



## ***ANÁLISIS DE BIOLOGÍA COMPUTACIONAL (GPO 850)***

Tarea colaborativa: Genes y genomas

***Profesora:***

*Dr<sup>a</sup> . Raquel Acuña González*

***Integrantes***

~

Fernanda Ponce Maciel	A01799293
Manuel Villalpando Linares	A01352033

**Fecha de entrega:**

23 de enero del 2024

**PARTE 1. Contesten y justifiquen las siguientes preguntas:**

**a) En un organismo con un genoma de ADN doble hélice, si  $G=C$  y  $A=T$ , entonces, ¿ $G+C = A+T$ ? ¿Por qué?**

No,  $G+C$  no es igual a  $A+T$ . La regla fundamental de apareamiento de las bases en el ADN establece que la adenina (A) se empareja con la timina (T) mediante dos enlaces de hidrógeno, y la guanina (G) se empareja con la citosina (C) mediante tres enlaces de hidrógeno. Sin embargo, la suma de  $G+C$  no es igual a la suma de  $A+T$  debido al número diferente de enlaces de hidrógeno entre estas bases.

**b) Si un organismo tiene 38% de A, ¿cuál es el porcentaje de G?**

Si  $A = T$  y  $C = G$ , por lo tanto, al sumar A y T, que son los genomas que van enlazados, esto resultaría 76%, por ende, el restante para llegar a 100% son 24%, el cual proviene de la relación entre C y G, y debido a que estos dos también son iguales y están siendo sumados, al dividir el 24% entre 2, la respuesta sería 12%.

**c) ¿Cuáles son las consecuencias de un alto contenido de GC?**

Podría aumentar la estabilidad de la doble hélice debido a los tres enlaces de hidrógeno entre G y C. Esto puede afectar la temperatura de fusión y la estructura tridimensional del ADN, lo que puede influir en la función génica y en la interacción con proteínas.

**d) ¿A qué refiere el superenrollamiento del ADN?**

Es la torsión adicional que puede ocurrir en la doble hélice de ADN, puede ser positivo o negativo. El superenrollamiento es esencial para la compactación del ADN en la célula y para procesos como la replicación y la transcripción.

**e) ¿Cómo se desnaturaliza el ADN? Es decir, ¿qué métodos se usan para separar las hebras?**

El ADN se desnaturaliza mediante la aplicación de calor o mediante la exposición a condiciones extremas de pH. Para separar las hebras, se puede utilizar la técnica de la reacción en cadena de la polimerasa (PCR), donde se calienta el ADN para desnaturalizarlo y luego se enfría para permitir que se apareen los cebadores.

**f) ¿Qué es la  $T_m$  del ADN de doble cadena?**

La temperatura de fusión ( $T_m$ ) es la temperatura a la cual la mitad de las moléculas de ADN de doble cadena están desnaturalizadas. Es un indicador de la estabilidad de la doble hélice y varía según la composición de bases.

**g) ¿Por qué es importante saber la  $T_m$  de un segmento de ADN de doble cadena? (desde el punto de vista biotecnológico)**

Es crucial para diseñar experimentos tales como la PCR, donde la desnaturalización y la hibridación son pasos esenciales. También es importante en la síntesis de sondas de ADN y en la manipulación de ADN en técnicas como la secuenciación.

**h) ¿Qué factores afectan la  $T_m$ ? Menciona algunos factores que afectan la  $T_m$ .**

Estos pueden ser la longitud de la secuencia, mientras tenga más longitud, también la  $T_m$  aumentará, la composición de bases, la concentración de las sales y la presencia de laguna modificación en las bases.

**i) Ejercicio: Observa la siguiente secuencia y encuentra la secuencia complementaria, la reversa y la reversa complementaria:**

- a. Secuencia original: 5´-ATGCTTGACGCTCAAACCATCGC-3´
- b. Secuencia complementaria: 3´-TACGAACTGCGAGTTTGGTAGCG-5´
- c. Secuencia reversa: 5´-CGCTACCATTGAGCGTCAAGCAT-3´
- d. Secuencia reversa complementaria: 3´GCATTGCGACTCGTCAAAGGTAGCG5´

