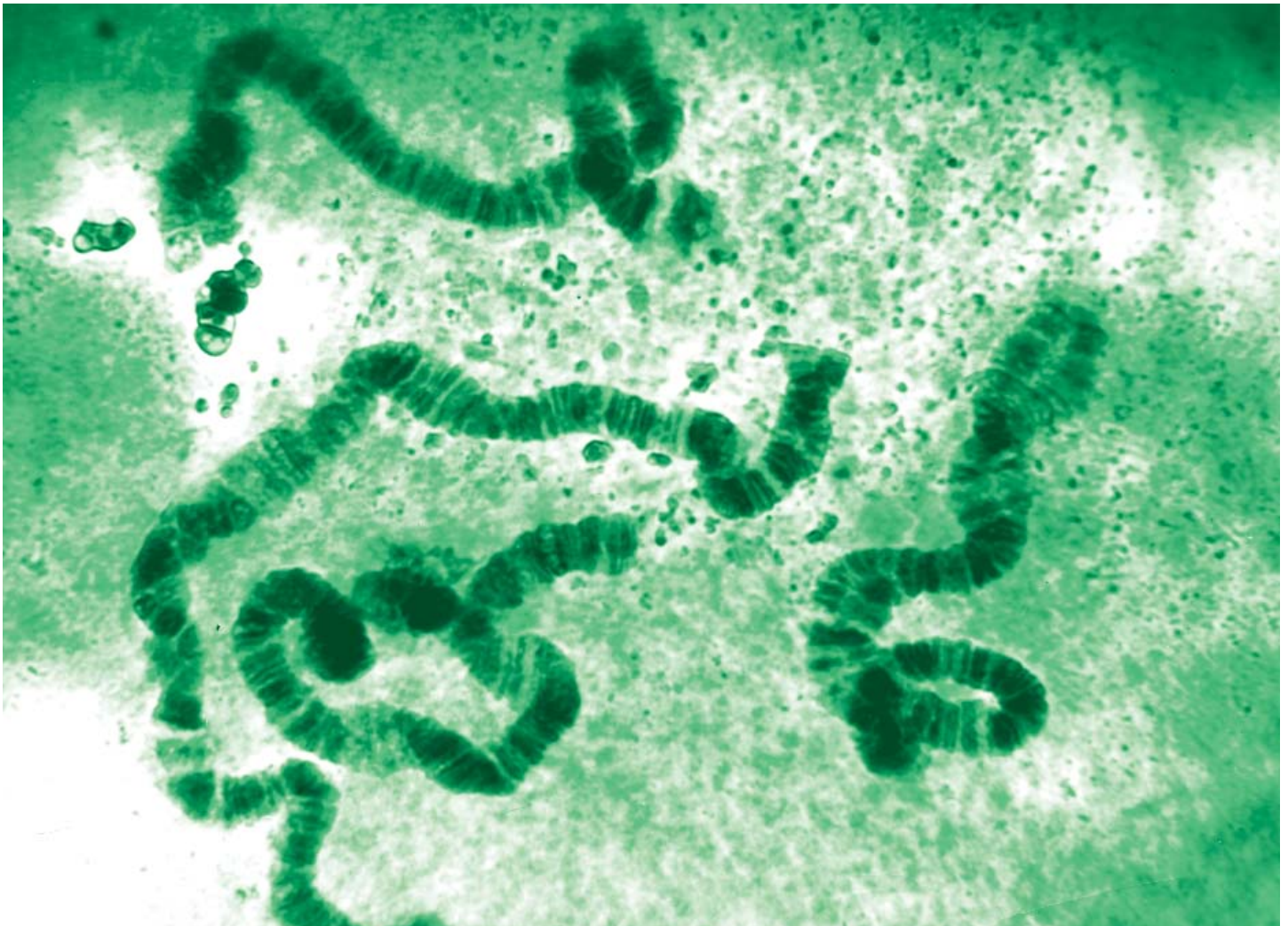


CIENCIAS NATURALES

LOS GENES



Introducción. ¿Qué dicen los genes? | **ADN: el material genético** | **El genoma eucariota** | ¿Cómo se expresa la información contenida en el ADN? | El proceso de transcripción | Topología de un gen eucariota típico | ¿Cómo se traduce el mensaje? | **Regulación de la expresión de genes eucariotas** | Transcripción diferencial de genes | Procesamiento alternativo de intrones | Traducción selectiva de ARNm | Modificación diferencial de proteínas

Autora: Dra. Nora B. Calcaterra (UNR y CONICET) | **Coordinación Autoral:** Dr. Alberto Kornblihtt (UBA y CONICET)

INTRODUCCIÓN. ¿QUÉ DICEN LOS GENES?

A partir de los experimentos de Gregor Mendel fue posible postular la existencia de factores heredables, responsables de transmitir las características de una especie de generación en generación. Estos factores heredables fueron posteriormente identificados como *genes*, secuencias específicas de nucleótidos de ADN ubicadas en los cromosomas de las células. Pero ¿cuál es la naturaleza de la conexión entre los cromosomas y los caracteres heredables? ¿Cuáles son los mecanismos que una célula utiliza

para especificar cómo será su descendencia? ¿Qué dicen los genes realmente? ¿Cómo es su mensaje traducido por la célula en características específicas, tales como el color de ojos o la altura de una persona, la forma de las hojas, el tamaño, el color y el sabor de los frutos?

En este fascículo analizaremos algunos mecanismos moleculares que llevan a cabo las células eucariotas y que permiten responder las preguntas planteadas.



Institut Pasteur / Francia

Cromosoma de una célula de insecto visto con microscopio óptico.

ADN: EL MATERIAL GENÉTICO

Una vez demostrado que el material genético es el ADN, fue necesario conocer la estructura tridimensional de esta molécula para poder responder a dos preguntas: cómo se replica y de qué manera dirige la síntesis de proteínas. En 1953, los científicos James Watson y Francis Crick propusieron un modelo que estableció las principales características de la estructura del ADN y permitió comenzar a entender las bases moleculares de la herencia y la evolución. El modelo estructural propone que el ADN es una hélice de doble cadena de desoxirribonucleótidos, de diámetro uniforme, que se tuerce hacia la derecha, antiparalela (las dos cadenas corren en sentido contrario).

¿Qué quiere decir que las dos cadenas son antiparalelas? La dirección de un polinucleótido puede definirse mirando los enlaces fosfodiéster entre nucleótidos adyacentes. En el esqueleto de azúcar-fosfato, un grupo fosfato unido al carbono 5' de una desoxirribosa se conecta al grupo hidroxilo del

carbono 3' de otra molécula de desoxirribosa, uniendo azúcares sucesivos entre sí. Por ello, los extremos de un polinucleótido son distintos: en uno de ellos el carbono 5' está libre (extremo 5') mientras que en el otro lo está el carbono 3' (extremo 3'). Las dos cadenas de polinucleótidos de una molécula de ADN corren en sentidos opuestos dado que mientras una de ellas lo hace en sentido 5'→3', la cadena complementaria corre en sentido 3'→5'.

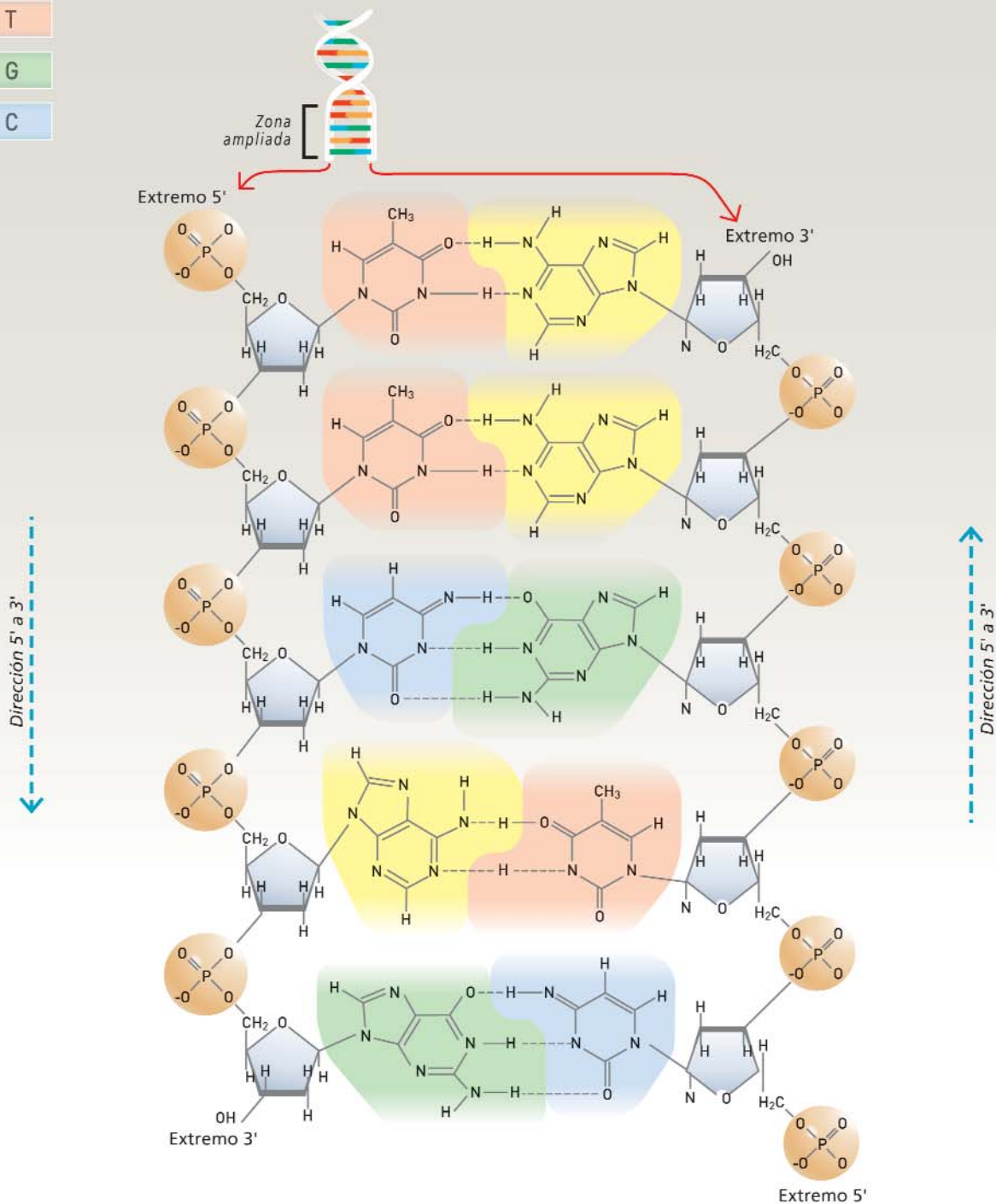
En la molécula de ADN, los esqueletos de azúcar-fosfato de las cadenas de polinucleótidos se enrollan alrededor de la parte externa de la hélice y las bases nitrogenadas se ubican en el centro. Las cadenas se mantienen unidas por apareamiento de las bases nitrogenadas purinas (A o G) o pirimidinas (T o C). En esta estructura, siempre una A (adenina) se enfrenta a una T (timina), y una C (citocina) se enfrenta a una G (guanina). Esta complementariedad de bases A-T y C-G hace que se mantenga constante el diámetro de la doble hélice.

