

EL DOGMA CENTRAL DE LA BIOLOGÍA MOLECULAR

Pablo J. Patiño¹, José Robinson Ramírez Pineda²

¹ Profesor Asociado, Grupo de Inmunodeficiencias Primarias – Corporación Biogénesis

² Profesor Asociado, PECET

Facultad de Medicina, Universidad de Antioquia

Uno de los principales interrogantes que pretende responder la biología molecular tiene que ver con los mecanismos responsables de la transmisión y expresión de la información genética en las células, pues esto es un factor fundamental para determinar la estructura y función de todo un organismo. La expresión de esta información genética en todas las células tiene lugar, la gran mayoría de las veces, como un sistema de una sola dirección: el ADN determina la síntesis de RNA y el RNA especifica la síntesis de polipéptidos, los que a su vez forman las proteínas. Puesto que este sistema es prácticamente universal en los sistemas biológicos ha sido denominado “**dogma central de la biología molecular**” (Figura 1). La primera fase de este proceso consiste en la síntesis de RNA usando una enzima que utiliza el ADN como molde para producir el polímero de RNA, fenómeno conocido como transcripción. La etapa siguiente, denominada traducción, consiste en la síntesis del polipéptido a partir de la molécula de RNA, evento que tiene lugar en los ribosomas. La molécula de RNA que contiene esta información se conoce como RNA mensajero (RNAm). Finalmente, en muchos casos los polipéptidos deben acoplarse unos con otros para dar origen a las proteínas que cumplirán un papel funcional o estructural dentro de la misma célula o que actuarán en un sitio distante del lugar de producción. Pero todo este proceso de generación de información no puede ocurrir sin que exista un proceso de copiado (replicación) del ADN que asegure la multiplicación de las

células y del organismo completo. De acuerdo a lo anterior, la expresión de la información genética sigue un principio de colinearidad: la secuencia lineal de nucleótidos que constituye el ADN es decodificada para generar una secuencia lineal de nucleótidos que forman el RNA, el cual a su vez puede ser decodificado, mediante la lectura en grupos de tres nucleótidos (codones), para producir una secuencia lineal de aminoácidos que conforman el polipéptido final.

Sin embargo, este dogma central de la biología molecular no se cumple siempre de manera absoluta, pues muchos virus, en los cuales su información genética está codificada en forma de RNA. En este caso, los procesos de replicación del genoma y expresión de los genes, se inicia a partir de la molécula de RNA. Además, en las células eucarióticas la transcripción no ocurre solamente en el núcleo, pues organelas como las mitocondrias y los cloroplastos poseen ADN que tiene la capacidad de replicar y expresar información necesaria para la función de las células.

Aunque en las células de los organismos que conocemos hoy en día el ADN es la molécula encargada de transmitir la información genética (herencia), las evidencias que se tienen indican que el RNA fue el responsable de esa función durante las primeras etapas de la evolución de los sistemas biológicos. Como se discutirá más adelante, exis-

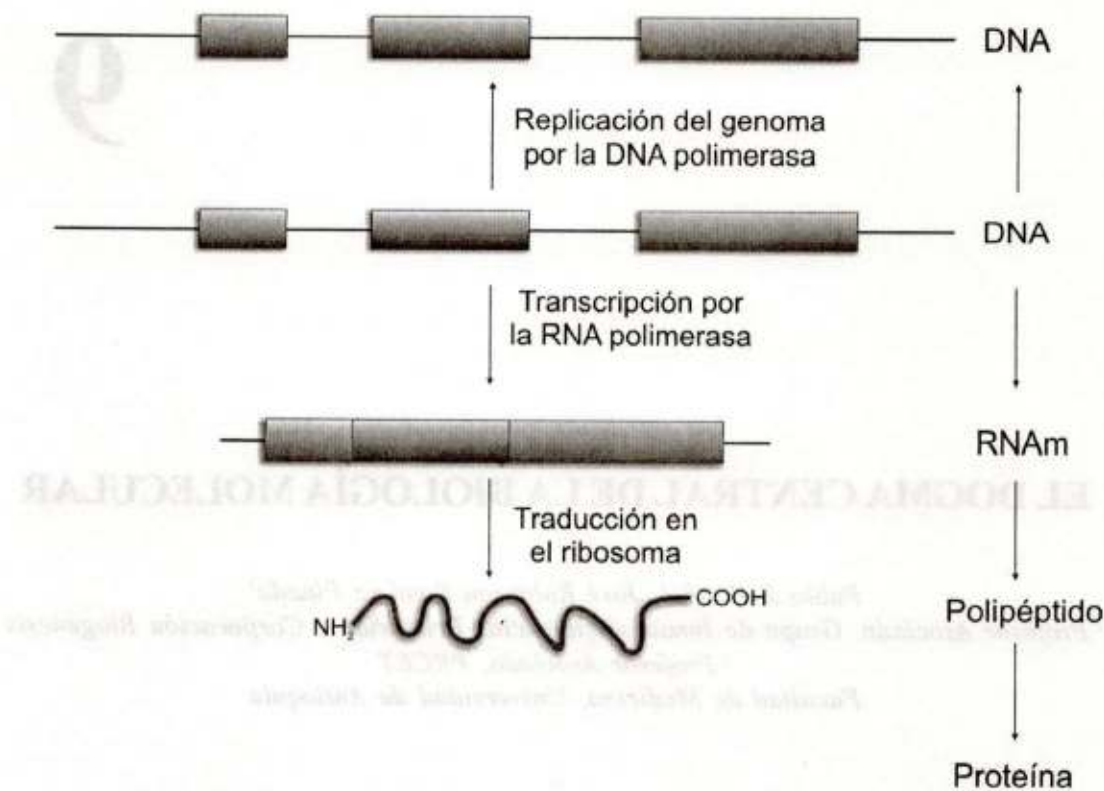


Figura 1. El dogma central de la biología molecular establece que la expresión de la información genética inicia mediante la síntesis de RNA a partir de la molécula de ADN (transcripción). Luego, la molécula de RNA (RNAm) es leída en los ribosomas para permitir la síntesis de polipéptidos (traducción). A su vez, los polipéptidos deben acoplarse entre sí para dar origen a las proteínas. De manera paralela existe un proceso de copiado del ADN (replicación) que asegure la multiplicación de las células y del organismo completo. Se debe tener en cuenta que no todos los RNA que se transcriben dan origen a proteínas, pues éstos pueden tener funciones diferentes (RNAt, RNAr).

ten varias moléculas de RNA que tiene capacidad catalítica (ribozimas), lo cual indica que moléculas primitivas de este ácido nucleico podrían ser autorreplicativas. Posiblemente el ADN reemplazó posteriormente al RNA como transportador de la información genética debido a que la ausencia del grupo hidroxilo en el carbono 2' de la desoxirribosa permite que el ADN sea una molécula con mucha menor reactividad química, lo cual es esencial para mantener la fidelidad de la herencia. Sin embargo, como se mencionó previamente, el genoma de algunos virus está constituido por RNA. Por ejemplo, en los retrovirus el RNA que constituye su genoma se replica en las células infectadas utilizando al ADN como molécula intermediaria, para lo cual utiliza la enzima transcriptasa reversa, una ADN polimerasa que utiliza el RNA como molde.

En este capítulo y en los siguientes se hará una descripción de las bases moleculares de los genomas así como de los mecanismos que nos permiten comprender los fenómenos de replicación y reparación del ADN, la transcripción del ADN a RNA y la traducción final de la información codificada en el RNA a las proteínas, las moléculas que en última instancia determinan la estructura y función de todos los sistemas biológicos conocidos.

GENOMA, ESTRUCTURA Y FUNCION

El material genético

Una de las propiedades fundamentales de todos los seres vivos es la posibilidad de reproducirse. Todos los organismos heredan de los progenitores la información genética que determina su estructura y función, lo que significa que los organismos, sean uni o multicelulares, provienen de células preexistentes, para lo cual el material genético debe replicarse y pasar a cada célula hija durante la división celular.

