

Ejercicios2_algoritmos

Tania Gonzalo y Daniel Parra

WORKSHEET

Ejercicio 1 (4 puntos)

En este ejercicio probarás el algoritmo Needleman-Wunsch en una secuencia corta de partes de hemoglobina (código PDB 1AOW) y mioglobina 1 (código PDB 1AZI). Aquí alineará la secuencia HGSAQVKGHG con la secuencia KTEAEMKASEDLKKHGT.

Las dos secuencias están dispuestas en una matriz en la Tabla 1. Las secuencias comienzan en la esquina superior derecha, y las penalizaciones por desfase inicial se enumeran en cada posición inicial de desfase. La penalización por desfase se considera -8. Las puntuaciones de similitud $S_{i,j}$ procedentes de la búsqueda de coincidencias proceden de la tabla BLOSUM40.

Para resolver este ejercicio, hemos aplicado el algoritmo de Needleman-Wunsch para alinear dos secuencias de proteínas que nos indican: una parte de la hemoglobina (HGSAQVKGHG) y una parte de la mioglobina (KTEAEMKASEDLKKHGT). Este algoritmo de alineamiento global nos permite encontrar la mejor correspondencia entre las secuencias considerando inserciones, eliminaciones y sustituciones de aminoácidos.

Para llevar a cabo la alineación, seguimos los siguientes pasos:

1. **Inicialización de la matriz:** Se construyó una matriz en la que las secuencias se dispusieron en los ejes horizontal y vertical. Se asignaron penalizaciones de apertura de brecha (-8) en la primera fila y la primera columna.
2. **Cálculo de los valores en la matriz:** Cada celda se llenó utilizando la fórmula de Needleman-Wunsch:

$$F(i, j) = \max \begin{cases} F(i-1, j-1) + S(i, j) & \text{(match o mismatch según la matriz BLOSUM40)} \\ F(i-1, j) + \text{gap penalty} & \text{(inserción)} \\ F(i, j-1) + \text{gap penalty} & \text{(eliminación)} \end{cases}$$

Donde $S(i, j)$ es el puntaje de sustitución obtenido de la matriz BLOSUM40.

La tabla obtenida se muestra en Figura 1 donde aparecen todas las celdas rellenas además de las flechas indicativas de la procedencia de cada uno de los números para posteriormente poder llevar a cabo la reconstrucción del alineamiento.

3. **Obtención del score final:** Una vez completada la matriz, el score final de alineación se encuentra en la casilla inferior derecha (Figura 1). Como podemos observar, en nuestro caso el score obtenido es de -21.
4. **Retroceso para encontrar las alineaciones óptimas:** Siguiendo las flechas que indicaban la mejor elección en cada celda, realizamos un recorrido inverso para reconstruir los alineamientos óptimos. Encontramos dos posibles caminos (alternativas), lo que dio lugar a dos alineaciones diferentes, ambos mostrados en la parte inferior de la Figura 1.

Este procedimiento nos permitió obtener una alineación óptima de las secuencias, teniendo en cuenta tanto coincidencias como penalizaciones por inserciones y eliminaciones.

Algoritmos Ejercicio 1

| | H | G | S | A | Q | V | K | G | H | G | |
|---|------|------|------|------|-----|-----|-----|-----|-----|-----|-----|
| Q | 0 | -8 | -16 | -24 | -32 | -40 | -48 | -56 | -64 | -72 | -80 |
| K | -8 | -1 | -9 | -16 | -24 | -31 | -39 | -42 | -50 | -58 | -66 |
| T | -16 | -9 | -3 | -7 | -15 | -23 | -30 | -38 | -44 | -52 | -60 |
| E | -24 | -16 | -11 | -3 | -8 | -13 | -21 | -29 | -37 | -44 | -52 |
| A | -32 | -24 | -15 | -10 | -2 | -6 | -13 | -21 | -28 | -36 | -43 |
| E | -40 | -32 | -23 | -15 | -6 | 4 | -4 | -12 | -20 | -28 | -36 |
| M | -48 | -39 | -31 | -23 | -14 | -4 | 5 | -3 | -11 | -19 | -27 |
| K | -56 | -47 | -39 | -31 | -22 | -12 | -3 | 11 | 3 | -5 | -13 |
| A | -64 | -55 | -46 | -38 | -26 | -20 | -11 | 3 | 12 | 4 | -4 |
| S | -72 | -63 | -54 | -41 | -34 | -25 | -19 | -5 | 4 | 11 | 4 |
| E | -80 | -71 | -62 | -49 | -42 | -32 | -27 | -13 | -4 | 4 | 8 |
| D | -88 | -79 | -70 | -57 | -50 | -40 | -35 | -21 | -12 | -4 | 2 |
| L | -96 | -87 | -78 | -65 | -58 | -48 | -38 | -29 | -20 | -12 | -6 |
| K | -104 | -95 | -86 | -73 | -66 | -56 | -46 | -32 | -28 | -20 | -14 |
| K | -112 | -103 | -94 | -81 | -74 | -64 | -54 | -40 | -34 | -28 | -22 |
| H | -120 | -111 | -102 | -89 | -82 | -72 | -62 | -48 | -42 | -21 | -29 |
| G | -128 | -119 | -110 | -97 | -88 | -80 | -70 | -56 | -40 | -29 | -13 |
| T | -136 | -127 | -119 | -106 | -96 | -88 | -78 | -64 | -48 | -37 | -21 |

Obtenemos dos alineamientos con puntuación -21

* Alineamiento 1

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| - | - | H | G | S | - | - | A | - | Q | - | V | K | G | H | G | - |
| K | T | E | A | E | M | K | A | S | E | D | L | K | K | H | G | T |

* Alineamiento 2

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| - | - | H | G | - | - | S | A | - | Q | - | V | K | G | H | G | - |
| K | T | E | A | E | M | K | A | S | E | D | L | K | K | H | G | T |

Figura 1: Solución manual del ejercicio 1. En la parte superior se muestra la matriz obtenida tras la aplicación del algoritmo de Needleman-Wunsch. En la parte inferior se muestran los dos alineamientos óptimos obtenidos.