En este fascículo no analizaremos en detalle los mecanismos moleculares mediante los que se replica una molécula de ADN. Sólo mencionaremos algunas de las principales características de este proceso. La bioquímica de la replicación es semejante en células procariontas y eucariotas. En todos los casos ocurre de manera finamente regulada y con la participación de una gran cantidad de proteínas. *In vivo*, la replicación ocurre una sola vez durante el ciclo de vida de una célula, durante la fase S de la interfase. Sin embargo, *in vitro* es posible lograr que una molécula de ADN se replique varias veces en un tubo de ensayo, sin la presencia de células. Para ello se requiere ADN, una enzima (ADN polimerasa) y una mezcla de los cuatro desoxinucleótidos trifosfato. Esto se debe a que el ADN sirve de molde para su propia síntesis, actuando como guía para la ubicación exacta de los nucleótidos complementarios en la nueva cadena. La posibilidad de replicación del ADN *in vitro* ha resultado de importancia fun-

ESTRUCTURA MOLECULAR DEL ADN

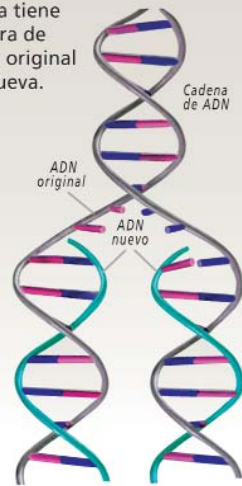
Se observa la doble hélice antiparalela del ADN desdoblada, integrada por eslabones moleculares de azúcar-fosfato. Las cadenas se mantienen unidas y con diámetro constante mediante bases nitrogenadas purina-pirimidina apareadas (AT y CG).

- A
- T
- G
- C



REPLICACIÓN DEL ADN

De acuerdo con la hipótesis semiconservativa, la molécula replicada tiene una hebra de material original y otra nueva.



damental para el desarrollo de la biotecnología.

El ADN tiene toda la información necesaria para el funcionamiento celular y el establecimiento de los caracteres específicos de una especie o individuo. Por ello, es importante que pueda ser copiado con fidelidad para ser repartido entre la progenie, ya sean células que se regeneran o nuevos individuos de una especie. Se han propuesto muchas hipótesis sobre cómo se replica el ADN. En 1953, Watson y Crick formularon la hipótesis semiconservativa que fue posteriormente demostrada por Matthew Meselson y Franklin Stahl en 1957. En la replicación semiconservativa, la doble hélice se separa y cada una de las cadenas sirve de molde para la síntesis de una nueva cadena complementaria.

El resultado final son dos moléculas idénticas a la original, cada una conteniendo una cadena original y otra sintetizada *de novo*.

El reconocimiento de que el ADN podía copiarse de esta forma sugirió cómo esta molécula podía poseer una de las características necesarias del material genético: la capacidad de mutar. Sin mutaciones no habría variabilidad genética ni diversidad de especies. Las mutaciones representan cambios en la secuencia de bases en el ADN. Si el ADN es copiado por un mecanismo que implica apareamiento de bases complementarias, todo cambio en una cadena será reflejado en una nueva secuencia durante un nuevo ciclo de replicación, y la secuencia mutada pasará a las moléculas hijas mediante el mismo mecanismo.

EL GENOMA EUCARIOTA

El genoma de un organismo se define como la totalidad del ADN de sus cromosomas y los genes que este contiene. Las organelas de los eucariotas (mitocondrias y plastidios) también contienen ADN. Por lo tanto, se puede hablar de genoma mitocondrial o genoma plastídico para diferenciarlos del genoma nuclear.

Todos los individuos de una misma especie tienen distribuido su genoma en un número característico de cromosomas. Por ejemplo, las células somáticas de los humanos tienen 46 cromosomas. Sin embargo, no somos la única especie en tener esa cantidad de cromosomas; algunas especies de animales y vegetales también tienen 46 cromosomas. Por otro lado, el tamaño del genoma no presenta una relación directa con la complejidad del organismo. Las células humanas, por ejemplo, tienen un genoma unas 700 veces más grande que el de la bacteria *E. coli*, pero las células de

algunos anfibios y plantas poseen un genoma 30 veces mayor que el de las humanas. Todo esto indica que no es el tamaño del genoma o el número de cromosomas lo que hace única a cada especie, sino la información especificada en los genes de esos cromosomas. Es el orden particular de las bases purinas o pirimidinas, es decir, la secuencia de ADN, la que especifica la exacta instrucción para generar un organismo particular.

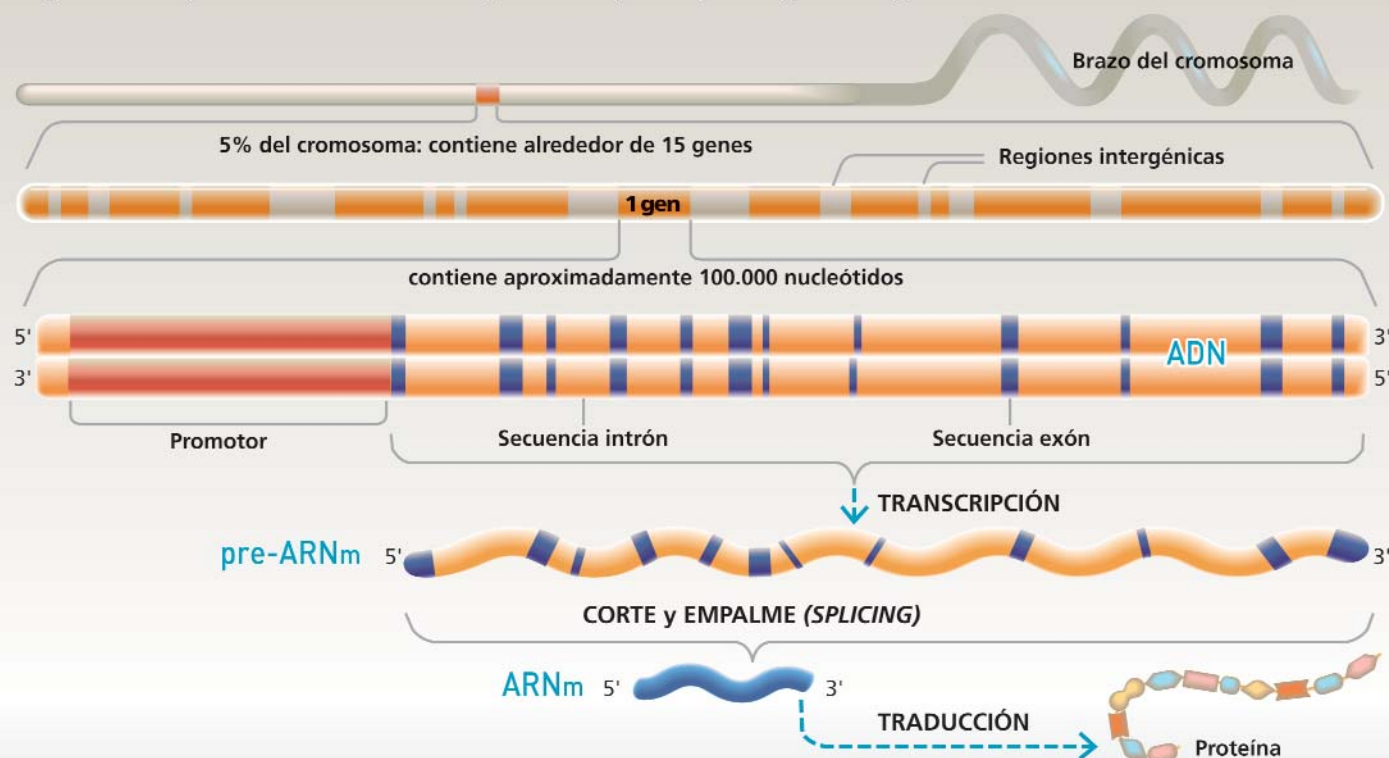
Como se analizará en detalle luego, no todo el genoma de un organismo lleva información para sintetizar una molécula de ARN funcional. Así, en la mayoría de los genomas de los vertebrados, los genes no se encuentran yuxtapuestos, sino que están separados por largos segmentos de ADN que constituyen las regiones intergénicas. Estas regiones representan alrededor de un 70 % del total del genoma. Por otro lado, las secuencias de genes eucariotas que codifican proteínas ha-

bitualmente no son continuas, sino que están interrumpidas por secuencias no codificantes. Las secuencias no codificantes se denominan intrones y las secuencias codificantes, exones. En consecuencia, los intrones, presentes en el genoma, están ausentes en el ARNm y, por lo tanto, no se traducen en proteínas. Los exones son los segmentos que están presentes en el genoma, en el ARNm precursor nuclear (pre-ARNm), permanecen en el ARNm maduro citoplasmático y son traducidos en proteínas. En la página siguiente se puede ver la organización típica de los genes en un cromosoma de vertebrado.

El conocimiento de la organización general del genoma de varias especies demostró las similitudes y diferencias estructurales existentes entre ellas, pero también planteó el desafío de conocer la secuencia de nucleótidos completa del genoma de una especie en particular. Estos emprendi-

ORGANIZACIÓN DE LOS GENES EN UN CROMOSOMA DE VERTEBRADO

Los genes están dispuestos ordenadamente a lo largo del ADN separados por las regiones intergénicas.



mientos científicos son conocidos como Proyectos Genoma y sus aplicaciones tienden a resolver problemas de la medicina, la biotecnología, las fuentes de energía, etc. En la actualidad se han completado los Proyectos Genoma de varias especies, entre ellas el de la mostaza silvestre (*Arabidopsis thaliana*), el de la mosca de la fruta (*Drosophila melanogaster*), el del gusano (*Caenorhabditis elegans*), el del ratón (*Mus musculus*) y el del humano (*Homo sapiens*).