Como se mencionó en el capítulo 3, las bases para la comprensión de la transmisión de la información genética fueron dadas por Gregor Mandel en la segunda mitad del siglo XIX; gracias a sus trabajos con cruces entre plantas con características fenotípicas diferentes. Mendel realizó una observación cuidadosa y metódica de los rasgos que presentaban generaciones sucesivas de estos organismos, lo que le permitió deducir que cada característica era determinada por un par de factores heredados, cada uno proveniente de un progenitor. Por ejemplo, el color de los guisantes estaba definido por factores hereditarios,

dentro de los cuales había algunos que eran dominantes sobre otros. Estos factores hereditarios fueron los que posteriormente recibieron la denominación de genes.

Una vez se determinó que las características de un individuo podían ser transmitidas genéticamente, se inició la búsqueda del factor o factores responsable de perpetuar esta información. Este material genético o genoma debería tener varias características:

- Capacidad de replicarse de manera fiel, generación tras generación.
- Capacidad de codificar una secuencia de proteínas. Para esa época, primera mitad del siglo XX, ya se sabía que las proteínas determinaban la estructura y función de las células.
- Capacidad de modificar su expresión ante las condiciones cambiantes del ambiente.

Solo hasta principios del siglo XX, cuando fue posible visualizar el material que se encontraba en el interior del núcleo de las células, se empezaron a conocer las bases estructurales que explicaban las observaciones de Mendel. Gracias al desarrollo de microscopios con cada vez mayor resolución y al descubrimiento de colorantes que teñían selectivamente diferentes componentes celulares, fue posible identificar las estructuras nucleares denominadas cromosomas, que se caracterizaban por presentar diferentes formas y tamaños durante la metafase en la división celular.

Se evidenció que el número y estructura de los cromosomas variaba de un organismo a otro y que incluso en un mismo organismo podían existir diferencias. En la gran mayoría de las células nucleadas de los organismos multicelulares, se encontraban dos copias de cada uno de los cromosomas, por lo que estas células se denominaron diploides. Por su parte, en las células sexuales o

germinales de estos organismos se encontró exactamente la mitad del número de cromosomas identificado en las células somáticas de cada organismo, por lo que se denominaron células haploides. Por su parte los procariotes no evidenciaban una estructura definida, lo que más adelante se explicó porque generalmente tienen un cromosoma único de estructura circular.

Estos hallazgos permitieron redescubrir los resultados de Mendel y proponer que los factores descritos por él estaban en los cromosomas. En 1905 se comprobó inequívocamente que los cromosomas eran los factores hereditarios de Mendel al observarse que todas las células somáticas, tomadas de mujeres, contenían dos cromosomas sexuales idénticos (llamados X), mientras las células somáticas masculinas siempre contenían dos cromosomas sexuales diferentes (uno X y otro llamado Y), lo que llevó concluir que el sexo de un organismo estaba relacionado directamente con el tipo de cromosomas presentes en las células de dicho organismo. De esta manera el sexo fue el primer fenotipo (característica física) que se asignó a una ubicación cromosómica determinada.

El ADN es el material genético

Como se describió en el Capítulo 3, el ADN fue identificado inicialmente por el biólogo suizo Frederick Miescher en 1868. La sustancia, que denominó nucleína, la separó en una porción básica (que ahora sabemos es el ADN) y una porción ácida (proteínas ácidas que se unen al ADN). Luego por estudios químicos se demostró que el ADN era un polímero de unidades, que se denominaron nucleótidos, cada una de las cuales estaba constituida por un azúcar, una base nitrogenada y un grupo fosfato (Figura 2), de las que existían cuatro diferentes gracias a la diferencia en la base (adenina, guanina, citosina, timina).

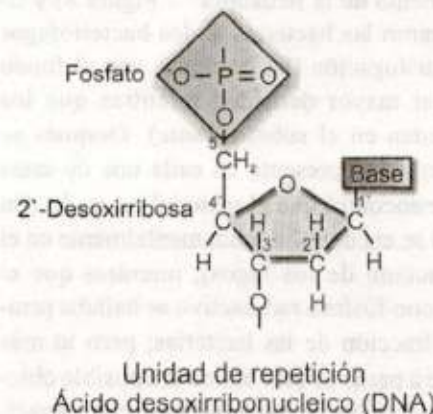
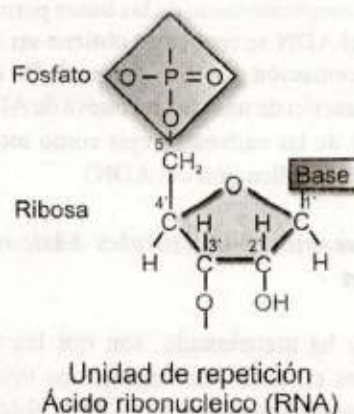


Figura 2. Los nucleótidos son las unidades básicas (monómeros) que constituyen los ácidos nucleicos (ADN y RNA). Cada uno de estos nucleótidos está constituida por un azúcar, una base nitrogenada y un grupo fosfato. La diferencia fundamental entre los nucleótidos de ADN y de RNA reside en la presencia de un grupo hidroxilo en la posición 2' de la pentosa que constituye el RNA.

Análisis químicos de los cromosomas han demostrado que cerca del 60% de su constitución corresponde a proteínas, mientras el 40% restante es ADN. Esta proporción parecía indicar en un principio que las proteínas eran las que contenían la información hereditaria. Además las proteínas están compuestas de 20 subunidades diferentes, mientras que el ADN está compuesto de solo cuatro, indicando que una molécula de proteína codificaría mucha más información. El concepto de que las proteínas eran las responsables de codificar la información genética perduró durante varias décadas, hasta que en 1944 Avery, Mcleod y McCarty publicaron experimentos demostrando la capacidad que tenía el ADN para transformar las características fenotípicas de un organismo. Sin embargo, la mayoría de los científicos aún consideraba que las proteínas eran el material genético y no dieron mucha credibilidad a los resultados.

Por medio del experimento realizado por Alfred Hershey y Martha Chase, publicado en 1952, se demostró definitivamente que el ADN era el material genético. Dicho experimento se basó en la capacidad que tienen los bacteriófagos (virus bacterianos) para infectar las bacterias. Inicialmente se cultivaron estos virus por separado en dos medios de cultivo diferentes, uno que contenían azufre radioactivo y otro con fósforo radioactivo. El azufre radioactivo se incorpora específicamente en las proteínas gracias a que el azufre hace parte de los aminoácidos metionina y cisteína. Por su lado, el fósforo marcado con radioactividad se incorpora únicamente en el ADN al hacer parte del esqueleto de fosfatos de este ácido nucleico. De esta manera resultan unos bacteriófagos que están marcados sólo en las proteínas y otros marcados sólo en el ADN. Luego se permitió que los bacteriófagos marcados infectaran a las bacterias; después de un tiempo, las mezclas se sometieron a una agitación fuerte en una licuadora ("experimento de la licuadora" – Figura 3) y finalmente se separaron las bacterias de los bacteriófagos por medio de centrifugación (las bacterias van al fondo del tubo por tener mayor densidad mientras que los bacteriófagos quedan en el sobrenadante). Después se analizó la radioactividad presente en cada una de estas dos fracciones. Se encontró que el material marcado con azufre radioactivo se encontraba fundamentalmente en el sobrenadante (fracción de los fagos), mientras que el material marcado con fósforo radioactivo se hallaba principalmente en la fracción de las bacterias; pero lo más importante era que a partir de esta última era posible obtener una nueva generación de bacteriófagos con capacidad infecciosa. Estos resultados fueron contundentes para demostrar que el material genético correspondía al ADN.

Estructura del ADN – La doble hélice

Como se mencionó antes, para inicios del siglo XX se aceptaba el concepto de los genes como unidades de transmisión de las características genéticas y su relación con los cromosomas, la cuestión fundamental para los años subsiguientes consistió en conocer cual era su naturaleza química y su estructura. Aunque los experimentos descritos previamente permitieron demostrar que el ADN era la sustancia que mediaba la herencia, fue solo hasta 1953 que los resultados de muchos investigadores llevaron a proponer el modelo estructural que sigue vigente en nuestros días y que representó el inicio de una era en la que era posible la manipulación genética.

