

# Introducción a las bases de datos de secuencia genómica y análisis de secuencias relacionadas con el polimorfismo de globina beta S

## (Práctica en laboratorio virtual Cibertorio)

### Fundamento:

La anemia drepanocítica (anemia de células falciformes) se origina por la mutación de un solo nucleótido en el gen de globina beta humana. Esto genera en la población un polimorfismo, definido por la presencia del alelo normal o el mutado, en cada uno de los dos cromosomas 11 homólogos.

En el contexto del estudio molecular de esta enfermedad y su diagnóstico genético, vamos a estudiar las secuencias depositadas en las bases de datos del genoma correspondientes al gen de globina beta y su punto polimórfico.

A lo largo de la práctica tendrás que conservar información para pasarla de una ventana a otra del programa; utiliza para ello un editor de texto plano como el “Bloc de Notas” de Windows. Conviene que añadas tus propios comentarios para identificar la información que vas guardando; puedes imprimirlo al acabar, para llevarte toda la información e incluirla en tu cuaderno de prácticas.

Ayuda para copiar y pegar texto: utiliza el menú que aparece al pulsar el botón derecho del ratón o, con el teclado, pulsa Ctrl+C para copiar y Ctrl+V para pegar.

### Bibliografía:

- A. Herráez (2012) *Texto Ilustrado e Interactivo de Biología Molecular e Ingeniería Genética*, 2ª ed., págs. 418-420. Elsevier España, Madrid.
- J. Luque y A. Herráez (2001) *Texto Ilustrado de Biología Molecular e Ingeniería Genética*, págs. 381-383. Ediciones Harcourt / Elsevier España, Madrid.

### Cómo comenzar:

El laboratorio virtual funciona dentro de una página *web*, por lo que debes comenzar abriendo el navegador de internet en tu ordenador o tableta. Cualquier navegador moderno debería funcionar, siempre que tenga activado JavaScript (se recomienda Firefox o Chrome).

Visita la página de Cibertorio en <http://biomodel.uah.es/lab/cibertorio/>

### Ensayo: análisis de secuencias

#### Objetivo:

Obtener información sobre el polimorfismo del gen de globina beta que da lugar a la hemoglobina S y causa la drepanocitosis. Familiarizarse con las bases de datos de secuencia, diferencias entre DNA genómico y cDNA, conceptos de intrón y exón.

#### Materiales virtuales:

- Las herramientas simples de la sección “Análisis de secuencia” dentro del laboratorio virtual de biología molecular “Cibertorio”, [biomodel.uah.es/lab/cibertorio/](http://biomodel.uah.es/lab/cibertorio/)
- Secuencias de la base de datos internacional GenBank, incluidas dentro de la “Base de datos CygnusLab™” de Cibertorio

#### Procedimiento:

- 1) Desde el menú de Cibertorio, inicia el módulo Análisis de secuencias y entra en la Base de datos CygnusLab™.
- 2) La primera sección de esta base de datos se titula “*HSBGLX mRNA - GenBank file for human beta globin mRNA*”. GenBank es una de las principales bases de datos internacionales que guardan las secuencias de genomas y proteínas, a las que se accede por internet. Esta entrada está identificada en la base de datos con el código HSBGLX. Observa cómo está organizada la información e intenta comprender lo que significa cada parte del registro; concretamente, localiza la siguiente información:

- tamaño del mRNA (nº de pares de bases)
- palabras clave (sirven para ayudar en la búsqueda de una secuencia dentro de la base de datos)
- especie de la que se ha obtenido la secuencia
- composición del DNA
- explica cómo está organizada en columnas la secuencia de nucleótidos
- secuencia de la proteína

Observa las secuencias del mRNA (en realidad, expresada en formato DNA) y de la proteína; ¿encuentras algún símbolo que no corresponda a un nucleótido o a un aminoácido?

(pista: las abreviaturas de aminoácidos son:

A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr

Debido a estas ambigüedades, trabajaremos con otra secuencia:

- Pasa a la segunda sección de la base de datos, titulada *"HUMHBB genomic DNA - Partial GenBank Human Beta Globin Genomic Sequence"* (se ha rotulado *"partial"* porque para abreviar se ha borrado parte de la información, que no necesitamos; puedes ver la entrada completa pulsando en el enlace de abajo). En este caso se trata de una secuencia genómica, lo que significa que es DNA y además incluye los intrones (¿recuerdas lo que son los intrones?).
- Busca el identificador "CDS" en el margen izquierdo, que indica las porciones codificantes (los exones). Anota esas cifras.
- Copia la secuencia completa, abre el Bloc de Notas de Windows y pégala. Ve copiando las bases que corresponden a cada exón anotado antes y pégalas aparte (no te preocupes por los números, los espacios y los saltos de línea). Copia la secuencia del conjunto de exones
 

Los extremos correctos que debes obtener para los exones son:  
 atgg...gcag      gctg...cagg      ctcc...ctaa
- Vuelve a la ventana principal de Cibertorio, que contiene las herramientas de análisis de secuencia, y abre el módulo Masajista de secuencias (se abrirá en la misma ventana en que tenías la base de datos). Pega la secuencia (exones combinados) en el recuadro blanco.
- Pulsa los botones eliminar saltos de línea, eliminar números y eliminar espacios en blanco. Luego pulsa el botón mostrar tripletes. Si lo has hecho bien, debes obtener tripletes completos. Si no, repasa las tareas anteriores (en especial, donde borraste las bases de los intrones). Vuelve a eliminar espacios en blanco.
- A este conjunto de secuencia que se traducirá se le llama marco de lectura abierto (ORF, *open reading frame*). Selecciona el texto (¡no debe tener espacios entre los tripletes!) y cópialo. Para estudiar cuál será la secuencia de la proteína, abre el módulo Traductor de proteínas desde la página de herramientas para el Análisis de Secuencia de Cibertorio. Pega la secuencia que acabas de copiar, en el recuadro titulado secuencia de DNA. Pulsa los botones correspondientes para obtener primero la secuencia del mRNA y luego la del polipéptido. Anota las secuencias del mensajero y del polipéptido.
- Compara la secuencia de aminoácidos obtenida con las presentes en los archivos de GenBank HUMHBB y HSBGLX, utilizados antes. Discútelo. Identifica las diferencias encontradas y ofrece una interpretación.