**Unidad 7 El dogma central de la biología molecular**

* 7.1 Genes y genomas

Genome Structure

Chromatin

Nucleosome

* 7.2 La replicación del DNA
* 7.3 Manipulación del DNA

**Unidad 9 Expresión del genoma**

* 9.1 Mecanismos de transcripción

El primer paso en la transferencia de información del gen a la proteína es producir una hebra de ARN cuya secuencia base coincida con la secuencia base de un segmento de ADN, a veces seguida de la modificación de ese ARN para prepararla para sus funciones celulares específicas. Por lo tanto, el ARN se produce mediante un proceso que copia la secuencia de nucleótidos del ADN. Debido a que este proceso recuerda a la transcripción (copia) de palabras escritas, la síntesis de ARN se llama transcripción. Se dice que el ADN se transcribe en ARN, y el ARN se llama transcripción.

**Descripción general:** El ADN como plantilla de transcripción ¿Cómo se codifica la información en la molécula de ADN se transfiere a la transcripción de ARN? La transcripción se basa en el emparejamiento complementario de bases. Considere la transcripción de un segmento cromosómico que constituye un gen. En primer lugar, las dos hebras de la doble hélice de ADN se separan localmente, y una de las hebras separadas actúa como una plantilla para la síntesis de ARN. En el cromosoma en general, ambas hebras de ADN se utilizan como plantillas, pero, en cualquier gen, solo se utiliza una hebra y, en ese gen, siempre es la misma hebra, comenzando en el extremo 3′ del gen de la plantilla (Figura 8-3).

A continuación, los ribonucleótidos que se han sintetizado químicamente en otras partes de la célula forman pares estables con sus bases complementarias en la plantilla. El ribonucleótido Se Empareja Con El ADN, GwithC, CwithG Y Uwith A. Cada ribonucleótido se coloca frente a su base completa por la enzima ARN polimerasa. Esta enzima se une al ADN y se mueve a lo largo de él, vinculando los ribonucleótidos alineados para hacer una molécula de ARN en constante crecimiento, como se muestra en la Figura 8-4a. Por lo tanto, ya vemos los dos principios de complementariedad de la base y la unión nucleico-ácido-proteína en acción (en este caso, la unión de la ARN polimerasa). Hemos visto que el ARN tiene un extremo de 5' y un extremo de 3'. Durante la síntesis, el crecimiento del ARN siempre está en la dirección de 5′-a-3′; en otras palabras, los nucleótidos siempre se añaden en una punta de crecimiento de 3′, como se muestra en la Figura 8-4b. Debido a que las hebras de ácido nucleico complementarias están orientadas de manera opuesta, el hecho de que el ARN se sintetiza de 5′ a 3′ significa que la hebra de la placa de temperatura debe estar orientada de 3′ a 5′. A medida que una molécula de ARN polimerasa se mueve a lo largo del gen, desenrolla la doble hélice de ADN por delante de ella y rebobina el ADN que ya ha sido transcrito. A medida que la molécula de ARN se alarga progresivamente, el extremo 5′ del ARN se desplaza de la plantilla y la burbuja de transcripción se cierra detrás de la polimerasa. Los "trenes" de ARN polimerasas, cada uno sintetizando un mol de ARN, se mueven a lo largo del gen (Figura 8-5).

También hemos visto que las bases en la transcripción y la plantilla son completas. En consecuencia, la secuencia de nucleótidos en el ARN debe ser la misma que en la hebra no modelo del ADN, excepto que las T son reemplazadas por U, como Se muestra en la Figura 8-6. Cuando las secuencias de base de ADN se citan en la literatura científica, la secuencia de la hebra que no es de plantilla se da convencionalmente porque esta secuencia es la misma que la que se encuentra en el ARN. Por esta razón, la hebra no de plantilla del ADN se conoce como la hebra de codificación. Esta distinción es extremadamente importante tener en cuenta cuando se discute la transcripción.

La transcripción es asimétrica: solo se utiliza una hebra del ADN de un gen como plantilla para la transcripción. Esta hebra está en la orientación de 3′-a-5′, y el ARN se sintetiza en la dirección de 5′-a-3′.

**Etapas de la transcripción**

La secuencia de codificación de proteínas en un gen es un segmento relativamente pequeño de ADN incrustado en una molécula de ADN mucho más larga (el cromosoma). ¿Cómo se transcribe el segmento apropiado en una molécula de ARN de una sola cadena de longitud correcta y secuencia de nucleótidos? Debido a que el ADN de un cromosoma es una unidad continua, la maquinaria de transcripción debe dirigirse al comienzo de un gen para comenzar a transcribir en el lugar correcto, continuar transcribiendo la longitud del gen y, finalmente, dejar de transcribir en el otro extremo. Estas tres etapas distintas de transcripción se denominan iniciación, alargamiento y terminación. Aunque el proceso general de transcripción es notablemente similar en procariotas y eucariotas, hay diferencias importantes. Por esta razón, seguiremos las tres etapas primero en los procariotas (usando la bacteria intestinal E. coli como ejemplo) y luego en los eucariotas. Iniciación en procariotas ¿Cómo encuentra la ARN polimerasa el punto de partida correcto para la transcripción? En los procariotas, la ARN polimerasa generalmente se une a una secuencia de ADN específico llamada promotor, ubicada cerca del inicio de la región transcrita. Un promotor es una parte importante de la región reguladora de un gen. Recuerde que, debido a que la síntesis de un transcripción de ARN comienza en su extremo de 5′ y continúa en la dirección de 5′ a 3′, la convención es dibujar y referirse a la orientación del gen en la dirección de 5′ a 3′, también. Por esta razón, generalmente se muestra la hebra de ADN de la placa no temática. Por lo general, el extremo de 5 pulgadas se dibuja a la izquierda y el de 3 pulgadas a la derecha. Desde este punto de vista, debido a que el promotor debe estar cerca del final del gen donde comienza la transcripción, se dice que está en el extremo 5′ del gen; por lo tanto, la región promotora también se llama la región reguladora 5′ (Figura 8-7a).

La primera base transcrita está siempre en el mismo lugar, designado como el sitio de iniciación. El promotor se conoce como aguas arriba del sitio de iniciación porque se encuentra delante del sitio de iniciación (5′ del gen), en la dirección opuesta a la dirección de la transcripción. Un sitio aguas abajo se ubicaría más adelante en la dirección de la transcripción. Por convención, la primera base de ADN que se transcribe es Numerado +1. Las posiciones de los nucleótidos aguas arriba del sitio de iniciación están indicadas por un signo negativo (−) y las aguas abajo por un signo positivo (+). La figura 8-7b muestra las secuencias promotoras de siete genes diferentes en el genoma de E. coli. Debido a que la misma ARN polimerasa se une a las secuencias promotoras de estos diferentes genes, las similitudes entre los promotores no son sorprendentes. En particular, dos regiones de gran similitud aparecen en prácticamente todos los casos. Estas regiones se han denominado regiones -35 (menos 35) y -10 porque se encuentran en 35 pares de bases y 10 pares de bases, respectivamente, aguas arriba de la primera base transcrita. Se muestran en amarillo en la Figura 8-7b. Como puede ver, las regiones -35 y -10 de diferentes genes no tienen que ser idénticas para realizar una función similar. Sin embargo, es posible llegar a una secuencia de nucleótidos, llamada secuencia de consenso, que está de acuerdo con la mayoría de las secuencias. La secuencia de consenso del promotor de E. coli se muestra en la parte inferior de la Figura 8-7b. Una holoenzima de ARN polimerasa (ver siguiente párrafo) se une al ADN en este punto, luego desenrolla la doble hélice del ADN y comienza la síntesis de una molécula de ARN. Tenga en cuenta en la Figura 8-7a que la parte del gen que codifica la proteína generalmente comienza en una secuencia ATG, pero el sitio de iniciación, donde comienza la transcripción, suele estar muy arriba de esta secuencia. La parte intermedia se conoce como la región no traducida de 5′ (5′ UTR).

La ARN polimerasa bacteriana que escanea el ADN en busca de una secuencia promotora se llama holoenzima ARN polimerasa (Figura 8-8). Este complejo multisubunitario se compone de las cinco subunidades de la enzima central (dos subunidades de a, una de β, una de β′ y una de ω) más una subunidad llamada factor sigma (σ). Las dos subunidades a ayudan a ensamblar la enzima y promover las interacciones con las proteínas reguladoras, la subunidad β es activa en la catálisis, la subunidad β′ se une al ADN y La subunidad ω tiene un papel en el ensamblaje de enzimas y la regulación de la expresión génica. La subunidad σ se une a las regiones -10 y -35, posicionando así la holoenzima para iniciar la transcripción correctamente en el sitio de inicio (ver Figura 8-8a). La subunidad σ también tiene un papel en la separación (fusión) de las hebras de ADN alrededor de la región -10 para que la enzima central pueda unirse firmemente al ADN en preparación para la síntesis de ARN. Después de que la enzima central se une, comienza la transcripción y la subunidad σ se disocia del resto del complejo (ver Figura 8-8b).

