

# 26

## CAPÍTULO

### Οργάνωση del genoma eucariótico

En todos los organismos celulares el material genético está constituido por ADN de doble banda; aun cuando su estructura general es muy similar y está compuesto por las mismas bases nitrogenadas (A, T, C, G) existen notables diferencias entre las organizaciones particulares de los genomas en procariontes y eucariontes.

Una primera diferencia reside en el contenido de ADN de los diferentes organismos. Una bacteria típica como la *E. coli* posee un material genético equivalente a  $4 \times 10^6$  bases. En un insecto como la *Drosophila melanogaster* es de  $1,6 \times 10^7$  y en los mamíferos, de  $4,8 \times 10^9$ .

En segundo lugar, en los procariontes el ADN ocupa un lugar central en la célula, sin que exista una separación física entre él y el resto del organismo. En las células eucariontes el material genético (ADN) se encuentra separado físicamente del resto de la célula por una envoltura constituida de una doble membrana (capítulo 23).

Pero tal vez la diferencia más sobresaliente radica en la forma característica que está organizado el material genético en los eucariontes. Como muchas de esas particularidades influyen notablemente en los mecanismos de expresión de la información genética y su regulación, se estudiarán primero las particularidades fundamentales de la organización del genoma en los organismos eucariontes.

### Genes y cromosomas

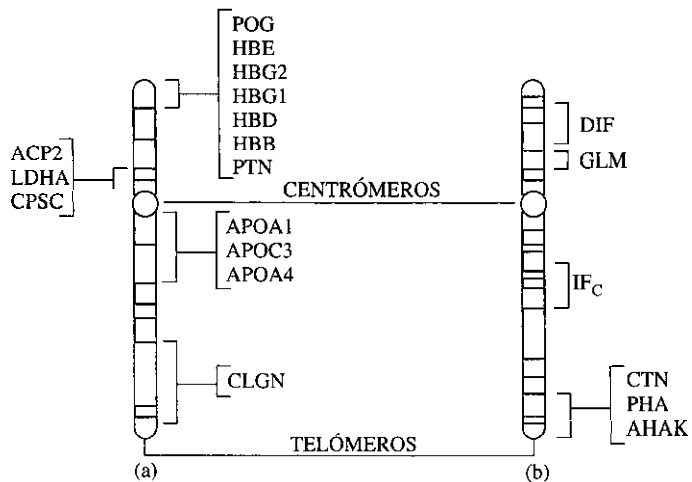
En las células de eucariontes el ADN total se encuentra dividido en un número variable de moléculas, característico de cada especie, y formando con proteínas, estructuras muy complejas denominadas cromosomas: el número y la forma de los cromosomas son típicos de cada especie.

En un organismo existen 2 tipos fundamentales de células en cuanto al número de cromosomas que contienen: las somáticas y las sexuales. Las primeras contienen el doble de cromosomas que las segundas, por eso se dice que aquéllas son diploides y éstas últimas, haploides. En las células somáticas los cromosomas se presentan por pares homólogos, los 2 miembros de la pareja son equivalentes, por lo que contienen una dotación genética doble con respecto a las sexuales.

El ADN que forma parte de los cromosomas es una sola molécula, que al parecer está en forma lineal y no circular. A todo su largo está asociado con proteínas que son fundamentalmente histonas, aunque existen otras no muy bien caracterizadas que junto con las histonas determinan las formas organizativas peculiares de los

cromosomas. Como se vio en el capítulo 23, el grado de “empaquetamiento” del complejo ADN + proteínas varía durante el ciclo celular, y las estructuras más compactas -los cromosomas propiamente dichos- existen sólo durante el período de división celular. Durante la interfase esta estructura se hace mucho menos compacta y se presenta en forma de cromatina. Como la mayoría de los estudios relacionados con la organización del genoma ha tomado como punto de referencia los cromosomas, en este texto se utilizará este término para referirse a estos complejos ADN + proteínas, sin tener en cuenta su grado de “empaquetamiento”. Cada cromosoma contiene un número característico de genes ordenados de forma lineal a lo largo de su estructura. Por procedimientos múltiples se ha podido conocer la localización de numerosos genes en los cromosomas humanos (Fig. 26.1). Durante la maduración de las células sexuales (gametogénesis) se produce la reducción del número de cromosomas, y cada gameto maduro (óvulo y espermatozoide) contiene solamente una sola copia de cada par de cromosoma, por lo tanto en las células sexuales cada gen está representado sólo una vez, a diferencia de las células somáticas que por ser diploides tienen una representación doble de cada gen

Fig. 26.1. Localización de algunos genes en cromosomas humanos. (a) Se representa la localización de algunos genes del cromosoma 11 como son algunos protooncogenes (POG), las cadenas de las globinas  $\epsilon$  (HBE),  $\gamma$  (HBG2 y HBG1),  $\delta$  (HBD) y  $\beta$  (HBB); paratohormona (PTH), fosfatasa ácida 2 (ACP2), lactato deshidrogenasa A (LDHA), genes de las catepsinas (CPSC), de las apolipoproteínas A1 (APOA1), C3 (APOC3) y A4 (APOA4), así como de la colagenasa (CLGN) (b) La localización de algunos genes cuyas alteraciones producen enfermedades como la deficiencia de interferón (DIF), galactosemia (GLM), porfiria hepática aguda (PHA) y anemias hemolíticas por deficiencia de la adenilato quinasa (AHAK), todos localizados en el cromosoma 9.



Aun cuando existen numerosos enfoques para tratar este problema, y de cada uno de ellos puede derivar una concepción distinta de gen, de acuerdo con las características de éste que más se pretenda destacar y como consecuencia de la profundización de los conocimientos sobre la naturaleza del gen, para los objetivos de este capítulo se utilizará la definición más empleada en la actualidad. Se da el nombre de gen a uno o varios sectores de una molécula de ADN, que contiene en su secuencia de bases nitrogenadas la información necesaria para llevar a cabo la síntesis de una cadena polinucleotídica específica. Este ADN puede codificar la síntesis de todos los tipos de ARN, especialmente los ribosomales, los de transferencia y el mensajero. En este último caso, la secuencia de bases es traducida en secuencia de aminoácidos durante la traducción.

Al conjunto de todos los genes contenidos en una célula se le da el nombre de genoma. Las secuencias, que pueden ser exactamente iguales en ambos cromosomas o diferir en algo, siguen originando en esencia el mismo producto, aunque con pequeñas diferencias, esto hace que en una población puedan existir numerosas formas del mismo gen. Uno de los ejemplos más sobresalientes en seres humanos es el gen que codifica la cadena  $\beta$  de la hemoglobina, este gen está localizado en el cromosoma 11 y de él se han reportado más de 200 formas que difieren por lo general en el cambio de alguna base nitrogenada que, aunque en muchos casos determina cambios en la secuencia aminoacídica de la proteína, sigue siendo esencialmente la  $\beta$ -globina (tabla 26.1).