Un paso fundamental para conocer la estructura y los mecanismos que explicaban la función del ADN fue la demostración hecha por Erwin Chargaff de que independientemente del tejido que analizara de un mismo animal, el porcentaje de los cuatro nucleótidos que constituían esta molécula era siempre el mismo, aunque los porcentajes variaban de especie a especie. Además este investigador encontró que en todas las células provenientes de animales diferentes el porcentaje del contenido de guanina siempre era igual al de citosina y el de adenina era igual al de timina.

Los resultados de Chargaff, así como los datos de cristalografía de rayos X que Rosalind Franklin obtuvo de moléculas de ADN, permitieron que en 1953 James Watson y Francis Crick presentaran su modelo para explicar la estructura del ADN (Ver Capítulo 3 y Figura 4). De acuerdo a este modelo, el ADN es una molécula en forma de doble hélice conformada por dos cadenas complementarias que tienen direcciones opuestas, las cuales se unen entre sí gracias a los puentes de hidrógeno que se forman entre A y T y entre G y C (Figura 5). El apareamiento complementario de las bases permitió concluir que cuando el ADN se replica se obtiene un duplicado exacto de la información genética parental. De tal manera que la polimerización de una cadena nueva de ADN ocurre usando cada una de las cadenas viejas como molde (Ver el capítulo sobre Replicación del ADN).

Los elementos estructurales básicos de los ácidos nucleicos

Como se ha mencionado, son dos los ácidos nucleicos esenciales en todos los organismos vivos conocidos, el ácido desoxirribonucleico (ADN) y el ácido ribonucleico (RNA). Químicamente hablando, los ácidos nucleicos son polisacáridos fosforilados o fosfopolisacáridos, en forma

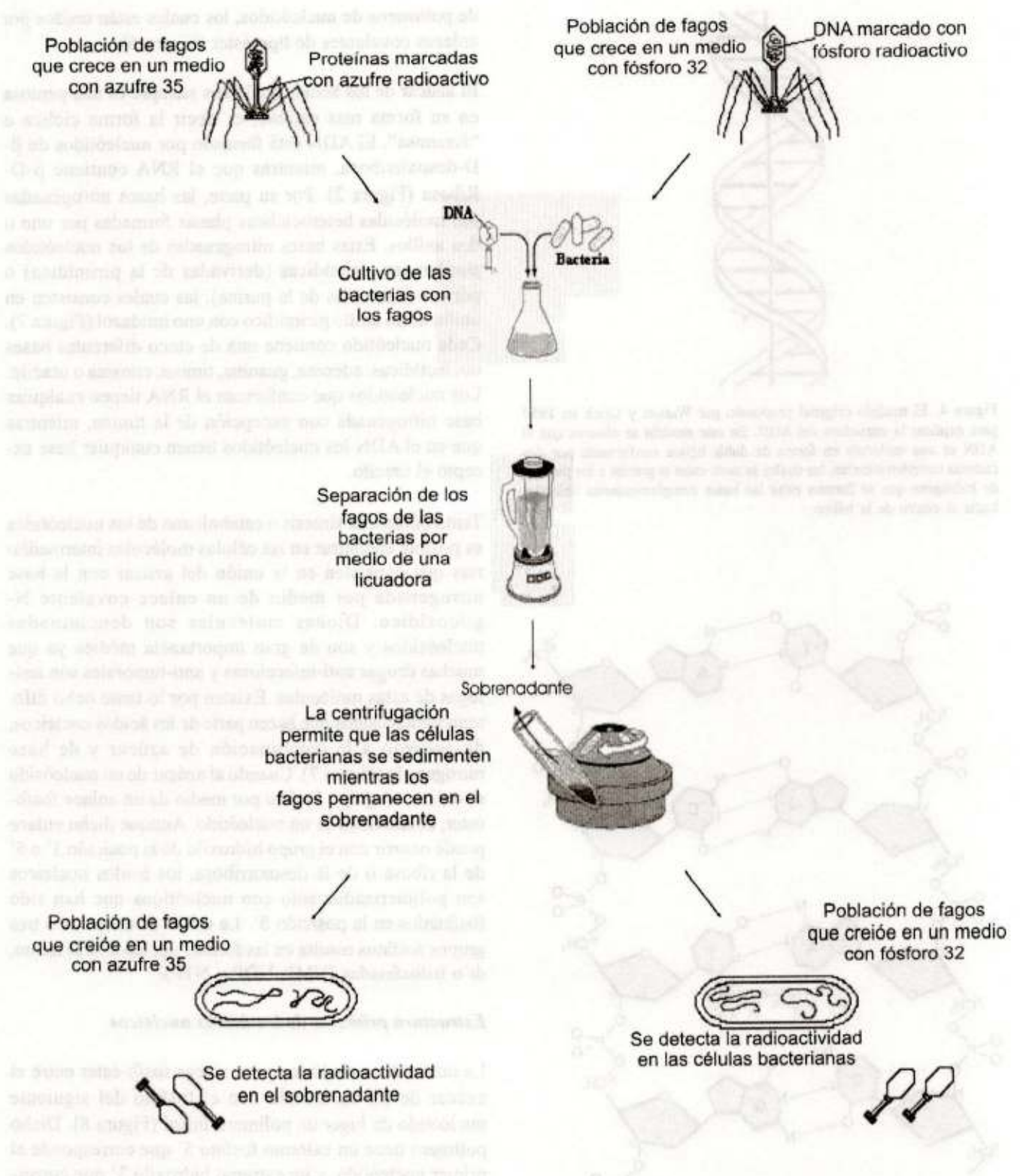


Figura 3. El experimento de Hershey y Chase demostró que el ADN era el material genético. Se cultivaron bacteriofagos por separado, unos en un medio de cultivo que contenía azufre radioactivo y otros en un medio con fósforo radioactivo, lo que permite producir bacteriofagos que están marcados en las proteínas y otros marcados en el ADN. Luego se permitió que los bacteriofagos marcados infectaran a las bacterias; después de un tiempo, las mezclas se sometieron a una agitación fuerte en una licuadora y finalmente se separaron las bacterias de los bacteriofagos por medio de centrifugación. Al medir la radioactividad se encontró que el material marcado con azufre radioactivo se encontraba fundamentalmente en el sobrenadante (fracción de los fagos), mientras que el material marcado con fósforo radioactivo se hallaba principalmente en la fracción de las bacterias.



Figura 4. El modelo original propuesto por Watson y Crick en 1953 para explicar la estructura del ADN. En este modelo se observa que el ADN es una molécula en forma de doble hélice conformada por dos cadenas complementarias, las cuales se unen entre sí gracias a los puentes de hidrógeno que se forman entre las bases complementarias ubicadas hacia el centro de la hélice.