E. coli, como la mayoría de las otras bacterias, tiene varios factores σ diferentes. Uno, llamado σ70 porque su masa en kilodaltones es 70, es la subunidad primaria σ utilizada para iniciar la transcripción de la gran mayoría de los genes de E. coli. Otros factores σ reconocen diferentes secuencias promotoras. Por lo tanto, al asociarse con diferentes factores σ, la misma enzima central puede reconocer diferentes secuencias promotoras y transcribir diferentes conjuntos de genes. Alargamiento A medida que la ARN polimerasa se mueve a lo largo del ADN, desenrolla el ADN por delante de él y rebobina el ADN que ya ha sido transcrito. De esta manera, mantiene una región de ADN de una sola cadena, llamada burbuja de transcripción, dentro de la cual se expone la hebra de la plantilla. En la burbuja, la polimerasa monitorea la unión de un trifosfato de ribonucleósido libre a la siguiente base expuesta en la plantilla de ADN y, si hay una coincidencia complementaria, la añade a la cadena.

La figura 8-9a da una imagen física de la elongación. Dentro de la burbuja, los últimos ocho o nueve nucleótidos añadidos a la cadena de ARN forman un híbrido ARN-ADN mediante el emparejamiento complementario de la base con la hebra de la plantilla. A medida que la longitud de la cadena de ARN en su extremo de 3′, el extremo de 5′ se extruye aún más de la polimerasa. Los pares de bases completas se rompen en el punto de salida, dejando la hebra de extrusión de una sola hebra.

**Terminación** La transcripción de un gen individual continúa más allá del segmento de codificación de proteínas del gen, creando una región no traducida de 3′ (3′ UTR) al final de la transcripción. La alargación continúa hasta que la ARN polimerasa reconoce secuencias especiales de nucleótidos que actúan como una señal para la terminación de la cadena. El encuentro con los nucleótidos de señal inicia la liberación del ARN naciente y la enzima de la plantilla (Figura 8-9b). Los dos mecanismos principales para la terminación en E. coli (y otras bacterias) se denominan intrínsecos y dependientes de rho. En el mecanismo intrínseco, la terminación es directa. Las secuencias de terminadores contienen alrededor de 40 pares de bases, que terminan en un tramo rico en GC que es seguido por una cadena de seis o más A. Debido a que G y C en la plantilla darán C y G, respectivamente, en la transcripción, el ARN en esta región también es rico en GC. Estas bases C y G son capaces de formar enlaces de hidrógeno complementarios entre sí, lo que resulta en un bucle de tallo de alfiler de pelo de ARN (Figura 8-10). Recuerde que el par de bases G-C es más estable que el par A-T porque está unido a hidrógeno en tres sitios, mientras que el par A-T (o A-U) se mantiene unido por solo dos enlaces de hidrógeno. Las horquillas de ARN con tallos que son en gran medida pares G-C son más estables que las horquillas con tallos que son en gran medida pares A-U. La estructura de la horquilla va seguida de una cadena de aproximadamente ocho U que son complementarias a los residuos de A en la plantilla de ADN.

Normalmente, en el curso del alargamiento de la transcripción, la ARN polimerasa se detendrá si el híbrido corto de ADN-ARN en la burbuja de transcripción es débil y dará marcha atrás para estabilizar el híbrido. Al igual que la de las horquillas, la fuerza del híbrido está determinada por el número relativo de pares de bases G-C en comparación con los pares de bases A-U (o los pares de bases A-T en los híbridos de ARN-ADN). En el mecanismo intrínseco, se cree que la polimerasa se detiene después de sintetizar la U (A-U forma un híbrido ADN-ARN débil). Sin embargo, la polimerasa de retroceso se encuentra con el bucle de alfiler del cabello. Este obstáculo desencadena la liberación de ARN de la polimerasa y la liberación de la polimerasa de la plantilla de ADN. El segundo tipo de mecanismo de terminación requiere la ayuda de una proteína llamada factor rho. Esta proteína reconoce las secuencias de nucleótidos que actúan como señales de terminación para la ARN polimerasa. Los ARN con señales de terminación dependientes de rho no tienen la cadena de residuos de U en su extremo de 3 ′ y generalmente no tienen bucles de horquilla. En su lugar, tienen una secuencia de aproximadamente 40 a 60 nucleótidos que es rica en residuos de C y pobre en residuos de G e incluye un segmento aguas arriba llamado sitio de rutina (utilización de rho). Rho es un hexámero que consta de seis subunidades idénticas que se unen a una cadena de ARN naciente en el sitio de la rutina. Estos sitios se encuentran justo aguas arriba de (recuerde que aguas arriba significa 5′ de) secuencias en las que la ARN polimerasa tiende a pausarse. Después de unirse, rho facilita la liberación del ARN de la ARN polimerasa. Por lo tanto, la terminación rho-dependiente implica la unión de rho a la rutina, la pausa de la polimerasa y la disociación mediada por rho del ARN de la ARN polimerasa.

Las transcripciones procariotas se inician a 5 ′ de la región codificante de los genes cuando la ARN polimerasa se une a una secuencia promotora de consenso o cuando se asocia con un factor σ que la guía a una secuencia promotora sin consenso. La terminación de la transcripción se produce en secuencias especiales de 3 ′ de la región de codificación que son intrínsecas o dependientes de rho.

* 9.2 RNA splicing

Pequeños ARN nucleares (snRNA): el mecanismo de empalme de exones Después del descubrimiento de exones e intrones, los científicos centraron su atención en el mecanismo de empalme de ARN. Debido a que los intrones deben eliminarse con precisión y los exones unidos con precisión, el primer enfoque fue comparar las secuencias de pre-ARNm en buscas de pistas sobre cómo se reconocen los intrones y los exones. La figura 8-15 muestra las uniones exón-intron de los pre-ARNm. Estas uniones son los sitios en los que tienen lugar las reacciones de empalme. En estas uniones, se encontró que ciertos nucleótidos específicos eran casi idénticos entre genes y especies; han sido altamente conservados porque participan en las reacciones de empalme. Cada intron se corta en cada extremo, y estos extremos de intron casi siempre tienen GU en el extremo de 5′ y AG en el extremo de 3′ (la regla GU-AG). Otro sitio invariante es un residuo A (el punto de rama A) entre 15 y 45 nucleótidos aguas arriba del sitio de empalme de 3′. Los nucleótidos que flanquean los altamente conservados también se conservan, pero en menor grado. La existencia de Las secuencias de nucleótidos conservados en las uniones de empalme sugirieron que debe haber maquinaria celular que reconozca estas secuencias y lleve a cabo el empalme. Como suele ocurrir en la investigación científica, la maquinaria de empalme se encontró por accidente y el mecanismo de empalme fue completamente inesperado. Un hallazgo fortuito en el laboratorio de Joan Steitz llevó al descubrimiento de componentes de la maquinaria de empalme. Los pacientes con una variedad de enfermedades autoinmunes, incluida la eritematosis lúpica sistémica, producen anticuerpos contra sus propias proteínas. En el curso del análisis de muestras de sangre de pacientes con lupus, Steitz y sus colegas identificaron anticuerpos que podrían unirse a un gran complejo molecular de pequeños ARN y proteínas. Debido a que este complejo de riboproteínas se localizó en el núcleo, los componentes de ARN se llamaron pequeños ARN nucleares. Se encontró que los snRNAs eran complementarios a las secuencias de consenso en las uniones de empalme, lo que llevó a los científicos a plantear la hipótesis de un papel de los ARNn en la reacción de empalme. Ahora se sabe que los nucleótidos conservados en la transcripción son reconocidos por cinco pequeñas ribonucleoproteínas nucleares (snRNP), que son complejos de proteínas y uno de los cinco ARNs sn (U1, U2, U4, U5 y U6). Estos snRNP y más de 100 proteínas adicionales son parte del spliceosoma, la gran máquina biológica que elimina los intrones y une los exones. Los componentes del spliceosoma interactúan con el CTD, como se sugiere en la Figura 8-13b.