Tabla 26.1. Diferentes tipos de alelos en la cadena  $\beta$  de la hemoglobina

Tipo	Defecto molecular	Fenotipo
Hb S	Cambia en $\beta 6$ glu-val	Sickleemia
Hb C	Cambia en $\beta 6$ glu-lis	
Hb E	Cambia en $\beta 26$ glu-lis	
Hb M <sub>Boston</sub>	Cambia en $\beta 53$ his-tir	Meta Hb
Hb M <sub>Saskatoon</sub>	Cambia en $\beta 63$ his-tir	Meta Hb

Al conjunto de formas alternativas del mismo gen que pueden ocupar el mismo sitio en el cromosoma (*locus*) se le da el nombre de alelos. Pueden existir numerosas formas alélicas del mismo gen, pero una célula diploide sólo puede contener 2 como máximo, una en cada uno de los cromosomas homólogos. Si el par está representado por genes idénticos, y por tanto sus productos son indistinguibles, el organismo es homocigótico para esa pareja, pero si se trata de formas alélicas, entonces es heterocigótico. Es muy difícil encontrar algunas células u organismos pluricelulares que no sean heterocigóticos para uno u otro par de genes; esta diferencia en la estructura de los genes se refleja en su producto, o sea, en la proteína y su función. Se utiliza el término de genes de tipo silvestre para referirse al gen que supuestamente conserva su forma original, a sus formas alternativas se les denomina mutantes, por cuanto es la mutación el mecanismo fundamental en la formación de los alelos. En ocasiones algunos de los alelos mutantes codifican un producto que no es funcional. Suponga que existe un gen "A" que es de tipo silvestre y su mutante "a" que da un producto anormal, incapaz de realizar cualquier función. Existen 3 combinaciones posibles para esta pareja de genes (Fig. 26.2):

1. Los 2 genes son de tipo silvestre, por lo que el organismo es homocigótico (AA) y todo el producto génico es normal.
2. Un gen es de tipo silvestre y el otro es mutante (Aa), el organismo es heterocigótico y sólo contiene aproximadamente la mitad del producto génico normal.
3. Los 2 genes son mutantes (aa), el organismo es homocigótico, por lo que todo el producto génico es anormal. Si ese producto realiza una función importante para la vida, la célula en esas condiciones no sobrevivirá o lo hará en condiciones precarias.

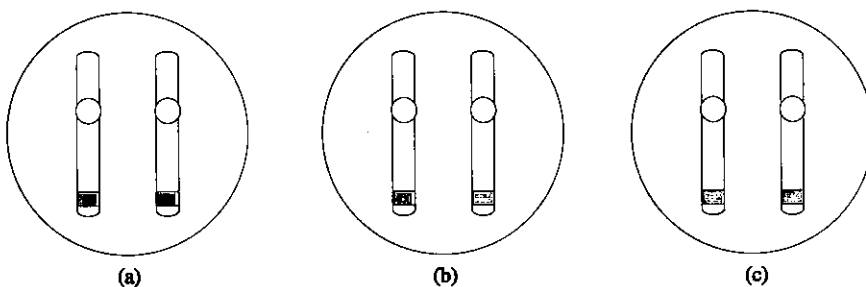


Fig. 26.2. Posibles combinaciones de alelos. Pueden existir numerosos alelos de un mismo gen, pero en un organismo dado sólo pueden haber 2, uno por cada cromosoma. Si se representa en azul el gen de tipo silvestre que es dominante y en rojo el mutante recesivo, podemos encontrar 3 combinaciones: (a) Los 2 genes son tipo silvestre y el organismo es homocigótico dominante. (b) Un gen es tipo silvestre y el otro mutado, así como el organismo es heterocigótico. (c) Los 2 genes son mutantes y el organismo es homocigótico recesivo. Dependiendo de las características del producto génico el heterocigótico puede manifestar o no un fenotipo anormal.

El conjunto de genes presente en una célula constituye el genoma, pero el par de ellos que determina un carácter particular recibe el nombre de genotipo. La manifestación externa del genotipo recibe el nombre de fenotipo; éste puede definirse desde muchos puntos de vista, desde la posibilidad de metabolizar una sustancia o reconocer una señal química, hasta los rasgos más externos del organismo como el

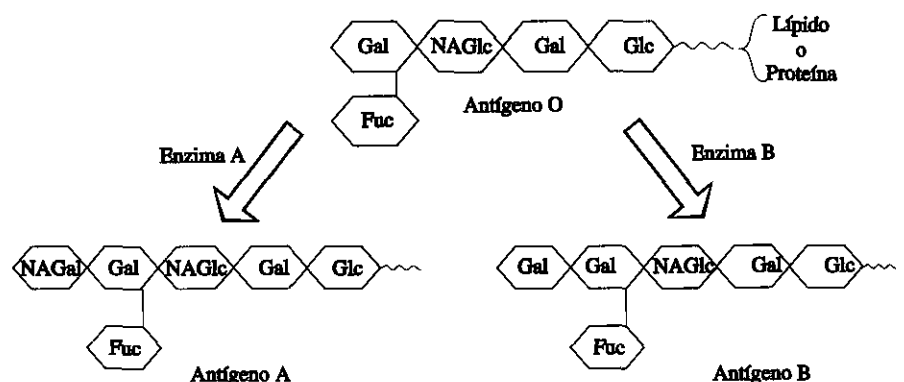
color de los ojos, la forma de las extremidades, etcétera. Los caracteres acostumbran a clasificarse en dominantes o recesivos, de acuerdo con los patrones de transmisión génica. En ocasiones se comete el error de hablar de genes dominantes y recesivos, pero en realidad estos términos sólo se refieren al carácter determinado por el genotipo.

Estos conceptos son relativos y contrarios, o sea, la existencia de uno determina la existencia del otro. En un inicio se decía que el carácter era dominante cuando siempre se expresaba en el fenotipo, mientras que el recesivo sólo lo hacía si estaba en estado homocigótico; aquí se concebía el fenotipo como las características más externas del organismo. En la medida que se han ido perfeccionando el conocimiento sobre la expresión de los genes y los procedimientos que permiten localizar y cuantificar el producto génico, la concepción del fenotipo y con ella la del carácter recesivo se han modificado; es de esperar que con las nuevas técnicas introducidas que permiten detallar la estructura del gen, la diferencia esencial inicial entre estos 2 conceptos tienda a desaparecer. La distinción entre caracteres dominantes y recesivos es muy importante en la práctica médica, pues hay enfermedades que se transmiten como un carácter dominante y otras como recesivo, todo lo cual el médico debe tener muy presente para hacer los asesoramientos genéticos del paciente y de la familia.