## **PARTE 2: Análisis de la secuencia del gen hmp de Escherichia coli**

Analicen la siguiente secuencia de una hebra del gene hmp (flavohemoglobina) de Escherichia coli, una bacteria mesófila, no patógena, enterobacteria y respondan: ¿cuál será la secuencia de aminoácidos de la flavohemoglobina?

```
AUGCUTGACGCUCAAACCAUCGCACAGUAAAAGCCACCAUCCCUUUACUGGUGGAAACGGGGCC
AAAGUUAACCGCCCAUUUCUACGACCGUAUGUUUACUCAUAACCCAGAACUCAAGAAAUUUUUA
ACAUGAGUAACCAGCGUAAUGGCGAUCAACGUGAAGCCUGUUUAACGCUAUUGCCGCCUACGC
CAGUAAUAUUGAAAACCGCUCUGCGCUGCUGCAGCGGUAGAAAAAUCGCGCAGAAGCACACC
AGCUUCCAGAUCAAACCGGAACAGUACAACAUCGUCGGUGAACACCUUGUUGGCAACGCUAGACG
AAAUGUUCAGCCCGGGGCGAGGAAGUGCUGGACGCGUGGGGUAAAGCCUAUGGUGUACUGGCUA
AUGUAUUUAUCAAUUCGCGAGGCGGAAAUCAUAAUCGAAAACGCCAGCAAAGCCGGUGGUUGGA
AGGUACUCGCGAUUUCGCAUUGUGGCUAAAACACCGCGCAGCGCGCUUAUCACCAGCUUCGA
```

ACUGGAGCCGGUCGACGGUGGCGCAGUGGCAGAAUACCGUCCGGGGCAAUAUCUCGGCGUCU  
GGCUGAAGCCGGAGGGUUUCCCACAUCAGGAAAUUCGUCAGUACUCUUUGACUCGCAAACCGG  
AUGGCAAAGGCUAUCGUUUUGCGGUGAAACGCGAAGAGGGUGGGCAGGUAUCCAAUGGUGUCA  
CAAUCACGCCAAUGUUGGCGAUGUCGUGAAACUGGUCGCUCCGGCAGGUGAUUUUCUUUAUGGC  
UUGUCGCAGAUGACACACCAGUGACGUUAAUCUCUGCCGGUGUUGGUCAAACGCCAAUGCUGG  
CAGUGCUCGACACGCUGGCAAAGCAGGCCACACAGCACAAAGUGAACUGGUUCCAUGCGGCAGA  
AAUGGCGAUGUUCACGCCUUUGCCGAUGAAGUUAAGGAACUGGGGCAGUCACUGCCGCGCUUU  
ACCGCGCACACCUUGGUAUCGUCAGCCGAGCGAAGCCGAUCGCGCUAAAGGUCAGUUUGAUAG  
CGAAGGUCUGAUGGAUUUGAGCAAUUCGGAAAGGUGCGUUCAGCGAUCCGACAUGCAGUUCUA  
UCUCUGCGGCCCCGGUUGGCUUCAUGCAGUUUACCGCGAAACAGUUAGUGGAUCUGGGCGUGA  
AGCAGGAAAACAUUCAUUACGAUUGCUUUGCCCCGCAUAAGGUGCUGUAA-3 ´

Met Leu Asp Ala Gln Thr Ile Ala Thr Tyr Lys Phe Gly Gly Asn Gly His Pro Phe Thr Gly Trp Gly Asn Gly  
Ala Lys Val Asn Arg Pro Pro Phe Leu Tyr Thr Thr Val Leu Phe Thr His Asn Pro Glu Leu Gln Arg Ile Phe  
Leu Asn Met Ser Lys Pro Ala Asn Gly Asp Gln Thr Ile Arg Val Asn Thr Leu Leu Ala Thr Leu Asp Glu Met  
Phe Ser Pro Gly Ala Gly Ser Ala Leu Asp Ala Trp Gly Leu Lys Pro Tyr Gly Val Leu Ala Asn Val Phe Ile  
Asn Arg Glu Ala Glu Ile Tyr Asn Glu Lys Arg Pro Ala Lys Ala Gly Gly Trp Glu Gly Tyr Ser Asp Phe Pro  
His Cys Gly Leu Lys His Arg Arg Ala Ala Leu Ile Thr Ser Phe Glu Leu Glu Pro Val Asp Gly Gly Ala Ser  
Gly Leu Asp Gly Ala Ser Trp Ala Ala Asn Thr Arg Pro Gly Gln Tyr Leu Gly Ser Trp Leu Lys Pro Glu Arg  
Phe Pro His Ile Arg Lys Phe Val Ser Thr Leu Leu Asp Ser Gln Thr Gly Trp Gln Arg Leu Ser Tyr His Pro  
Met Leu Ala Met Leu Arg Thr Ala Gly Lys Ser Gly His Thr Ala Gln Val Asn Trp Phe His Ala Ala Glu Lys  
Met Ala Met Phe Thr Pro Phe Ala Asp Glu Val Lys Glu Leu Gly Gln Ser Thr Ala Ala Phe Thr Ala His Thr  
Trp Tyr Arg Gln Pro Ser Glu Glu Ala Asp Arg Ala Lys Val Gln Phe Asp Ser Glu Gly Leu Met Asp Phe Glu  
Gln Thr Gly Arg Trp Ala Cys Lys Lys Gln Glu Asn Ile His Tyr Asn Ala Phe Gly Pro His Lys Val Leu

### PARTE 3. Investigación y Respuestas

#### a) ¿Qué significa el formato FASTA de una secuencia? Describe sus características más relevantes que lo separan de otros formatos.