¿CÓMO SE EXPRESA LA INFORMACIÓN CONTENIDA EN EL ADN?

Unos años después de haber propuesto el modelo estructural del ADN, Francis Crick propuso "el dogma central de la biología molecular", que dice que la información fluye en el sentido: ADN → ARN → proteína. Según sus palabras, "una vez que la informa-

ción pasó a una proteína no puede salir nuevamente".

Este flujo de la información consta de dos etapas: primero, en la transcripción, se copia la información de la secuencia de ADN en información correspondiente a una secuencia de ARN. Luego, durante la traducción, esta información se traduce en una secuencia de aminoácidos, es decir, una proteína.

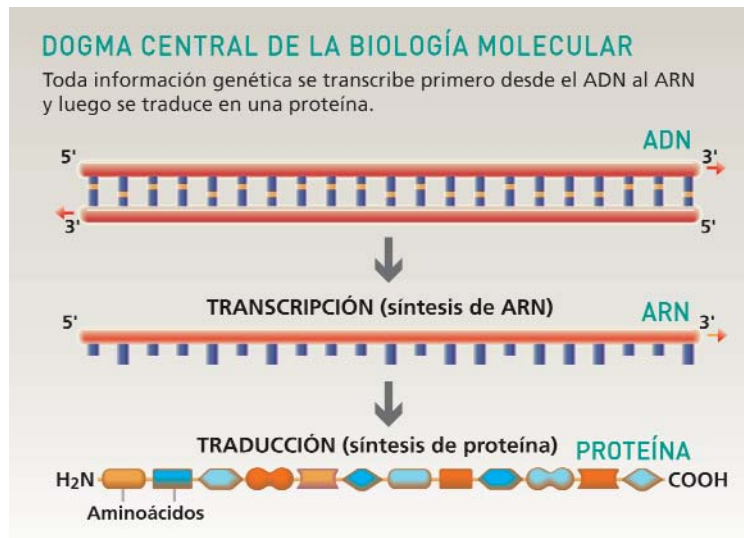
Si bien el flujo de la información es en un solo sentido, en algunos casos la transcripción puede ser invertida: ARN → ADN. Esta reacción ocurre mediante enzimas específicas llamadas transcriptasas reversas, presentes, por ejemplo, en los retrovirus como el VIH. Sin embargo, la traducción nunca puede ser invertida. Es decir, la información contenida en una proteína no puede pasar al ADN. Esto explica por qué no es posible heredar los caracteres adquiridos de los progenitores.

¿Cómo pasa la información del nú-

cleo (donde está el ADN) a las proteínas (que son sintetizadas en el citoplasma)? ¿Cuál es la relación entre una secuencia de nucleótidos específica y la secuencia de aminoácidos en una proteína? Para responder a estos interrogantes, Crick postuló la existencia de moléculas mensajeras y moléculas adaptadoras. Hoy sabemos que esas moléculas son distintas clases de ARN. El ARN es un ácido nucleico muy semejante al ADN, pero entre ambas moléculas hay diferencias:

- El azúcar que compone el ARN es ribosa, en lugar de desoxirribosa.
- En lugar de timina, el ARN contiene una pirimidina relacionada, el uracilo (U).
- El ARN se encuentra como cadena simple y no forma una estructura helicoidal regular como el ADN.

A pesar de esto último, algunas moléculas de ARN presentan una estructura secundaria, es decir que ciertas zonas de la molécula pueden formar



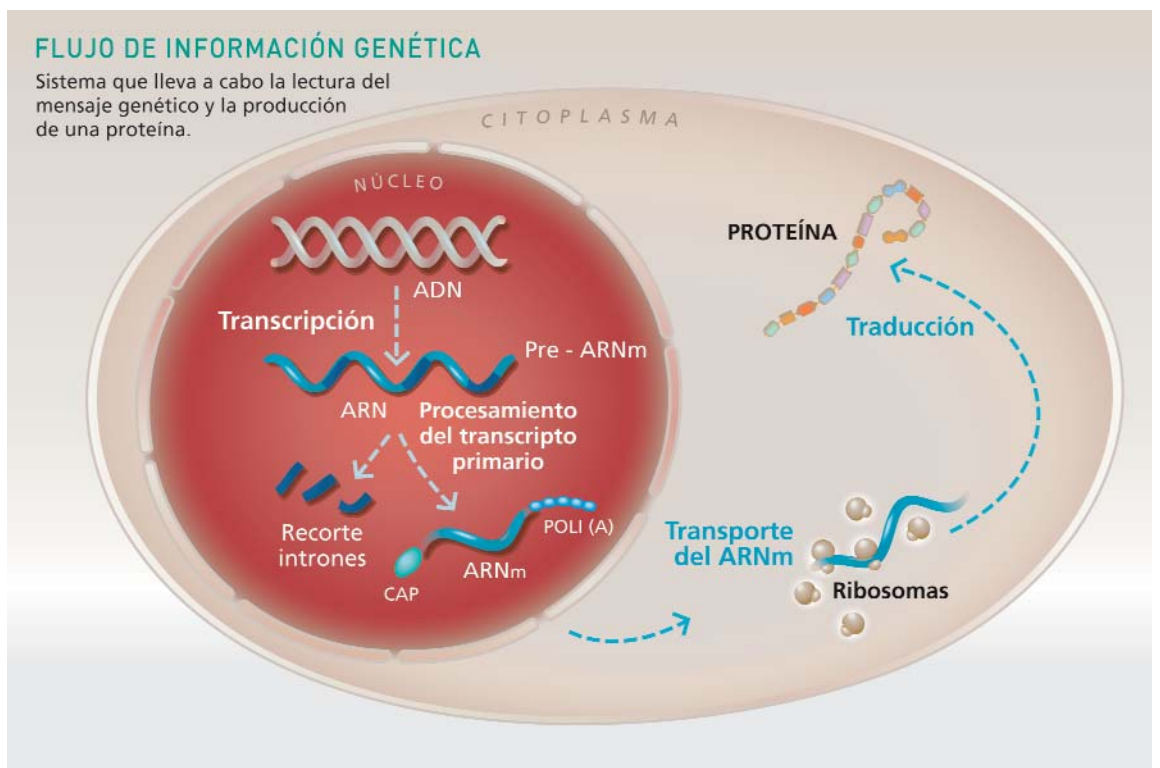
EL PROCESO DE TRANSCRIPCIÓN

Todas las moléculas de ARN son sintetizadas por el proceso de transcripción a partir de una secuencia de ADN, en reacciones catalizadas por ARN polimerasas. Estas enzimas operan moviéndose en dirección 3'→5' a lo largo de la cadena molde de ADN, sintetizando una nueva cadena de ribonucleótidos en la dirección 5'→3'. Esa cadena de ARN recién sintetizada se llama transcrito primario o pre-ARNm. Como resultado, la secuencia de nucleótidos del transcrito primario es antiparalela y complementaria de la secuencia de ADN de la que es transcrita. Es importante destacar que una molécula de ARN se copia sólo de una de las dos cadenas de ADN. Entonces, si sólo una hebra de ADN se transcribe, ¿cómo selecciona la ARN polimerasa la hebra correcta de ADN? La enzima reconoce en cada gen una secuencia determinada de nucleótidos, el promotor, y se une a ella. Esta unión establece el sitio donde comienza la transcripción y el sen-

bucles o doble cadena por apareamiento de bases dentro de la misma molécula.

Todas las células tienen diferentes clases de ARN. Los tres más importantes son el ARN ribosomal (ARNr) –se encuentra en los ribosomas, que es el lugar físico donde ocurre la traducción–, el ARN mensajero (ARNm) –pasa del núcleo al citoplasma llevando la información desde los cromosomas

hacia el ribosoma para dirigir la síntesis proteica– y el ARN de transferencia (ARNt) –transporta los aminoácidos al ribosoma para ser utilizados durante la traducción; son las moléculas adaptadoras entre nucleótidos y aminoácidos–. Estos tres tipos de ARN, junto con proteínas ribosomales y ciertas enzimas, constituyen el sistema que lleva a cabo la lectura del mensaje genético y la producción de una proteína.



TRANSCRIPCIÓN DE ADN

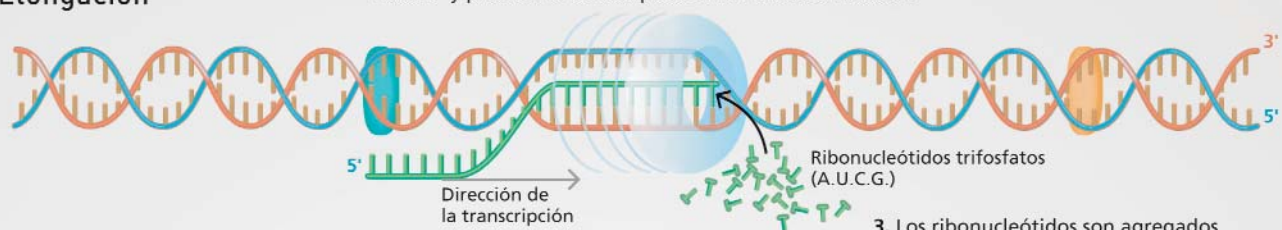
Iniciación

1. La ARN polimerasa se une al promotor y comienza a desenrollar las cadenas de ADN.



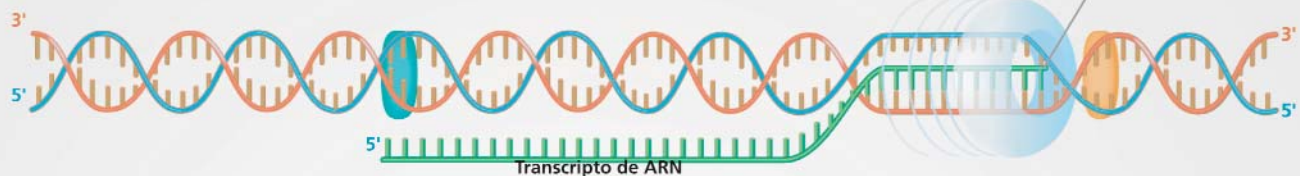
Elongación

2. La ARN polimerasa lee la cadena molde de ADN desde 3' hacia 5' y produce el transcrito de ARN desde 5' hasta 3'.



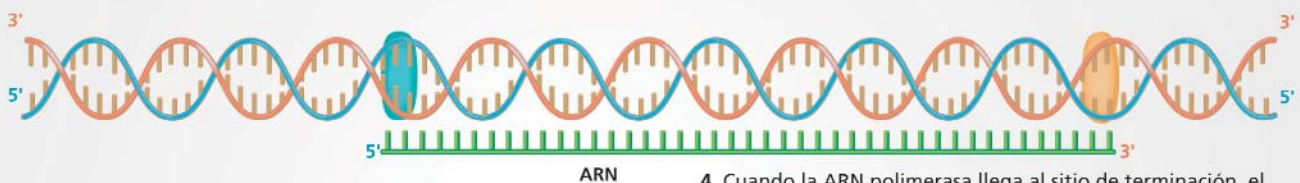
Prolongación

3. Los ribonucleótidos son agregados en el extremo 3' del ARN en



Terminación

4. Cuando la ARN polimerasa llega al sitio de terminación, el transcrito de ARN se libera completamente de la maquinaria de transcripción y la ARN polimerasa se separa del ADN.