Un ejemplo permitirá esclarecer los conceptos definidos hasta el momento. Para los grupos sanguíneos ABO existen básicamente 3 alelos. El grupo sanguíneo A es dominante y está determinado por el alelo A. Nótese que aquí el fenotipo no es algo externo, pues por la sola observación del sujeto no es posible determinar su grupo sanguíneo. El grupo sanguíneo B es también dominante y lo determina el alelo B. El grupo O tiene carácter recesivo y se origina por el alelo O. Una persona con la combinación génica AA será del grupo sanguíneo A, en tanto la del genotipo BB será del B. Las combinaciones genotípicas AO y BO originan fenotipos A y B, respectivamente. El genotipo OO corresponde al grupo sanguíneo O. Ahora bien, como A y B son dominantes, la persona que tenga un genotipo AB expresará ambos genes por igual y será del grupo sanguíneo AB. Cuando 2 genes que determinan el mismo carácter se expresan de igual forma en el fenotipo, se dice que los caracteres son codominantes.

Esto tiene su explicación molecular. Las sustancias que determinan el grupo sanguíneo ABO son pequeños heteropolisacáridos que forman parte de la membrana de los eritrocitos. Como se muestra en la figura 26.3, los sujetos del grupo sanguíneo O

Fig. 26.3. Determinación genética de los grupos sanguíneos ABO. Los eritrocitos poseen un polisacárido unido a lípidos o proteínas de sus membranas, éste posee carácter antigénico y está formado como se muestra en la figura por glucosa (Glu), galactosa (Gal), fucosa (Fuc) y N-acetilglucosamina (NAGlu), y es denominado antígeno O. El gen A codifica la enzima A que añade al polisacárido un resto de N-acetilgalactosamina (NAGal) y lo convierte en el antígeno A. Por su parte el gen B codifica la enzima B que añade al antígeno O un resto de galactosa y lo convierte en el antígeno B. Si la persona posee los 2 genes poseerá también las 2 enzimas y aproximadamente la mitad de sus antígenos serán A y la otra mitad será B, por lo que ella será del grupo AB. De no poseer ninguno de estos genes, entonces será del grupo O.



tienen en sus eritrocitos el polisacárido mostrado en (a). El gen A codifica una enzima que añade la N- acetil-galactosamina al extremo del polisacárido, cambiando su especificidad reaccional; en tanto el gen B codifica la enzima que añade galactosa, otorgando una especificidad diferente. En los sujetos AA ó AO la estructura resultante será como se muestra en (b), en los BB ó BO es como en (c). En los OO es como en (a), mientras que en los AB se encontrarán como en (b) y (c) aproximadamente a partes iguales.

## Genes y ADN

De acuerdo con numerosos estudios el ADN contenido en una célula diploide humana es aproximadamente  $7,3 \times 10^{-12}$  g. Si el peso molecular promedio de un par de nucleótidos es de  $1,025 \times 10^{-21}$  g se deduce que la célula diploide humana contiene alrededor de  $(7,3 \times 10^{-12}/1,025 \times 10^{-21}) 7,1 \times 10^9$  pares de bases por genoma diploide, ó  $3,5 \times 10^9$  por genoma haploide.

Si todo este ADN codificara proteínas de aproximadamente 500 aminoácidos cada una, podría producirse más de un millón de ellas. Como para cada aminoácido se requieren 3 bases harán falta  $(500 \times 3) 1 500$  pares de bases. A partir del genoma total serían  $(3,5 \times 10^9/1,5 \times 10^3) 2,3 \times 10^6$  proteínas.

Por varios métodos genéticos se ha determinado que la célula humana contiene aproximadamente de 70 000 a 100 000 genes. Si se mantienen los mismos "supuestos" anteriores, para la codificación de 100 000 proteínas serían necesarias  $[(7 \times 10^4) \times (3 \times 10^3)] 2,1 \times 10^8$  pares de bases, lo cual representa  $[(2,1 \times 10^8)/(3 \times 10^9)$  aproximadamente  $10^{-2}$ ], es decir, que sólo aproximadamente el 10 % del ADN contiene información para la síntesis de moléculas específicas (proteínas, ARNr y ARNt). De aquí se deduce que la mayor parte del ADN está implicada en los mecanismos de regulación de la expresión genética o en otras funciones aún desconocidas.

Si se aísla el ADN de una bacteria y se corta mediante enzimas en unos 500 a 1 000 fragmentos y se centrifugan en un gradiente de densidad de CsCl, se observan que se reúnen en una zona única y estrecha (Fig. 26.4).

Esto se debe a que la densidad del ADN es proporcional a su contenido en G + C y la composición promedio de cada fragmento, que contiene varios cientos de bases, no varía mucho de uno a otro. Si se aplica el mismo procedimiento al ADN de un mamífero se obtienen 2 zonas, la mayor con una composición aproximada de A + T del 60 %, pero la menor con más del 90 % de A + T, lo cual constituye un hecho inusual. Al ADN contenido en esta zona se le denomina ADN satélite, éste ha sido observado en muchos organismos y en el ser humano constituye del 0,5 al 1 % ( $10^5$  pares de bases) del genoma.

Una característica importante del ADN satélite es la de estar formado por secuencias repetidas organizadas en tándem, una a continuación de la otra. Estas secuencias generalmente son cortas, de 5 a 10 bases, pero pueden llegar a tener de 100 a 200 bases. Las secuencias características de la *D. melanogaster* son las siguientes:

5'-A-C-A-A-A-T-T-3'

5'-A-C-A-A-A-C-T-3'

5'-A-T-A-A-A-C-T-3'

Se ha podido determinar la localización de este ADN satélite, para ello se procedió de la forma siguiente: se tomaron muestras de ADN satélite y a partir de él se formó ARN marcado con  $[^3\text{H}]$ -uridina. Se fijaron células cultivadas sobre un portaobjetos y se trataron con álcalis para desnaturalizar el ADN. Con este tratamiento el cromosoma queda prácticamente intacto. Se añadió el  $[^3\text{H}]$ -ARN y se dio tiempo suficiente para permitir la hibridación. Se lavó la preparación para eliminar el exceso de reactivo y se sometió a un proceso de autorradiografía. Al revelar la placa se encontró que el ADN satélite se localizaba fundamentalmente en los centrómeros y con menor frecuencia en los telómeros; se ha pensado que esté relacionado con la fijación del huso durante la división celular, pero esto no ha sido comprobado. No debe confundirse al ADN satélite con los satélites descritos en la estructura de algunos cromosomas, pues aunque la palabra puede tener en ambos casos la misma connotación (una estructura asociada con otra) se refiere en cada caso a conceptos diferentes.