Ejercicio 2 (6 puntos)

Dado el conjunto de secuencias múltiples:

- *S1: PPGVKSDCAS*
- *S2: PADGVKDCAS*
- *S3: PPDGKSDS*
- *S4: GADGKDCCS*
- *S5: GADGKDCAS*

Utilice el popular método de alineación progresiva para alinear globalmente el conjunto anterior de secuencias. Genere el árbol guía por unión de vecinos. Compare su resultado (alineamiento) con el de Clustal-Omega.

Con el alineamiento final representa el logo. Para este proposito los caracteres nulos o gap son ignorados y no cuentan para el número de observaciones de una columna.

Para la resolución de este ejercicio nos piden llevar a cabo un alineación global de secuencias mediante alineación progresiva, donde el conjunto de secuencias con las que se trabaja es:

| Secuencia | Cadena |
|-----------|------------|
| S1 | PPGVKSDCAS |
| S2 | PADGVKDCAS |
| S3 | PPDGKSDS |
| S4 | GADGKDCCS |
| S5 | GADGKDCAS |

Metodología

Para este ejercicio se utilizó el método de alineación progresiva para realizar el alineamiento global de las secuencias. Se llevaron a cabo los siguientes pasos:

- **Alineamientos por pares:** Se utilizaron las herramientas EMBOSS Needle y Pairwise Sequence Alignment (PSA) para generar alineamientos dos a dos entre las secuencias.
- **Construcción de la matriz de distancias:** A partir de los alineamientos, se calculó una matriz de distancias basada en los valores de similitud entre las secuencias de forma manual.
- **Generación del árbol guía de forma manual:** Se empleó el método de unión de vecinos (Neighbor Joining) para construir el árbol guía a partir de la matriz de distancias.

- **Alineamiento progresivo manual:** Utilizando el árbol guía, se realizó el alineamiento progresivo de las secuencias.
- **Comparación con Clustal Omega:** Se comparó el alineamiento obtenido con el generado por Clustal Omega (1.2.4).
- **Generación del logo de secuencias:** Finalmente, se representó el alineamiento en formato de logo ignorando los caracteres nulos o gaps mediante el uso de las siguientes librerías en Python para el análisis y visualización de secuencias biológicas: **Biopython** (módulo AlignIO) para la manipulación de alineamientos de secuencias, **Logomaker** para la generación de logotipos de secuencia, **Matplotlib** para la visualización gráfica y **NumPy** para el procesamiento numérico.

1. Alineamientos por pares

Para los alineamientos por pares se utilizaron los siguientes parámetros:

- **Matriz de sustitución:** BLOSUM40
- **Penalización por gap:** -10 (aproximado a -8, usado en el ejercicio anterior)

Estos también se muestran en la figura Figura 2.

| | | |
|---------------------------|------------------------|------------------------|
| OUTPUT FORMAT ⓘ pair ▼ | MATRIX ⓘ BLOSUM40 ▼ | GAP OPEN ⓘ 10 ▼ |
| GAP EXTEND ⓘ 10.0 ▼ | END GAP ⓘ false ▼ | END GAP OPEN ⓘ 10 ▼ |
| END GAP EXTEND ⓘ 0.5 ▼ | | |

Figura 2: Parámetros para el alineamiento dos a dos.

Se calcularon los alineamientos y sus respectivos scores, definidos como la relación entre el número de coincidencias y la cantidad total de residuos (excluyendo gaps). Los resultados obtenidos fueron los siguientes:

| Secuencias comparadas | Longitud 1 | Longitud 2 | Coincidencias | Score |
|-----------------------|------------|------------|---------------|-------|
| S1 vs S2 | 10 | 10 | 8 | 8/10 |
| S1 vs S3 | 10 | 8 | 5 | 5/9 |
| S1 vs S4 | 10 | 9 | 4 | 8/19 |

| Secuencias comparadas | Longitud 1 | Longitud 2 | Coincidencias | Score |
|-----------------------|------------|------------|---------------|-------|
| S1 vs S5 | 10 | 9 | 5 | 10/19 |
| S2 vs S3 | 10 | 8 | 4 | 4/9 |
| S2 vs S4 | 10 | 9 | 7 | 14/19 |
| S2 vs S5 | 10 | 9 | 8 | 16/19 |
| S3 vs S4 | 8 | 9 | 4 | 8/17 |
| S3 vs S5 | 8 | 9 | 3 | 6/17 |
| S4 vs S5 | 9 | 9 | 8 | 8/9 |

Los resultados de los alineamientos obtenidos los podemos observar a continuación:

```

S1          1 P-PGVKSDCAS      10
              | .||| |||
S2          1 PADGVK-DCAS      10

```

Figura 3: Alineamiento dos a dos de S1 vs S2, con 10 aa cada secuencia y 8 coincidencias, tiene un score 8/10.

```

S1          1 PPGVKSDCAS      10
              ||..|||.
S3          1 PPDGKSDS--      8

