAGCCGACCAGGGGAAGTCCAAGCTCCTGGTTTGAAAGGA
TCCGCCCCCTGGGTATCCGTGCCACCTCCTTCGCCCTGGACTGGATCAAGCGGCACTACGCCCCCTGATCCAGG
```

La primera línea inicia con el símbolo mayor que (>), e indica un encabezado, generalmente contiene información sobre la secuencia en cuestión. La segunda línea es la secuencia per se. Es importante mencionar que, por convención la cadena que se maneja en los archivos .fasta es la que tiene dirección 5'-3'.

1. ¿De qué tamaño es la cadena de DNA contenida?
2. Sólo con la secuencia, ¿Podríamos saber si es un genoma completo?
3. ¿Hay otras posibilidades?, ¿Qué opinan? Justifica tus respuestas

#### Solución

1. Si nos referimos al ejemplo visto anteriormente, para determinar la longitud de la cadena de DNA contenida podemos contar el número de bases presentes en la secuencia, nos da un total de 188 bases de nucleótidos.
2. No es posible determinar si la secuencia proporcionada representa un genoma completo solo con la información dada. La secuencia proporcionada es relativamente corta y no contiene suficiente información para determinar si representa un genoma completo. Los genomas completos suelen ser mucho más largos.
3. Sí, hay otras posibilidades. La secuencia proporcionada podría ser una porción de un gen, un fragmento de un genoma más grande, una secuencia codificante específica, etc. Pero sin más contexto o información, es difícil determinar con certeza su función o su relación con un genoma completo.

## 2. El dogma central de la biología molecular

Después de mucho trabajo de investigación, se descubre que hay un gen importante que se encuentra desde la posición 618,862 hasta la 621,360. Por lo que tu equipo desea modelar la proteína y requiere obtener la cadena de aminoácidos, por esto les solicitan generar un programa en Python que puedan correr desde la terminal, que acepte como dato de entrada el archivo FASTA en cuestión y arroje como dato de salida un archivo FASTA con la cadena de aminoácidos en nomenclatura de una sola letra del gen de interés.

### **i Tip**

No pueden utilizar ninguna paquetería/módulo de Python que ya tenga implementadas las funciones para hacer esto.

1. Resulta que el gen de interés codifica para la enzima involucrada en la síntesis de la cadena complementaria de DNA durante la replicación, ¿De qué proteína se trata?
2. ¿De qué tamaño es la cadena de aminoácidos?
3. Con la información que tenemos hasta ahora, ¿Podríamos saber si se trata de un eucarionte, procarionte o virus? ¿Si, no, por qué?
4. Si tu respuesta anterior fue **no**, ¿Qué podríamos hacer para saber ésta información?

### **Solución**

1. Dado que el gen codifica para una enzima involucrada en la síntesis de la cadena complementaria de ADN durante la replicación, podríamos estar hablando de una ADN polimerasa.
2. La cadena de aminoácidos proporcionada contiene 441 aminoácidos en total. Este es el tamaño de la proteína generada por el algoritmo propuesto.
3. Basándonos únicamente en la secuencia de aminoácidos proporcionada, no podemos determinar con certeza si se trata de un eucarionte, procarionte o virus. Esto se debe a que muchas proteínas pueden estar presentes en diferentes tipos de organismos y la información proporcionada es insuficiente para hacer una determinación precisa.

### **! Consideraciones importantes**

Para poder traducir recuerda tomar en cuenta el código genético que vimos en clase. El resumen de la [Clase 2: Ácidos nucleicos y dogma central](#) contiene el código genético, y [aquí](#) se encuentra la nomenclatura de los aminoácidos en una y tres letras. Recuerda que todos los genes tienen un codón de inicio y uno de alto.

## Solución

**Implementacion del generador de cadena del aminoácido:**

```

import sys

def translate_codon(codon):
    codon_table = {
        'ATA': 'I', 'ATC': 'I', 'ATT': 'I', 'ATG': 'M',
        'ACA': 'T', 'ACC': 'T', 'ACG': 'T', 'ACT': 'T',
        'AAC': 'N', 'AAT': 'N', 'AAA': 'K', 'AAG': 'K',
        'AGC': 'S', 'AGT': 'S', 'AGA': 'R', 'AGG': 'R',
        'CTA': 'L', 'CTC': 'L', 'CTG': 'L', 'CTT': 'L',
        'CCA': 'P', 'CCC': 'P', 'CCG': 'P', 'CCT': 'P',
        'CAC': 'H', 'CAT': 'H', 'CAA': 'Q', 'CAG': 'Q',
        'CGA': 'R', 'CGC': 'R', 'CGG': 'R', 'CGT': 'R',
        'GTA': 'V', 'GTC': 'V', 'GTG': 'V', 'GTT': 'V',
        'GCA': 'A', 'GCC': 'A', 'GCG': 'A', 'GCT': 'A',
        'GAC': 'D', 'GAT': 'D', 'GAA': 'E', 'GAG': 'E',
        'GGA': 'G', 'GGC': 'G', 'GGG': 'G', 'GGT': 'G',
        'TCA': 'S', 'TCC': 'S', 'TCG': 'S', 'TCT': 'S',
        'TTC': 'F', 'TTT': 'F', 'TTA': 'L', 'TTG': 'L',
        'TAC': 'Y', 'TAT': 'Y', 'TAA': '*', 'TAG': '*',
        'TGC': 'C', 'TGT': 'C', 'TGA': '*', 'TGG': 'W',
    }
    return codon_table.get(codon, 'X')

