



Genes y genomas

Luis Alejandro
Rodríguez González
A01637333@itesm.mx

Ian
García González
A01706892@itesm.mx

Said Guadalupe
Ortigoza Trujillo
A01707430@itesm.mx

BT1013 – 829
Análisis de biología computacional
Profesor Edgar Acuña

Escuela de Ingeniería y Ciencias
Instituto Tecnológico y de Estudios Superiores de Monterrey

Jueves, 15 de abril de 2021

Parte 1.

- 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 es posible, debido a que los nucleótidos pueden estar en diferentes razones, ya que puede haber un 89% de GC y 11% de AT

- b) Si un organismo tiene 38% de A, ¿cuál es el porcentaje de G?
12%

- c) ¿Cuáles son las consecuencias de un alto contenido de GC?
Los pares GC tienen 3 enlaces de hidrógeno (vs. 2 que tienen los pares AT), por lo que la cadena será más fuerte y, por lo tanto, más estable (Tock, s.f.).

- d) ¿A qué refiere el superenrollamiento del ADN?
Es el retorcimiento sobre sí misma de una hélice siguiendo otra hélice que se denomina superhélice en lugar de línea recta (García Mondéjar, 2005).

- e) ¿Cómo se desnaturaliza el ADN? Es decir, ¿qué métodos se usan para separar las hebras?
De acuerdo con Wang et al. (2014), algunos métodos para separar la doble hélice del ADN son los siguientes:

- *Calor: aplicado directamente o mediante choques de frío*
- *Molino de cuentas: centrifugado*
- *Sonicación directa o indirecta*
- *Alcalino: mediante NaOH*
- *Formamida*
- *Dimetilsulfóxido*

- f) ¿Qué es la T_m del ADN de doble cadena?
Se define como la temperatura a la que se pierde la mitad de la estructura helicoidal (Tock, s.f.).

- g) ¿Por qué es importante saber la T_m de un segmento de ADN de doble cadena?
Conocer el valor de T_m de un segmento de ADN es importante porque es necesario para llevar a cabo experimentos de biología molecular como las pruebas PCR y qPCT. Esto es relevante para la situación problema del curso, ya que uno de los métodos de diagnóstico de covid-19 (Whitman, 2012).

- h) ¿Qué factores afectan la T_m ?
La longitud de la molécula y la proporción de pares GC. Mientras más haya, mayor es la temperatura de desnaturalización, porque tienen enlaces más fuertes (Kotrla, 2015). La fórmula para calcularla es la siguiente (para cadenas cortas de 14 a 20 pares):

$$T_m = (2^\circ\text{C})(A + T) + (4^\circ\text{C})(G + C)$$

- i) Observa la siguiente secuencia y encuentra la secuencia complementaria, la reversa y la reversa complementaria: 5'-ATGCTTGACGCTCAAACCATCGC-3'

Complementaria - TACGAACTGCGAGTTTGGTAGCG

Inversa - CGCTACCAAACCTCGCAGTTCGTA

Inversa complementaria - GCGATGGTTTGAGCGTCAAGCAT

Parte 2. 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?

Secuencia traducida

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

Parte 3.

a) ¿Qué significa el formato FASTA de una secuencia?

En el formato FASTA la secuencia comienza con un nombre y una descripción. Esta línea se distingue porque siempre comienza con el signo '>'. A continuación sigue la secuencia propiamente dicha con el formato en texto plano. Pueden incluirse varias secuencias en un mismo fichero.

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

- Flybase (Drosophila)
- SGD (Levadura)
- TAIR (Arabidopsis)
- ENSEMBL (Hombre, ratón y otros)

(COMAV, 2013)