```

Figura 4: Alineamiento dos a dos de S1 vs S3, con 10 aa en S1 y 8 aa en S3; y 5 coincidencias, tiene un score 5/9.

```

S1          1 PPGVK-SDCAS      10
              |.. .||.
S4          1 --GADGKDCCS      9

```

Figura 5: Alineamiento dos a dos de S1 vs S4, con 10 aa en S1 y 9 aa en S4; y 4 coincidencias, tiene un score 8/19.

| | | |
|----|---------------|----|
| S1 | 1 PPGVK-SDCAS | 10 |
| | .. . | |
| S5 | 1 --GADGKDCAS | 9 |

Figura 6: Alineamiento dos a dos de S1 vs S5, con 10 aa en S1 y 9 aa en S5; y 5 coincidencias, tiene un score 10/19.

| | | |
|----|--------------|----|
| S2 | 1 PADGVKDCAS | 10 |
| | | |
| S3 | 1 PPDGKSDS-- | 8 |

Figura 7: Alineamiento dos a dos de S2 vs S3, con 10 aa en S2 y 8 aa en S3; y 4 coincidencias, tiene un score 4/9.

| | | |
|----|--------------|----|
| S2 | 1 PADGVKDCAS | 10 |
| | . . | |
| S4 | 1 GADG-KDCCS | 9 |

Figura 8: Alineamiento dos a dos de S2 vs S4, con 10 aa en S2 y 9 aa en S4; y 7 coincidencias, tiene un score 14/19.

| | | |
|----|--------------|----|
| S2 | 1 PADGVKDCAS | 10 |
| | . | |
| S5 | 1 GADG-KDCAS | 9 |

Figura 9: Alineamiento dos a dos de S2 vs S5, con 10 aa en S2 y 9 aa en S5; y 8 coincidencias, tiene un score 16/19.

| | | |
|----|--------------|---|
| S3 | 1 PPDGKSDS-- | 8 |
| | .. . | |
| S4 | 1 GADGK-DCCS | 9 |

Figura 10: Alineamiento dos a dos de S3 vs S4, con 8 aa en S3 y 9 aa en S4; y 4 coincidencias, tiene un score 8/17.

| | | |
|----|-------------|---|
| S3 | 1 PPDGKSDS- | 8 |
| |: | |
| S5 | 1 GADGKDCAS | 9 |

Figura 11: Alineamiento dos a dos de S3 vs S5, con 8 aa en S3 y 9 aa en S5; y 3 coincidencias, tiene un score 6/17.

| | | |
|----|-------------|---|
| S4 | 1 GADGKDCCS | 9 |
| | .I | |
| S5 | 1 GADGKDCAS | 9 |

Figura 12: Alineamiento dos a dos de S4 vs S5, con 9 aa cada secuencia y 8 coincidencias, tiene un score 8/9.

2. Construcción de la matriz de distancias

La matriz de distancias se construyó a partir del score de similitud, aplicando la transformación:

$$d = 1 - \text{score}$$

Obteniendo la siguiente matriz de distancias que se puede observar en la parte superior de Figura 13:

| | S1 | S2 | S3 | S4 | S5 |
|----|------|------|------|------|------|
| S1 | 0.00 | 0.20 | 0.44 | 0.58 | 0.47 |
| S2 | 0.20 | 0.00 | 0.56 | 0.26 | 0.16 |
| S3 | 0.44 | 0.56 | 0.00 | 0.53 | 0.65 |
| S4 | 0.58 | 0.26 | 0.53 | 0.00 | 0.11 |
| S5 | 0.47 | 0.16 | 0.65 | 0.11 | 0.00 |

3. Construcción del árbol guía de forma manual

A partir de la matriz de distancias, se generó el árbol guía utilizando el método de unión de vecinos (*Neighbor Joining*). El árbol guía obtenido muestra la relación evolutiva entre las secuencias y nos sirve como base para la alineación progresiva. El proceso de generación del árbol guía se hizo de forma manual y se muestra en Figura 13.

El árbol guía obtenido finalmente fue el mostrado en Figura 14.

Ejercicio 2

• Matriz de distancias tras los alineamientos Pánuire

| | | | | | |
|----|------|------|------|------|----|
| S1 | 0 | | | | |
| S2 | 0,2 | 0 | | | |
| S3 | 0,44 | 0,56 | 0 | | |
| S4 | 0,58 | 0,26 | 0,53 | 0 | |
| S5 | 0,47 | 0,16 | 0,65 | 0,11 | 0 |
| | S1 | S2 | S3 | S4 | S5 |

$$S2-S5 = 1 - \frac{16}{19} = 0,16$$

$$S3-S5 = 1 - \frac{6}{17} = 0,65$$

$\left[\begin{matrix} S4 \\ S5 \end{matrix} \right] A$

$$A-S1 = \frac{0,58 + 0,47}{2} = 0,525$$

$$A-S2 = \frac{0,26 + 0,16}{2} = 0,21$$

$$A-S3 = \frac{0,53 + 0,65}{2} = 0,59$$

| | | | |
|----|--------|------|---|
| B | 0 | | |
| S3 | 0,5 | 0 | |
| A | 0,3675 | 0,59 | 0 |
| | B | S3 | A |

$\left[\begin{matrix} A \\ B \end{matrix} \right] S3$

$$S1-S2 = 1 - \frac{8}{10} = 0,2$$

$$S1-S3 = 1 - \frac{5}{9} = 0,44$$

$$S1-S4 = 1 - \frac{8}{14} = 0,58$$

$$S1-S5 = 1 - \frac{10}{19} = 0,47$$

$$S2-S3 = 1 - \frac{4}{9} = 0,56$$

$$S2-S4 = 1 - \frac{14}{19} = 0,26$$

$$S3-S4 = 1 - \frac{8}{17} = 0,53$$

$$S4-S5 = 1 - \frac{8}{9} = 0,11$$

| | | | | |
|----|-------|------|------|---|
| S1 | 0 | | | |
| S2 | 0,2 | 0 | | |
| S3 | 0,44 | 0,56 | 0 | |
| A | 0,525 | 0,21 | 0,59 | 0 |
| | S1 | S2 | S3 | A |