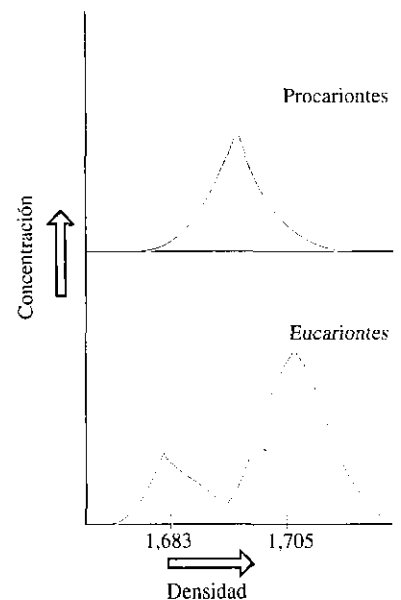
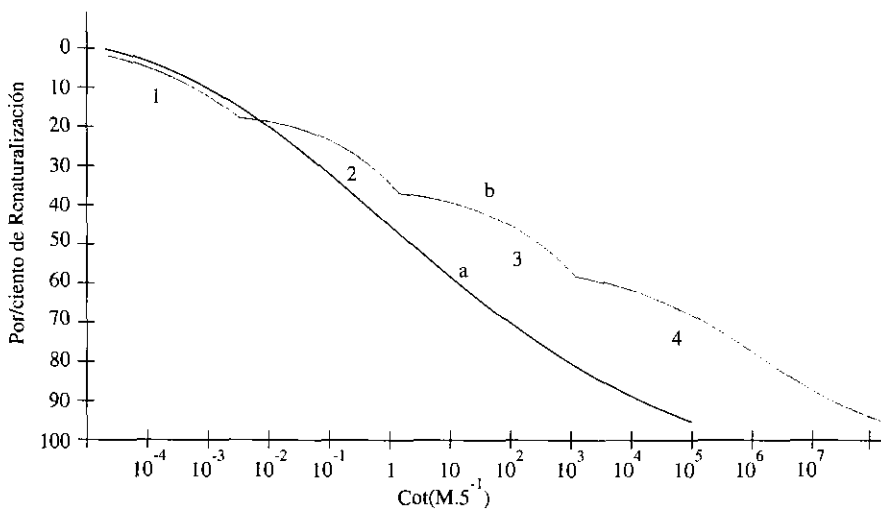


Fig. 26.4. Detección del ADN satélite. Si el ADN de los procariotes se corta con nucleasas específicas en un grupo reducido de fragmentos de gran tamaño y después se centrifugan en gradiente de densidad, estos fragmentos se agrupan en una zona estrecha del tubo de centrifugación donde la densidad de ellos sea igual a la del medio. El ADN de eucariotes, al ser sometido al mismo procedimiento, se concentra en 2 zonas de densidad diferente. La de menor densidad corresponde al ADN satélite.

Estos hallazgos impulsaron a realizar el estudio más a fondo acerca de las características de los ADN de eucariontes. Un procedimiento que ha dado muchos resultados es el estudio de la cinética del proceso de renaturalización. Para ello se extrae el ADN de un organismo y sin fraccionarlo se desnaturaliza, después se coloca en condiciones que permiten la renaturalización, siguiendo el proceso en función de la concentración inicial de ADN ( $C_0$ ) y el tiempo ( $t$ ). Cuando esto se hace con el ADN de la *E. coli* se obtiene una curva típica que se muestra en la figura 26.5 (a); se interpreta como si el proceso ocurriera por la presencia de una sola especie molecular. Cuando se hace con ADN de mamíferos se obtiene una curva que aparece en la figura 26.5. (b), que presenta varias mesetas como si en el proceso intervinieran más de una especie molecular, una de renaturalización rápida, otras intermedias y otra lenta; esto se debe a la existencia de secuencias con mayor o menor grado de repetitividad.

Fig. 26.5. Curva Cot. Si se desnaturaliza el ADN de la *E. coli* (azul) y se sigue el proceso de renaturalización, se obtiene una línea continua como corresponde a una sola especie molecular. El ADN de mamíferos (rojo) presenta varias curvaturas, lo cual indica que éste se comporta de forma heterogénea. Una fracción (1) presenta una reasociación rápida y otra (4) es lenta, a la vez que se observan otras 2 (2 y 3) que presentan una velocidad de reasociación intermedia; esto significa que el ADN de mamíferos se comporta como si estuviera formado por 4 especies moleculares.



El análisis detallado de estas curvas y el uso de otros métodos complementarios permiten distinguir 4 tipos de secuencias en el ADN de eucariontes: las altamente repetidas, que aparecen por varios cientos o millones de copias y se corresponden con las de renaturalización rápida; las medianamente repetidas, que se presentan en un número de copias que va de 10 a varios cientos; las moderadamente repetidas con un número de copias de 2 a 10 y junto con las anteriores representan la fracción de velocidad intermedia de renaturalización; por último, las secuencias de copia única que son las de menor velocidad de renaturalización. En forma de copia única se encuentra la mayoría de los genes de la célula, pero en realidad constituye la menor fracción del ADN celular. Ejemplos de secuencias moderadamente repetidas son los genes de las histonas y de los ARNt. En algunos casos estas secuencias no son exactamente repetidas y los productos génicos presentan pequeñas diferencias en su composición. Las secuencias medianamente repetidas en general no son codificantes y se cree que tengan alguna función reguladora. Las altamente repetidas incluyen al ADN satélite y algunos genes, como los que codifican los precursores de los ARNr, para los que se han calculado que existen unas 24 000 copias para el de 5 S y unas 500 copias para el precursor de los otros 3 tipos.

### Estructura del gen

Una vez analizadas las características generales de los ADN de eucariontes, es conveniente detenerse a examinar la estructura interna del gen. Como ya se señaló, un gen está constituido por uno o más segmentos específicos del ADN, cuya secuencia de bases contiene la información necesaria para la síntesis de moléculas específicas, o sea, proteínas, ARNt y ARNr.

Si se compara la longitud de algunos genes de eucariontes con la de sus productos se observa que éste es generalmente mucho menor que aquél. Así, de acuerdo con la

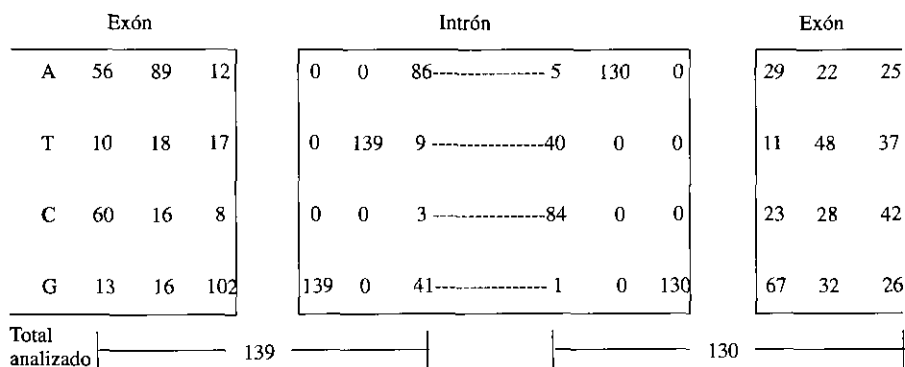
longitud de su gen, la cadena  $\beta$  de la hemoglobina debía poseer unos 500 aminoácidos cuando en realidad sólo contiene 146 aminoácidos.