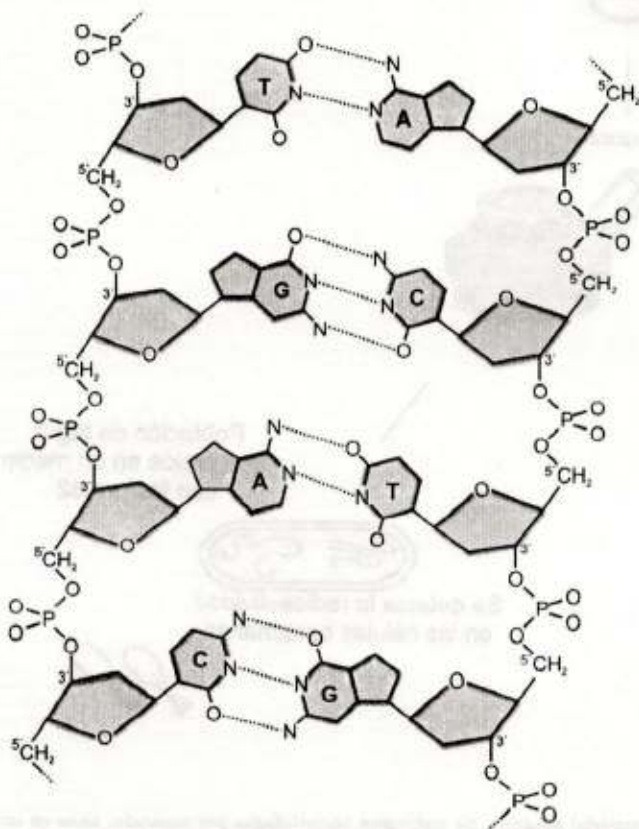


Figura 5. El ADN se forma gracias a la unión de dos cadenas complementarias que tienen direcciones opuestas. Esta complementariedad se obtiene mediante los puentes de hidrógeno que se forman entre Adenina y Timina y entre Guanina y Citosina.

de polímeros de nucleótidos, los cuales están unidos por enlaces covalentes de tipo éster (Figura 6).

El azúcar de los ácidos nucleicos siempre es una pentosa en su forma más estable, es decir la forma cíclica o "furanosa". El ADN está formado por nucleótidos de β -D-desoxirribose, mientras que el RNA contiene β -D-Ribosa (Figura 2). Por su parte, las bases nitrogenadas son moléculas heterocíclicas planas formadas por uno o dos anillos. Estas bases nitrogenadas de los nucleótidos pueden ser pirimídicas (derivadas de la pirimidina) o púricas (derivadas de la purina), las cuales consisten en unión de un anillo pirimídico con uno imidazol (Figura 7). Cada nucleótido contiene una de cinco diferentes bases nucleotídicas: adenina, guanina, timina, citosina o uracilo. Los nucleótidos que conforman el RNA tienen cualquier base nitrogenada con excepción de la timina, mientras que en el ADN los nucleótidos tienen cualquier base excepto el uracilo.

Tanto durante la síntesis o catabolismo de los nucleótidos es posible encontrar en las células moléculas intermedias que consisten en la unión del azúcar con la base nitrogenada por medio de un enlace covalente N-glicosídico. Dichas moléculas son denominadas nucleósidos y son de gran importancia médica ya que muchas drogas anti-infecciosas y anti-tumorales son análogas de estas moléculas. Existen por lo tanto ocho diferentes nucleósidos que hacen parte de los ácidos nucleicos, de acuerdo a la combinación de azúcar y de base nitrogenada (Figura 7). Cuando al azúcar de un nucleósido se enlaza un grupo fosfato por medio de un enlace fosfoéster, el resultado es un nucleótido. Aunque dicho enlace puede ocurrir con el grupo hidroxilo de la posición 3' o 5' de la ribosa o de la desoxirribose, los ácidos nucleicos son polimerizados solo con nucleótidos que han sido fosfatados en la posición 5'. La unión de uno, dos o tres grupos fosfatos resulta en las formas nucleosídicas mono, di o trifosfatadas (NMP, NDP y NTP).

Estructura primaria de los ácidos nucleicos

La unión de nucleótidos por enlaces fosfoéster entre el azúcar de un nucleótido con el fosfato del siguiente nucleótido da lugar un polímero lineal (Figura 8). Dicho polímero tiene un extremo fosfato 5' que corresponde al primer nucleótido, y un extremo hidroxilo 3' que corresponde al último nucleótido. La característica que diferencia un polímero de ADN de cualquier otro, es el orden en el que se encuentren sus nucleótidos, el cual es definido a su vez por las bases nitrogenadas que hacen parte de dichos nucleótidos. La molécula resultante tiene un "esqueleto" conformado por unidades alternantes de azúcar-

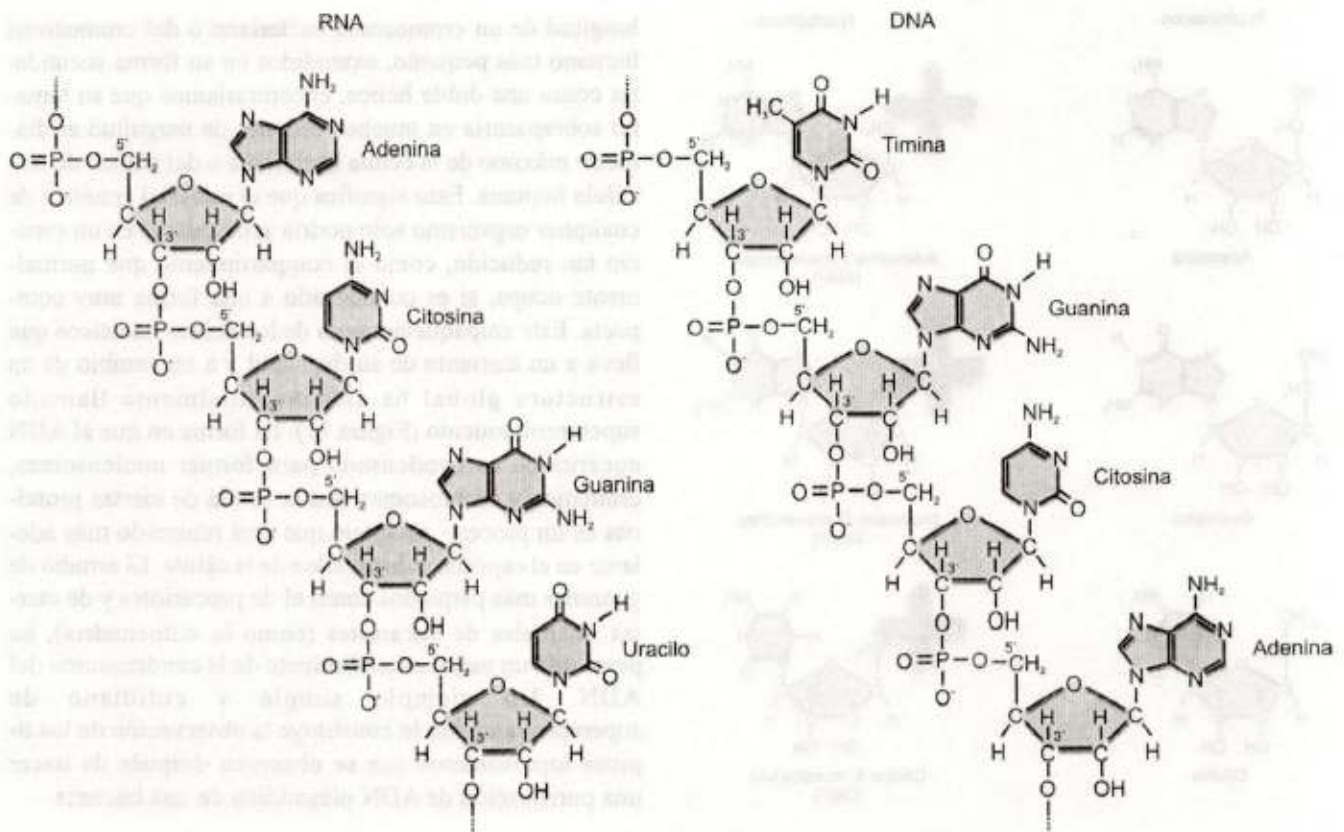


Figura 6. Los ácidos nucleicos son polímeros de nucleótidos, que se unen mediante enlaces covalentes de tipo éster. Además de la diferencia en el grupo OH presente en la ribosa, el RNA se distingue del ADN porque en esta última molécula se encuentra el nucleótido timina, mientras en la primera se encuentra el uracilo.

fosfato que es similar a lo largo de la secuencia, con las bases nitrogenadas unidas al azúcar y dispuestas de manera lateral al esqueleto, que como ya se dijo son las que determinan una secuencia particular. Cuando el polímero es formado por la unión de nucleótidos de ribosa el resultado es el RNA y cuando se trata de nucleótidos de desoxirribosa se conforma el ADN (ver Figura 6).