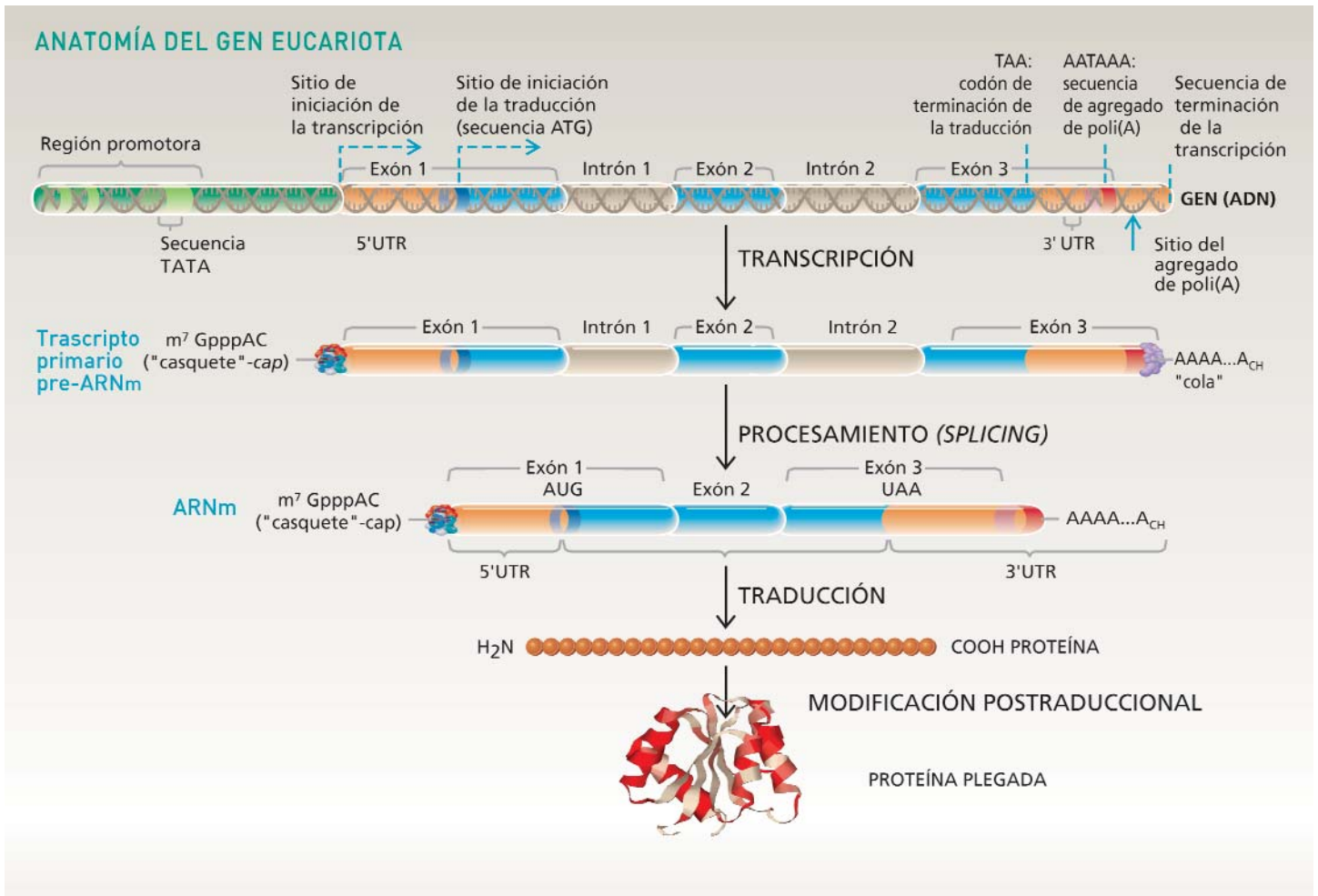


tido de avance de la ARN polimerasa. Dado que esta enzima sólo puede fabricar ARN en sentido 5'→3', una vez fijado el sentido de avance, sólo puede transcribirse la hebra de ADN que corre en sentido 3'→5'. La otra hebra, que posee la secuencia complementaria y corre en sentido 5'→3', no es reconocida por la enzima y, por lo tanto, no es transcrita a ARN. La ARN polimerasa reconoce y se une al promotor y esto hace que la doble hélice de ADN se abra y comience la transcripción. Luego, la enzima va

añadiendo ribonucleótidos, moviéndose a lo largo de la cadena molde, desenrollando la hélice y exponiendo nuevas regiones con las que se aparearán los ribonucleótidos complementarios. La cadena de ARN en crecimiento permanece brevemente unida por puentes de hidrógeno al ADN molde (sólo 10 o 12 ribonucleótidos están unidos al ADN en cualquier instante dado) y luego se desprende como una cadena simple, como puede verse en la imagen de esta página.

TOPOLOGÍA DE UN GEN EUCARIOTA TÍPICO

Antes de seguir avanzando en el análisis de los mecanismos por los que la información contenida en el ADN es traducida en una proteína, es necesario examinar las características que determinan que una secuencia dada de nucleótidos sea un gen. Sabemos que el ADN es una cadena continua de nucleótidos, donde no existen espacios ni "marcas". Entonces, ¿qué es lo que diferencia a un gen de una



secuencia intergénica? Típicamente, un gen eucariota presenta los siguientes elementos.

- **El promotor y las regiones reguladoras** (potenciadores y silenciadores). Como se verá más adelante, proteínas llamadas factores de transcripción se unen a ellas y regulan la actividad de un gen.

- **El sitio de iniciación de la transcripción.** Aquí se determina qué nucleótido será el extremo 5' del ARN. En el pre-ARNm, este extremo se modifica por el agregado de un nucleótido de 7-metilguanósina, llamado caperuza o *cap*, tan pronto ha sido transcrito. Esta modificación del extremo 5' del pre-ARNm se llama *capping*. La caperuza es necesaria para la unión del ARNm al ribosoma y la subsiguiente traducción.

- **El sitio de iniciación de la traducción (ATG).** Esta secuencia (que se convier-

te en AUG en el ARN) se localiza a distancias variables del sitio de iniciación de la transcripción, según el gen de que se trate. Entre el sitio de iniciación de la transcripción y el triplete ATG queda comprendida una secuencia que no se traduce, llamada región no traducida del extremo 5' o 5'UTR (del inglés *5' untranslated region*).

- **Exones e intrones.** Secuencias presentes en el gen y en el pre-ARNm que serán conservadas o eliminadas, respectivamente, durante el procesamiento del transcrito. Los genes transcriben pre-ARNm, que son más largos que los ARNm maduros porque contienen intrones que deben ser eliminados en el núcleo antes de pasar al citoplasma. Este mecanismo se conoce como "procesamiento por corte y empalme" o *splicing*. El procesamiento del pre-ARNm es mediado por un complejo

de proteínas y pequeños ARN nucleares (*small nuclear RNA* o *snRNA*). Este complejo, llamado *espliceosoma*, se ensambla reconociendo secuencias de nucleótidos específicas en los intrones. Una vez ensamblado el *espliceosoma*, el pre-ARNm es cortado, se liberan las secuencias reconocidas como intrones y se empalman los extremos de las secuencias reconocidas como exones, para producir un ARNm maduro.

- **El codón de terminación.** Puede tener tres diferentes conformaciones: TAA, TGA y TAG.

- **La región 3'UTR** que, aunque es transcripta, no es traducida a proteínas. Esta región incluye la secuencia AATAAAA, necesaria para la poliadenilación, una modificación del ARN que consiste en el agregado de una "cola" de unos 200 o 300 nucleótidos de adenina. Estas A son agregadas enzimática-

CÓDIGO GENÉTICO

Cada codón (3 nucleótidos del ARNm) codifica un aminoácido específico.

PRIMERA BASE		SEGUNDA BASE				TERCERA BASE	
	U	C	A	G			
U	UUU Fenil-alanina UUC UUA Leucina UUG	UCU Serina UCC UCA UCG	UAU Tirosina UAC UAA Codón stop UAG	UGU Cisteína UGC UGA Codón stop UGG	U	C	A
C	CUU Leucina CUC CUA CUG	CCU Prolina CCC CCA CCG	CAU Histidina CAC CAA Glutamina CAG	CGU Arginina CGC CGA CGG	U	C	A
A	AUU Isoleucina AUC AUA AUG Metionina Codón inicio	ACU Treonina ACC ACA ACG	AAU Asparagina AAC AAA Lisina AAG	AGU Serina AGC AGA Arginina AGG	U	C	A
G	GUU Valina GUC GUA GUG	GCU Alanina GCC GCA GCG	GAU Ácido aspártico GAC GAA Ácido glutámico GAG	GGU Glicina GGC GGA GGG	U	C	A

La síntesis de proteínas ocurre en los ribosomas, que consisten en dos subunidades, una grande y otra pequeña, cada una formada por ARNr y proteínas. La subunidad grande contiene una depresión en una de sus superficies, en la que se ajusta la subunidad pequeña. El ARNm se inserta en un surco formado entre las superficies de contacto de las dos subunidades.