Estos componentes del spliceosoma se unen a las secuencias de intron y exones, como se muestra en la Figura 8-16. Los snRNP U1 y U2 ayudan a alinear los sitios de empalme en cualquiera de los extremos de un intrón formando enlaces de hidrógeno con las secuencias de intron y exon conservadas. Luego, los snRNP reclutan a U4, U5 y U6 para formar el espli- ceosoma, que cataliza la eliminación del intrón a través de dos pasos de empalme consecutivos (ver Figura 8-16). El primer paso une un extremo del intrón a la adenina interna conservada, formando una estructura de bucle, la forma de la lariat de un niño de vaca. El segundo paso libera el lariat y se une a los dos exones adyacentes. La figura 8-17 representa la química detrás de la eliminación de intron y el empalme de exones. Químicamente, los dos pasos son reacciones de transesterificación entre los nucleótidos conservados. Los grupos de hidroxilo en las posiciones de 2′ y 3′ de los ribonucleótidos son participantes clave de la reacción.

Los intrones autoensalizables y el mundo del ARN. Dos casos excepcionales de procesamiento de ARN llevaron a un descubrimiento considerado por algunos como tan importante como el de la estructura de doble hélice del ADN. En 1981, Tom Cech y sus compañeros de trabajo informaron que, en un tubo de ensayo, la transcripción primaria de un ARNr del protozoo ciliado Tetrahymena podría extirpar un intrón de 413 nucleótidos de sí mismo sin la adición de ninguna proteína (Figura 8-18). Posteriormente, se ha demostrado que otros intrones tienen esta propiedad y han llegado a ser conocidos como intrones autoensalantes. Unos años antes, mientras estudiaba el procesamiento del ARNt en bacterias, Sidney Altman identificó una ribonucleoproteína (llamada RNasa P) responsable de cortar la molécula de pre-ARNt en un sitio específico. La gran sorpresa llegó cuando determinaron que la actividad catalítica de la RNasa P residía en el componente de ARN de la enzima en lugar de en el componente proteico. Los hallazgos de Cech y Altman se consideran descubrimientos históricos porque marcaron la primera vez que se demostró que las moléculas biológicas distintas de la proteína catalizaban las reacciones. Como tal, era apropiado que recibieran el Premio Nobel de Química en 1989. El descubrimiento de trones autoempalmados ha llevado a un reexamen del papel de los snRNA en el spliceosoma. Los estudios más recientes indican que la eliminación de intrones está catalizada por los snRNA y no por el componente proteico del spliceosoma. Como verá en el capítulo 9, los ARN en el ribosoma (los ARNr), no las proteínas ribosómicas, tienen el papel central en la mayoría de los eventos importantes de la síntesis de proteínas. Los numerosos ejemplos de ribozimas han proporcionado evidencia sólida para una teoría llamada el mundo del ARN, que sostiene que el ARN debe tener Fue el material genético en las primeras células porque solo se sabe que el ARN codifica la información genética y cataliza las reacciones biológicas.

La eliminación de Intron y la unión de exones son catalizadas por moléculas de ARN. En los eucariotas, los snRNAs del spliceosoma catalizan la eliminación de los trones del pre-ARNm. Algunos intrones son autoensalmes; en estos casos, el propio intrón cataliza su propia eliminación. Los ARN capaces de catálisis se denominan ribozimas.

* 9.3 Traducción y el código genético

Código genético La hipótesis de un gen-un polipéptido de Beadle y Tatum (ver Capítulo 6) fue la fuente de la primera visión emocionante de las funciones de los genes: los genes eran de alguna manera responsables de la función de las enzimas, y cada gen aparentemente controlaba una enzima. Esta hipótesis se convirtió en uno de los grandes conceptos unificadores de la biología porque proporcionó un puente que reunió los conceptos y las técnicas de investigación de la genética y la bioquímica. Cuando se dedujo la estructura del ADN en 1953, parecía probable que hubiera una correspondencia lineal entre la secuencia de nucleótidos en el ADN y la secuencia de aminoácidos en una proteína. Pronto se dedujo que la secuencia de ácido nucleico en el ARNm que va de 5′ a 3′ corresponde a la secuencia de aminoácidos que va de N-terminal a C-terminal. Si los genes son segmentos de ADN y si una hebra de ADN es solo una cadena de nucleo-tareas, entonces la secuencia de nucleótidos debe dictar de alguna manera la secuencia de aminoácidos en las proteínas. ¿Cómo dicta la secuencia de ADN la secuencia de proteínas? La analogía con un código me viene a la mente de inmediato. La lógica simple nos dice que, si los nucleótidos son las "letras" de un código, entonces una combinación de letras puede formar "palabras" que representan diferentes aminoácidos. En primer lugar, debemos preguntar cómo se lee el código. ¿Se superpone o no se superpone? Entonces debemos preguntar cuántas letras en el ARNm componen una palabra, o codón, y qué codón o codones representan cada aminoácido. El desciframiento del código genético es la historia que se cuenta en esta sección.

Códigos superpuestos versus no superpuestos La figura 9-4 muestra la diferencia entre un código superpuesto y un código no superpuesto. El ejemplo muestra un código de tres letras o triplete. Para un código que no se superponga, los aminoácidos consecutivos se especifican mediante palabras clave consecutivas (codones), como se muestra en la parte inferior de la Figura 9-4. Para un código superpuesto, los aminoácidos consecutivos se especifican mediante codones que tienen algunas bases consecutivas en común; por ejemplo, las dos últimas bases de un codón también pueden ser las dos primeras bases del siguiente codón. Los codones de over-lapping se muestran en la parte superior de la Figura 9-4. Por lo tanto, para la secuencia AUUGCUCAG en un código que no se superpone, los tres trillizos AUU, GCU y CAG codifican los tres primeros aminoácidos, respectivamente. Sin embargo, en un código superpuesto, los trillizos AUU, UUG y UGC codifican los tres primeros aminoácidos si la superposición es de dos bases, como se muestra en la Figura 9-4. En 1961, ya estaba claro que el código genético no se superponía. Los análisis de proteínas alteradas por la mutación mostraron que solo un solo aminoácido cambia a la vez en una región de la proteína. Este resultado se predice mediante un código que no se superpone. Como se puede ver en la Figura 9-4, un código superpuesto predice que un solo cambio de base alterará hasta tres aminoácidos en posiciones adyacentes de la proteína. Número de letras en el codón Si se lee una molécula de ARNm de un extremo a otro, solo se puede encontrar una de las cuatro bases diferentes, A, U, G o C, en cada posición. Por lo tanto, si las palabras que codifican aminoácidos fueran de una letra, solo serían posibles cuatro palabras. Este vocabulario no puede ser el código genético porque debemos tener una palabra para cada uno de los 20 aminoácidos que se encuentran comúnmente en las proteínas celulares. Si las palabras fueran dos letras, entonces 4 × 4 = 16 palabras sería posible; por ejemplo, AU, CU o CC. Este vocabulario todavía no es lo suficientemente grande. Si las palabras tienen tres letras, entonces 4 × 4 × 4 = 64 palabras son posibles; por ejemplo, AUU, GCG o UGC. Este vocabulario proporciona palabras más que suficientes para describir los aminoácidos. Podemos concluir que la palabra clave debe consistir en al menos tres nucleótidos. Sin embargo, si todas las palabras son "trillizos", entonces las posibles palabras están en un exceso considerable de las 20 necesarias para nombrar los aminoácidos comunes. Volveremos a estos codónes en exceso más adelante en el capítulo. Uso de supresores para demostrar un código de triplete. La prueba convincente de que un codón tiene, de hecho, tres letras (y no más de tres) proviene de hermosos experimentos genéticos reportados por primera vez en 1961 por Francis Crick, Sidney Brenner y sus compañeros de trabajo. Estos experimentos utilizaron mutantes en el locus rII del fago T4. El uso de mutaciones rII en el análisis de recombinación se discutió en el capítulo 5. Phage T4 suele ser capaz de crecer en dos cepas diferentes de E. coli, llamadas B y K. Sin embargo, las mutaciones en el gen rII cambian el rango de huéspedes del fago: los fagos mutantes todavía pueden crecer en un huésped de E. coli B, pero no pueden Crecer en un huésped de E. coli K. Las mutaciones que causan este fenotipo de rII se indujeron mediante el uso de una sustancia química llamada proflavina, que se pensaba que actuaba mediante la adición o eliminación de pares de un solo nucleótido en el ADN. (Esta suposición se basa en evidencia experiencial que no se presenta aquí). Los siguientes ejemplos ilustran la acción de la proflavina sobre el ADN de doble cadena.