Estructura secundaria y organización superior de los ácidos nucleicos

El modelo de Watson y Crick para la molécula de ADN establece que ésta es una estructura homogénea en forma de espiral, conformada por dos cadenas polinucleotídicas dispuestas de forma anti-paralela e interaccionando entre sí por medio de puentes de hidrógeno. Dichos puentes de hidrógeno ocurren entre las bases nitrogenadas siguiendo las leyes de complementariedad de Chargaff: guanina solo se aparea con citosina y adenina se aparea solo con timina. El esqueleto azúcar-fosfato se dispone en la parte externa de la hélice mientras que las bases nitrogenadas se proyectan al interior para interaccionar con las bases nitrogenadas de la cadena complementaria (Figuras 5 y 9). El giro de las dos cadenas alrededor de la otra forma

una hélice en cuya superficie se generan dos surcos (uno mayor y otro menor), los cuales, al igual que el esqueleto, van girando de forma helicoidal (Figura 10).

La estructura tridimensional descrita por Watson y Crick es de tipo dextrógiro (el giro de la hélice progresa en el mismo sentido de las agujas de un reloj) y ha sido denominado forma B del ADN (Figura 10). Estudios posteriores permitieron establecer que existen conformaciones alternativas del ADN, aunque algunas, como la forma A, solo se presentan en condiciones experimentales poco probables de ocurrir *in vivo* y otras, como la forma Z, se presentan en ciertas regiones del genoma. Estas se diferencian de la forma B en cuanto al sentido del giro de la hélice (la Z tiene giro izquierdo), al igual que en su forma y tamaño, en la amplitud y profundidad de sus surcos y en muchos otros parámetros estructurales. Aunque se sabe que en condiciones fisiológicas la forma B es la más estable y abundante, se especula que las formas alternativas particularmente la Z, podrían ser biológicamente relevantes.

Los ácidos nucleicos y en particular el ADN no existen en la naturaleza como una simple estructura secundaria definida como una doble hélice extendida. Si calculáramos la

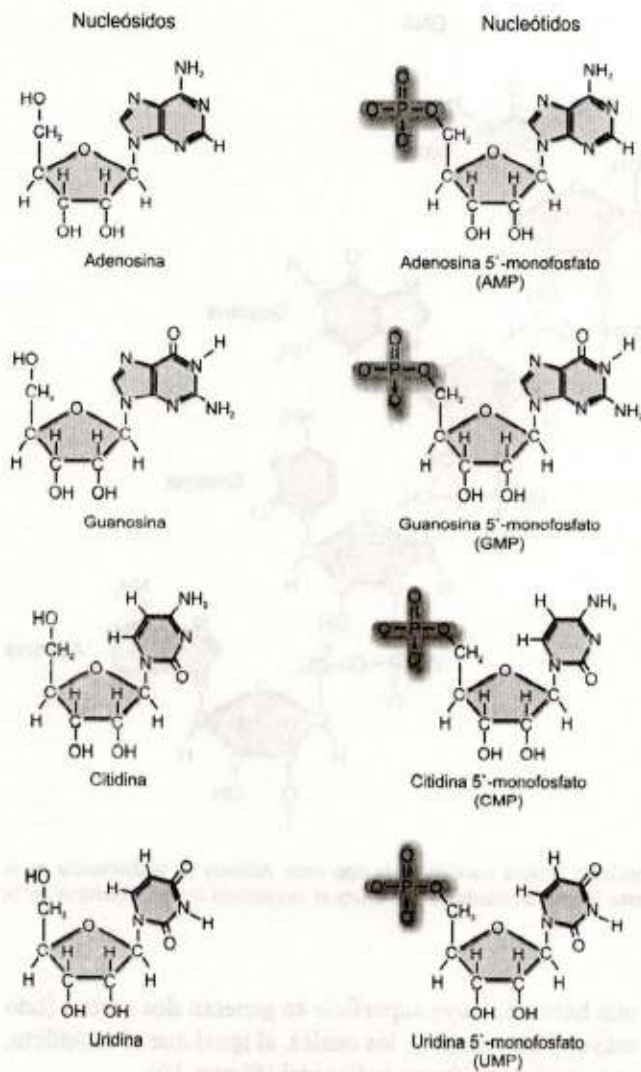


Figura 7. Las bases nitrogenadas que hacen parte de los nucleótidos son estructuras heterocíclicas planas formadas por uno o dos anillos, pueden ser pirimídicas (tienen un solo anillo) o púricas (tienen un anillo pirimidico y uno imidazol). Existen moléculas precursoras de los nucleótidos denominadas nucleósidos que consisten en la unión del azúcar con la base nitrogenada, las cuales se convierten en los primeros cuando la pentosa se enlaza un grupo fosfato por medio de un enlace fosfo-éster.

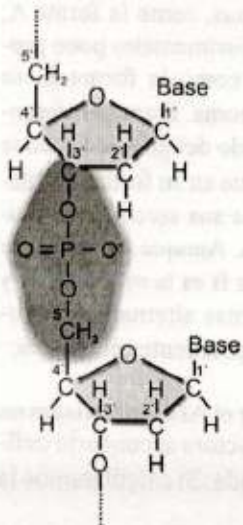


Figura 8. La polimerización de los ácidos nucleicos ocurre mediante la unión de los nucleótidos por enlaces fosfo-éster entre el azúcar de un nucleótido con el fosfato del siguiente nucleótido, lo cual origina un polímero lineal.

longitud de un cromosoma bacteriano o del cromosoma humano más pequeño, extendidos en su forma secundaria como una doble hélice, encontraríamos que su tamaño sobrepasaría en muchos órdenes de magnitud el diámetro máximo de la célula bacteriana o del núcleo de una célula humana. Esto significa que el material genético de cualquier organismo solo podría acomodarse en un espacio tan reducido, como el compartimento que normalmente ocupa, si es condensado a una forma muy compacta. Este empaquetamiento de los ácidos nucleicos que lleva a un aumento de su densidad y a un cambio de su estructura global ha sido habitualmente llamado superenrollamiento (Figura 11). La forma en que el ADN eucariótico es condensado para formar nucleosomas, cromatina y cromosomas con la ayuda de ciertas proteínas es un proceso complejo que será retomado más adelante en el capítulo sobre núcleo de la célula. El estudio de genomas más pequeños como el de procariotes y de ciertas organelas de eucariotes (como la mitocondria), ha permitido un mejor entendimiento de la condensación del ADN. Un ejemplo simple y cotidiano de superenrollamiento lo constituye la observación de los típicos topoisómeros que se observan después de hacer una purificación de ADN plasmídico de una bacteria.

Los genes son los responsables de codificar la información en los sistemas vivos

Como ya se ha mencionado los ácidos nucleicos, en particular el ADN, son las moléculas responsables de almacenar la información genética en los sistemas biológicos; sin embargo, no toda la secuencia de una molécula de ácido nucleico codifica información útil para el organismo, dicha información se encuentra organizada en regiones discretas con una estructura definida que se conoce como gen.

Un gen es una secuencia de nucleótidos del ADN (o RNA en algunos virus) que se transcribe para generar un RNA funcional. Existen tres tipos de RNA que se pueden transcribir a partir del ADN: el RNA mensajero (RNAm), el cual lleva la información que luego se va a traducir en las proteínas; el RNA de transferencia (RNAt), que es el encargado de transportar los aminoácidos hacia el sitio del ribosoma donde ocurre la síntesis del polipéptido y del cual existen en el ser humano 60 diferentes moléculas; y el RNA ribosomal (RNAr), del cual existen cuatro moléculas diferentes y que interactúan con las proteínas ribosomales para conformar las dos subunidades de los ribosomas. Como se puede deducir de lo anterior todos estos tipos de RNA son esenciales para la síntesis de las proteínas, proceso que se explicará en detalle en el capítulo sobre traducción del RNAm.