Para la traducción también se requieren los ARNt, que están plegados en una estructura tridimensional que se asemeja a la letra "L". Tienen un triplete de bases, el anticodón, en el plegamiento central de la estructura secundaria. Una enzima específica, la aminoacil ARNt sintetasa, agrega un aminoácido determinado en el extremo 3' de cada ARNt, convirtiéndolos en aminoacil-ARNt. El apareamiento de bases complementarias codón-anticodón permite trasladar la información especificada en el ARNm a un aminoácido específico.

La traducción se realiza en varias etapas. La iniciación ocurre cuando la subunidad ribosómica más pequeña se une al extremo 5' de una molécula de ARNm y se mueve en sentido 5'→3' hasta encontrar el codón de iniciación AUG. La primera molécula de aminoacil-ARNt, que lleva el aminoácido metionina (Met), se acopla con el codón de iniciación. El complejo de

camente al transcripto; no son parte de la secuencia génica. La cola de poli(A) confiere estabilidad al ARNm, permite que este salga del núcleo y su ausencia inhibe la traducción.

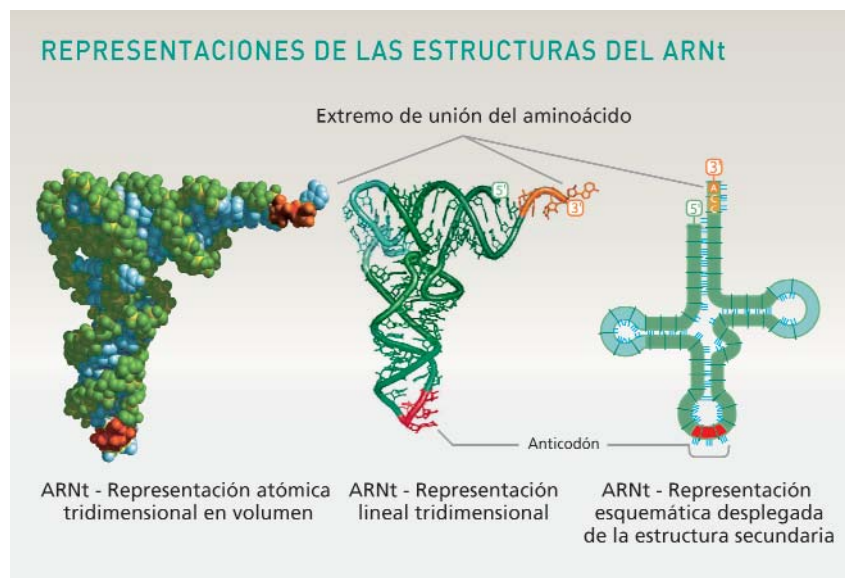
A partir de lo explicado sobre los elementos característicos de un gen eucariota se desprende que sólo podemos hablar de ARNm una vez que el transcripto de ARN (pre-ARNm) ha sufrido las modificaciones en sus extremos 5' (*capping*) y 3' (*poliadenilación*) y hayan sido procesadas sus secuencias intrónicas (*splicing*).

señales de finalización del proceso de traducción. Teniendo en cuenta que las proteínas están compuestas por 20 aminoácidos diferentes y existen 61 codones para especificarlos, es obvio que más de un triplete puede codificar el mismo aminoácido. Por eso decimos que el código genético es redundante o degenerado. Otra característica del código genético es su universalidad: el mismo código es utilizado por todos los organismos.

¿CÓMO SE TRADUCE EL MENSAJE?

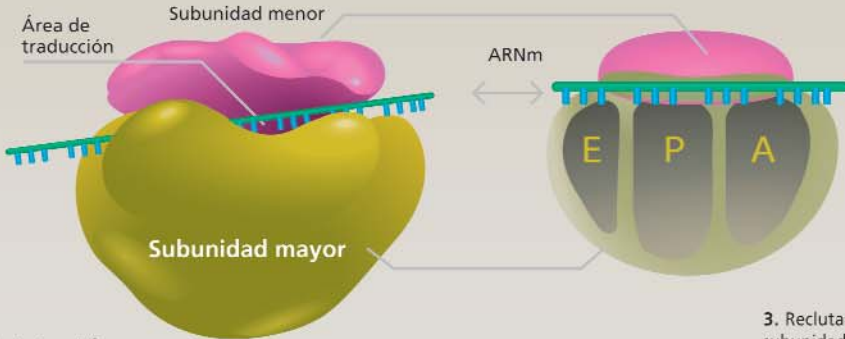
Finalizadas la transcripción, las modificaciones de los extremos 5' y 3' y el *splicing*, el ARNm es exportado al citoplasma donde se traduce el mensaje. Durante la traducción, el orden de nucleótidos del ARNm especifica el orden de aminoácidos en un polipéptido. De acuerdo con el código genético, cada triplete o codón en el ARNm codifica un aminoácido en particular.

El código genético está compuesto por 64 codones: 61 de ellos especifican aminoácidos y los otros 3 son



SÍNTESIS PROTEICA

La representación tridimensional del ribosoma exhibe el sitio donde se sintetiza la nueva cadena peptídica. El corte esquemático permite representar los sitios internos E, P y A, y el surco donde se aloja el ARNm.

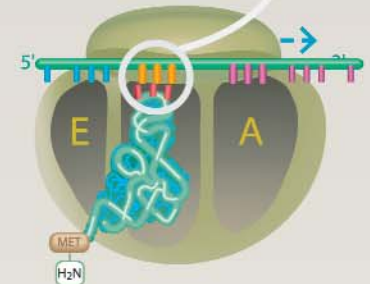


Iniciación

1. Una subunidad menor del ribosoma se une al extremo 5' del ARNm y se mueve en dirección: 5' → 3'.
2. La subunidad menor encuentra el codón de iniciación.

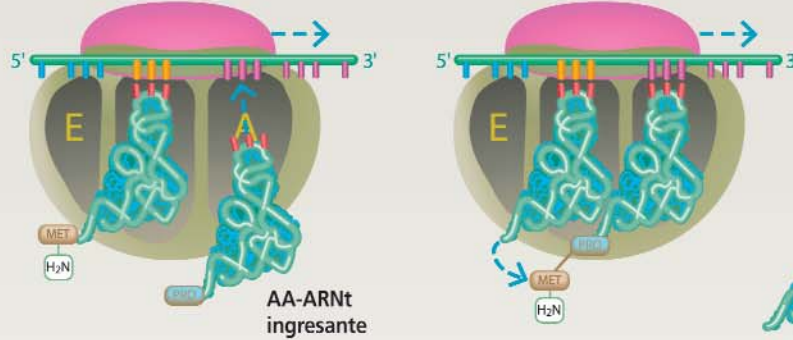


3. Reclutamiento de la subunidad mayor. La iniciación termina al completarse el armado del ribosoma. El ARNt cargado con MET ocupa el sitio P.

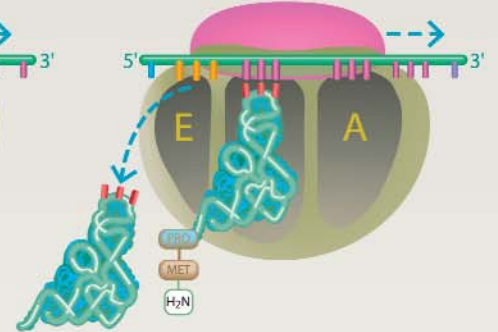


Elongación

1. Otro AA-ARNt se une al ARNm a través del anticodón y ocupa el sitio A del ribosoma.
2. Se produce el enlace peptídico entre los dos aminoácidos.

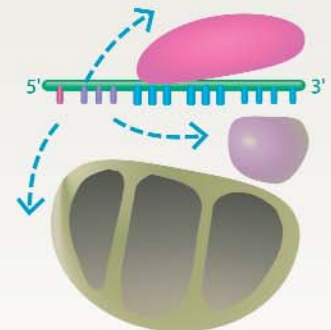
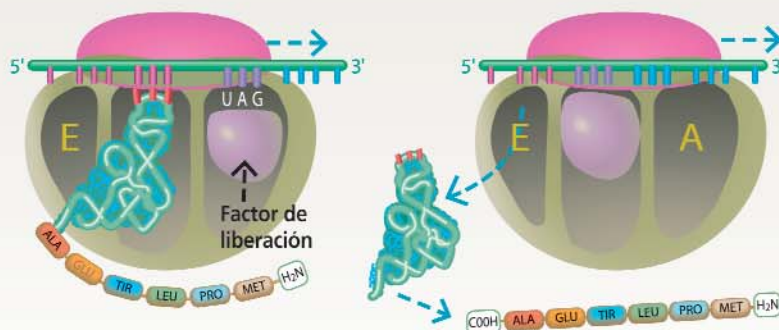


3. El ARNt libre se ubica en el sitio E y es posteriormente liberado.



Terminación

1. El ribosoma alcanza a un codón de terminación. A ese codón se une un factor de liberación.
2. Se libera el polipéptido completo y el último ARNt se desprende del ribosoma.



iniciación se completa cuando la subunidad ribosómica grande se une a estas estructuras. El ribosoma armado contiene ahora 3 sitios: A, P y E; en este momento, el complejo ARNt-MET ocupa el sitio P y los sitios A y E quedan vacantes.

La elongación comienza cuando un segundo aminoacil-ARNt se coloca en el sitio A y su anticodón se acopla con el ARNm. Se forma un enlace peptídico entre los dos aminoácidos. Lo interesante es que esta reacción es catalizada por un pequeño ARN presente en la subunidad mayor del ribosoma. A estos ARN con actividad enzimática se los denomina *ribozimas*. Al mismo tiempo que ocurre la unión de dos aminoácidos se rompe el enlace entre el

primer aminoácido y su ARNt. El ribosoma se mueve a lo largo de la cadena de ARNm en una dirección 5'→3', y el segundo ARNt, con el dipéptido unido (peptidil-ARNt), se mueve desde el sitio A al sitio P. El primer ARNt se mueve hacia el sitio E y se desprende del ribosoma, dejando el aminoácido que transportaba en el péptido en formación. En suma, durante la elongación del péptido, el sitio A une el aminoacil-ARNt correspondiente al aminoácido que deberá incorporarse al péptido en formación, el sitio P une el peptidil-ARNt y el sitio E es el sitio de salida para el ARNt que estuvo previamente unido al sitio P. Un tercer aminoacil-ARNt se coloca en el sitio A y se forma otro enlace peptídico. La

cadena peptídica naciente siempre está unida al ARNt que se está moviendo del sitio A al sitio P, y el aminoacil-ARNt entrante siempre ocupa el sitio A. Este paso se repite una y otra vez incorporando los distintos aminoácidos hasta que se completa el polipéptido.