Comenzando con una mutación particular inducida por la proflavina llamada FCO, Crick y sus colegas encontraron "reversiones" (reversiones de la mutación) que fueron capaces de crecer en la cepa K de E. coli. El análisis genético de estas placas reveló que los "revertientes" no eran idénticos a los verdaderos tipos salvajes. De hecho, se descubrió que la reversión se debía a la presencia de una segunda mutación en un sitio diferente al de la FCO, aunque en el mismo gen.}

Esta segunda mutación "suprimió" la expresión mutante de la FCO original. Recuerde del Capítulo 6 que una mutación supresora contrarresta o suprime los efectos de otra mutación para que la bacteria se parece más al tipo salvaje. ¿Cómo podemos explicar estos resultados? Si asumimos que el gen se lee solo desde un extremo, entonces la adición o eliminación original inducida por la proflavina podría resultar en una mutación porque interrumpe un mecanismo de lectura normal que establece el grupo de bases que se leen como palabras. Por ejemplo, si cada grupo de tres bases en el ARNm resultante hace una palabra, entonces el "marco de lectura" podría establecerse tomando las tres primeras bases del final como la primera palabra, las siguientes tres Introducción al Análisis Genético, 11e como la segunda palabra, y así sucesivamente. En ese caso, una adición inducida por proflavina o la eliminación de la Figura 09UN2 #928 de a05si/n12g/l1e4pair en el ADN cambiaría el marco de lectura en el mRNA Dragonfly Media Group de ese punto correspondiente, causando que todas las palabras siguientes se lean mal. Tal mutación por cambio de marco podría reducir la mayor parte del mensaje genético a galimatías. Sin embargo, el marco de lectura adecuado podría restaurarse mediante una inserción o eliminación compensatoria en otro lugar, dejando solo un corto tramo de galimatías entre los dos.

Degeneración del código genético Como ya se ha dicho, con cuatro letras para elegir en cada posición, un codón de tres letras podría hacer 4 × 4 × 4 = 64 palabras. Con solo 20 palabras necesarias para los 20 aminoácidos comunes, ¿para qué se usan las otras palabras, si es que se usan alguna? El trabajo de Crick sugirió que el código genético es degenerado, lo que significa que cada uno de los 64 trillizos debe tener algún significado dentro del código. Para que el código degenere, algunos de los aminoácidos deben estar especificados por al menos dos o más trillizos diferentes. El razonamiento es así. Si solo se usaran 20 trillizos, entonces los otros 44 serían una tontería, ya que no codificarían ningún aminoácido. En ese caso, se podría esperar que la mayoría de las mutaciones de cambio de marco produzcan palabras sin sentido, que presumiblemente detienen el proceso de construcción de proteínas, y la supresión de las mutaciones de cambio de marco rara vez, si es que alguna vez, funcionaría. Sin embargo, si todos los trillizos especificaban algún aminoácido, entonces las palabras cambiadas simplemente darían lugar a la inserción de aminoácidos incorrectos en la proteína. Por lo tanto, Crick razonó que muchos o todos los aminoácidos deben tener varios nombres diferentes en el código de par de bases; esta hipótesis se confirmó más tarde bioquímicamente.

La discusión hasta ahora demuestra que 1. La secuencia lineal de nucleótidos en un gen determina la secuencia lineal de aminoácidos en una proteína. 2. El código genético no se superpone. 3. Tres bases codifican un aminoácido. Estos trillizos se denominan codones. 4. El código se lee desde un punto de partida fijo y continúa hasta el final de la secuencia de codificación. Sabemos que el código se lee secuencialmente porque una sola mutación de cambio de fotogramas en cualquier lugar de la secuencia de codificación altera la alineación del codón para el resto de la secuencia. 5. El código está degenerado en el sentido de que algunos aminoácidos están especificados por más de un codón.

Rompiendo el código El desciframiento del código genético, determinando el aminoácido especificado por cada triplete, fue uno de los avances genéticos más emocionantes de los últimos 50 años. Después de que las técnicas experimentales necesarias estuvieran disponibles, el código genético se rompió a toda prisa. Un gran avance fue el descubrimiento de cómo hacer ARNm sintético. Si los nucleótidos del ARN se mezclan con una enzima especial (polinucleótido fosforosa), se forma un ARN de una sola cadena en la reacción. A diferencia de la transcripción, no se necesita ninguna plantilla de ADN para esta síntesis, por lo que los nucleótidos se clasifican al azar. La capacidad de sintetizar ARN ofrecía la emocionante perspectiva de crear secuencias específicas de ARNm y luego ver qué aminoácidos especificarían. El primer mensajero sintético obtenido se hizo mezclando solo nucleótidos de uracilo con la enzima de síntesis de ARN, produciendo... UUUU . . . [poly(U)]. En 1961, Marshall Nirenberg y Heinrich Matthaei mezclaron poli(U) con la maquinaria de síntesis de proteínas de E. coli in vitro y observaron la formación de una proteína. La principal emoción se centró en la cuestión de la secuencia de aminoácidos de esta proteína. Resultó ser polifenilalanina, una cadena de fenilala, nueve moléculas unidas para formar un polipéptido.

Por este descubrimiento, Nirenberg fue galardonado con el Premio Nobel. A continuación, se sintetizaron los ARNm que contenían dos tipos de nucleótidos en grupos repetidos. Por ejemplo, el ARNm sintético que tiene la secuencia (AGA)n, que es una secuencia larga de AGAAGAAGAAGAGA, se utilizó para estimular la síntesis de polipéptidos in vitro (en un tubo de ensayo que también contenía un extracto celular con todos los componentes necesarios para la traducción). La secuencia de los polipéptidos resultantes se observó a partir de una variedad de tales pruebas, con el uso de diferentes trillizos que residen en otros ARN sintéticos. A partir de tales pruebas, se pudieron verificar muchas palabras de código. (Este tipo de experimento se detalla en el Problema 44 al final de este capítulo. Al resolverlo, puedes ponerte en el lugar de H. Gobind Khorana, que recibió un Premio Nobel por dirigir los experimentos.) Los enfoques experimentales adicionales llevaron a la asignación de cada aminoácido a uno o más codones. Recordemos que se propuso que el código fuera degenerado, lo que significa que algunos aminoácidos tenían más de una asignación de codón. Esta degeneración se puede ver claramente en la Figura 9-5, que muestra los codones y los aminoácidos que especifican. Prácticamente todos los organismos de la Tierra utilizan este mismo código genético. (Solo hay algunas excepciones en las que un pequeño número de codones tienen diferentes significados, por ejemplo, en los genomas mitocondriales).

Detener codones Es posible que haya notado en la Figura 9-5 que algunos codones no especifican un aminoácido en absoluto. Estos codones son codones de stop, o termi- nation. Se pueden considerar como similares a los puntos o comas que puntúan el mensaje codificado en el ADN. Una de las primeras indicaciones de la existencia de codones de parada vino en 1965 del trabajo de Brenner con el fago T4. Brenner analizó ciertas mutaciones (m1-m6) en un solo gen que controla la proteína de la cabeza del fago. Descubrió que la proteína de la cabeza de cada mutante era una cadena polipeptídica más corta que la del tipo salvaje. Brenner examinó los extremos de las proteínas acortadas y las comparó con la proteína de tipo salvaje. Para cada mutante, registró el siguiente aminoácido que se habría insertado para continuar la cadena de tipo salvaje. Los aminoácidos para las seis mutaciones fueron la glutamina, la lisina, el ácido glutámico, la tirosina, el triptófano y la serina. Estos resultados no presentan un patrón inmediatamente obvio, pero Brenner dedujo que ciertos codones para cada uno de estos aminoácidos son similares. Específicamente, cada uno de estos codones puede mutar al codón UAG mediante un solo cambio en un par de nucleótidos de ADN. Por lo tanto, postuló que UAG es un codón de parada (terminación), una señal al mecanismo de traducción de que la proteína ahora está completa. UAG fue el primer codón de parada descifrado; se llama codón ámbar (ámbar es la traducción al inglés del apellido del descubridor del codón, Bernstein). Los mutantes que son defectuosos debido a la presencia de un codón ámbar anormal se llaman mutantes ámbar. Otros dos codones de parada son UGA y UAA. De forma análoga al codón ámbar, y continuando con el tema de nombrar colores y gemas, UGA se llama codón de ópalo y UAA se llama codón ocre. Los mutantes que son defectuosos porque contienen códones de ópalo o ocre anormales se llaman mutantes de ópalo y ocre, respectivamente. Los codones de parada a menudo se llaman codones sin sentido porque no designan ningún aminoácido. Además de una proteína de cabeza más corta, los mutantes de fago de Brenner tenían otra característica interesante en común: la presencia de una mutación supresora (su-) en el cromosoma huésped haría que el fago desarrollara una proteína de la cabeza de longitud de cadena normal (de tipo salvaje) a pesar de la presencia de la mutación m. Consideraremos más a fondo los codones y sus supresores después de haber tratado el proceso de síntesis de proteínas.