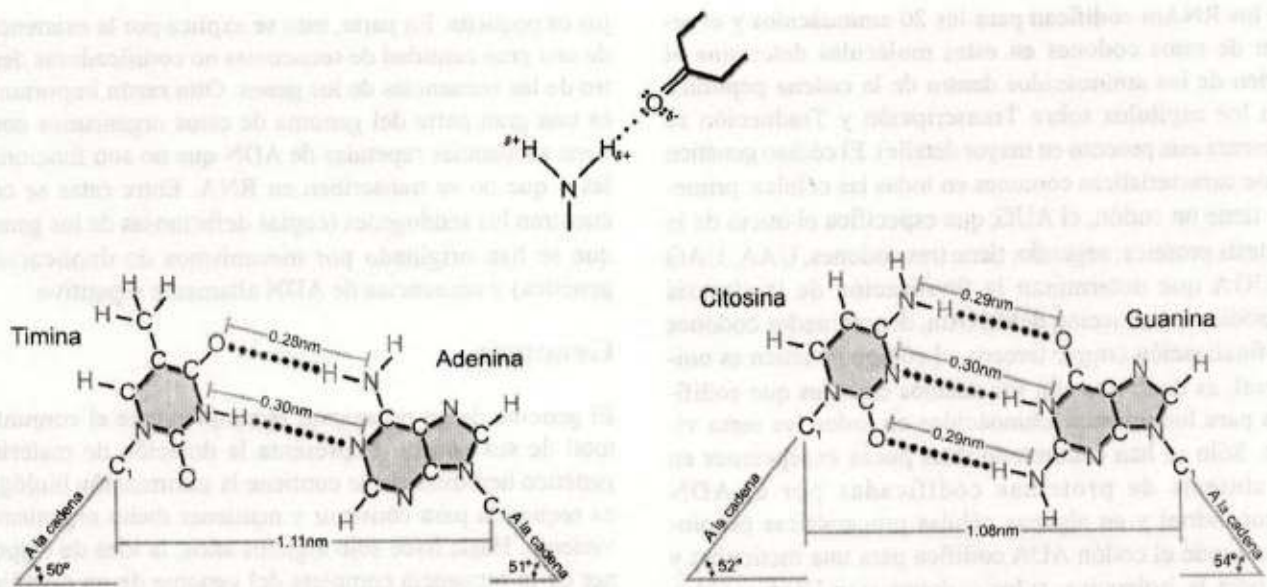


Figura 9. Los puentes de hidrógeno que permiten la conformación de la doble hélice de ADN se establecen entre las bases nitrogenadas al interior de esta estructura. La guanina forma tres puentes de hidrógeno con la citosina, mientras la adenina forma dos al aparearse con la timina. Estos puentes de hidrógeno tienen una distancia muy similar lo que permite una gran regularidad en esta molécula. Por su parte el esqueleto azúcar-fosfato se dispone en la parte externa de la hélice.

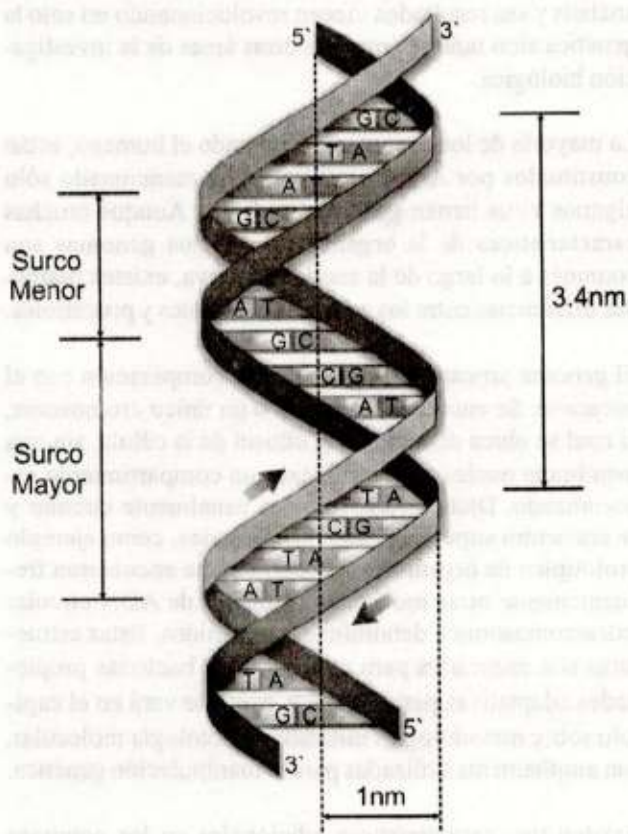


Figura 10. La doble hélice de Watson y Crick gira en el mismo sentido de las agujas de un reloj (dextrógira), la cual se denomina forma B del ADN. Este giro de las dos cadenas genera dos surcos (uno mayor y otro menor), los cuales, al igual que el esqueleto, van girando de forma helicoidal.

Nucleosoma

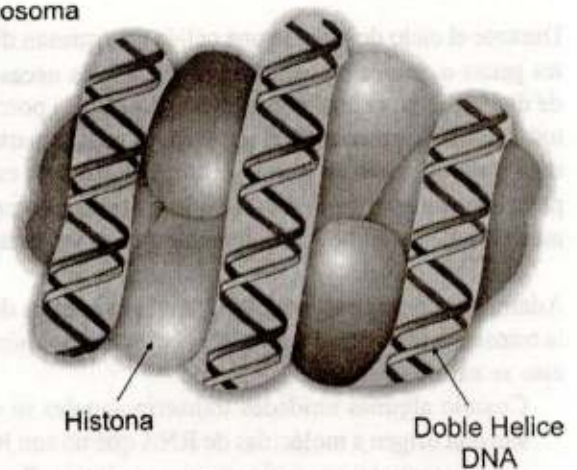


Figura 11. Debido a la enorme longitud de las moléculas de ADN necesarias para codificar la información genética, es necesario hacer un empaquetamiento de estas estructuras, lo cual se conoce como superenrollamiento. En los eucariotes el ADN se condensa para formar nucleosomas gracias a la presencia de las histonas, proteínas que forman un eje sobre el cual el ADN se enrolla.

La información genética para sintetizar una proteína está determinada por la secuencia de nucleótidos del ADN que se transcribe a RNAm. Tripletas de nucleótidos contiguos del RNAm codifican la información para los diferentes aminoácidos. Cada tripleta de nucleótidos es una letra del código genético que se conoce como **codón**. El diccionario completo del código genético tiene $4^3 = 64$ diferentes posibles combinaciones de codones; 61 codones

de los RNAm codifican para los 20 aminoácidos y el orden de estos codones en estas moléculas determina el orden de los aminoácidos dentro de la cadena peptídica (en los capítulos sobre Transcripción y Traducción se presenta este proceso en mayor detalle). El código genético tiene características comunes en todas las células: primero, tiene un codón, el AUG, que especifica el inicio de la síntesis proteica; segundo, tiene tres codones, UAA, UAG y UGA que determinan la finalización de la síntesis peptídica o traducción del RNAm, denominados codones de finalización (*stop*); tercero, el código genético es universal, es decir que son los mismos codones que codifican para los mismos aminoácidos en todos los seres vivos. Sólo se han encontrado unas pocas excepciones en la síntesis de proteínas codificadas por el ADN mitocondrial y en algunas células procarióticas primitivas, donde el codón AUA codifica para una metionina y no para la isoleucina, y los codones *stop* UGA y UAA codifican para el triptófano y la glutamina respectivamente y no son codones de finalización.

Solo una fracción pequeña del ADN en los organismos más complejos se expresa de manera funcional (da origen a RNA o proteínas)

Durante el ciclo de vida de una célula se expresan diferentes genes o grupos de genes de acuerdo a las necesidades de dicha célula, lo cual constituye una pequeña porción de todo el ADN presente en el genoma, lo cual es particularmente significativo en los organismos eucariotas; esto implica que la mayor parte del ADN de las células en organismos complejos como el ser humano nunca se transcribe.