La terminación ocurre cuando el ribosoma alcanza un codón de terminación, el polipéptido se escinde del último ARNt y el ARNt se desprende del sitio E. El sitio A es ocupado por un factor de liberación que produce la disociación de las dos subunidades del ribosoma.

A modo de resumen, la imagen de la página anterior ilustra los procesos descriptos.

REGULACIÓN DE LA EXPRESIÓN DE GENES EUCARIOTAS

La finalización de varios Proyectos Genoma mostró que el número de regiones que codifican proteínas no presenta una correlación con la complejidad del organismo ni con su origen evolutivo. *C. elegans*, *D. melanogaster* y *Homo sapiens* contienen aproximadamente 19.000, 14.000 y 25.000 genes, respectivamente. ¿Cómo puede el genoma de la mosca de la fruta contener menos genes que el del indudablemente más simple gusano? El conocimiento de la secuencia completa de todos los genes de un individuo no alcanza para responder esta pregunta. "La secuencia no es el final del día. Es sólo el comienzo", dice Sydney Brenner, un líder de estos proyectos. La respuesta está en los diferentes proteomas, es decir, en el conjunto de proteínas que cada uno de ellos expresa. No basta con conocer la secuencia del genoma completo, hay que caracterizar las proteínas que los genes ordenan elaborar, por eso el proteoma es el actual desafío de la biología.

¿Qué determina los tipos y cantidades de las diversas proteínas que caracterizan un organismo o una célula en particular? Todas las células de un organismo multicelular con tejidos especializados poseen los mismos genes. Entonces, ¿cómo es que en un mismo organismo una célula puede diferenciarse en un glóbulo blanco y otra en célula de la piel? Lo que ocurre es que, aunque los genes están siempre presentes en un determinado genoma, no están siempre "encendidos". Decimos que un gen se *expresa* (se enciende) cuando se transcribe y el ARNm se traduce a proteína. El control de cualquiera de los pasos de este proceso puede conducir a una expresión génica diferencial en tipos celulares específicos. Por lo tanto, los glóbulos blancos se diferencian de las células de la piel no porque tengan diferentes genes, sino porque expresan diferentes proteínas.

La regulación de la expresión génica puede ocurrir en distintos niveles:

- Transcripción diferencial de genes, regulando cuáles de los genes son trans-

criptos a ARNm en un momento o tipo celular determinado.

- Procesamiento alternativo de intrones, regulando cuál de los ARN transcritos (o qué parte de ese ARN nuclear) será transportado al citoplasma para ser traducido a proteína.

- Selección en el núcleo de los ARNm que deben ser exportados al citoplasma.

- Traducción selectiva del ARNm, regulando cuál de los ARNm será traducido a proteína en el citoplasma.

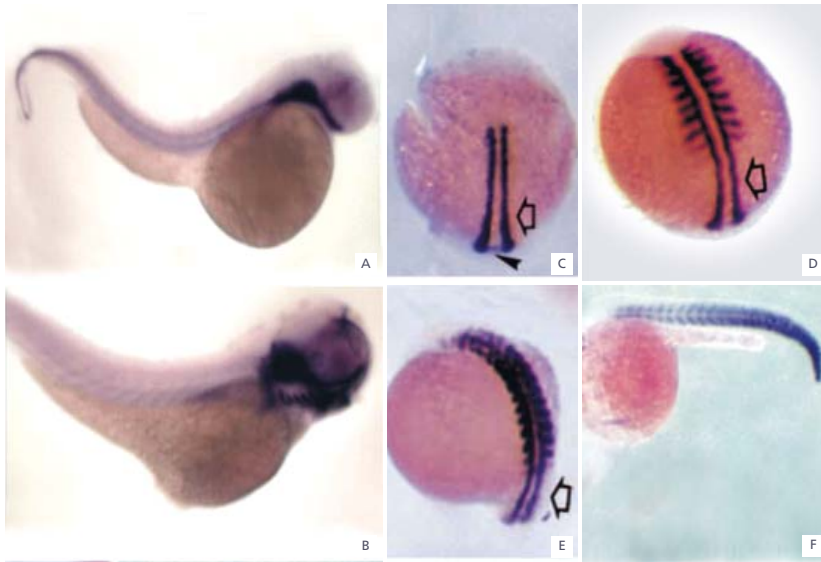
- Desestabilización selectiva de ARNm en el citoplasma.

- Modificación diferencial de proteínas, regulando qué proteínas serán funcionales en la célula.

TRANSCRIPCIÓN DIFERENCIAL DE GENES

Los genes eucariotas están contenidos en un complejo de ADN y proteínas llamado *cromatina*. La unidad básica de la cromatina es el nucleosoma, que está compuesto por un corazón proteico de ocho moléculas de la

EXPRESIÓN DIFERENCIAL DE GENES DURANTE EL DESARROLLO EMBRIONARIO

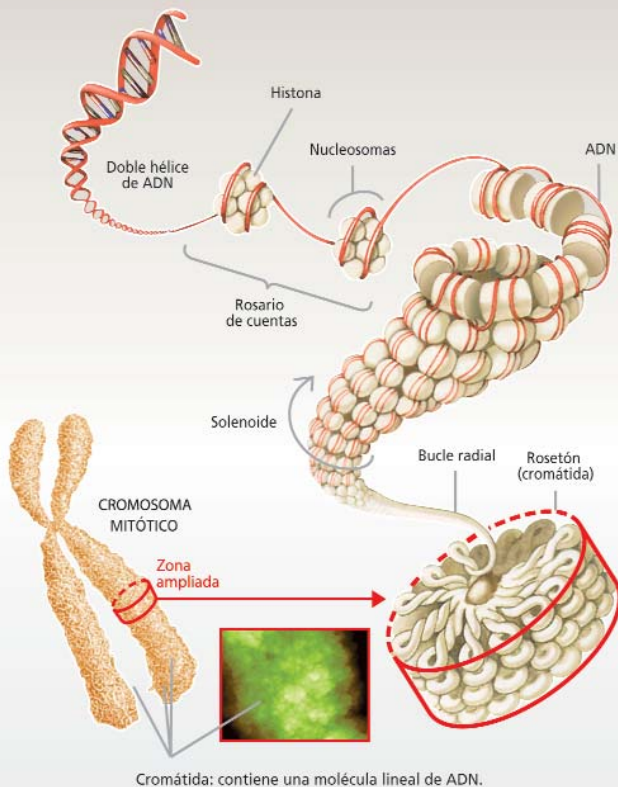


Luego de la fecundación, el huevo se divide por mitosis numerosas veces hasta formar un individuo completo y complejo formado por miles de células que cumplen funciones diferentes y que se organizan en estructuras anatómicas morfológicamente definidas. Durante el desarrollo embrionario no hay pérdida de material genético y el control de la expresión génica es esencial para lograr un desarrollo normal y armónico. En la figura se muestra

(en color violeta intenso) la expresión diferencial de dos genes marcadores típicos en distintos estadios del desarrollo embrionario del pez cebra. Las imágenes A y B corresponden a la expresión del gen *co/2A1*. Este gen está en el genoma de todas las células, pero sólo se expresa en las que van a dar lugar a cartilago. De manera similar, en las figuras C a F se muestra que el gen *myo D* sólo se transcribe en las que formarán músculo.

COMPACTACIÓN DE LA CROMATINA

La cromatina, constituida por ADN y proteínas, se compacta hasta alcanzar el mayor grado de compactación en el cromosoma mitótico.



proteína histona (dos moléculas de cada una de las histonas H2A-H2B e histonas H3-H4) alrededor del cual se enrollan dos vueltas de ADN de aproximadamente 140 pares de bases. Estas estructuras se disponen una a continuación de la otra, separadas por unos 40 pares de bases, asemejándose a un rosario de cuentas. Los nucleosomas se compactan formando solenoides que son estabilizados por la histona H1.

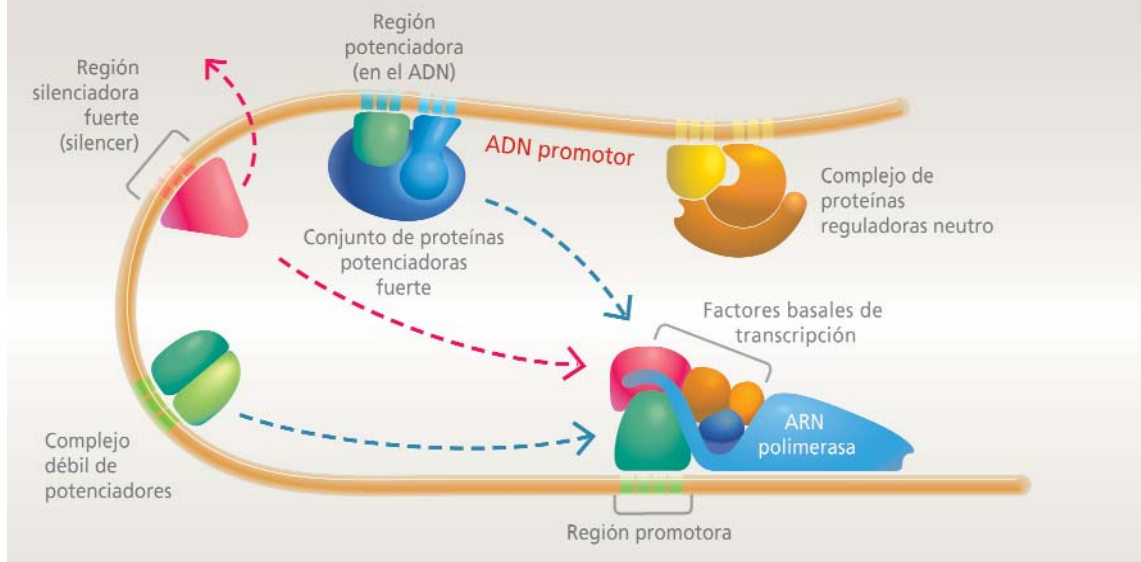
La compactación de la cromatina inhibe la transcripción porque impide el acceso de la maquinaria de transcripción (ARN polimerasa y otras proteínas) a los genes. Existen dos tipos de cromatina: la eucromatina (más laxa) y la heterocromatina (más condensada). La transcripción ocurre sólo durante la interfase y donde la eucromatina está laxa. Algunas regiones heterocromáticas son constantes en todas las células y nunca se expresan. Este tipo de heterocromatina, que se denomina heterocromatina constitutiva, no codifica proteínas. Otras regiones de cromatina condensada, por el contrario, varían de un tipo de célula a otro en un mismo organismo, y regulan la biosíntesis diferencial de proteínas.