Hechos como éste comenzaron a ponerse de manifiesto a mediados de la década de los años 60 y despertaron un vivo interés. Después de múltiples esfuerzos se ha podido comprobar que los genes eucariontes no son continuos, sino que se encuentran interrumpidos por secuencias de nucleótidos que no codifican para la proteína dada; por lo tanto la característica general del genoma de poseer zonas codificantes y no codificantes se repite a pequeña escala en la estructura misma del gen. Así cada gen está constituido por un número de secuencias codificantes que alternan con secuencias no codificantes. Las primeras reciben el nombre de exones (*exit*, salida, alude al hecho de que estas secuencias una vez transcritas abandonan el núcleo) y las segundas, el de intrones (*inside*, dentro, pues una vez transcritas quedan dentro del núcleo). El número de exones es variable para los diferentes genes; el de las globinas presenta 3 exones, las cadenas pesadas de las inmunoglobulinas 5, la ovoalbúmina 8, la conalbúmina 18 y la cadena  $\alpha$  del colágeno 53 (tabla 26.2).

**Tabla 26.2. Características estructurales de algunos genes humanos**

Proteína	Residuos	Cromosomas	Gen(pb)	Exones
$\alpha$ -globina	141	16	850	3
$\beta$ -globina	146	11	1 600	3
Colágeno I $\alpha$ 1	1 000	17	18 000	51
Colágeno I $\alpha$ 2	1 000	7	38 000	52
Insulina	51	11	1 430	3
Factor IX	415	X	34 000	8
Factor VIII	2 332	X	186 000	26
Receptor de LDL	839	19	45 000	18
$\alpha$ 1-anti tripsina	394	14	10 000	5

La secuencia de bases así como la longitud de los intrones no parecen presentar regularidad alguna, pero suelen ser mayores que los exones. La única regularidad que ha sido observada es una pequeña secuencia que se encuentra al inicio y al final de cada intrón, pues todos los conocidos comienzan por GT y terminan con AG; esto puede comprobarse en la figura 26.6, donde se muestran los resultados de un estudio de la secuencia de las uniones entre intrones y exones en más de 100 genes.

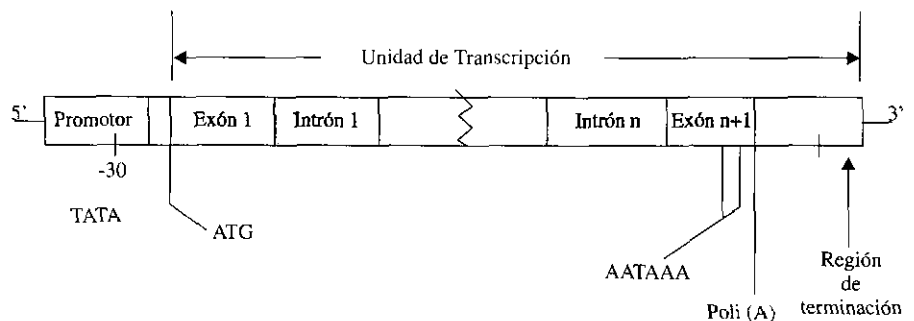


**Fig. 26.6. Características de los intrones.** La figura representa la secuencia de bases nitrogenadas alrededor de las uniones exón-intrón y la intrón-exón, así como demuestra la constancia del par GT en el inicio y del par AG en el final.

Todo parece indicar que la fragmentación de los genes no es al azar. Son muchos los ejemplos que demuestran que cada exón codifica para una zona específica de la proteína con características estructurales y funcionales muy definidas, o sea, un dominio. Esto ha hecho pensar que los intrones pudieran haber tenido una importante función en la evolución, que favorece el intercambio de segmentos génicos al formar nuevos arreglos que darían lugar a proteínas diferentes, a partir de bloques de otras proteínas. La presencia de los intrones hace más flexible al genoma de los eucariontes, lo cual aceleraría el proceso evolutivo. Por otra parte, estos exones pueden arreglarse de forma diferente, al ser transcritos en su ARNm o durante el procesamiento de estos últimos, dando lugar a la formación de proteínas diferentes a partir del mismo segmento de ADN. Una vez más se trata de conceptos relativos, pues en muchos casos los intrones de un gen representan otros genes. Así en el gen NFI, cuya alteración produce una enfermedad llamada neurofibromatosis 1, presenta un intrón de gran tamaño en el cual están ubicados otros 3 genes.

Por último, es bueno señalar que no toda la secuencia de bases que aparece en los exones se traduce en secuencias específicas de aminoácidos. Hacia el extremo 5' aparece una secuencia de bases, cuya función está relacionada con la iniciación de la traducción y facilita la incorporación del ARNm al ribosoma. Otra secuencia aparece hacia el extremo 3' y está relacionada con el mecanismo de terminación. En esta zona existe una señal que indica el lugar donde debe producirse la modificación del extremo 3' del ARNm, conocida como sitio de poliadenilación y que se estudiará con más detalles en el capítulo siguiente (Fig. 26.7).

Fig. 26.7. Estructura generalizada de un gen eucarionte. Si leemos el gen de izquierda a derecha, lo primero que encontramos es la región del promotor, con sus secuencias reguladoras especialmente TATA en la zona de -30. El sitio para el Cap corresponde al primer nucleótido que será transcrito y ocupará el extremo 5' -P del ARNm. El triplete ATG señala el sitio donde debe comenzar la traducción. Luego exones e intrones se alternan de acuerdo con el gen específico de que se trate. Después del último exón aparece una zona que contiene la secuencia AATAAA, que indica que unas 20 bases más adelante está el sitio donde debe incorporarse la cola de poli A. El punto exacto de la terminación no ha sido totalmente esclarecido.



Por su parte, los procariontes tienen generalmente su ADN formando una sola cadena, cuya asociación con proteínas no resulta tan regular como en los eucariontes. Casi siempre contienen sus genes en copia única y éstos no presentan secuencias intercaladas. Cuando en algunos casos los genes pueden ser mayores que su producto, las adiciones se encuentran hacia los extremos, de manera que la secuencia codificadora es casi siempre continua.

## Familias génicas

En procariontes, los genes cuyos productos están relacionados funcionalmente se organizan en operones (capítulo 34) y son transcritos en un ARNm policistrónico, que contiene información para más de una cadena polipeptídica. Todo el sistema se controla por medio de una secuencia específica de bases -el operador- y puede activarse o desactivarse regulando el funcionamiento de éste.

Esto no sucede en eucariontes. Muchos genes eucariontes funcionalmente relacionados pueden formar grupos (*cluster*) denominados familias génicas. La actividad de los miembros de la familia es habitualmente (pero no siempre) regulada de forma coordinada.