Además, solo una parte del RNA que se sintetiza durante la transcripción se traduce posteriormente a polipéptidos; esto se explica porque:

- Cuando algunas unidades transcripcionales se expresan dan origen a moléculas de RNA que no son RNAm y por lo tanto no especifican para un péptido. Por ejemplo, las moléculas de RNA que constituyen los RNAr, los RNAt y una serie de RNA nucleares pequeños y citoplasmáticos.
- El transcripto primario (la molécula de RNA que se transcribe) puede contener una porción importante de RNA que no codifica para el polipéptido, el cual es eliminado mediante un mecanismo de procesamiento para dar origen a un RNAm mucho más pequeño.
- Sólo una parte del RNAm maduro es traducido; una porción importante a ambos extremos de las moléculas de RNAm no son traducidas por la maquinaria ribosomal.

De acuerdo a lo anterior, la fracción de ADN que realmente codifica en los genomas de los eucariotes comple-

jos es pequeña. En parte, esto se explica por la existencia de una gran cantidad de secuencias no codificadoras dentro de las secuencias de los genes. Otra razón importante es una gran parte del genoma de estos organismos contiene secuencias repetidas de ADN que no son funcionales o que no se transcriben en RNA. Entre éstas se encuentran los pseudogenes (copias defectuosas de los genes que se han originado por mecanismos de duplicación genética) y secuencias de ADN altamente repetitivo.

Genomas

El genoma de un organismo es simplemente el conjunto total de sus genes. Representa la dotación de material genético hereditario que contiene la información biológica requerida para construir y mantener dicho organismo viviente. Hasta hace solo algunos años, la idea de disponer de la secuencia completa del genoma de un organismo parecía descabellada. Los grandes desarrollos en técnicas de secuenciación automática y en bioinformática han permitido que actualmente podamos disponer de la secuencia completa del ADN de una larga lista de especies vivientes, la cual sigue creciendo a pasos acelerados. Esta avalancha de información es actualmente objeto de análisis y sus resultados vienen revolucionando no solo la genética sino también muchas otras áreas de la investigación biológica.

La mayoría de los genomas, incluyendo el humano, están constituidos por ADN, como ya se ha mencionado sólo algunos virus tienen genomas de RNA. Aunque muchas características de la organización de los genomas son comunes a lo largo de la escala evolutiva, existen marcadas diferencias entre los genomas eucariotes y procariotes.

El genoma procarionte es pequeño en comparación con el eucariote. Se encuentra reducido a un único cromosoma, el cual se ubica disuelto en el citosol de la célula, sin una membrana nuclear que establezca un compartimiento especializado. Dicho cromosoma es usualmente circular y se encuentra superenrollado. En bacterias, como ejemplo prototípico de organismos procariotes, se encuentran frecuentemente otras moléculas pequeñas de ADN circular extracromosómico denominadas plásmidos. Estas estructuras son esenciales para conferir a las bacterias propiedades adaptativas particulares y, como se verá en el capítulo sobre metodologías utilizadas en biología molecular, son ampliamente utilizadas para la manipulación genética.

Existen tres características adicionales en los genomas procariotes claramente distintivas de los genomas eucariotes: primero, la densidad de genes, es decir, el número de secuencias que codifican una proteína por tramo de ADN, es muchísimo más alta que en eucariotes. En

algunos casos, la distancia entre un gen y otro puede ser tan sólo un nucleótido. Por ejemplo, como se verá en el capítulo sobre transcripción, las bacterias organizan sus genes en operones, es decir, en bloques de genes que se transcriben (unidades transcripcionales) físicamente ligados en una sola molécula. Esto permite la expresión rápida y simultánea de genes requeridos bajo ciertas circunstancias. Segundo, los genes de los procariotes son continuos, es decir, no presentan intrones (ver más adelante). Finalmente, con pocas excepciones, el genoma procariote carece de ADN repetitivo. Como se puede deducir, estas características establecen una diferencia fundamental entre los organismos procariotes y eucariotes.

Aunque los genomas de los diversos organismos eucariotes son altamente variables en su tamaño, estructura y or-

ganización, comparten varias características comunes. El genoma eucariote está compuesto por el genoma nuclear y por el genoma de las organelas (las mitocondrias y los cloroplastos). Mientras el genoma nuclear está organizado en cromosomas lineales independientes, los genomas de la mitocondria y del cloroplasto están compuestos por varias copias de un único cromosoma circular pequeño. Los tamaños del genoma nuclear eucariote pueden variar de varias megabases a cientos de miles de megabases (Tabla 1). Hasta cierto punto, el tamaño del genoma refleja la complejidad del organismo, ya que eucariotes unicelulares como la levadura tienen genomas más pequeños que los multicelulares como los de los vertebrados y las plantas. Sin embargo, parece no haber correlación alguna entre el número de cromosomas y el tamaño del genoma o entre la complejidad del organismo y el número de cromosomas.

Tabla I. El tamaño del genoma, el número de genes y la densidad génica de una amplia variedad de organismos ha sido caracterizado.

Organismo		Tamaño del genoma (Megabases)	Número de genes	Densidad génica (genes/Mb)
Nombre común o clase	Nombre científico			
Eucariotes				
Levadura del pan	<i>Saccharomyces cerevisiae</i>	12	6.241	480
Nematodo	<i>Caenorhabditis elegans</i>	97	18.424	190
Crucífera	<i>Arabidopsis thaliana</i>	125	25.498	204
Mosca del vinagre	<i>Drosophila melanogaster</i>	180	13.601	75
Pez globo	<i>Fugu rubripes</i>	400	35.000	100
Arroz	<i>Oryza sativa</i>	450		
Erizó de mar	<i>Strongylocentrotus purpuratus</i>	900	27.350	30
Maíz	<i>Zea mays</i>	2.400		
Ser humano	<i>Homo sapiens</i>	3.400	35.000	10
Cebolla	<i>Allium cep</i>	18.000		
Ameba	<i>Amoeba dubia</i>	686.000		
Arqueas				
Crenarchaeota	<i>Aeropyrum pernix</i>	1,55	1.522	981
Euryarchaeota	<i>Methanococcus jannaschii</i>	1,66	1.715	1033
Euryarchaeota	<i>Archaeoglobus fulgidus</i>	2,18	2.420	1110
Bacterias				
Proteobacteria	<i>Buchnera sp. CCE</i>	0,45		
Gram positiva	<i>Mycoplama genitalium</i>	0,58	479	831
Proteobacteria	<i>Buchnera sp. APS</i>	0,64	564	881
Gram negativa	<i>Haemophilus influenzae</i>	1,8	1.727	959
Cianobacteria	<i>Synechocystis sp.</i>	3,6	3.168	880
Gram positiva	<i>Bacillus subtilis</i>	4,2	4.100	976
Proteobacteria	<i>Escherichia coli</i>	4,6	4.288	932

Tomando como prototipo el genoma humano, se pueden anotar varias características de los genomas eucariotes. Una propiedad importante es que la mayoría de los genes humanos son discontinuos, es decir, están conformados por intrones, los cuales son segmentos del gen que no terminarán representados (secuencias no codificantes) en la proteína, y por exones que corresponden a los segmentos de los genes que realmente codifican aminoácidos en la proteína. Dichos exones tienen que sufrir un proceso de corte y empalme que ocurre durante los pasos del procesamiento que hacen parte de la expresión génica (ver capítulo sobre Transcripción). En promedio, existen nueve exones por gen humano aunque existe un gen cuyo número de exones es de 178. Los diferentes exones de un gen pueden ser empalmados de maneras diferentes dando origen a RNA mensajeros diferentes y por lo tanto a proteínas distintas a partir de la misma secuencia de ADN. Esto implica que del mismo gen se pueden generar más de una proteína diferente por medio de empalme alternativo. Otra característica del genoma humano es que presenta con frecuencia pseudogenes, que como se dijo, son copias de genes funcionales que acumularon mutaciones y por lo tanto no se convierten en proteínas. A diferencia de los genes procariotes, el genoma humano ha revelado una baja densidad de genes por tramo de ADN, se calcula que existe un gen cada 40.000 pares de nucleótidos. Las

vastas regiones intergénicas están constituidas por ADN no codificante cuya función sigue siendo enigmática. Estas regiones intergénicas están básicamente conformadas por ADN repetitivo de diferentes tipos, cuyas unidades de repetición pueden estar dispersas en el genoma o agrupadas en bloques.