c) Busca la secuencia de ADN de la Alcohol Deshidrogenasa de *Escherichia coli*, Proteína ADHE, gene adhE. Muestren la secuencia de doble cadena.

```
>NC_000913.3:c1298121-1295446 Escherichia coli str. K-12 substr. MG1655,
complete genome
ATGGCTGTTACTAATGTCGCTGAACTTAACGCACTCGTAGAGCGTGTAAGGAGCCAGCGTGAATATG
CCAGTTTCACTCAAGAGCAAGTAGACAAAATCTTCCGCGCCGCGCTCTGGCTGCTGCAGATGCTCGAAT
CCCACTCGCGAAAAATGGCCGTTGCCGAATCCGGCATGGGTATCGTCGAAGATAAAGTGATCAAAAACAC
TTTGCTTCTGAATATATCTACAACGCCTATAAAGATGAAAAACCTGTGGTGTCTGTCTGAAGACGACA
CTTTTGGTACCATCACTATCGCTGAACCAATCGGTATTATTTGCGGTATCGTTCCGACCACTAACCCGAC
TTCAACTGCTATCTTCAAATCGCTGATCAGTCTGAAGACCCGTAACGCCATTATCTTCTCCCCGCACCCG
CGTGCAAAAGATGCCACCAACAAAGCGGCTGATATCGTTCTGCAGGCTGCTATCGCTGCCGGTGCTCCGA
AAGATCTGATCGGCTGGATCGATCAACCTTCTGTTGAACTGTCTAACGCACTGATGCACCACCCAGACAT
CAACCTGATCCTCGCGACTGGTGGTCCGGGCATGGTTAAAGCCGCATACAGCTCCGGTAAACCAGCTATC
GGTGTAGGCGCGGGCAACACTCCAGTTGTTATCGATGAACTGCTGATATCAAACGTGCAGTTGCATCTG
TACTGATGTCCAAAACCTTCGACAACGGCGTAATCTGTGCTTCTGAACAGTCTGTTGTTGTTGTTGACTC
TGTTTATGACGCTGTACGTGAACGTTTTGCAACCCACGGCGGCTATCTGTTGCAGGGTAAAGAGCTGAAA
GCTGTTTCAAGATGTTATCCTGAAAAACGGTGCGCTGAACGCGGCTATCGTTGGTCAGCCAGCCTATAAAA
TTGCTGAACTGGCAGGCTTCTCTGTACCAGAAAAACCAAGATTCTGATCGGTGAAGTGACCGTTGTTGA
TGAAAGCGAACCCTTCGCACATGAAAAACTGTCCCCGACTCTGGCAATGTACCGCGCTAAAGATTTCGAA
GACGCGGTAGAAAAAGCAGAGAAACTGGTTGCTATGGGCGGTATCGGTCATACCTCTTGCTGTACACTG
ACCAGGATAACCAACCGGCTCGCGTTTCTTACTTCGGTCAGAAAATGAAAACGGCGGCTATCCTGATTAA
CACCCAGCGTCTCAGGGTGGTATCGGTGACCTGTATAACTTCAAACCTCGACCTTCCCTGACTCTGGGT
TGTGTTCTTGGGGTGGTAACTCCATCTCTGAAAACGTTGGTCCGAAACACCTGATCAACAAGAAAACCG
TTGCTAAGCGAGCTGAAAACATGTTGTGGCACAACCTTCCGAAATCTATCTACTTCCGCCGTGGCTCCCT
GCCAATCGCGCTGGATGAAGTGATTACTGATGGCCACAAACGTGCGCTCATCGTGACTGACCGCTTCCTG
TTCAACAATGGTTATGCTGATCAGATCACTTCCGTACTGAAAGCAGCAGGCGTTGAACTGAAGTCTTCT
TCGAAGTAGAAGCGGACCCGACCCTGAGCATCGTTTCGTAAAGGTGCAGAACTGGCAAACCTCCTTCAAACC
```

```
AGACGTGATTATCGCGCTGGGTGGTGGTTCCCCGATGGACGCCGCGAAGATCATGTGGGTATGTACGAA
CATCCGGAAACTCACTTCGAAGAGCTGGCGCTGCGCTTTATGGATATCCGTAAACGTATCTACAAGTTCC
CGAAAATGGGCGTGAAAGCGAAAATGATCGCTGTCAACCACACTTCTGGTACAGGTTCTGAAGTCACTCC
GTTTGCGGTTGTAAGTACGACGCTACTGGTCAGAAATATCCGCTGGCAGACTATGCGCTGACTCCGGAT
ATGGCGATTGTCGACGCCAACCTGGTTATGGACATGCCGAAGTCCCTGTGTGCTTTCGGTGGTCTGGACG
CAGTAACACAGCCATGGAAGCTTATGTTTCTGTACTGGCATCTGAGTTCTCTGATGGTCAGGCTCTGCA
GGCACTGAACTGCTGAAAGAATATCTGCCAGCGTCCTACCACGAAGGGTCTAAAAATCCGGTAGCGCGT
GAACGTGTTTACAGTGCAGCGACTATCGCGGGTATCGCGTTTGCGAACGCCTTCCTGGGTGTATGTCACT
CAATGGCGCACAACTGGGTTCAGTTCCATATTCCGCACGGTCTGGCAAACGCCCTGCTGATTTGTAA
CGTTATTGCTACAAATGCGAACGACAACCCGACCAAGCAGACTGCATTCAGCCAGTATGACCGTCCGCGAG
GCTCGCCGTCGTTATGCTGAAATTGCCGACCACCTTGGGTCTGAGCGCACCGGGCGACCGTACTGCTGCTA
AGATCGAGAACTGCTGGCATGGCTGGAACGCTGAAAGCTGAACTGGGTATTCCGAAATCTATCCGTGA
AGCTGGCGTTTACGGAAGCAGACTTCCTGGCGAACGTGGATAAACTGTCTGAAGATGCATTCGATGACCAG
TGCACCGGCGCTAACCCGCGTTACCCGCTGATCTCCGAGCTGAAACAGATTCTGCTGGATACCTACTACG
GTCGTGATTATGTAGAAGGTGAAACTGCAGCGAAGAAAGAAGCTGCTCCGGCTAAAGCTGAGAAAAAAGC
GAAAAATCCGCTTAA
```

Bibliografía

- COMAV. (2013). *Bases de datos biológicas*. Bioinformatics at COMAV.
https://bioinf.comav.upv.es/courses/intro_bioinf/bases_datos.html
- García Mondéjar, L. (2005). *Definición de superenrollamiento*. Universidad de Alcalá.
<http://biomodel.uah.es/an/super/inicio.htm>
- Kotrla, T. (2015). *Unit 9 Assignments- Nucleic Acid Amplification*. Austin Community College.
https://www.austincc.edu/mlt/mdfund/mdfund_unit9assignmentsMeltingTemperature.html
- NCBI (2021) *adhE aldehyde-alcohol dehydrogenase [Escherichia coli str. K-12 substr. MG1655]*. <https://www.ncbi.nlm.nih.gov/gene/945837>
- NCBI (2021) *Escherichia coli str. K-12 substr. MG1655, complete genome*.
https://www.ncbi.nlm.nih.gov/nuccore/NC_000913.3?report=fasta&from=2685835&to=2687025
- Tock, K. (s.f.). *Biopolymers*. Stanford University.
<https://web.stanford.edu/~kaleeg/chem32/biopol/>
- Wang, X.; Lim, H. J.; & Son, A. (2014). *Characterization of denaturation and renaturation of DNA for DNA hybridization*. NCBI.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168728/>
- Whitman, M. (2012). *The importance of melting temperature in molecular biology applications*. IDT.
<https://www.idtdna.com/pages/education/decoded/article/the-importance-of-tm-in-molecular-biology-applications>