**9.3 trna: el adaptador** Una vez que se descifró el código genético, los científicos comenzaron a preguntarse cómo la secuencia de aminoácidos de una proteína era determinada por los codones tripletes del ARNm. Un modelo temprano, rápidamente descartado como ingenuo e improbable, propuso que los codones de ARNm podrían plegarse y formar 20 cavidades distintas que se unen directamente a aminoácidos específicos en el orden correcto. En cambio, en 1958, Crick reconoció lo siguiente: Por lo tanto, es una hipótesis natural que el aminoácido es llevado a la plantilla por una molécula adaptadora, y que el adaptador es la parte que realmente se ajusta al ARN. En su forma más simple [esta hipótesis] requeriría veinte adaptadores, uno por cada aminoácido.1

Especuló que el adaptador "podría contener nucleótidos. Esto les permitiría unirse a la plantilla de ARN mediante el mismo "emparejamiento" de bases que se encuentra en el ADN". Además, "se requeriría una enzima separada para unir cada adaptador a su propio aminoácido". Ahora sabemos que la "hipótesis del adaptador" de Crick es en gran medida correcta. De hecho, los aminoácidos están unidos a un adaptador (recuerde que los adaptadores constituyen una clase especial de ARN estables llamados ARN de transferencia). Cada aminoácido se une a un ARNt específico, que luego lleva ese aminoácido al ribosoma, el complejo molecular que unirá el aminoácido a un polipéptido en crecimiento. Traslación de codones por ARNt La estructura del ARNt guarda el secreto de la especificidad entre un codón de ARNm y el aminoácido que designa. La molécula de ARNt de cadena única tiene una forma de hoja clo-ver que consiste en cuatro tallos de doble hélice y tres bucles de cadena simple (Figura 9-6a). El bucle medio de cada ARNt se llama bucle anticodón porque lleva un triplete de nucleótido llamado anticodón. Esta secuencia es complementaria al codón para el aminoácido transportado por el ARNt. El anticodón en el ARNt y el codón en el ARNm se unen mediante un emparejamiento específico de bases de ARN a ARN. (Una vez más, vemos el principio de complementariedad del ácido nucleico en el trabajo, esta vez en la unión de dos ARN diferentes). Debido a que los codones en el ARNm se leen en la dirección de 5′ × 3′, los anti-dons están orientados y escritos en la dirección de 3′ × 5′, como muestra la Figura 9-6a. Los aminoácidos se unen a los ARNt por enzimas llamadas aminoacil-ARNt sintetasas. Hay 20 de estas enzimas en la célula, una para cada una de las 20 aminoácidos. Cada aminoácido tiene una sintasa específica que lo vincula solo a aquellos ARNt que reconocen los codones de ese aminoácido en particular. Para catalizar esta reacción, las sintetasas tienen dos sitios de unión, uno para el aminoácido y el otro para su ARNt cognado (Figura 9-7). Un aminoácido está unido al extremo libre de 3′ de su ARNt, el aminoácido alanina en el caso que se muestra en las Figuras 9-6a y 9-7. Se dice que el ARNt con un aminoácido unido está cargado. Un ARNt normalmente existe como una hoja de trébol doblada en forma de L, como se muestra en la Figura 9-6b, en lugar de la hoja de trébol "aplanada" que se muestra en la Figura 9-6a. La estructura tridimensional del ARNt se determinó con el uso de la cristalografía de rayos X. En los años transcurridos desde que se utilizó esta técnica para deducir la estructura de doble hélice del ADN, se ha refinado para que ahora se pueda utilizar para determinar la estructura de macromoléculas muy complejas, como el ribosoma. Aunque los ARNt difieren en su secuencia de nucleótidos primarios, todos los ARNt se pliegan en prácticamente la misma conformación en forma de L, excepto por las diferencias en el bucle del anticodón y el extremo del aminoacilo. Esta similitud de la estructura se puede ver fácilmente en la Figura 9-8, que muestra dos ARNt diferentes superimesados. La conservación de la estructura nos dice que la forma es importante para la función del ARNt. ¿Qué pasaría si el aminoácido equivocado se uniera de forma covalente a un ARNt? Un experimento convincente respondió a esta pregunta. El experimento utilizó cisteinil-tRNA (tRNACys), el arNt específico de la cisteína. Este ARNt estaba "cargado" con cisteína, lo que significa que la cisteína estaba unida al ARNt. La proteína sintetizada con esta especie híbrida tenía alanina en cualquier lugar donde esperábamos cisteína. El experimento demostró que los aminoácidos son "illiterados"; se insertan en la posición correcta porque los "adaptadores" de ARNt reconocen los codones de ARNm e insertan sus aminoácidos unidos de manera adecuada. Por lo tanto, la unión del aminoácido correcto a su ARNt cognado es un paso crítico para garantizar que una proteína se sintetice correctamente. Si se une el aminoácido equivocado, no hay forma de evitar que se incorpore a una cadena de proteínas en crecimiento.

Degeneración revisada Como se puede ver en la Figura 9-5, el número de codones para un solo aminoácido varía, desde un codón (UGG para triptófano) hasta seis (UCC, UCU, UCA, UCG, AGC o AGU para la serina). No está exactamente claro por qué el código genético contiene esta variación, pero dos hechos lo explican: 1. La mayoría de los aminoácidos pueden ser llevados al ribosoma por varios tipos alternativos de ARNt. Cada tipo tiene un anticodón diferente que se empareja con un codón diferente en el ARNm. 2. Ciertas especies de ARNt cargadas pueden llevar sus aminoácidos específicos a cualquiera de los varios codones.

**wobble:** La oscilación es una situación en la que el tercer nucleótido de un anticodón (en el extremo 5') puede formar cualquiera de dos alineaciones (Figura 9-9). Este tercer nucleótido puede formar enlaces de hidrógeno, ya sea con su nucleótido complementario normal en la tercera posición del codón o con un nucleótido diferente en esa posición. Las "reglas de oscilación" dictan qué nucleótidos pueden y no pueden formar enlaces de hidrógeno con Nucleótidos alternativos a través de la oscilación (Tabla 9-1). En la Tabla 9-1, la letra I significa inosina, una de las raras bases que se encuentran en el ARNt, a menudo en el anticondón.

Se dice que el código genético es degenerado porque, en muchos casos, se asigna más de un codón a un solo aminoácido; además, varios codones pueden emparejarse con más de un anticodón (se oscilación).

Lecturas:

Chapter 8 “RNA: Transcription and Processing” Anthony J.F. Griffiths, Susan R. Wessler, Sean B. Carroll, John Doebley (2015). An Introduction to Genetic Analysis-W. H. Freeman.

Chapter 9 “Proteins and Their Synthesis” Anthony J.F. Griffiths, Susan R. Wessler, Sean B. Carroll, John Doebley (2015). An Introduction to Genetic Analysis-W. H. Freeman.

**Unidad 10 Regulación genética de la transcripción en Procariontes**

* 10.1 Regulación genética

Regulación de los genes A pesar de su simplicidad de forma, las bacterias tienen en común con los organismos más grandes y complejos la necesidad de regular la expresión de sus genes. Una de las principales razones es que son oportunistas nutricionales. Considere cómo las bacterias obtienen los muchos compuestos importantes, como los azúcares, los aminoácidos y los nucleótidos, necesarios para el metabolismo. Las bacterias nadan en un mar de nutrientes potenciales. Pueden adquirir los compuestos que necesitan del medio ambiente o sintetizarlos por vías enzimáticas. Pero sintetizar estos compuestos también requiere el gasto de energía y recursos celulares para producir las enzimas necesarias para estas vías. Por lo tanto, dada la opción, las bacterias tomarán compuestos del medio ambiente en su lugar. La selección natural favorece la eficiencia y se opone al desperdicio de recursos y energía. Para ser económicas, las bacterias sintetizarán las enzimas necesarias para producir compuestos solo cuando no hay otra opción, en otras palabras, cuando los compuestos no están disponibles en su entorno local. Las bacterias han desarrollado sistemas reguladores que acoplan la expresión de productos genéticos con sistemas de sensores que detectan el compuesto relevante en el entorno local de una bacteria. La regulación de las enzimas que participan en el metabolismo del azúcar proporciona un ejemplo. Las moléculas de azúcar se pueden descomponer para proporcionar energía o se pueden utilizar como bloques de construcción para una gran variedad de compuestos orgánicos. Sin embargo, hay muchos tipos diferentes de azúcar que las bacterias podrían usar, como la lactosa, la glucosa, la galactosa y la xilosa. Se requiere una proteína de importación diferente para permitir que cada uno de estos azúcares entre en la célula. Además, se requiere un conjunto diferente de enzimas para procesar cada uno de los azúcares. Si una célula sintetizara simultáneamente todas las enzimas que podría necesitar, la célula gastaría mucha más energía y materiales para Producir las enzimas de las que podría derivarse de la descomposición de posibles fuentes de carbono. La célula ha ideado mecanismos para apagar (reprimir) la transcripción de todos los genes que codifican las enzimas que no son necesarias en un momento dado y para activar (activar) aquellos genes que codifican las enzimas que son necesarias. Por ejemplo, si solo hay lactosa en el medio ambiente, la célula cerrará la transcripción de los genes que codifican las enzimas necesarias para la importación y el metabolismo de la glucosa, la galactosa, la xilosa y otros azúcares. Por el contrario, E. coli iniciará la transcripción de los genes que codifican las enzimas necesarias para la importación y el metabolismo de la lactosa. En resumen, las células necesitan mecanismos que cumplan con dos criterios: 1. Deben ser capaces de reconocer las condiciones ambientales en las que deben activar o reprimir la transcripción de los genes relevantes. 2. Deben ser capaces de activar o desactivar, como un interruptor, la transcripción de cada gen específico o grupo de genes. Obtengamos una vista previa del modelo actual para la regulación transcripcional procariota y luego usemos un ejemplo bien entendido, la regulación de los genes en el metabolismo de la lactosa de azúcar, para examinarlo en detalle. En particular, nos centraremos en cómo se disecó este sistema regulatorio con el uso de las herramientas de la genética clásica y la biología molecular.