De los 30.000 a 40.000 genes humanos que se propone existen después de la realización del proyecto genoma humano, se ha podido asignar una función a cerca del 50% de ellos. La mayoría de ellos codifican proteínas, mientras una minoría codifica para RNA que no se traduce a proteínas. El grupo mayoritario (24%) corresponden a genes involucrados en el mantenimiento y expresión génica. Un 20% de ellos codifican para proteínas involucradas en la transducción de señales, mientras otro 20% son genes asociados con funciones bioquímicas generales.

El genoma mitocondrial humano fue secuenciado hace más de dos décadas y sus 26.569 pares de bases codifican para 13 proteínas involucradas en la respiración celular y para 24 RNAs involucrados en la expresión de los primeros. Este es un ADN no repetitivo de alta densidad génica, cuyos genes no poseen intrones.

Gen	Exones	Intrones	Función
ADAMTS-1	1	0	Proteína de la matriz extracelular
ADAMTS-2	1	0	Proteína de la matriz extracelular
ADAMTS-3	1	0	Proteína de la matriz extracelular
ADAMTS-4	1	0	Proteína de la matriz extracelular
ADAMTS-5	1	0	Proteína de la matriz extracelular
ADAMTS-6	1	0	Proteína de la matriz extracelular
ADAMTS-7	1	0	Proteína de la matriz extracelular
ADAMTS-8	1	0	Proteína de la matriz extracelular
ADAMTS-9	1	0	Proteína de la matriz extracelular
ADAMTS-10	1	0	Proteína de la matriz extracelular
ADAMTS-11	1	0	Proteína de la matriz extracelular
ADAMTS-12	1	0	Proteína de la matriz extracelular
ADAMTS-13	1	0	Proteína de la matriz extracelular
ADAMTS-14	1	0	Proteína de la matriz extracelular
ADAMTS-15	1	0	Proteína de la matriz extracelular
ADAMTS-16	1	0	Proteína de la matriz extracelular
ADAMTS-17	1	0	Proteína de la matriz extracelular
ADAMTS-18	1	0	Proteína de la matriz extracelular
ADAMTS-19	1	0	Proteína de la matriz extracelular
ADAMTS-20	1	0	Proteína de la matriz extracelular
ADAMTS-21	1	0	Proteína de la matriz extracelular
ADAMTS-22	1	0	Proteína de la matriz extracelular
ADAMTS-23	1	0	Proteína de la matriz extracelular
ADAMTS-24	1	0	Proteína de la matriz extracelular
ADAMTS-25	1	0	Proteína de la matriz extracelular
ADAMTS-26	1	0	Proteína de la matriz extracelular
ADAMTS-27	1	0	Proteína de la matriz extracelular
ADAMTS-28	1	0	Proteína de la matriz extracelular
ADAMTS-29	1	0	Proteína de la matriz extracelular
ADAMTS-30	1	0	Proteína de la matriz extracelular
ADAMTS-31	1	0	Proteína de la matriz extracelular
ADAMTS-32	1	0	Proteína de la matriz extracelular
ADAMTS-33	1	0	Proteína de la matriz extracelular
ADAMTS-34	1	0	Proteína de la matriz extracelular
ADAMTS-35	1	0	Proteína de la matriz extracelular
ADAMTS-36	1	0	Proteína de la matriz extracelular
ADAMTS-37	1	0	Proteína de la matriz extracelular
ADAMTS-38	1	0	Proteína de la matriz extracelular
ADAMTS-39	1	0	Proteína de la matriz extracelular
ADAMTS-40	1	0	Proteína de la matriz extracelular
ADAMTS-41	1	0	Proteína de la matriz extracelular
ADAMTS-42	1	0	Proteína de la matriz extracelular
ADAMTS-43	1	0	Proteína de la matriz extracelular
ADAMTS-44	1	0	Proteína de la matriz extracelular
ADAMTS-45	1	0	Proteína de la matriz extracelular
ADAMTS-46	1	0	Proteína de la matriz extracelular
ADAMTS-47	1	0	Proteína de la matriz extracelular
ADAMTS-48	1	0	Proteína de la matriz extracelular
ADAMTS-49	1	0	Proteína de la matriz extracelular
ADAMTS-50	1	0	Proteína de la matriz extracelular
ADAMTS-51	1	0	Proteína de la matriz extracelular
ADAMTS-52	1	0	Proteína de la matriz extracelular
ADAMTS-53	1	0	Proteína de la matriz extracelular
ADAMTS-54	1	0	Proteína de la matriz extracelular
ADAMTS-55	1	0	Proteína de la matriz extracelular
ADAMTS-56	1	0	Proteína de la matriz extracelular
ADAMTS-57	1	0	Proteína de la matriz extracelular
ADAMTS-58	1	0	Proteína de la matriz extracelular
ADAMTS-59	1	0	Proteína de la matriz extracelular
ADAMTS-60	1	0	Proteína de la matriz extracelular
ADAMTS-61	1	0	Proteína de la matriz extracelular
ADAMTS-62	1	0	Proteína de la matriz extracelular
ADAMTS-63	1	0	Proteína de la matriz extracelular
ADAMTS-64	1	0	Proteína de la matriz extracelular
ADAMTS-65	1	0	Proteína de la matriz extracelular
ADAMTS-66	1	0	Proteína de la matriz extracelular
ADAMTS-67	1	0	Proteína de la matriz extracelular
ADAMTS-68	1	0	Proteína de la matriz extracelular
ADAMTS-69	1	0	Proteína de la matriz extracelular
ADAMTS-70	1	0	Proteína de la matriz extracelular
ADAMTS-71	1	0	Proteína de la matriz extracelular
ADAMTS-72	1	0	Proteína de la matriz extracelular
ADAMTS-73	1	0	Proteína de la matriz extracelular
ADAMTS-74	1	0	Proteína de la matriz extracelular
ADAMTS-75	1	0	Proteína de la matriz extracelular
ADAMTS-76	1	0	Proteína de la matriz extracelular
ADAMTS-77	1	0	Proteína de la matriz extracelular
ADAMTS-78	1	0	Proteína de la matriz extracelular
ADAMTS-79	1	0	Proteína de la matriz extracelular
ADAMTS-80	1	0	Proteína de la matriz extracelular
ADAMTS-81	1	0	Proteína de la matriz extracelular
ADAMTS-82	1	0	Proteína de la matriz extracelular
ADAMTS-83	1	0	Proteína de la matriz extracelular
ADAMTS-84	1	0	Proteína de la matriz extracelular
ADAMTS-85	1	0	Proteína de la matriz extracelular
ADAMTS-86	1	0	Proteína de la matriz extracelular
ADAMTS-87	1	0	Proteína de la matriz extracelular
ADAMTS-88	1	0	Proteína de la matriz extracelular
ADAMTS-89	1	0	Proteína de la matriz extracelular
ADAMTS-90	1	0	Proteína de la matriz extracelular
ADAMTS-91	1	0	Proteína de la matriz extracelular
ADAMTS-92	1	0	Proteína de la matriz extracelular
ADAMTS-93	1	0	Proteína de la matriz extracelular
ADAMTS-94	1	0	Proteína de la matriz extracelular
ADAMTS-95	1	0	Proteína de la matriz extracelular
ADAMTS-96	1	0	Proteína de la matriz extracelular
ADAMTS-97	1	0	Proteína de la matriz extracelular
ADAMTS-98	1	0	Proteína de la matriz extracelular
ADAMTS-99	1	0	Proteína de la matriz extracelular
ADAMTS-100	1	0	Proteína de la matriz extracelular