

Práctica de Laboratorio 1 – FASTA

Gisela Belmonte Cruz

Diego Marrero Ferrera

Octubre 2025

Índice

1. Ejercicio 1 – Análisis de secuencias FASTA	1
2. Ejercicio 2 – Obtención de proteínas y análisis	2

1. Ejercicio 1 – Análisis de secuencias FASTA

Esta primera parte consistió en desarrollar un programa en Python que trabaja con archivos **FASTA** que contienen varias secuencias de ADN.

El programa realiza los siguientes pasos:

1. Lee un fichero FASTA con múltiples secuencias de ADN.
2. Calcula el **contenido de GC (%)** de cada secuencia y lo muestra por pantalla.
3. Genera las correspondientes **cadena de aminoácidos** a partir de las secuencias de ADN.
4. Guarda las secuencias traducidas en un nuevo archivo FASTA, conservando los identificadores y descripciones originales.

A continuación se muestra el código desarrollado para este ejercicio:

Listing 1: Código del ejercicio 1 (ejercicio1.py)

```
from Bio import SeqIO
from Bio.SeqRecord import SeqRecord

in_fasta = "dna_seqs.fasta"
out_fasta = "proteins.fasta"

protein_records = []
```

```
registros = SeqIO.parse(in_fasta, "fasta")

try:
    while True:
        registro_secuencia = next(registros)
        print(f"\nID: {registro_secuencia.id}")

        seq = registro_secuencia.seq.upper()
        total = len(seq)
        count_GC = seq.count("G") + seq.count("C")
        porcentaje_GC = (count_GC / total) * 100 if
            total > 0 else 0

        protein_seq = registro_secuencia.seq.translate(
            table=1, to_stop=True)

        print(f"Cadena de aminoacidos: {protein_seq}")
        print(f"Longitud: {total}")
        print(f"GC count: {count_GC}")
        print(f"GC%: {porcentaje_GC:.2f}")

        prot_record = SeqRecord(
            protein_seq,
            id=registro_secuencia.id,
            description=registro_secuencia.description
        )
        protein_records.append(prot_record)

except StopIteration:
    print("\nNo hay mas secuencias en el fichero")

SeqIO.write(protein_records, out_fasta, "fasta")
print(f"\nSe escribieron {len(protein_records)}
    secuencias de proteinas en {out_fasta}")
```

2. Ejercicio 2 – Obtención de proteínas y análisis

En esta segunda parte se realizó una búsqueda de información y descarga de los archivos FASTA correspondientes a tres proteínas: **Oxitocina**, **Colágeno** y **Queratina**.

Descripción de las proteínas

Oxitocina es una hormona formada por una cadena corta de aminoácidos que actúa tanto en el cerebro como en el cuerpo. Es conocida popularmente como la *hormona del amor* porque está relacionada con la empatía, la confianza y las relaciones sociales. Además, desempeña un papel fundamental durante el parto, al estimular las contracciones uterinas, y en la lactancia, favoreciendo la liberación de leche. Más allá de sus funciones biológicas, la oxitocina ha sido objeto de numerosos estudios por su influencia en el comportamiento social y emocional.

Colágeno es la proteína más abundante del cuerpo humano y constituye la base estructural de la piel, los huesos, los tendones y los cartílagos. Su función principal es aportar resistencia y elasticidad a los tejidos. Con el paso del tiempo, la producción natural de colágeno disminuye, lo que se asocia al envejecimiento de la piel y la pérdida de firmeza. Existen diferentes tipos de colágeno, cada uno adaptado a un tejido específico, y su secuencia rica en glicina y prolina le da la estabilidad necesaria para formar fibras resistentes.

Queratina es una proteína fibrosa muy resistente e insoluble, que forma parte del pelo, las uñas y la capa externa de la piel. Actúa como una barrera protectora frente a agentes externos (calor, radiación, productos químicos, etc.). Por su capacidad para mantener la forma y resistencia del pelo, la queratina también se ha popularizado en tratamientos cosméticos que buscan reparar el daño capilar. En animales, las variaciones de queratina son las responsables de estructuras como plumas, cuernos o garras.

Script empleado

Los archivos FASTA de estas proteínas fueron procesados con el siguiente script:

Listing 2: Código del ejercicio 2 (ejercicio2.py)

```
from Bio import SeqIO
from pathlib import Path

protein_dir = Path("proteins")

proteins = {
    "Oxitocina": protein_dir / "oxytocin.fasta",
    "Colageno": protein_dir / "collagen.fasta",
    "Queratina": protein_dir / "keratin.fasta"
}
```

```
print("Se han tomado como proteínas las siguientes:")
print("- Oxitocina: hormona relacionada con el afecto,
      el parto y la lactancia.")
print("- Colageno: prote na estructural que da soporte
      y resistencia a tejidos.")
print("- Queratina: prote na fibrosa que forma parte
      del cabello, u as y piel.\n")

for name, filepath in proteins.items():
    record = SeqIO.read(filepath, "fasta")
    print(f"> {name}")
    print(record.seq)
    print()
```

Fuentes consultadas (Datos y Secuencias FASTA)

- UniProt – Universal Protein Resource
- NCBI – National Center for Biotechnology Information

Repositorio del proyecto

Enlace al código completo y resultados: https://github.com/giselabcruz/pl_1_bio