Los conceptos básicos de la regulación transcripcional procariota: interruptores genéticos La regulación de la transcripción depende principalmente de dos tipos de interacciones entre la proteína y el ADN. Ambos tienen lugar cerca del sitio donde comienza la transcripción de genes. Una de estas interacciones ADN-proteína determina dónde comienza la transcripción. El ADN que participa en esta interacción es un segmento de ADN llamado promotor (Capítulo 8, Sección 8.2), y la proteína que se une a este sitio es la ARN polimerasa. Cuando la ARN polimerasa se une al ADN del promotor, la transcripción puede comenzar a unas pocas bases de distancia del sitio del promotor. Cada gen debe tener un pro-moter o no se puede transcribir. El otro tipo de interacción ADN-proteína determina si se lleva a cabo la transcripción impulsada por el promotor. Los segmentos de ADN cerca del promotor sirven como sitios de unión para proteínas reguladoras específicas de la secuencia llamadas activadores y represores. En las bacterias, la mayoría de los sitios de unión para los represores se denominan operadores. Para algunos genes, una proteína activadora debe unirse a su sitio de ADN diana como requisito previo necesario para que comience la transcripción. Tales casos a veces se denominan regulación positiva porque se requiere la presencia de la proteína unida para la transcripción (Figura 11-2). Para otros genes, se debe evitar que una proteína represora se una a su sitio objetivo como requisito previo necesario para que comience la transcripción. Tales casos a veces se denominan regulación negativa porque la ausencia del represor vinculado permite que comience la transcripción. ¿Cómo regulan los activadores y represores la transcripción? A menudo, una proteína activadora unida al ADN ayuda físicamente a unir la ARN polimerasa a su promotor cercano para que la polimerasa pueda comenzar a transcribirse. Una proteína represora unida al ADN suele actuar interfiriendo físicamente con la unión de la ARN polimerasa a su pro-moter (bloqueando el inicio de la transcripción) o impidiendo el movimiento de la ARN polimerasa a lo largo de la cadena de ADN (bloqueando la transcripción). En conjunto, estas proteínas reguladoras y sus sitios de unión constituyen interruptores genéticos que controlan los cambios eficientes en la expresión génica que se producen en respuesta a las condiciones ambientales.

Los interruptores genéticos controlan la transcripción de genes. La función de encendido/apagado de los interruptores depende de las interacciones de varias proteínas con sus sitios de unión en el ADN. La ARN polimerasa interactúa con el promotor para comenzar la transcripción. Las proteínas activadoras o represoras se unen a sitios en las cercanías del promotor para controlar su accesibilidad a la ARN polimerasa.

Tanto las proteínas activadoras como las represoras deben ser capaces de reconocer cuándo las condiciones ambientales son apropiadas para sus acciones y actuar en consecuencia. Por lo tanto, para que las proteínas activadoras o represoras hagan su trabajo, cada una debe ser capaz de existir en dos estados: uno que puede unir sus objetivos de ADN y otro que no puede. El estado de unión debe ser apropiado para el conjunto de condiciones fisiológicas presentes en la célula y su entorno. Para muchas proteínas reguladoras, la unión al ADN se efectua a través de la interacción de dos sitios diferentes en la estructura tridimensional de la proteína. Un sitio es el dominio de unión al ADN. El otro sitio, el sitio alostérico, actúa como un sensor que establece el dominio de unión al ADN en uno de dos modos: funcional o no funcional. El sitio alostérico interactúa con pequeñas moléculas llamadas efectos alostéricos. En el metabolismo de la lactosa, en realidad es un isómero del azúcar lactosa (llamado alo lactosa) que es un efector alostérico: el azúcar se une a una proteína reguladora que inhibe la expresión de los genes necesarios para el metabolismo de la lactosa. En general, un efector alostérico se une al sitio alostérico de la proteína reguladora de tal manera que cambia su actividad. En este caso, la alolactosa cambia la forma y la estructura del dominio de unión al ADN de una proteína reguladora. Algunas proteínas activadoras o represisoras deben unirse a sus efectores alostéricos antes de que puedan unirse al ADN. Otros pueden unirse al ADN solo en ausencia de sus efectores alostéricos. Dos de estas situaciones se muestran en la Figura 11-3.

* 10.2 Control positivo y negativo

El control inductor-represor del operón lac es un ejemplo de represión, o control negativo, en el que la expresión normalmente se bloquea. Por el contrario, el sistema CAP-cAMP es un ejemplo de activación, o control positivo, porque actúa como una señal que activa la expresión; en este caso, la señal de activación es la interacción del complejo CAP-cAMP con el sitio de unión a CAP en el ADN. La figura 11-18 describe estos dos tipos básicos de sistemas de control.

El lac operon es un grupo de genes estructurales que especifican las enzimas que participan en el metabolismo de la lactosa. Estos genes están controlados por las acciones coordinadas de las regiones promotoras y operadoras de acción cis. La actividad de estas regiones está, a su vez, determinada por moléculas represoras y activadoras especificadas por genes reguladores separados.

Doble control positivo y negativo: el operón de arabinosa Al igual que con el sistema lac, el control de la transcripción en bac- teria no es ni puramente positivo ni puramente negativo; más bien, tanto la regulación positiva como la negativa pueden gobernar los operones individuales. La regulación del operón de arabinosa proporciona un ejemplo en el que una sola proteína de unión al ADN puede actuar como un represor o un activador, un giro en el tema general de la regulación transcripcional por las proteínas de unión al ADN. Los genes estructurales araB, araA y araD codifican las enzimas metabólicas que descomonen la arabinosa de azúcar. Los tres genes se transcriben en una unidad como un solo ARNm. La figura 11-19 muestra un mapa de la ópera ara. La transcripción se activa en araI, la región del iniciador, que contiene un sitio de unión para una proteína activadora. El gen araC, que mapea cerca, codifica una proteína activadora. Cuando se une a la arabinosa, esta proteína se une al sitio araI y activa la transcripción del ara operón, tal vez ayudando a la ARN polimerasa a unirse al promotor. Además, el mismo sistema de represión de catabolitos CAP-cAMP que evita lac operón La expresión en presencia de glucosa también impide la expresión del operón ara. En presencia de arabinosa, tanto el complejo CAP-cAMP como el complejo AraC-arabinosa deben unirse a araI para que la ARN polimerasa se una al promotor y transcriba el ara operón (Figura 11-20a). En ausencia de arabinosa, la proteína AraC asume una conformación diferente y reprime el operón ara uniéndose tanto a araI como a un segundo sitio distante, araO, formando así un bucle (Figura 11-20b) que impide la transcripción. Por lo tanto, la proteína AraC tiene dos conformaciones, una que actúa como activador y otra que actúa como represor. El interruptor de encendido/apagado del operón es "lanzado" por arabinosa. Las dos conformaciones, dependiendo de si el efector alostérico arabinosa se ha unido a la proteína, difieren en sus capacidades para unirse a un sitio objetivo específico en la región araO del operón.

La transcripción de Operon puede ser regulada tanto por la activación como por la represión. Los operones que regulan el metabolismo de compuestos similares, como los azúcares, se pueden regular de maneras muy diferentes.