def dna_to_protein(dna_sequence):
    protein_sequence = ""
    for i in range(0, len(dna_sequence), 3):
        codon = dna_sequence[i:i+3]
        amino_acid = translate_codon(codon)
        if amino_acid != '*':
            protein_sequence += amino_acid
        else:
            break
    return protein_sequence

if __name__ == "__main__":
    if len(sys.argv) != 3:
        print("Uso: python programa.py archivo_entrada.fasta archivo_salida.fasta")
        sys.exit(1)

    input_file = sys.argv[1]
    output_file = sys.argv[2]

    with open(input_file, "r") as f_in:
        dna_sequence = "".join(line.strip() for line in f_in.readlines()[1:])

    protein_sequence = dna_to_protein(dna_sequence)

    with open(output_file, "w") as f_out:
        f_out.write(f">{input_file}\n")
        f_out.write(protein_sequence)

```

Listado 1: Este es el código que traduce los .fasta .

### 3. Parte 3. Jurassic Park

- Lee el artículo [Jurassic Park: why we're still struggling to realise it 30 years on](#), de al menos 500 palabras.
- Adjunta a tu reporte un glosario de los conceptos que consideres importantes.
- [Aquí](#) hay un manual de cómo hacer un reporte de lectura si es que no conocen el término.

#### Solución

- A continuación, mostramos nuestro reporte de lectura sobre el artículo:

#### 3.1. Jurassic Park: why we're still struggling to realise it 30 years on

En el artículo se discute el impacto que tuvo Jurassic Park en la comunidad científica cuando llegó del cine, en la década de los noventas. En la película se nos presenta la idea de **extraer ADN de insectos fosilizados y conservados en ámbar** para **traer de vuelta a la vida a los dinosaurios**. En su tiempo de estreno, aquella idea fue respaldada por varias publicaciones y estudios hechos en ámbar fosilizado.

Desde entonces la ciencia ha avanzado y **se han creado múltiples técnicas que no estaban disponibles** en la fecha de estreno de la película, además de **numerosos reportes de evidencias de ADN y proteínas**, que brindan información genética, encontrados en fósiles. Sin embargo, estos reportes son causa de controversia entre varios científicos.

Es aquí cuando el artículo describe la posibilidad de recuperar ADN antiguo, señalando que aunque es extremadamente frágil y se descompone rápidamente después de la muerte del organismo, hay casos en donde el ADN ha logrado sobrevivir en climas polares y en fósiles relativamente jóvenes.

Dado que no hay evidencias de que el ADN puede sobrevivir miles de millones de años (como las fechas cuando los dinosaurios aún rondaban por la tierra), se ha cambiado el **enfoque hacía las proteínas**. Los científicos creen que las proteínas sobreviven más tiempo que el ADN, e incluso se han encontrado varios ejemplos de proteínas fosilizadas, particularmente secuencias de aminoácidos de colágeno, aunque estas solo son de hace pocos millones de años.

Además **existen varios reportes** que afirman que **se han encontrado proteínas degradadas y fibras de colágeno de hasta 130 millones de antigüedad**, aunque suelen bastante **criticados y controversiales** por la metodología de los equipos, la posibilidad de la contaminación de las muestras, entre otros aspectos.

Aunque parece ser que la idea de resucitar a los dinosaurios es cada vez más ficción que realidad, el autor nos presenta un estudio que realizó con su equipo que podría traer más optimismo a la mesa.

Se usaron rayos concentrados de luz más rayos X para irradiar muestras de plumas antiguas. Estas técnicas permitieron analizar a *Sinornithosaurus*, un dinosaurio emplumado de hace 125 millones de años. El análisis sugiere que las plumas antiguas

eran bastante similares químicamente a las plumas de hoy en día, además de que **rastros de proteínas sobreviven por milse de millones de años.**

Finalmente, es necesario obtener ADN intacto para poder resucitar a una especie. Si bien los científicos han hechos varios progresos, los datos de los experimentos sugieren que es poco probable que el ADN sobreviva por miles de millones de años. Aunque se encontraran fragmentos de ADN en fósiles de dinosaurios es muy probable que sean muy cortos, y no logren darnos información útil acerca de la especie o sean causa de la fosilización.