Además del control de la transcripción como consecuencia del grado de compactación de la cromatina, existe otro nivel de control del inicio de la transcripción. Se ha visto que aun en regiones donde la cromatina está laxa no siempre hay transcripción de genes. Es decir que, aunque el grado de compactación de la cromatina permita el acceso de la ARN polimerasa, no todos los genes se encenderán, y pueden permanecer apagados en un determinado momento o estadio metabólico de la célula. El control de dónde y cuándo un gen en particular es transcrito ocurre a nivel de ciertas secuencias del gen que no se transcriben y que, por lo tanto, no aparecen en el pre-ARNm ni en el producto polipeptídico. Estas secuencias son los promotores, los potenciadores (*enhancers*) y los silenciadores.

Los promotores son secuencias cortas de ADN (de seis a diez nucleótidos) a los que se une la ARN polimerasa para iniciar la transcripción. La ARN polimerasa de eucariotas no puede unirse directamente a esta región e iniciar la transcripción. Se une y actúa sólo después de que varias proteínas,

CONTROL DEL INICIO DE LA TRANSCRIPCIÓN

Las proteínas con funciones reguladoras se unen a secuencias reguladoras presentes en el ADN y modulan la probabilidad de inicio de la transcripción.



llamadas factores de transcripción basales, se han ensamblado sobre el promotor del gen.

Existen otros factores de transcripción, llamados reguladores, que se unen a las regiones potenciadoras o silenciadoras. Los potenciadores son secuencias de ADN a las que se unen factores de transcripción reguladores y activan la utilización de un promotor, controlando la eficiencia y la velocidad de la transcripción. De manera similar, las secuencias silenciadoras pueden bloquear específicamente la transcripción de un gen en un grupo de células o regular el momento en que un gen se expresa. Estas secuencias pueden estar incluso a distancias muy lejanas del promotor. Son muy variadas y específicas de cada gen, regulan su expresión dependiendo del tipo celular, de señales provenientes de otras células, del entorno, etc.

Una vez que una célula se ha diferenciado ocurre una serie de procesos celulares para mantener estable su patrón de transcripción y asegurar que su progenie conserve las mismas características y propiedades.

PROCESAMIENTO ALTERNATIVO DE INTRONES

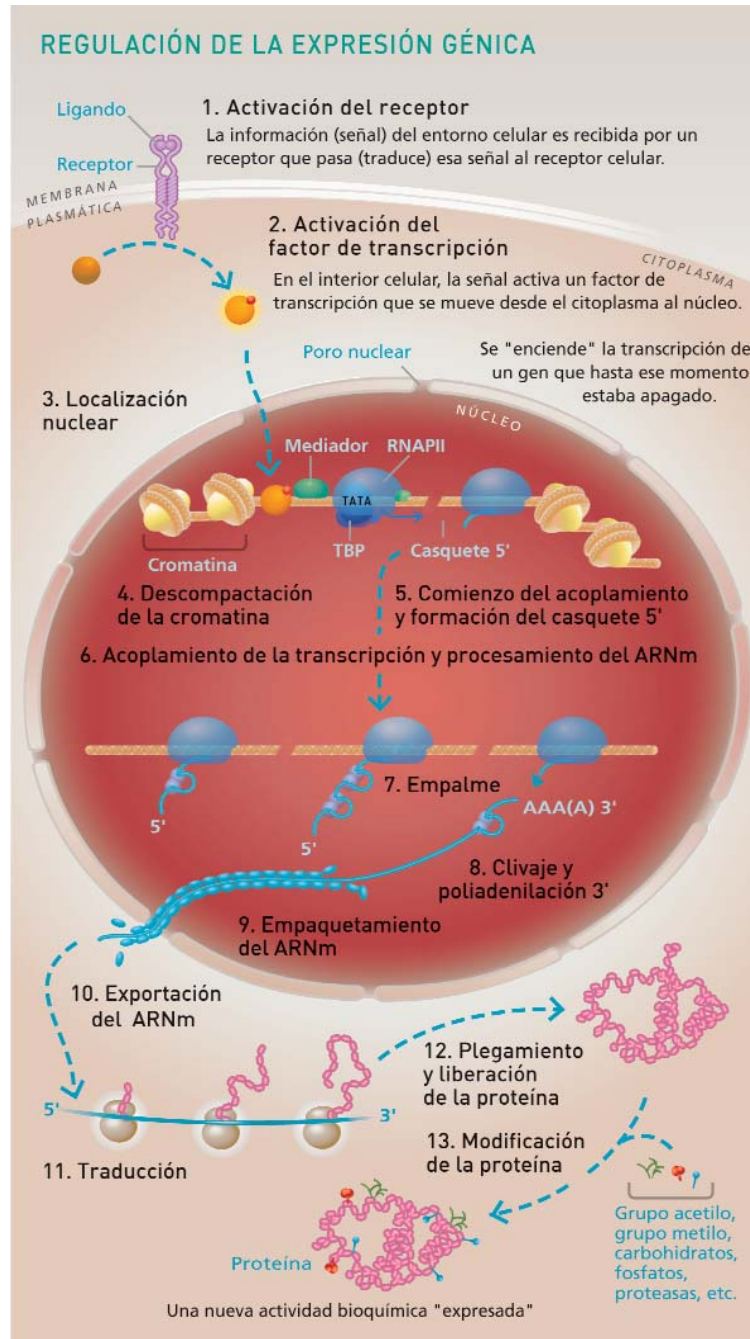
Originalmente se pensó que cualquiera fuera la composición de intrones y exones de un pre-ARNm determinado, este daría lugar al mismo ARNm. Sin embargo, se ha demostrado que diferentes tipos celulares que transcriben el mismo pre-ARNm, pero lo procesan de manera diferencial, generan diferentes ARNm y, por lo tanto, distintas proteínas.

En la ilustración del procesamiento diferencial del pre-ARNm se muestra cómo es posible generar familias de proteínas a partir del procesamiento diferencial de un único pre-ARNm. La eliminación de ciertos exones potenciales en una célula pero no en otras conduce a la formación de proteínas estrechamente relacionadas, pero distintas. Estas proteínas diferentes codificadas por el mismo gen se denominan *isoformas*.

El procesamiento alternativo del pre-ARNm determina qué secuencia presente en el pre-ARNm será eliminada como intrón. Lo que es un intrón en el núcleo de una célula puede no

serlo en el núcleo de otra. En los extremos de los intrones existen secuencias conservadas en todos los genes que pueden ser reconocidas o no por el espliceosoma de forma regulada. En consecuencia, la maduración por corte y empalme es un proceso regulado.

La existencia de un procesamiento alternativo de un pre-ARNm implica que genes con docenas de intrones pueden originar miles de proteínas relacionadas pero diferentes. A modo de ejemplo, describiremos dos genes para los que el procesamiento alternativo de intrones permite generar una extraordinaria diversidad molecular. Uno de ellos es el gen *slo*, que codifica una proteína de membrana presente en las células del oído interno. Esta proteína es, al menos en parte, la responsable de la percepción del amplio rango de frecuencias que el oído puede percibir. Existen al menos ocho sitios de procesamiento alternativo en el pre-ARNm de *slo*, lo que posibilita que más de 500 ARNm diferentes puedan ser finalmente generados. Este procesamiento alternativo es funcionalmente relevante ya que distintas isoformas del gen *slo* participan en la percepción



de diferentes frecuencias de sonido. De esta manera, un mismo gen genera diferentes productos proteicos con distintas funciones fisiológicas.

El otro gen al que nos referiremos es el gen *Dscam* de *Drosophila*. Este gen codifica una proteína de la superficie celular responsable de marcar el camino a los axones de las neuronas en crecimiento hasta alcanzar su adecuada localización en el

cuerpo de la mosca. El pre-ARNm de este gen es procesado alternativamente y, en teoría, puede generar 38.016 isoformas diferentes. Esto es entre dos y tres veces el número de genes que contiene *Drosophila*. Diferentes isoformas del gen *Dscam* presentes en la superficie de una neurona permiten reconocer distintos poblaciones de axones y determinar su destino en el individuo.

Estos ejemplos muestran claramente que el número de genes en un genoma entero puede ser mucho menor que la cantidad de proteínas que ese genoma expresa. Eso podría explicar por qué no hay una correlación directa entre el número de genes y el grado de complejidad de un organismo.

TRADUCCIÓN SELECTIVA DE ARNm

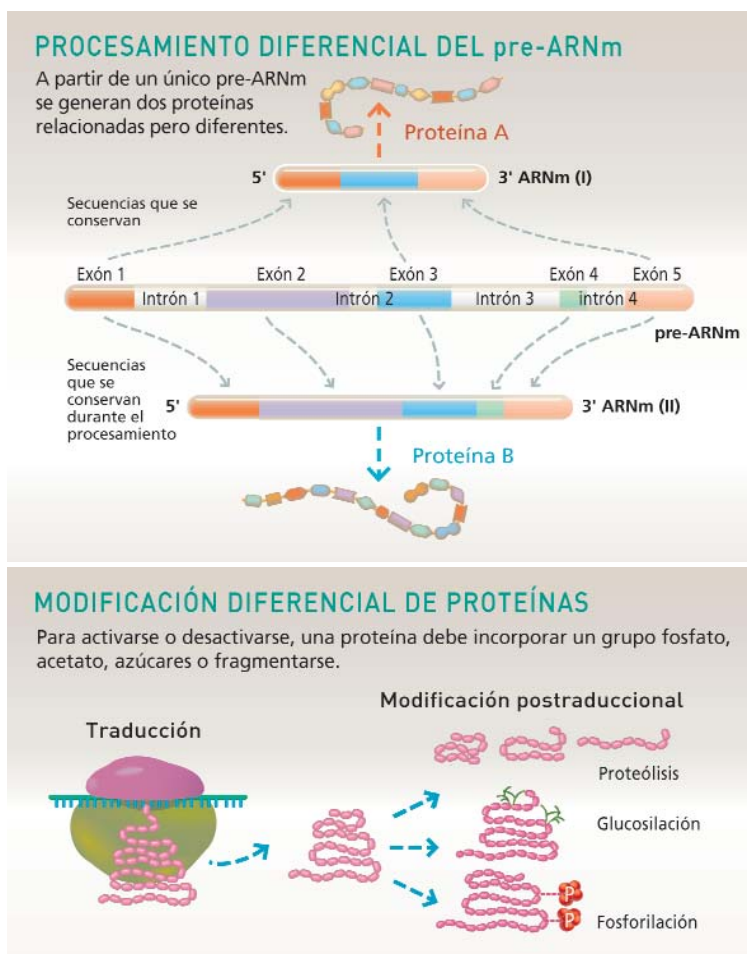
Que el ARNm haya alcanzado el citoplasma no es garantía de que será traducido. La traducción es controlada principalmente en dos niveles: control de la vida media del ARNm e inhibición selectiva de la traducción.