El formato FASTA es un formato basado en texto para representar secuencias de nucleótidos o aminoácidos mediante códigos que emplean letras. El formato permite describir nombres de secuencias y comentarios de dichas secuencias. Sus características incluyen una línea de encabezado que comienza con el símbolo ">" seguido de información descriptiva, seguida de la secuencia en sí. Es un formato simple y ampliamente utilizado en bioinformática.

#### b) ¿Qué bancos de datos existen para obtener secuencias de ADN de genomas de organismos?

GenBank, el European Nucleotide Archive (ENA), y el DNA Data Bank of Japan (DDBJ). Estos almacenan y proporcionan acceso a secuencias de genomas de diferentes organismos.

**c) Busca la secuencia de ADN de la Alcohol Deshidrogenasa de *Escherichia coli* (gen *adhE*) y obtengan la secuencia de nucleótidos.**

>NC\_000913.3:c1298121-1295446 *Escherichia coli* str. K-12 substr. MG1655, complete genome

```
ATGGCTGTTACTAATGTCGCTGAACCTAACGCACTCGTAGAGCGTGTAAGGAGCCAGCGTGAATATG
CCAGTTTCACTCAAGAGCAAGTAGACAAAATCTTCCGCGCCGCCGCTCTGGCTGCTGCAGATGCTCGAAT
CCCCTCGCGAAAATGGCCGTTGCCGAATCCGGCATGGGTATCGTCGAAGATAAAGTGATCAAAAACAC
TTTGCTTCTGAATATATCTACAACGCCTATAAAGATGAAAAACCTGTGGTGTCTGTCTGAAGACGACA
CTTTTGGTACCATCACTATCGCTGAACCAATCGGTATTATTTGCGGTATCGTTCCGACCACTAACCCGAC
TTCAACTGCTATCTTCAAATCGCTGATCAGTCTGAAGACCCGTAACGCCATTATCTTCTCCCCGACCCG
CGTGCAAAAGATGCCACCAACAAAGCGGCTGATATCGTTCTGCAGGCTGCTATCGCTGCCGGTGCTCCGA
AAGATCTGATCGGCTGGATCGATCAACCTTCTGTTGAACTGTCTAACGCACTGATGCACCACCCAGACAT
CAACCTGATCTCGCGACTGGTGGTCCGGGCATGGTTAAAGCCGCATACAGCTCCGGTAAACCAGCTATC
GGTGTAGGCGCGGGCAACACTCCAGTTGTTATCGATGAAACTGCTGATATCAAACGTGCAGTTGCATCTG
TACTGATGTCCAAAACCTTCGACAACGGCGTAATCTGTGCTTCTGAACAGTCTGTTGTTGTTGACTC
TGTTTATGACGCTGTACGTGAACGTTTTGCAACCCACGGCGGCTATCTGTTGCAGGGTAAAGAGCTGAAA
GCTGTTCAAGGATGTTATCCTGAAAAACGGTGCGCTGAACGCGGCTATCGTTGGTCAGCCAGCCTATAAAA
TTGCTGAACTGGCAGGCTTCTCTGTACCAGAAAACACCAAGATTCTGATCGGTGAAGTGACCGTTGTTGA
TGAAAGCGAACCGTTTCGCACATGAAAACTGTCCCGACTCTGGCAATGTACCGCGCTAAAGATTTCGAA
GACGCGGTAGAAAAAGCAGAGAACTGGTTGCTATGGGCGGTATCGGTCATACCTCTTGCCTGTACACTG
ACCAGGATAACCAACCGGCTCGCGTTTCTTACTTCGGTCAGAAAATGAAACGGGCGGTATCCTGATTAA
CACCCAGCGTCTCAGGGTGGTATCGGTGACCTGTATAACTTCAAACCTCGACCTTCCCTGACTCTGGGT
TGTGGTTCTTGGGGTGGTAACTCCATCTCTGAAAACGTTGGTCCGAAACACCTGATCAACAAGAAAACCG
TTGCTAAGCGAGCTGAAAACATGTTGTGGCACAACCTCCGAAATCTATCTACTTCCGCCGTGGTCCCT
GCCAATCGCGCTGGATGAAGTGATTACTGATGGCCACAACGTGCGCTCATCGTACTGACCGCTTCCCTG
TTCAACAATGGTTATGCTGATCAGATCACTTCCGTAAGAGCAGCAGGCGTTGAACTGAAGTCTTCT
TCGAAGTAGAAGCGGACCCGACCCTGAGCATCGTTGTAAGGTGCAGAACTGGCAAACTCCTTCAAACC
AGACGTGATTATCGCGCTGGGTGGTGGTTCCCGATGGACGCGCGAAGATCATGTGGTTATGTACGAA
CATCCGGAACCTCACTTCAAGAGCTGGCGCTGCGCTTTATGGATATCCGTAACGTATCTACAAGTTCC
CGAAAATGGGCGTGAAAGCGAAAATGATCGCTGTCAACCACTTCTGGTACAGGTTCTGAAGTCACTCC
GTTTGCGGTTGTAACGACGACGCTACTGGTCAGAAATATCCGCTGGCAGACTATGCGCTGACTCCGGAT
ATGGCGATTGTCGACGCCAACCTGGTTATGGACATGCCGAAGTCCCTGTGTGCTTTCCGGTGGTCTGGACG
CAGTAACTCACGCCATGGAAGCTTATGTTTCTGTACTGGCATCTGAGTTCTCTGATGGTCAGGCTCTGCA
GGCACTGAACTGCTGAAAGAATATCTGCCAGCGTCTACCACGAAGGGTCTAAAAATCCGGTAGCGGT
GAACGTGTTACAGTGCAGCGACTATCGCGGTATCGCGTTTGCGAACGCCTTCTGGGTGTATGTCACT
CAATGGCGCACAAAACCTGGGTCCCAGTTCCATATTCGCGACGGTCTGGCAAACGCCCTGCTGATTGTAA
CGTTATTCGCTACAATGCGAACGACAACCCGACCAAGCAGACTGCATTACGCCAGTATGACCGTCCGCAG
GCTCGCGCTGTTATGCTGAAATTGCCGACCACTTGGGTCTGAGCGCACCGGGCGACCGTACTGCTGCTA
AGATCGAGAACTGCTGGCATGGCTGGAAACGCTGAAAGCTGAACTGGGTATTCCGAAATCTATCCGTGA
AGCTGGCGTTGAGGAAGCAGACTTCTGGCGAACGTGGATAAACTGTCTGAAGATGCATTGATGACCAG
TGCACCGGCGCTAACCCGCTTACCCGCTGATCTCCGAGCTGAAACAGATTCTGCTGGATACCTACTACG
GTCGTGATTATGTAGAAGGTGAACTGCAGCGAAGAAAGAAGCTGCTCCGGCTAAAGCTGAGAAAAAAGC
GAAAAAATCCGCTTAA
```