De cualquier modo, es necesarios preguntarnos ¿que avances se lograrán en los siguientes años?, ¿que nuevas técnicas surgirán para ayudarnos en el cometido?, ¿encontraremos fósiles que cambiaran o ayudarán a cambiar el curso de estas teorías?. A final de cuentas, aún quedan muchas cosas por descubrir.

— Los bípedos

- También presentamos nuestro glosario con los términos que consideramos importantes:
- **DNA:** El ácido desoxirribonucleico es el material genético de la célula, contenido en los cromosomas dentro del núcleo celular y las mitocondrias.
- **fossil amber:** Es una resina fósil, producto residual de algunos árboles prehistóricos.
- **palaeontologists:** Una persona que aprende de las formas de vida pasadas al excavar fósiles de seres vivos como huesos de dinosaurio, y los estudia
- **proteins:** Son las moléculas formadas por aminoácidos, que están unidos por un tipo de enlaces conocidos como enlaces peptídicos.
- **molecules:** Es un conjunto de átomos (de un mismo elemento químico o de muchos diferentes) que están organizados e interrelacionados mediante enlaces químicos.
- **genetic material:** Cualquier material de origen vegetal, animal o microbiano u otro que tenga información genética y que la transmita de una generación a la siguiente.
- **chromosomes:** Un cromosoma es la estructura que alberga al ADN en la célula
- **cell:** Es la unidad más pequeña que puede vivir por sí sola. Forma todos los organismos vivos y los tejidos del cuerpo.
- **X-rays:** Son una forma de radiación electromagnética, similar a la luz visible.
- **alpha-keratin:** Es un tipo de queratina. Esta proteína es el principal componente de los pelos, cuernos, uñas y la capa epidérmica de la piel de muchos animales, especialmente mamíferos.
- **fossilisation:** Es una transformación físico-química que a lo largo de cientos de miles de años va sufriendo el organismo (ya sea este un animal o una planta) hasta convertirlo en un fósil.
- **haemoglobin:** Proteína del interior de los glóbulos rojos que transporta oxígeno desde los pulmones a los tejidos y órganos del cuerpo
- **red blood cells:** Es un tipo de glóbulo sanguíneo (célula de la sangre) que se produce en la médula ósea y se encuentra en la sangre. Los glóbulos rojos contienen una proteína llamada hemoglobina, que transporta oxígeno desde los pulmones a todas las partes del cuerpo.
- **amino acids:** Son moléculas que se combinan para formar proteínas.

## Bibliografía

- [1] «Figure: Estructura del DNA - Manual MSD versión para profesionales». [En línea]. Disponible en: <https://www.msdmanuals.com/es-mx/professional/multimedia/figure/estructura-del-dna>
- [2] «– Ámbar». [En línea]. Disponible en: <https://amberworldmuseum.com/ambar/#:~:text=El%20Ambar%20es%20una%20resina,en%20material%20estable%20y%20hermoso>
- [3] «PNC Crezca con Éxito– Lecciones de los ayudantes comunitarios». [En línea]. Disponible en: <https://www.pnc.com/es/about-pnc/corporate-responsibility/grow-up-great/lesson-center/community-helpers/dig-it.html#:~:text=paleont%C3%B3logo%3A%20una%20persona%20que%20aprende,en%20la%20tierra%20o%20roca>
- [4] «Wikiwand - Alfa-Queratina». [En línea]. Disponible en: <https://www.wikiwand.com/es/Alfa-queratina>
- [5] E. Equipo editorial, «20 ejemplos de proteínas». [En línea]. Disponible en: <https://www.ejemplos.co/20-ejemplos-de-proteinas/#:~:text=Con%20el%20nombre%20de%20prote%C3%ADnas,%2C%20elastina%2C%20actina%2C%20insulina>
- [6] «Glosario: Material genético - Comisión Europea». [En línea]. Disponible en: [https://ec.europa.eu/health/scientific\\_committees/opinions\\_layman/security-scanners/es/glosario/mno/material-genetico.htm](https://ec.europa.eu/health/scientific_committees/opinions_layman/security-scanners/es/glosario/mno/material-genetico.htm)
- [7] «Diccionario de cáncer del NCI». [En línea]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/celula>
- [8] «Rayos X». [En línea]. Disponible en: <https://www.nibib.nih.gov/espanol/temas-cientificos/rayos-x#:~:text=Los%20rayos%20X%20son%20una,los%20objetos%2C%20incluso%20el%20cuerpo>
- [9] M. M. de Fósiles y Minerales, «Fósiles qué son | cómo se forman | tipos de fosilización - Mufomi». [En línea]. Disponible en: <https://mufomi.org/educacion/fosiles-que-son/>
- [10] «Diccionario de cáncer del NCI». [En línea]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/globulo-rojo>
- [11] «Aminoácidos: MedlinePlus enciclopedia médica». [En línea]. Disponible en: <https://medlineplus.gov/spanish/ency/article/002222.htm>