* 10.3 Vías metabólicas y niveles adicionales de regulación: atenuación

El control coordinado de los genes en las bacterias está muy extendido. En la sección anterior, vimos ejemplos que ilustran la regulación de las vías para la descomposición de azúcares específicos. De hecho, la mayoría de los genes coordinados en las bacterias se coordinan a través de mecanismos de operón. En muchas vías que sintetizan moléculas esenciales a partir de bloques de construcción inorgánicos simples, los genes que codifican las enzimas se organizan en operones, completos con ARNm multigénicos. Además, en los casos en los que se conoce la secuencia de actividad catalítica, hay una notable congruencia entre el orden de los genes operones en el cromosoma y el orden en el que sus productos actúan en la vía metabólica. Esta congruencia se ilustra notablemente con la organización del operón triptófano en E. coli (Figura 11-21). El operón de triptófano contiene cinco genes (trpE, trpD, trpC, trpB, trpA) que codifican enzimas que contribuyen a la síntesis del aminoácido triptófano.

En las bacterias, los genes que codifican enzimas que están en las mismas vías metabólicas generalmente se organizan en operones.

Hay dos mecanismos para regular la transcripción del operón triptófano y algunos otros operones que funcionan en la biosíntesis de aminoácidos. Uno proporciona un control global de la expresión del ARNm del operón, y el otro proporciona un control ajustado. El nivel de expresión del gen del operón trp está gobernado por el nivel de triptófano. Cuando el triptófano está ausente en el medio de crecimiento, la expresión del gen trp es alta; cuando los niveles de triptófano son altos, el operón trp se reprime. Un mecanismo para controlar la transcripción del operón trp es similar al mecanismo que ya hemos visto que controla el operón lac: una proteína represora Une a un operador, lo que impide el inicio de la transcripción. Este represor es el represor Trp, el producto del gen trpR. El represor Trp se une al triptófano cuando están presentes niveles adecuados del aminoácido, y solo después de la unión al triptófano, el represor Trp se unirá al operador y apagará la transcripción del operón. Este simple mecanismo garantiza que la célula no desperdicie energía produciendo triptófano cuando el aminoácido es lo suficientemente abundante. Las cepas de E. coli con mutaciones en trpR continúan expresando el ARNm trp y, por lo tanto, continúan produciendo triptófano cuando el aminoácido es abundante. Al estudiar estas cepas mutantes de trpR, Charles Yanofsky descubrió que, cuando se eliminó el triptófano del medio, la producción de ARNm de trp se vuplicó más por varias veces. Este hallazgo fue evidencia de que, además del represor Trp, existía un segundo mecanismo de control para regular negativamente la transcripción. Este mecanismo se llama atenuación porque la producción de ARNm no está mal atenuada, lo que significa "disminuida", cuando el triptófano es abundante. A diferencia de los otros mecanismos de control bacteriano descritos hasta ahora, la atenuación actúa en un paso después del inicio de la transcripción. Los mecanismos que rigen la atenuación se descubrieron identificando mutaciones que redujeron o abolieron la atenuación. Las cepas con estas mutaciones producen el ARNm trp a niveles máximos, incluso en presencia de triptófano. Yanofsky mapeó las mutaciones a una región entre el operador de trp y el gen trpE; esta región, denominada la secuencia líder, se encuentra en el extremo 5′ del ARNm del operón trp antes del primer codón del gen trpE (Figura 11-22). La secuencia líder trp es inusualmente larga para un ARNm procariota, 160 bases, y los análisis detallados han revelado cómo una parte de esta secuencia funciona como un atenuador que rige la transcripción adicional del ARNm trp.

Las observaciones clave son que, en ausencia de la proteína represora de TrpR, la presencia de triptófano detiene la transcripción después de las primeras 140 bases más o menos, mientras que, en ausencia de triptófano, la transcripción del operón continúa. El mecanismo para terminar o continuar la transcripción consta de dos elementos clave. En primer lugar, la secuencia líder de ARNm trp codifica una marea corta de pep- de 14 aminoácidos que incluye dos codones de triptófano adyacentes. El triptófano es uno de los aminoácidos menos abundantes en las proteínas, y está codificado por un solo codón. Por lo tanto, este par de codones de triptófano es una característica inusual. En segundo lugar, las partes del líder de ARNm trp forman estructuras de tallo y bucle que son capaces de alternar entre dos conformaciones. Una de estas conformaciones favorece la terminación de la transcripción (Figura 11-23a). La lógica reguladora del operón gira en torno a la abundancia de triptófano. Cuando el triptófano es abundante, hay un suministro suficiente de aminoacil-tRNATrp para permitir la traducción del péptido de 14-aminoácidos. Recuerde que la transcripción y la traducción en las bacterias están acopladas; por lo que los ribosomas pueden involucrar las transcripciones de ARNm e iniciar la traducción antes de que se complete la transcripción. El compromiso del ribosoma altera la conformación del ARNm trp a la forma que favorece la terminación de la transcripción (donde los segmentos 3 y 4 forman pares de bases; Figura 11-23b). Sin embargo, cuando el triptófano es escaso, el ribosoma se detiene en los codones de triptófano, el par de bases de los segmentos 2 y 3, y la transcripción puede continuar (Figura 11-23c). Otros operones para enzimas en vías biosintéticas tienen controles de atenuación similares. Una firma de los operones de biosíntesis de aminoácidos es la presencia de múltiples codones para el aminoácido que se sintetiza en un péptido separado codificado por la secuencia líder de 5′. Por ejemplo, el operón phe tiene siete phe-nylalanina de Codones en un péptido líder y su operón tiene siete codones histi-dine en tándem en su péptido líder (Figura 11-24).

Un segundo nivel de regulación en los operones de biosíntesis de aminoácidos es la atenuación de la transcripción mediada por la abundancia del aminoácido y la traducción de un péptido líder.

* 10.4 Ciclo de vida de los bacteriófagos.
* 10.5 Reguladores y operones complejos.

11.6 Ciclos de vida de los bacteriófagos: más reguladores, operones complejos En ese cine de París, François Jacob tuvo un destello de visión de que el fenómeno de la inducción de profagos podría ser estrechamente análogo a la inducción de la síntesis de β-galactosidasa. Tenía razón. Aquí, vamos a ver cómo se regula el ciclo de vida del bacteriófago λ. Aunque su regulación es más compleja que la de los operones individuales, está controlada por modos de regulación génica ahora familiares. El bacteriófago λ es un llamado fago templado que tiene dos ciclos de vida alternativos (Figura 11-25). Cuando una bacteria normal está infectada por un fago λ de tipo salvaje, pueden seguir dos posibles resultados: (1) el fago puede replicarse e igualar la célula (el ciclo lítico) o (2) el genoma del fago puede integrarse en el cromosoma bacteriano como un profago inerte (el ciclo lisogénico). En el estado lítico, la mayoría de los 71 genes del fago se expresan en algún momento, mientras que en el estado lisogénico, la mayoría de los genes están inactivos. ¿Qué decide cuál de estos dos caminos se toma? El control fisiológico de la decisión entre la vía lítica o la lisogénica depende de los recursos disponibles en la bacteria huésped. Si los recursos son abundantes, se prefiere el ciclo lítico porque entonces hay suficientes nutrientes para hacer muchas copias del virus. Si los recursos son limitados, se toma la vía lisogénica. El virus permanece presente como profago hasta que las condiciones mejoren. El prófago inerte puede ser inducido por la luz ultravioleta para entrar en el ciclo lítico, el fenómeno estudiado por Jacob. Los estados lítico y lisogénico se caracterizan por programas muy distintos de expresión génica que deben ser regulados. El estado alternativo que se selecciona está determinado por un complejo cambio genético que comprende varias proteínas reguladoras de unión al ADN y un conjunto de sitios de operadores. Al igual que para el lac y otros sistemas reguladores, los análisis genéticos de los mutantes fueron fuentes de conocimientos cruciales sobre los componentes y la lógica del cambio genético λ. Jacob utilizó pantallas fenotípicas simples para aislar mutantes que eran defectuosos en la vía lítica o lisogénica. Los mutantes de cada tipo podrían ser reconocidos por la aparición de placas infectadas en un césped de bacterias.