$\left[\begin{matrix} S1 \\ S2 \end{matrix} \right] B$

$$B-S3 = \frac{0,44 + 0,56}{2} = 0,5$$

$$B-A = \frac{0,525 + 0,21}{2} = 0,3675$$

Figura 13: Construcción de matrices de distancias.

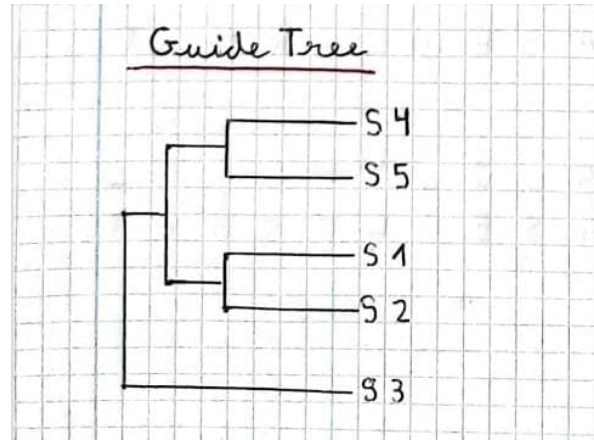


Figura 14: Guide Tree obtenido a partir de alineamientos 2 a 2 con parámetros similares al ejercicio 1.

4. Alineamiento progresivo manual

Utilizando el árbol guía, se llevó a cabo el alineamiento progresivo de las secuencias. Este alineamiento se obtuvo manualmente y posteriormente se comparará con el resultado de Clustal Omega.

El alineamiento final obtenido se muestra en Figura 15.

Alineamiento progresivo

| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|
| A | G | A | D | G | K | D | C | C | S | | G | A | D | G | - | K | - | D | C | C | S | | |
| | G | A | D | G | K | D | C | A | S | | G | A | D | G | - | K | - | D | C | A | S | | |
| + | | | | | | | | | | ⇒ | | | | | | | | | | | | | |
| B | P | - | P | G | V | K | S | D | C | A | S | | P | A | D | G | V | K | - | D | C | A | S |
| | P | A | D | G | V | K | - | D | C | A | S | | P | - | P | G | V | K | S | D | C | A | S |
| | | | | | | | | | | | | + | | | | | | | | | | | |
| | | | | | | | | | | | | S3 | P | P | D | G | - | K | S | D | - | S | |

Alineamiento final

| | | | | | | | | | | | |
|-----|---|---|---|---|---|---|---|---|---|---|---|
| S4: | G | A | D | G | - | K | - | D | C | C | S |
| S5: | G | A | D | G | - | K | - | D | C | A | S |
| S2: | P | A | D | G | V | K | - | D | C | A | S |
| S1: | P | - | P | G | V | K | S | D | C | A | S |
| S3: | P | P | D | G | - | K | S | D | - | - | S |

Figura 15: Alineamiento progresivo obtenido a partir de alineamientos 2 a 2 con parámetros similares al ejercicio 1.

Cálculo con parámetros por defecto

Además del cálculo utilizando BLOSUM40, se decidió repetir el proceso usando la matriz BLOSUM62 y los parámetros por defecto, para evaluar si el alineamiento final variaba en función de los parámetros. Los parámetros utilizados fueron los que se muestran en Figura 16.

| | | |
|------------------|------------|----------------|
| OUTPUT FORMAT ⓘ | MATRIX ⓘ | GAP OPEN ⓘ |
| pair ▼ | BLOSUM62 ▼ | 10 ▼ |
| GAP EXTEND ⓘ | END GAP ⓘ | END GAP OPEN ⓘ |
| 0.5 ▼ | false ▼ | 10 ▼ |
| END GAP EXTEND ⓘ | | |
| 0.5 ▼ | | |

Figura 16: Parámetros para el alineamiento dos a dos por defecto.

Se realizaron los alineamientos por pares con estos parámetros y se generó una nueva matriz de distancias, un nuevo árbol guía y el alineamiento progresivo correspondiente.

Los scores obtenidos con estos parámetros fueron se muestran en la siguiente tabla:

| Secuencias comparadas | Longitud 1 | Longitud 2 | Coincidencias | Score |
|-----------------------|------------|------------|---------------|-------|
| S1 vs S2 | 10 | 10 | 7 | 7/10 |
| S1 vs S3 | 10 | 8 | 5 | 5/9 |
| S1 vs S4 | 10 | 9 | 4 | 8/19 |
| S1 vs S5 | 10 | 9 | 5 | 10/19 |
| S2 vs S3 | 10 | 8 | 4 | 4/9 |
| S2 vs S4 | 10 | 9 | 7 | 14/19 |
| S2 vs S5 | 10 | 9 | 8 | 16/19 |
| S3 vs S4 | 8 | 9 | 3 | 6/17 |
| S3 vs S5 | 8 | 9 | 3 | 6/17 |
| S4 vs S5 | 9 | 9 | 8 | 8/9 |