Las familias génicas se clasifican habitualmente en 3 grupos:

1. Familias multigénicas simples.
2. Familias multigénicas complejas.
3. Familias multigénicas complejas controladas por el desarrollo.

**1. Familias multigénicas simples.** Uno o varios genes están repetidos en tándem. Los más notables son los genes de ARNr en el xenopus. El ADNr de 5 S se organiza de forma que cada gen está separado del siguiente por una secuencia espaciadora (no codificante), con una longitud de 2 a 6 veces la del gen (Fig. 26.8 (a)). Los genes de los ARNr de 5,8 S; 18 S y 28 S en seres humanos constituyen otra familia multigénica simple, en la cual la unidad de transcripción contiene los 3 genes separados por secuencias espaciadoras, como se muestra en la figura 26.8(b).

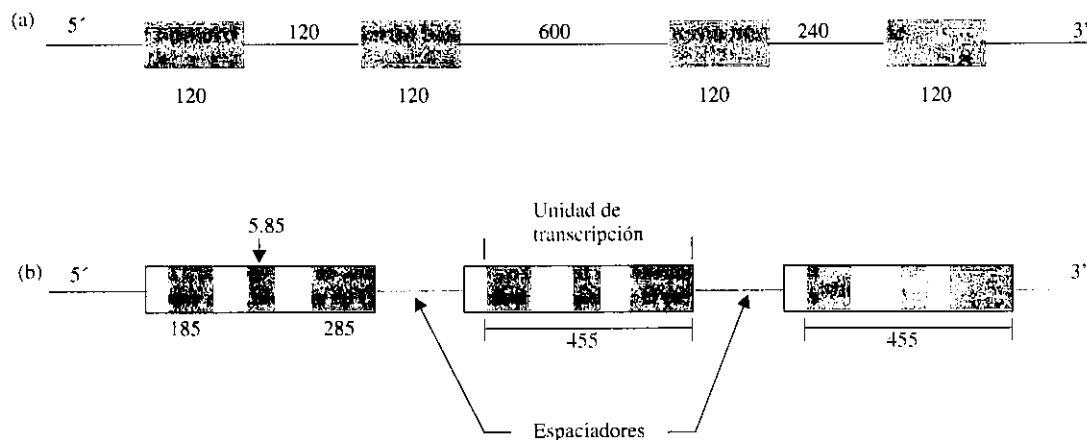


Fig. 26.8. Familias multigénicas simples. Uno o varios genes se encuentran organizados de forma lineal y separados por secuencias espaciadoras. (a) Se muestra la organización de la familia de los genes para el ARNr de 5 S en *X. laevis*. (b) La familia del gen para el precursor de 45 S de los ARNr de 18; 5,8 y 28 S en el ser humano.

**2. Familias multigénicas complejas.** Están formadas por grupos de genes relacionados, cada uno con transcripción independiente, y separados por secuencias espaciadoras. El ejemplo más sobresaliente es la familia de las histonas. En el erizo de mar consta de 5 genes en un pequeño fragmento (7 kb) de ADN separado por espaciadores; esta unidad pentagénica aparece repetida en tándem unas 1 000 veces. Cada gen es transcrito en forma independiente en un ARNm monocistrónico. Aunque se desconoce el mecanismo existe un control estricto de la velocidad de transcripción y traducción de estos genes, de forma tal que las histonas se sintetizan en el momento necesario para la síntesis de la cromatina (período S) y en la relación molecular necesaria, o sea, por cada molécula de H1 se producen 2 de cada una de las restantes (H2A, H2B, H3 y H4), que es la proporción en que entran a formar parte de la cromatina (capítulo 23). La existencia de múltiples genes permite disponer de una gran cantidad de estas proteínas en el momento preciso. En otros organismos la organización de esta familia difiere ligeramente, como puede observarse en la figura 26.9.

**3. Familias multigénicas complejas controladas por el desarrollo.** Se trata de grupos de genes que se activan e inactivan en diferentes etapas del desarrollo

ontogénico del organismo. El ejemplo mejor conocido es el de los genes de la hemoglobina. Esta proteína es un tetrámero formado por 2 subunidades de tipo  $\alpha$  y 2 de tipo  $\beta$ . Hay varias formas de estas subunidades que se diferencian en unos pocos aminoácidos y la forma presente depende de la etapa del desarrollo del organismo.

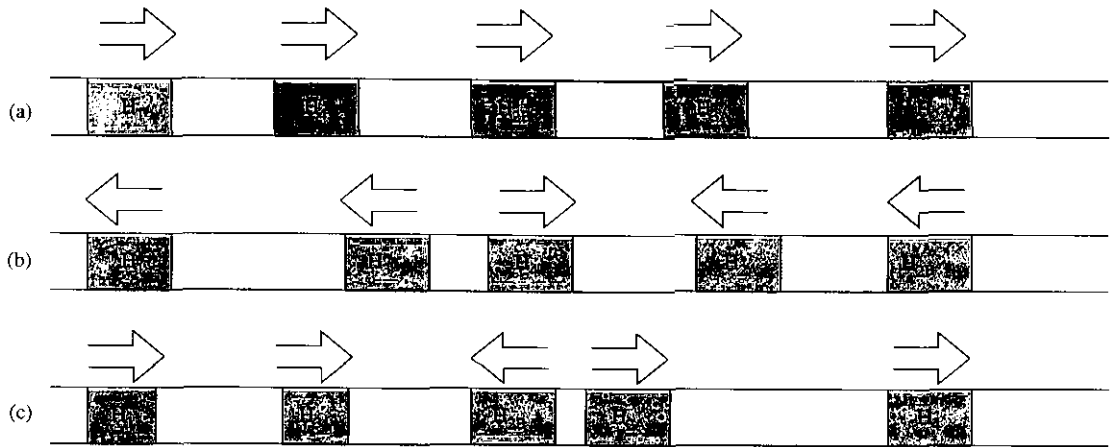
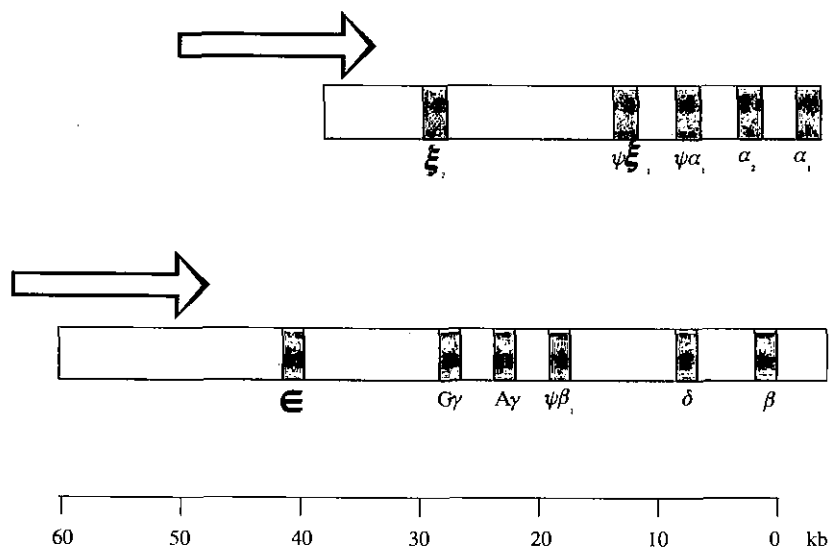


Fig. 26.9. Familias multigénicas complejas. El grupo está formado por genes que se transcriben de forma independiente. La figura representa la familia de los genes para las histonas en (a) el erizo de mar, (b) la *Drosophila melanogaster* y (c) en el tritón, una especie de salamandra marina. La flecha indica el sentido de la transcripción.