Las partículas de fagos de tipo salvaje se colocan en un césped de bacterias sensibles, los claros (llamados "placas") aparecen donde las bacterias están infectadas y lisadas, pero estas placas son turbias porque las bacterias que están lisogenizadas crecen dentro de ellas (Figura 11-26). Los fagos mutantes que forman placas claras son incapaces de lisogenizar las células. Tales mutantes claros (designados por c) resultan ser análogos a los mutantes I y O del sistema lac. Estos mutantes a menudo se aislaban como mutantes sensibles a la temperatura que tenían fenotipos claros a temperaturas más altas, pero fenotipos de tipo salvaje a temperaturas más bajas. Tres clases de mutantes llevaron a la identificación de las características regulatorias clave del fago λ. En la primera clase, los mutantes para los genes cI, cII y cIII forman placas claras; es decir, son incapaces de establecer lysog- eny. Se aisló una segunda clase de mutantes que no lisogenizan las células, pero que pueden replicarse y entrar en el ciclo lítico en una célula lisogenizada. Estos mutantes resultan ser análogos a los mutantes constitutivos del operador del sistema lac. Un tercer mutante clave puede lisogenizar, pero no puede lizar las células. El gen mutado en este caso es El gen cro (para el control del represor y otras cosas). La decisión entre las vías lítica y las lisogénicas depende de la actividad de las proteínas codificadas por los cuatro genes cI, cII, cIII y cro, tres de los cuales son proteínas de unión al ADN. Primero nos centraremos en los dos genes cI y cro y las proteínas que codifican (Tabla 11-4). El gen cI codifica un represor, a menudo conocido como λ represor, que reprime el crecimiento lítico y promueve la lisogenia. El gen cro codifica un represor que represa la lisogenia, permitiendo así el crecimiento lítico. El interruptor genético que controla los dos ciclos de vida de los fagos λ tiene dos estados: en el estado lisogénico, cI está encendido, pero cro está apagado, y en el ciclo lítico, cro está encendido, pero cI está apagado. Por lo tanto, el represor λ y Cro están en competencia, y cualquier represor que prevalezca determina el estado del interruptor y de la expresión del genoma λ. La carrera entre λ represor y Cro se inicia cuando el fago λ infecta una bacteria normal. La secuencia de eventos en la raza está determinada críticamente por la organización de los genes en el genoma λ y de los promotores y operadores entre los genes cI y cro. El genoma de aproximadamente 50 kb λ codifica las proteínas que tienen papeles en la replicación del ADN, la recombinación, el ensamblaje de la partícula de fago y la lisis celular (Figura 11-27). Estas proteínas se expresan en una secuencia lógica de tal manera que primero se hacen copias del genoma, estas copias se empaquetan en partículas virales y, finalmente, se lisa la célula huésped para liberar el virus y comenzar la infección de otras células huésped (ver Figura 11-25). El orden de expresión génica viral fluye a partir del inicio de la transcripción en dos promotores, PL y PR (para promotores hacia la izquierda y hacia la derecha con respecto al mapa genético). En la infección, la ARN polimerasa inicia la transcripción en ambos promotores. Mirando el mapa genético (ver Figura 11-27), vemos que de PR, cro es el primer gen transcrito, y de PL, N es el primer gen transcrito. El gen N codifica un regulador positivo, pero el mecanismo de esta proteína difiere de los de otros reguladores que hemos considerado hasta ahora. La proteína N funciona al permitir que la ARN polimerasa continúe transcribiéndose a través de regiones del ADN que de otro modo harían que la transcripción terminara. Una proteína reguladora como la N que actúa al prevenir la terminación de la transcripción se llama antiterminador. Por lo tanto, N permite la transcripción de cIII y otros genes a la izquierda de N, así como cII y otros genes a la derecha de cro. El gen cII codifica una proteína activadora que se une a un sitio que promueve la transcripción hacia la izquierda de un promotor diferente, PRE (para el promotor del establecimiento represor), que activa la transcripción del gen cI. Recuerde que el gen cI codifica el represor λ, lo que evitará el crecimiento lítico. Antes de que se lleve a cabo la expresión del resto de los genes virales, se debe tomar una "decisión", ya sea para continuar con la expresión del gen viral y lisar la célula o para reprimir esa vía y lisogenizar la célula. La decisión de lisar o lisogenizar una célula gira en torno a la actividad de la proteína cII. La proteína cII es inestable porque es sensible a las proteasas bacterianas, enzimas que degradan las proteínas.

Estas proteasas responden a las condiciones ambientales: son más activas cuando los recursos son abundantes, pero menos activas cuando las células están muertas de hambre. Echemos un vistazo a lo que le sucede a cII cuando los recursos son abundantes y no abundantes. Cuando los recursos son abundantes, cII se degrada y se produce poco represivo λ. Los genes transcritos de PL y PR continúan expresándose, y prevalece el ciclo lítico. Sin embargo, si los recursos son limitados, cII es más activo y se produce más λ represor. En este caso, los genes transcritos de PL y PR son reprimidos por el represor λ y se introduce el ciclo lisogénico. La proteína cII también es responsable de activar la transcripción de int, un gen que codifica una proteína adicional requerida para la lisogenia, una integrasa requerida para que el genoma λ se integre en el cromosoma huésped. La proteína cIII protege a cII de la degradación; por lo tanto, también contribuye a la decisión lisogénica. Vamos a recapitular brevemente la secuencia de eventos y los puntos de decisión en el ciclo de vida del bac- teriófago λ: Sobre la infección: La ARN polimerasa huésped inicia la transcripción en PL y PR, expresando los genes cro y N. Entonces: La proteína antiterminadora N permite la transcripción del gen cIII y los genes de la recombinación (ver Figura 11-27, izquierda), y el gen cII y otros genes (ver Figura 11-27, derecha).

Entonces: La proteína cII, protegida por la proteína cIII, activa cI e int activando la transcripción en PRE. Entonces: Si los recursos y las proteasas son abundantes, cII se degrada, Cro reprime cI y el ciclo lítico continúa. Si los recursos y las proteasas no son abundantes, cII está activo, la transcripción de cI se pro- cede a un alto nivel, la proteína Int integra el cromosoma del fago y cI (λ represor) cierra todos los genes excepto a sí mismo.

El interruptor genético del fago λ ilustra cómo unas pocas proteínas reguladoras de unión al ADN, que actúan a través de unos pocos sitios, controlan la expresión de un número mucho mayor de genes en el virus en un mecanismo de "cascada". Al igual que en los sistemas lac, ara, trp y otros, los estados alternativos de expresión génica están determinados por señales fisiológicas.

La unión específica de la secuencia de las proteínas reguladoras al ADN ¿Cómo reconocen el represor λ y el Cro diferentes operadores con diferentes vínculos de afinidad? Esta pregunta dirige nuestra atención a un principio fundamental en el control de la transcripción de genes: las proteínas reguladoras se unen a secuencias de ADN específicas. Para que las proteínas individuales se unan a ciertas secuencias y no a otras, se requiere especificidad en las interacciones entre las cadenas laterales de los aminoácidos de la proteína y los grupos químicos de las bases de ADN. Los estudios estructurales detallados de λ represor, Cro y otros reguladores bacterianos han revelado cómo interactúan las estructuras tridimensionales de los reguladores y el ADN y cómo la disposición de aminoácidos particulares les permite reconocer secuencias de bases específicas. El análisis cristalográfico ha identificado una característica estructural común de los dominios de unión al ADN de λ y Cro. Ambas proteínas hacen contacto con el ADN a través de un dominio de hélice-giro-hélice que consiste en dos hélices α unidas por una región de enlace flexible corta (Figura 11-29). Una hélice, la hélice de reconocimiento, encaja en el surco principal del ADN. En esa posición, los aminoácidos en la cara externa de la hélice son capaces de interactuar con grupos químicos en las bases de ADN. Los aminoácidos específicos en la hélice de reconocimiento determinan la afinidad de una proteína por una secuencia de ADN específica. Las hélices de reconocimiento del represor λ y Cro tienen estructuras similares y algunos residuos de aminoácidos idénticos. Las diferencias entre las hélices en los residuos de aminoácidos clave determinan sus propiedades de unión al ADN. Por ejemplo, en las proteínas represoras λ y Cro, las cadenas laterales de glutamina y serina entran en contacto con las mismas bases, pero un residuo de alanina en el represor λ y la lisina y la asparagina resi- Las cuotas en la proteína Cro imparten diferentes afinidades de unión para las secuencias en OR1 y OR3 (Figura 11-30). Los represores Lac y TrpR, así como el activador AraC y muchas otras proteínas, también se unen al ADN a través de motivos de hélice de giro y hélice de diferentes especificidades, dependiendo de las secuencias de aminoácidos primarios de sus hélices de reconocimiento. En general, otros dominios de estas proteínas, como los que se unen a sus efectos alostéricos, son diferentes.

Chapter 11 “Regulation of Gene Expression in Bacteria and Their Viruses” in: Anthony J.F. Griffiths, Susan R. Wessler, Sean B. Carroll, John Doebley (2015). An Introduction to Genetic Analysis-W. H. Freeman.