Y los alineamientos obtenidos se muestran a continuación:

| | | |
|----|---------------|----|
| S1 | 1 -PPGVKSDCAS | 10 |
| | .. | |
| S2 | 1 PADGVK-DCAS | 10 |

Figura 17: Alineamiento dos a dos con parámetros por defecto, de S1 vs S2, con 10 aa cada secuencia y 7 coincidencias, tiene un score 7/10.

| | | |
|----|--------------|----|
| S1 | 1 PPGVKSDCAS | 10 |
| | .. . | |
| S3 | 1 PPDGKSDS-- | 8 |

Figura 18: Alineamiento dos a dos con parámetros por defecto, de S1 vs S3, con 10 aa en S1 y 8 aa en S3; y 5 coincidencias, tiene un score 5/9.

| | | |
|----|---------------|----|
| S1 | 1 PPGVK-SDCAS | 10 |
| | .. . | |
| S4 | 1 --GADGKDCCS | 9 |

Figura 19: Alineamiento dos a dos con parámetros por defecto, de S1 vs S4, con 10 aa en S1 y 9 aa en S4; y 4 coincidencias, tiene un score 8/19.

| | | |
|----|---------------|----|
| S1 | 1 PPGVK-SDCAS | 10 |
| | .. . | |
| S5 | 1 --GADGKDCAS | 9 |

Figura 20: Alineamiento dos a dos con parámetros por defecto, de S1 vs S5, con 10 aa en S1 y 9 aa en S5; y 5 coincidencias, tiene un score 10/19.

| | | |
|----|--------------|----|
| S2 | 1 PADGVKDCAS | 10 |
| | . . . | |
| S3 | 1 PPDGKSDS-- | 8 |

Figura 21: Alineamiento dos a dos con parámetros por defecto, de S2 vs S3, con 10 aa en S2 y 8 aa en S3; y 4 coincidencias, tiene un score 4/9.

| | | |
|----|--------------|----|
| S2 | 1 PADGVKDCAS | 10 |
| | . . | |
| S4 | 1 GADG-KDCCS | 9 |

Figura 22: Alineamiento dos a dos con parámetros por defecto, de S2 vs S4, con 10 aa en S2 y 9 aa en S4; y 7 coincidencias, tiene un score 14/19.

| | | |
|----|--------------|----|
| S2 | 1 PADGVKDCAS | 10 |
| | . | |
| S5 | 1 GADG-KDCAS | 9 |

Figura 23: Alineamiento dos a dos con parámetros por defecto, de S2 vs S5, con 10 aa en S2 y 9 aa en S5; y 8 coincidencias, tiene un score 16/19.

| | | |
|----|-------------|---|
| S3 | 1 PPDGKSDS- | 8 |
| | | |
| S4 | 1 GADGKDCCS | 9 |

Figura 24: Alineamiento dos a dos con parámetros por defecto, de S3 vs S4, con 8 aa en S3 y 9 aa en S4; y 3 coincidencias, tiene un score 6/17.

| | | |
|----|-------------|---|
| S3 | 1 PPDGKSDS- | 8 |
| |: | |
| S5 | 1 GADGKDCAS | 9 |

Figura 25: Alineamiento dos a dos con parámetros por defecto, de S3 vs S5, con 8 aa en S3 y 9 aa en S5; y 3 coincidencias, tiene un score 6/17.

| | | |
|----|-------------|---|
| S4 | 1 GADGKDCCS | 9 |
| | . | |
| S5 | 1 GADGKDCAS | 9 |

Figura 26: Alineamiento dos a dos con parámetros por defecto, de S4 vs S5, con 9 aa cada secuencia y 8 coincidencias, tiene un score 8/9.

Con ello, contruimos la matriz de distancias:

| | S1 | S2 | S3 | S4 | S5 |
|----|------|------|------|------|------|
| S1 | 0.00 | 0.30 | 0.44 | 0.58 | 0.47 |
| S2 | 0.30 | 0.00 | 0.56 | 0.26 | 0.16 |
| S3 | 0.44 | 0.56 | 0.00 | 0.53 | 0.65 |
| S4 | 0.58 | 0.26 | 0.53 | 0.00 | 0.11 |
| S5 | 0.47 | 0.16 | 0.65 | 0.11 | 0.00 |

Y se llevó a cabo el alineamiento progresivo de forma análoga al anterior (Figura 27)

Finalmente, construimos el guide tree:

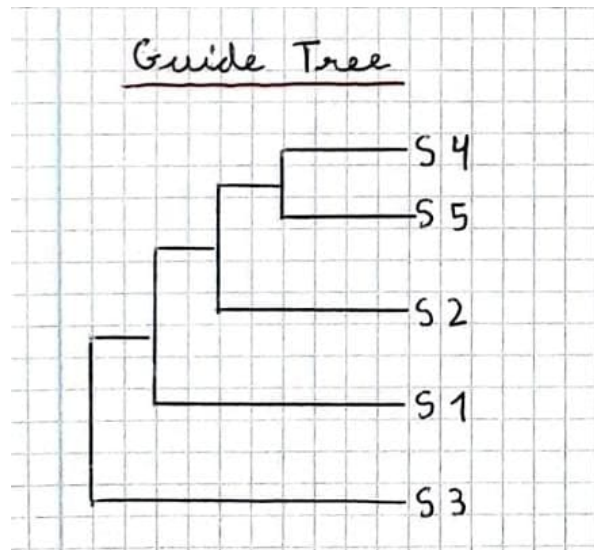


Figura 28: Guide Tree obtenido a partir de alineamientos 2 a 2

Y el alineamiento progresivo se muestra en la figura Figura 29

Alineamiento múltiple

Para llevar a cabo el alineamiento múltiple de secuencias, se utilizaron los siguientes parámetros:

• Matriz de distancias tras los alineamientos Pairwise (BLOSUM 62):

| | | | | | |
|----|------|------|------|------|----|
| S1 | 0 | | | | |
| S2 | 0,3 | 0 | | | |
| S3 | 0,44 | 0,56 | 0 | | |
| S4 | 0,58 | 0,26 | 0,65 | 0 | |
| S5 | 0,47 | 0,16 | 0,65 | 0,11 | 0 |
| | S1 | S2 | S3 | S4 | S5 |

$$S2-S5 = 1 - \frac{16}{19} = 0,16$$

$$S3-S5 = 1 - \frac{6}{17} = 0,65$$

$$\left\{ \begin{matrix} S4 \\ S5 \end{matrix} \right\} A \rightarrow$$

$$A-S1 = \frac{0,58+0,43}{2} = 0,525$$

$$A-S2 = \frac{0,26+0,16}{2} = 0,21$$

$$A-S3 = \frac{0,65+0,65}{2} = 0,65$$

| | | | | | |
|----|-------|------|------|---|--|
| S1 | 0 | | | | |
| S2 | 0,3 | 0 | | | |
| S3 | 0,44 | 0,56 | 0 | | |
| A | 0,525 | 0,21 | 0,65 | 0 | |
| | S1 | S2 | S3 | A | |

| | | | | |
|----|--------|-------|---|--|
| S1 | 0 | | | |
| S3 | 0,44 | 0 | | |
| B | 0,4125 | 0,605 | 0 | |
| | S1 | S3 | B | |

$$\left\{ \begin{matrix} A \\ S2 \end{matrix} \right\} B \leftarrow$$

$$B-S1 = \frac{0,3+0,525}{2} = 0,4125$$

$$B-S3 = \frac{0,56+0,65}{2} = 0,605$$

$$\left\{ \begin{matrix} B \\ S1 \\ S3 \end{matrix} \right\}$$

Figura 27: Construcción de matrices de distancias.

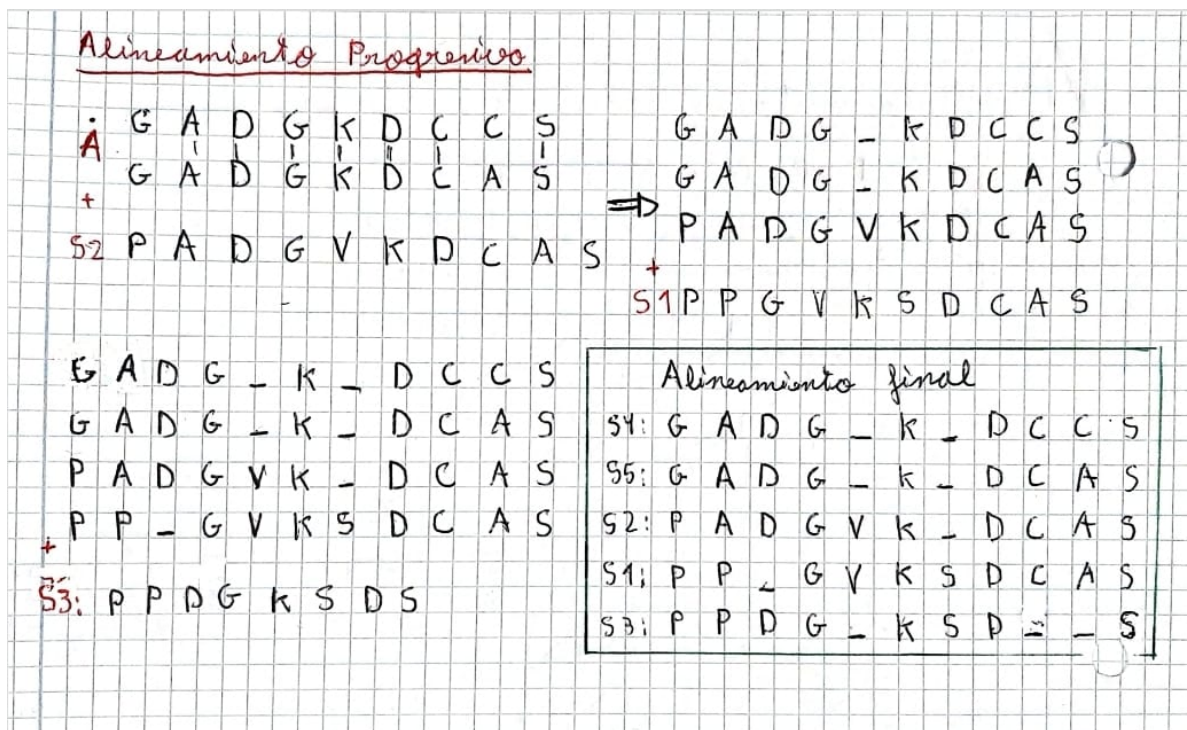


Figura 29: Alineamiento progresivo obtenido a partir de alineamientos 2 a 2 con matriz BLO-SUM62

| | | | |
|-----------------------------------|----------------------------------|-----------------------|-------------------|
| OUTPUT FORMAT ⓘ | | DEALIGN INPUT ⓘ | |
| ClustalW with character counts ▼ | | no ▼ | |
| MBED-LIKE CLUSTERING GUIDE-TREE ⓘ | MBED-LIKE CLUSTERING ITERATION ⓘ | COMBINED ITERATIONS ⓘ | |
| no ▼ | yes ▼ | default(0) ▼ | |
| MAX GUIDE TREE ⓘ | MAX HMM ITERATIONS ⓘ | ORDER ⓘ | DISTANCE MATRIX ⓘ |
| default ▼ | default ▼ | aligned ▼ | yes ▼ |
| OUTPUT GUIDE TREE ⓘ | | | |
| yes ▼ | | | |

Figura 30: Parámetros para el alineamiento múltiple.