En el caso de las subunidades de tipo  $\beta$ , existen 4 variantes ( $\epsilon$ ,  $\gamma$ ,  $\delta$  y  $\beta$ ). Durante el período embrionario (menos de 8 semanas después de la fecundación) sólo se produce la cadena  $\epsilon$ , de la semana 8 a la 41 se sustituye por la  $\gamma$  y a partir del nacimiento por la  $\beta$  y una pequeña fracción de la  $\delta$ ; igual sucede con las subunidades de tipo  $\alpha$ . La familia  $\alpha$  está localizada en el cromosoma 16 y la  $\beta$  en el 11 y curiosamente están organizadas en el mismo orden en que se activan (Fig. 26.10).

Fig. 26.10. Familias multigénicas complejas controladas por el desarrollo. Este grupo está compuesto por genes que se activan e inactivan de acuerdo con el momento del desarrollo. (a) La familia génica de la  $\alpha$ -globina en el cromosoma 16,  $\xi$  es una cadena embrionaria, la  $\alpha_1$  y  $\alpha_2$  se producen desde la vida fetal hasta la adultez.  $\psi\xi$  y  $\psi\alpha_1$  representan pseudogenes. (b) La familia génica de la  $\beta$ -globina localizada en el cromosoma 11;  $\epsilon$  es una cadena embrionaria,  $\gamma$  y  $\delta$  son fetales y las  $\beta$  y  $\delta$  comienzan en la vida fetal y se mantienen hasta el estado adulto.  $\psi\beta_1$  representa un pseudogen. Como se indica en las flechas, estos genes se organizan en el mismo orden en que son activados, lo cual no siempre sucede.



## Genoma humano

El estudio del genoma humano es uno de los campos más importantes y más complejos de la biología celular contemporánea. El tamaño y la organización cromosómica de éste son similares al de otros animales, pero los métodos empleados para su estudio resultan más complicados. El desarrollo de la tecnología del ADN recombinante ha permitido encontrar procedimientos directos para continuar, profundizar y rectificar resultados encontrados en décadas anteriores por estudios indirectos, a los cuales mucho contribuyó el estudio de la correlación entre las enfermedades de origen genético y los hallazgos bioquímicos y citogenéticos de los pacientes. A manera de conclusión de este capítulo se mencionarán las características más notables del genoma de los seres humanos.

El genoma humano está constituido por moléculas de ADN organizadas en 22 pares de cromosomas autosómicos, el par de cromosomas sexuales (XY) y el mitocondrial (también llamado cromosoma M ó 25). Sus dimensiones pueden expresarse en términos de número total de pares de bases o de genes. El genoma nuclear haploide humano contiene  $3.3 \times 10^9$  pb y el mitocondrial 16 569 pb.

Probablemente el hombre tiene entre 50 000 y 100 000 genes estructurales, que codifican la secuencia aminoacídica de proteínas o la nucleotídica de ARN ribosomales y de transferencia. A este estimado se ha llegado por medio de numerosos procedimientos y resulta apoyado por el número de genes localizados en un segmento de ADN de longitud conocida, que ha sido estudiado en gran detalle. No debe confundirse el número de genes con el de proteínas, pues como se podrá ver en capítulos posteriores un mismo gen puede dar lugar a más de una proteína y, además, existen otros mecanismos que producen la formación de un gran número de proteínas a partir de un segmento génico limitado.

El cromosoma mitocondrial presenta una mayor densidad de genes (por ausencia de intrones y espaciadores mucho más cortos) y presenta 13 genes para proteínas, 22 para ARNt y 2 para ARNr (12S y 16S).

Se han localizado unos 800 loci, cerca de 120 en el cromosoma X y los restantes en los autosómicos. Cada uno contiene como mínimo 1 200 pares de bases en información codificada, pero son mucho mayores debido a los intrones y las secuencias que los flanquean, tanto hacia el extremo 5' como hacia el 3'. Hacia la zona adyacente al extremo 5' se encuentran las regiones reguladoras, y hacia la 3' la secuencia AATAAA conocida como sitio de poliadenilación.

Entre uno y otro gen se hallan secuencias espaciadoras de varias kb cuyas funciones, si las tienen, son desconocidas. Éstas consisten en secuencias repetidas, de las cuales la más frecuente es la llamada familia Alu, pues contiene la cuarteta AGCT/TCGA que es reconocida por la enzima de restricción Alu I (de *Arthrobacter luteus*); estas secuencias repetidas están presentes cientos de miles de veces en el genoma humano, especialmente en los espaciadores, pero en ocasiones en los intrones.

De todos los estudios realizados hasta el presente pueden derivarse algunas características generales en cuanto a la organización de los genes humanos en los cromosomas y que pueden resumirse en las siguientes:

1. En general los genes cuyos productos están relacionados evolutiva o funcionalmente aparecen agrupados en los cromosomas, sin embargo no existen agrupaciones génicas para estructuras y funciones de órganos particulares, como oídos, corazón, pulmón u organelos subcelulares como lisosomas, mitocondrias, etcétera.
2. Los genes que codifican las enzimas que catalizan reacciones sucesivas en una vía metabólica particular no se encuentran por lo general agrupados en el mismo cromosoma. Así, los genes para las enzimas galactoquinasa, galactosa-1-fosfato uridil-transferasa y la galactosa-4-epimerasa, que propician la entrada de la galactosa

a la vía glucolítica están localizados en los cromosomas 17, 9 y 1, respectivamente. Igual sucede con las enzimas del ciclo de la urea, pues la carbamilfosfato sintetasa I, ornitina transcarbamilasa, argininosuccínico sintasa, argininosuccínico liasa y la arginasa se encuentran codificadas en los cromosomas 2, X, 9, 7 y 6, respectivamente. Un hecho curioso es que la triosa fosfato isomerasa, gliceraldehído-3-fosfato deshidrogenasa, enolasa y lactato deshidrogenasa, relacionadas todas con la vía glucolítica, presentan sus genes agrupados en el cromosoma 12.