## Referencias Bibliográficas:

*ADHE Fused acetaldehyde-COA dehydrogenase and Iron-dependent alcohol*

*dehydrogenasealdehyde/alcohol dehydrogenase ADHE [Escherichia coli Str. K-12*

*substr. MG1655] - Gene - NCBI. (s. f.). <https://www.ncbi.nlm.nih.gov/gene/945837>*

Arsalan, A. (2023, 31 diciembre). What is denaturation of DNA: 9 Answers You should know - .

*Lambda Geeks. <https://es.lambdageeks.com/what-is-denaturation-of-dna-thermal-denaturation/>*

Bonamore, A., & Boffi, A. (2007). Flavohemoglobin: structure and reactivity. *IUBMB Life*,

60(1), 19-28. <https://doi.org/10.1002/iub.9>

Created with Hot Potatoes by Half-Baked Software, registered to UNREGISTERED. (s. f.). *La*

*molécula de ADN. [https://www.juntadeandalucia.es/averroes/centros-](https://www.juntadeandalucia.es/averroes/centros-tic/14002984/helvia/aula/archivos/repositorio/750/820/html/moleculaadn/ejercicio3.htm)*

*[tic/14002984/helvia/aula/archivos/repositorio/750/820/html/moleculaadn/ejercicio3.htm](https://www.juntadeandalucia.es/averroes/centros-tic/14002984/helvia/aula/archivos/repositorio/750/820/html/moleculaadn/ejercicio3.htm)*

*Doble hélice. (s. f.). Genome.gov. <https://www.genome.gov/es/genetics-glossary/Doble-helice>*

*Escherichia Coli Str. K-12 substr. MG1655, Complete Genome - Nucleotide - NCBI. (s. f.).*

*[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_000913.3?from=1295446&to=1298121&repo](https://www.ncbi.nlm.nih.gov/nuccore/NC_000913.3?from=1295446&to=1298121&report=fasta&strand=true)*  
*[rt=fasta&strand=true](https://www.ncbi.nlm.nih.gov/nuccore/NC_000913.3?from=1295446&to=1298121&report=fasta&strand=true)*

Nhgri. (2019, 9 marzo). *Secuenciación del ADN*. Genome.gov.

*<https://www.genome.gov/es/about-genomics/fact-sheets/Secuenciacion-del-ADN>*