

## Guión de la práctica

# 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.

Pretendemos estudiar el posible diagnóstico genético de esta enfermedad a través de un polimorfismo RFLP, es decir, mediante restricción y electroforesis.

Como material de partida disponemos de muestras de DNA (virtuales) correspondientes a una parte de ese gen, para las dos variantes (normal, AA o WW, y drepanocítica, SS).

Por lo tanto, vamos a estudiar un locus que corresponde a una parte del gen de globina beta y que contiene el punto polimórfico.

#### Más información (no necesaria para la práctica):

Concretamente, las muestras comprenden el exón 1, el intrón A, el exón 2 y parte del intrón B.

Recuerda que el polimorfismo se sitúa en el exón 1.

#### Aclaración de la nomenclatura:

A la hemoglobina normal del adulto,  $\alpha_2\beta_2$ , se la llama habitualmente **hemoglobina A**; de ahí el nombre A para el alelo normal de la globina beta aquí estudiado. También se representa como W, de *wild type* (tipo salvaje o silvestre).

La hemoglobina de las células falciformes (*sickle cells*),  $\alpha_2\beta_2^S$ , se denomina **hemoglobina S**.

A lo largo de la práctica tendrás que conservar información para pasarla de una ventana a otra del programa; utiliza para ello 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:

- 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 (Internet Explorer o Firefox).

Una vez que se abra el programa, debes buscar la página de Cibertorio:

Puede estar en los "Favoritos" o "Marcadores"; si no, visita <http://biomodel.uah.es/lab/cibertorio/>

## Ensayo: análisis de secuencias.

### Materiales virtuales:

- Las herramientas simples de la sección “Análisis de secuencia” dentro de Cibertorio.
- Secuencias de la base de datos internacional GenBank, incluidas dentro de la “Base de datos BobCo™” de Cibertorio.

### Procedimiento

- 1) Desde el menú de Cibertorio, inicia el módulo “**Análisis de secuencias**” y entra en la “**Base de datos BobCo**”.
- 2) La segunda sección de la base de datos se titula “*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:
  - a) tamaño del mRNA (nº de pares de bases)
  - b) palabras clave (sirven para ayudar en la búsqueda de una secuencia dentro de la base de datos)
  - c) especie de la que se ha obtenido la secuencia
  - d) composición del DNA
  - e) explica cómo está organizada en columnas la secuencia de nucleótidos
  - f) 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:

- 3) Baja a la tercera sección de la base de datos, titulada “*Partial GenBank Human Beta Globin Genomic Sequence*” (código HUMHBB) (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?).
- 4) Busca el identificador “CDS” en el margen izquierdo, que indica las porciones codificantes (los exones). Anota esas cifras.
- 5) 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

- 6) 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.
- 7) 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**”.
- 8) 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.
- 9) Compara la secuencia de aminoácidos obtenida con las presentes en los archivos de GenBank HUMHBB y HSBGLX, utilizados antes. Discútelos. Identifica las diferencias encontradas y ofrece una interpretación.