Estos parámetros fueron configurados en Clustal Omega, una herramienta bioinformática para la alineación de secuencias.

Como resultado del proceso, se obtuvo el siguiente árbol guía, que representa la relación filogenética entre las secuencias alineadas:

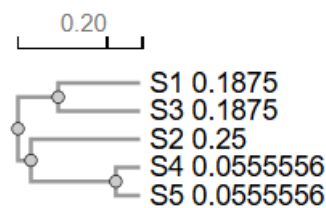


Figura 31: Guide tree del alineamiento múltiple por Clustal Omega.

Finalmente, se generó el alineamiento múltiple de las secuencias, donde se pueden observar las similitudes y diferencias entre ellas:

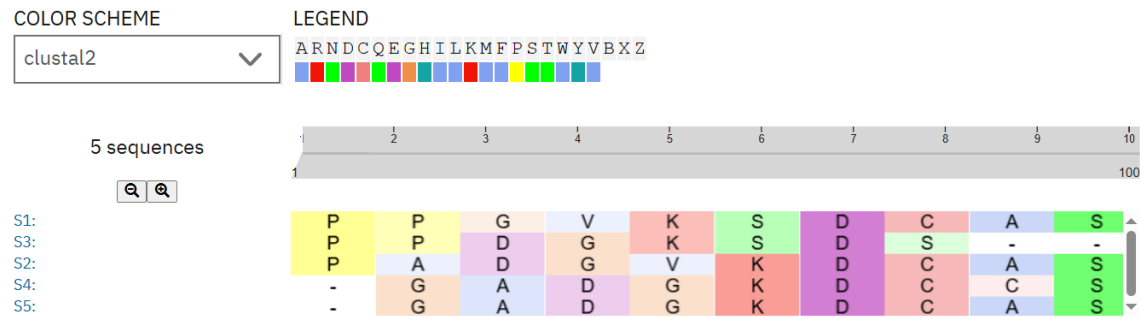


Figura 32: Alineamiento múltiple por Clustal Omega.

Este alineamiento nos permite analizar la conservación de regiones específicas, identificar mutaciones y establecer relaciones evolutivas entre las secuencias comparadas.

Generación del logo de secuencias

Observamos que los guide tree obtenidos por ambos alineamientos dos a dos, difieren entre sí y con el guide tree obtenido por el Clustal Omega; siendo común entre ellos únicamente el grupo S4-S5.

Respecto a los alineamientos progresivos obtenidos en los alineamientos 2 a 2 con la BLOSUM40 y BLOSUM62, son muy semejantes, variando únicamente por la posición de la segunda prolina (P) de la S1, debido a la utilización de distintas matrices de sustitución y los diferentes valores utilizados para penalizar los gaps.

Ambas varían mucho del alineamiento múltiple obtenido mediante el Clustal Omega, debido a que este método es heurístico y no garantiza encontrar el alineamiento globalmente óptimo, teniendo dificultad frente a deleciones e inserciones, en las que incluye gaps; siendo crucial la importancia de los parámetros introducidos para que esta se ajuste a la realidad. En el peor de los casos se utilizan valores predeterminados (como ocurre con Clustal Omega (1.2.4)), que pueden no ser ideales para todo el conjunto de datos.

Logo del alineamiento progresivo del alineamiento por pares mediante la BLOSUM40

Cargamos las librerías necesarias, cargamos los documentos FASTA de los alineamientos progresivos y obtuvimos el logo de ambos alineamientos dos a dos y del Clustal Omega.

```
#Importación de librerías
from Bio import AlignIO
import logomaker
import matplotlib.pyplot as plt
import numpy as np
```

```

# Cargar y procesar el alineamiento
alignment = AlignIO.read("Alineamiento_final_BLOSUM40.fa", "fasta")
sequences = [str(record.seq) for record in alignment]

# Crear matriz de conteos (sin gaps)
counts_matrix = logomaker.alignment_to_matrix(sequences, to_type='counts', \
    characters_to_ignore='-')

# Convertir a frecuencias relativas (ignorando gaps)
frequency_matrix = counts_matrix.div(counts_matrix.sum(axis=1), axis=0)

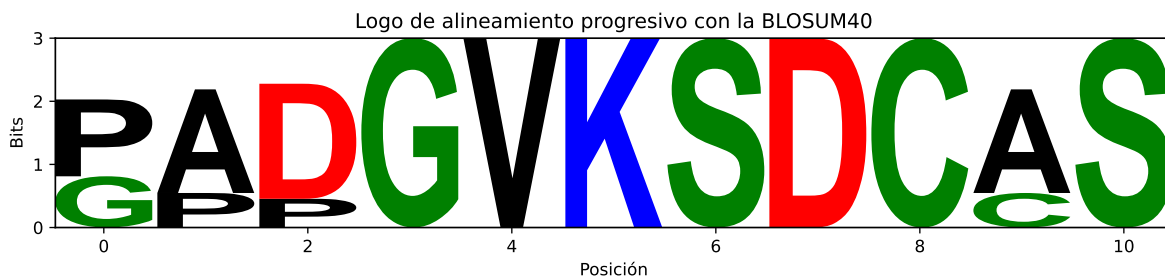
# Calcular información (en bits) para cada posición
information_matrix = logomaker.transform_matrix(frequency_matrix, \
    from_type='probability', to_type='information')

# Crear el logo
logo = logomaker.Logo(
    information_matrix, # Usamos la matriz de información directamente
    color_scheme='chemistry',
    font_name='Arial',
    show_spines=True,
    stack_order='big_on_top'
)

# Personalización adicional
logo.ax.set_title("Logo de alineamiento progresivo con la BLOSUM40")
logo.ax.set_xlabel("Posición")
logo.ax.set_ylabel("Bits")
logo.ax.set_ylim(0, 3)

plt.tight_layout()
plt.show()