Una misma molécula de ARNm puede ser traducida muchas veces en un polipéptido. Por eso, cuanto más tiempo persista una molécula de ARNm, más proteína se podrá fabricar a partir de ella.

Una inhibición selectiva de la traducción ocurre, por ejemplo, en el ovocito. Esta célula almacena grandes cantidades de ARNm que permanecen "dormidos" hasta la fecundación y son traducidos durante las primeras etapas de la embriogénesis. El control de la inhibición puede ocurrir por unión de proteínas a los extremos 5'UTR o 3'UTR de los mensajeros, dificultando el acceso a los ribosomas, o por eliminación de la cola de poli(A) de los ARNm mientras son almacenados en el ovocito. El ARNm se sintetiza adecuadamente con su agregado de poli(A), pero luego esos nucleótidos de adenina son removidos. En el primer caso, una vez ocurrida la fecundación, las proteínas asociadas pueden ser eliminadas o modificadas, permitiendo el reconocimiento por el ribosoma y la traducción del mensaje. En el segundo caso, una enzima presente en el citoplasma de las células del embrión cataliza el agregado de una cola de poli(A) que capacita ese ARNm para ser traducido. Por llevarse a cabo en el citoplasma, este tipo de poliadenilación se denomina poliadenilación citoplasmática.

MODIFICACIÓN DIFERENCIAL DE PROTEÍNAS

Una vez que un polipéptido ha sido sintetizado pueden ocurrir diferentes cambios para determinar si será biológicamente activo o no. Algunas proteínas no son activas a menos que sean modificadas por agregado covalente de un grupo fosfato (fosforilación) o acetato (acetilación), y también por agregado de azúcares (glicosilación). Otras deben unir iones o incorporar moléculas orgánicas pequeñas llamadas grupos prostéticos o cofactores. Por ejemplo, la hemoglobina, además de la correcta síntesis de los cuatro polipéptidos que la componen, debe incorporar a su estructura un grupo prostético "hemo" para poder cumplir con su función biológica de transportar el oxígeno a través de la sangre. Todos estos procesos ocurren luego de que la traducción finalizó, y por ello se conocen como "modificaciones postraduccionales": son reguladas y sólo en el caso de llevarse a cabo adecuadamente la proteína será funcional y la expresión del gen se hará evidente.



¿QUÉ ES UN GEN?

Según el contexto en que se analiza, un gen puede ser definido de diferentes maneras. Desde el punto de vista mendeliano, un gen es una unidad de herencia que determina una característica fenotípica. De acuerdo con la teoría cromosómica de la herencia, es posible asignar cada uno de los genes de un organismo a un *locus* específico en los cromosomas, por lo que algunas veces se usa el término *locus* como sinónimo de gen. La biología celular define gen como una unidad de información determinada por una secuencia específica de nucleótidos que se halla en un sitio particular de un cromosoma. Todas estas defini-

ciones son útiles, pero no abordan el aspecto funcional. En 1909, el físico británico Archibald Garrod fue el primero en sugerir que los genes dictan los fenotipos por medio de enzimas que catalizan procesos químicos específicos en una célula. Posteriormente, en 1940, Tatum sugirió entonces la hipótesis *un gen: una enzima*. Para explicar el origen de las proteínas que no son enzimas y las formadas por más de un tipo de unidad, fue necesario adoptar una expresión más exacta y global de lo que es un gen y así se propuso la hipótesis *un gen: un polipéptido*. Aun esta definición debe ser refinada y ampliada. Por un lado, la mayoría de los genes eucariotas contienen intrones que pueden ser

procesados de manera regulada mediante *splicing* alternativo; por otro lado, existen largas porciones de esos genes que no poseen un segmento correspondiente a un polipéptido –por ejemplo, las regiones 5' y 3' UTR, la cola de poli(A)–. También es aceptado que las regiones promotoras y las secuencias reguladoras, que no son transcritas a ARN, son parte de un gen funcional ya que deben estar presentes para que ocurra la transcripción. Además, una definición molecular debería ser lo suficientemente amplia como para incluir el ADN que lleva la información para sintetizar ribozimas, ARNr, ARNt y snARN. Una definición más adecuada de gen sería "una región del ADN requerida para la síntesis de un ARN funcional".

EPISTEMOLOGÍA

Agustín Adúriz-Bravo

Al inicio del fascículo se mencionan dos enunciados de la biología molecular de diferente naturaleza: *la hipótesis semiconservativa* y *el dogma central*. En primer lugar tenemos la hipótesis, que es, desde su sentido más literal, una suposición o conjetura: un enunciado provisorio que tiene la capacidad de ser sometido a pruebas empíricas mediante la observación, la experimentación u otro tipo de intervenciones sobre el mundo.

El conjunto de pruebas empíricas que atraviesa una hipótesis se conoce como *contrastación*; la *contrastación* de una hipótesis admite dos resultados extremos y muchos "grises" intermedios. Esos resultados extremos son la *validación*, que significa que la hipótesis pasa la prueba con éxito y es aceptada provisionalmente para seguir avanzando en la investigación científica, y la *refutación*, que refiere a que la

hipótesis no soporta las pruebas y es descartada totalmente como errónea. Los puntos intermedios entre estas dos posturas, que son muy frecuentes en las ciencias naturales, incluyen intervenciones no totalmente concluyentes, interpretaciones provisionales de los datos o validaciones parciales.

Antiguamente se hablaba de "verificación" o de "comprobación" de una hipótesis, queriendo significar que los datos empíricos pueden proporcionar una base segura e incuestionable para los enunciados de la ciencia, y que de ellos cabe decir que son "verdaderos". Pero la historia de la ciencia nos ha mostrado que esta es una mirada epistemológica ingenua. Los resultados de la ciencia, por su naturaleza, son siempre provisorios, perfectibles y sujetos a revisión y crítica.

Buena parte de la provisoriedad de la ciencia se debe a la importante carga teórica que posee la exploración del mundo natural; esta está mediada por las ideas,

las finalidades, los métodos y los valores de los científicos. Con un cambio en cualquiera de todos esos elementos se pueden introducir modificaciones en el conocimiento científico "establecido". Esta idea queda ilustrada en la plaqueta final que recorre los diferentes sentidos que se dio al concepto de gen a lo largo de la historia de la biología.

En segundo lugar, se menciona en el fascículo un dogma, que también podríamos llamar principio o axioma. Este tipo de enunciado es mucho más fundamental, abstracto y estructurante que las hipótesis. No significa esto que no pueda ser reemplazado en la historia de la ciencia, sino que un principio participa del núcleo de una teoría, dando sentido a muchísimos modelos enlazados entre sí. Por tanto, un cuestionamiento a los principios científicos asume la forma de una auténtica "revolución", en la cual muchos elementos de la ciencia son alterados profundamente. Se puede producir,

incluso, lo que llamaríamos un "cambio de paradigma", es decir, una nueva elección de presupuestos teóricos y metodológicos.

Ahora bien, todos los enunciados de la ciencia (hipótesis, dogmas y muchos otros a los que no daremos nombres específicos), independientemente de las diferencias en su grado de validación, son luego puestos en marcha para seguir avanzando y, al mismo tiempo, quedan sujetos a constante refinamiento, revisión y crítica. Esto se puede ver muy bien en el fascículo, en donde enunciados más establecidos coexisten con nuevos hallazgos que son hoy objeto de investigación "puntera".

El estilo del texto científico, como hemos comentado en otro fascículo, parece dar estatuto de "descripción verdadera" a todas sus afirmaciones. Esta es una característica retórica de las ciencias, que cae lejos de la propia metodología autocorrectiva y revisionista que ellas tienen, basada en el uso de los modelos y las inferencias.

Bibliografía

- Alberts, B., D. Brays, J. Lewis, K. Raff, K. Roberts y J. D. Watson: *Biología molecular de la célula*, Barcelona, Omega, 1996; 3ª ed.
- Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts y P. Walter: *Biología molecular de la célula*, Barcelona, Omega, 2002; 4ª ed.
- Curtis, H., N. S. Barnes, A. Schnek y G. Flores: *Biología*, Buenos Aires, Panamericana, 2000.
- De Robertis, E., J. Hib y R. Ponzio: *Biología celular y molecular*, Buenos Aires, El Ateneo, 1996.
- Lordish, H., A. Berk, S. L. Zipurskey, P. Matsudaira, D. Baltimore y J. Darnell: *Biología celular y molecular*, Buenos Aires, Panamericana, 2002; 4ª ed.

Purves, W. K., D. Sadava, G. H. Orians y H. C. Heller: *Vida. La ciencia de la biología*, Buenos Aires, Panamericana, 2002.

Agradecimientos

El equipo de Publicaciones de la Dirección Nacional de Gestión Curricular y Formación Docente agradece a las siguientes instituciones y personas por permitirnos reproducir material fotográfico y colaborar en la documentación de imágenes: Institut Pasteur (Francia); Chantal Policieux.



Ministro de Educación, Ciencia y Tecnología, Lic. Daniel Filmus
Secretario de Educación, Lic. Juan Carlos Tedesco
Subsecretaria de Equidad y Calidad, Lic. Alejandra Birgin
Directora Nacional de Gestión Curricular y Formación Docente,
 Lic. Laura Pitman

Coordinadora del Área de Ciencias Naturales, Lic. Nora Bahamonde
 Coordinadora del Área de Desarrollo Profesional, Lic. Silvia Storino
 Coordinadora del Programa de Capacitación Explora, Lic. Viviana Celso
 Coordinadora de Publicaciones,
 Lic. Raquel Franco

Coordinación y documentación,
 Lic. Rafael Blanco
 Edición, Lic. Gonzalo Blanco
 Diseño y diagramación,
 DG María Eugenia Más
 Corrección, Norma A. Sosa Pereyra

www.me.gov.ar