3. Los genes que codifican para las subunidades de proteínas heteroméricas se hallan usualmente en cromosomas diferentes; mientras el gen para la subunidad  $\alpha$  de la hemoglobina está en el cromosoma 16, la subunidad  $\beta$  está codificada en el 11. Lo mismo sucede con las formas isoenzimáticas de la lactato deshidrogenasa cuya subunidad M se codifica en el cromosoma 11 y la H en el 12. Una excepción notable es el fibrinógeno, cuyas 3 subunidades se codifican por genes localizados en el cromosoma 4.
4. En contraste con lo anterior, algunos genes codifican más de una cadena polipeptídica; esto se debe fundamentalmente al procesamiento posterior del ARNm; así sucede con la insulina cuyas 2 cadenas son codificadas por un gen único ubicado en el cromosoma 11. Es también el caso del gen de la proopiomelanocortina localizado en el cromosoma 2 que da origen a 7 polipéptidos con funciones diferentes.
5. Los genes que codifican la forma citosólica y mitocondrial de la misma enzima se encuentran en cromosomas diferentes, como sucede con la transaminasa glutámico-oxalacética, cuya forma citosólica se codifica en el cromosoma 10 y la otra en el 16 y la isocitrato deshidrogenasa en el 2 y el 15, respectivamente.
6. Una característica sobresaliente es la existencia de los pseudogenes, o sea, secuencias de bases similares a la de genes funcionales, pero que no se expresan al parecer por haber perdido algún sector indispensable para su transcripción; estos pseudogenes ocupan una porción considerable del genoma. Los más conocidos son los de las cadenas  $\alpha$  y  $\beta$  de la hemoglobina en los cromosomas 16 y 11, respectivamente.

El desarrollo que han alcanzado las diferentes técnicas, para el estudio de los ácidos nucleicos, hace tener esperanzas de que en los próximos años se conocerán muchos más elementos acerca de la estructura y la organización del genoma humano. A este objetivo debe contribuir de forma sobresaliente el Proyecto del Genoma Humano, que pretende establecer la secuencia de bases de todas las moléculas de ADN contenidas en los cromosomas humanos y realizar estudios comparativos con secuencias de otros organismos, para determinar la función de cada uno de los segmentos de ADN. Esta investigación es trascendental para la biología contemporánea y debe estar terminada para los primeros años del próximo siglo.

## Resumen

**El material genético de todos los organismos celulares está constituido por el ADN, que puede existir en una sola molécula como en los procariontes, o en varias como en los eucariontes, en los cuales, además, está separado del resto de la célula por una doble membrana. El ADN de eucariontes se encuentra asociado con proteínas, formando el par cromatina-cromosoma. El número y la forma de los cromosomas son característicos de cada especie. En los cromosomas se hallan secuencias de bases nitrogenadas que contienen la información necesaria para la síntesis de las proteínas, los ARNr y los ARNt: éstos son los genes. El conjunto de todos los genes en un organismo define el genoma. Un gen puede tener muchas formas alélicas pero una célula sólo puede tener 2. Si son iguales, la célula es homocigótica, y heterocigótica si son desiguales. La pareja de genes que determina**

un carácter forman el genotipo y su manifestación externa es el fenotipo. Un ejemplo característico de la relación genotipo-fenotipo en seres humanos lo constituyen los grupos sanguíneos ABO.

En la célula eucarionte existen distintos tipos de secuencias nucleotídicas en el ADN: las altamente repetidas, mediana y moderadamente repetidas y las secuencias únicas. Los genes casi siempre se encuentran en secuencias de copia única, aunque existen notables excepciones como los de las histonas y los ARNr. Las secuencias altamente repetidas se localizan en los centrómeros y con menor frecuencia en los telómeros; su función es desconocida. Los genes eucariontes son discontinuos, pues presentan secuencias intercaladas no codificantes (intrones), separando las codificantes (exones). Los genes cuentan además con sus promotores y con sus secuencias funcionales -aunque no codificantes- tanto hacia el extremo 5'-P como hacia el 3'-OH.

Muchos genes relacionados funcionalmente se agrupan, formando familias génicas que pueden ser simples, complejas o reguladas por el desarrollo ontogenético del organismo.

El genoma humano constituye un caso especial dentro de los organismos eucariontes por la importancia extraordinaria que tiene su conocimiento profundo. Se han localizado más de 800 genes en los cromosomas humanos y a partir de numerosos estudios se han podido precisar algunas de sus características más sobresalientes: los genes que codifican proteínas específicas de un órgano o tejido no están agrupados en el mismo cromosoma, así como tampoco los relacionados con una vía metabólica; las subunidades de proteínas heteroméricas se codifican por genes localizados en cromosomas diferentes; algunos genes codifican más de una cadena polipeptídica; también se encuentran separados los genes que codifican formas diferentes de la misma enzima; por último, la presencia de pseudogenes en una proporción notable del genoma es un hecho apreciable en el genoma humano.

Estos conocimientos tienen no sólo una importancia teórica, sino también práctica sobre todo en el campo de las ciencias médicas.

## Ejercicios

1. Un gen A tiene un alelo A'. Si el genotipo femenino es AA y el masculino A'A' ¿cuáles serán los posibles genotipos de los descendientes?
2. ¿Cuál sería el resultado de la pregunta anterior si los genotipos masculino y femenino hubieran sido AA'?
3. Se ha calculado que en el ADN de la *E. coli* existen 1 500 genes. Si cada uno codifica una proteína de 500 aminoácidos y el ADN total es de  $4 \times 10^6$  pb. ¿Cuál será la proporción codificante de ese ADN?
4. ¿Cuáles son los tipos de secuencias de ADN que encontramos en las células eucariontes y qué relación tienen con las funciones del ADN?
5. A un cultivo de células se le añade [<sup>3</sup>H]-uridina durante 10 minutos. Las células se lavan para eliminar el exceso de reactivo. Se sigue entonces la marca radiactiva por los compartimentos celulares. En los primeros minutos después de la exposición toda la radiactividad se localiza en el núcleo de todas las células. A medida que pasa el tiempo se observa que las células se comportan de forma diferente, mientras en algunas de ellas la radiactividad permanecía en el núcleo, en otras se localizaba en el citoplasma ¿Cómo pueden explicarse estos resultados?
6. Se sabe que la ARN polimerasa tarda aproximadamente 5 minutos en sintetizar el ARNr de 28 S. Calcule cuántas moléculas pueden formarse en 8 horas si:
  - a) Existe un sólo gen para el ARNr y se transcribe con una sola polimerasa cada vez.
  - b) Existen 100 copias del gen y cada uno es transcrito por una sola molécula de la enzima.
  - c) Existen 100 copias del gen y cada uno es transcrito por 100 polimerasas simultáneamente.