```



Podemos observar la representación gráfica del logo de alineamiento progresivo basado en la matriz de sustitución BLOSUM40. Este tipo de visualización permite evaluar la conservación y variabilidad de secuencias de aminoácidos en alineaciones múltiples, proporcionando información relevante sobre la evolución y estructura de las proteínas.

En el eje vertical aparecen representados los bits que representan la cantidad de información en cada posición de la secuencia. Valores altos indican una mayor conservación de los aminoácidos en esa posición. En el eje horizontal tenemos las posiciones de la secuencia donde podemos observar la ubicación específica de los aminoácidos en la alineación. En el logo en sí aparecen representadas las letras correspondientes a un aminoácido concreto, y su tamaño refleja la frecuencia con la que aparece en una posición determinada.

explicar colores!!

Logo del alineamiento progresivo del alineamiento por pares mediante la BLOSUM62

```
# Cargar y procesar el alineamiento
alignment = AlignIO.read("Alineamiento_final_BLOSUM62.fa", "fasta")
sequences = [str(record.seq) for record in alignment]

# Crear matriz de conteos (sin gaps)
counts_matrix = logomaker.alignment_to_matrix(sequences, to_type='counts', \
    characters_to_ignore='-')

# Convertir a frecuencias relativas (ignorando gaps)
frequency_matrix = counts_matrix.div(counts_matrix.sum(axis=1), axis=0)

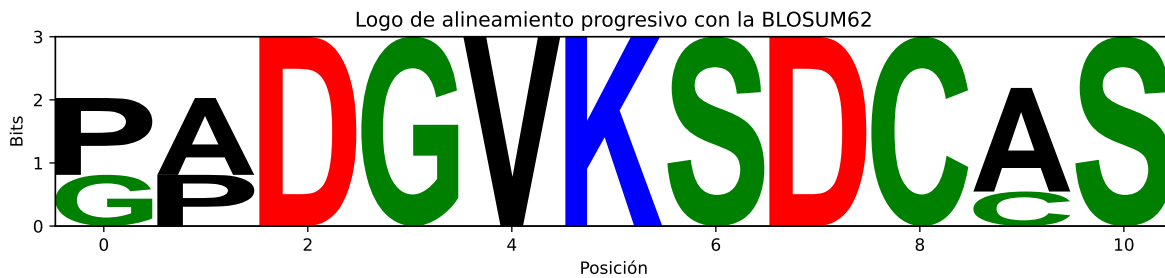
# Calcular información (en bits) para cada posición
information_matrix = logomaker.transform_matrix(frequency_matrix, \
    from_type='probability', to_type='information')

# Crear el logo
logo = logomaker.Logo(
    information_matrix, # Usamos la matriz de información directamente
    color_scheme='chemistry',
    font_name='Arial',
    show_spines=True,
    stack_order='big_on_top'
)

# Personalización adicional
logo.ax.set_title("Logo de alineamiento progresivo con la BLOSUM62")
logo.ax.set_xlabel("Posición")
```

```
logo.ax.set_ylabel("Bits")
logo.ax.set_ylim(0, 3)

plt.tight_layout()
plt.show()
```



Logo del alineamiento múltiples (MSA) por Clustal Omega

```
# Cargar y procesar el alineamiento
alignment = AlignIO.read("clustalo-I20250321-165928-0743-92365021-p1m.fa", \
    "fasta")
sequences = [str(record.seq) for record in alignment]

# Crear matriz de conteos (sin gaps)
counts_matrix = logomaker.alignment_to_matrix(sequences, to_type='counts', \
    characters_to_ignore='-')

# Convertir a frecuencias relativas (ignorando gaps)
frequency_matrix = counts_matrix.div(counts_matrix.sum(axis=1), axis=0)

# Calcular información (en bits) para cada posición
information_matrix = logomaker.transform_matrix(frequency_matrix, \
    from_type='probability', to_type='information')

# Crear el logo
logo = logomaker.Logo(
    information_matrix, # Usamos la matriz de información directamente
    color_scheme='chemistry',
    font_name='Arial',
    show_spines=True,
    stack_order='big_on_top'
)
```

```
# Personalización adicional
logo.ax.set_title("Logo de alineamiento Clustal Omega")
logo.ax.set_xlabel("Posición")
logo.ax.set_ylabel("Bits")
logo.ax.set_ylim(0, 3)

plt.tight_layout()
plt.show()
```



De los tres logos obtenidos, podemos destacar que aquel con más variedad en las posiciones es el obtenido por el alineamiento de Clustal Omega y el que posee más similitudes es el resultante del alineamiento progresivo dos a dos mediante la BLOSUM62.