

# El CÃ³digo GenÃ©tico

Alumbac - 24/03/2010

El cÃ³digo genÃ©tico es el conjunto de normas por las que la informaciÃ³n codificada en el material genÃ©tico (secuencias de ADN o ARN) se traduce en proteÃ­nas (secuencias de aminoÃ¡cidos) en las cÃ©lulas vivas. El cÃ³digo define la relaciÃ³n entre secuencias de tres nucleÃ³tidos, llamadas codones, y aminoÃ¡cidos. Un codÃ³n se corresponde con un aminoÃ¡cido especÃ­fico.

La secuencia del material genÃ©tico se compone de cuatro bases nitrogenadas distintas, que tienen una funciÃ³n equivalente a letras en el cÃ³digo genÃ©tico: adenina (A), timina (T), guanina (G) y citosina (C) en el ADN y adenina (A), uracilo (U), guanina (G) y citosina (C) en el ARN.

Debido a esto, el nÃºmero de codones posibles es 64, de los cuales 61 codifican aminoÃ¡cidos (siendo ademÃ¡s uno de ellos el codÃ³n de inicio, AUG) y los tres restantes son sitios de parada. La secuencia de codones determina la secuencia aminoacÃ­dica de una proteÃ­na en concreto, que tendrÃ¡ una estructura y una funciÃ³n especÃ­ficas.

La transferencia de la informaciÃ³n consiste en que el genoma de un organismo se encuentra en el ADN o, en el caso de algunos virus, en el ARN la porciÃ³n de genoma que codifica una proteÃ­na o un ARN se conoce como gen. Esos genes que codifican proteÃ­nas estÃ¡n compuestos por unidades de trinucleÃ³tidos llamadas codones, cada una de los cuales codifica un aminoÃ¡cido. Cada subunidad nucleotÃ­dica estÃ¡ formada por un fosfato, una desoxirribosa y una de las cuatro posibles bases nitrogenadas en la configuraciÃ³n en doble hÃ©lice ademÃ¡s, estos puentes siempre se forman entre una adenina de una cadena y una timina de la otra y entre una citosina de una cadena y una guanina de la otra. Esto quiere decir que el nÃºmero de residuos A y T serÃ¡ el mismo en una doble hÃ©lice y lo mismo pasarÃ¡ con el nÃºmero de residuos de G y C. En el ARN, la timina (T) se sustituye por uracilo (U), y la desoxirribosa por una ribosa.

Cada gen codificante de proteÃ­na se transcribe en una molÃ©cula plantilla, que se conoce como ARN mensajero o ARNm. Este, a su vez, se traduce en el ribosoma, en una cadena aminoacÃ­dica o polipeptÃ­dica. En el proceso de traducciÃ³n se necesita un ARN de transferencia especÃ­fico para cada aminoÃ¡cido con el aminoÃ¡cido unido a Ã©l covalentemente, guanosina trifosfato como fuente de energÃ­a y ciertos factores de traducciÃ³n. Los ARNt tienen anticodones complementarios a los codones del ARNm y se pueden "cargar" covalentemente en su extremo 3' terminal CCA con

aminoácidos. Los ARNt individuales se cargan con aminoácidos específicos por las enzimas llamadas aminoacil ARNt sintetetasas, que tienen alta especificidad tanto por aminoácidos como por ARNt. La alta especificidad de estas enzimas es motivo fundamental del mantenimiento de la fidelidad de la traducción de proteínas.

Hay  $4^3 = 64$  combinaciones diferentes de codones que sean posibles con tripletes de tres nucleótidos: los 64 codones están asignados a aminoácido o a señales de parada en la traducción. Si, por ejemplo, tenemos una secuencia de ARN, UUUAACCC, y la lectura del fragmento empieza en la primera U (convenio 5' a 3'), habrá tres codones que serán UUU, AAA y CCC, cada uno de los cuales especifica un aminoácido. Esta secuencia de ARN se traducirá en una secuencia aminoacídica de tres aminoácidos de longitud.

El código genético es compartido por todos los organismos conocidos, incluyendo virus y organulos, aunque pueden aparecer pequeñas diferencias. Así, por ejemplo, el código UUU codifica para el aminoácido fenilalanina tanto en bacterias, como en arqueas, como en eucariotes. Este hecho indica que el código genético ha tenido un origen único en todos los seres vivos conocidos.

Gracias a la genética molecular, se han distinguido 22 códigos genéticos, que se diferencian del llamado código genético estándar por el significado de uno o más codones. La mayor diversidad se presenta en las mitocondrias, orgánulos de las células eucariotas que se originaron evolutivamente a partir de miembros del dominio Bacteria a través de un proceso de endosimbiosis. El genoma nuclear de los eucariotas sólo suele diferenciarse del código estándar en los codones de iniciación y terminación.

#### Especificidad y continuidad

Ningún código codifica más de un aminoácido, ya que, de no ser así, conllevaría problemas considerables para la síntesis de proteínas específicas para cada gen. Tampoco presenta solapamiento: los tripletes se hallan dispuestos de manera lineal y continua, de manera que entre ellos no existan comas ni espacios y sin compartir ninguna base nitrogenada. Su lectura se hace en un solo sentido (5' - 3'), desde el código de iniciación hasta el código de parada. Sin embargo, en un mismo ARNm pueden existir varios codones de inicio, lo que conduce a la síntesis de varios polipéptidos diferentes a partir del mismo transcrito.

#### Degeneración

El código genético tiene redundancia pero no ambigüedad (ver las tablas de codones). Por ejemplo, aunque los codones GAA y GAG especifican los dos el aminoácido glutámico (redundancia), ninguno especifica otro aminoácido (no ambigüedad). Los

codones que codifican un aminoácido pueden diferir en alguna de sus tres posiciones.

De una posición de un codón se dice que es cuatro veces degenerada si con cualquier nucleótido en esta posición se especifica el mismo aminoácido. Por ejemplo, la tercera posición de los codones de la glicina (GGA, GGG, GGC, GGU) es cuatro veces degenerada, porque todas las sustituciones de nucleótidos en este lugar son sinónimos; es decir, no varían el aminoácido. Sólo la tercera posición de algunos codones puede ser cuatro veces degenerada. Se dice que una posición de un codón es dos veces degenerada si sólo dos de las cuatro posibles sustituciones de nucleótidos especifican el mismo aminoácido. Por ejemplo, la tercera posición de los codones del ácido glutámico (GAA, GAG) es doble degenerada. En los lugares dos veces degenerados, los nucleótidos equivalentes son siempre dos purinas (A/G) o dos pirimidinas (C/U), así que sólo sustituciones transversales (purina a pirimidina o pirimidina a purina) en dobles degenerados son antónimas. Se dice que una posición de un codón es no degenerada si una mutación en esta posición tiene como resultado la sustitución de un aminoácido. Sólo hay un sitio triple degenerado en el que cambiando tres de cuatro nucleótidos no hay efecto en el aminoácido, mientras que cambiando los cuatro posibles nucleótidos aparece una sustitución del aminoácido. Que el código genético sea degenerado es lo que determina la posibilidad de mutaciones sinónimas.

La degeneración aparece porque el código genético designa 20 aminoácidos y la señal de parada. Debido a que hay cuatro bases, los codones en triplete se necesitan para producir al menos 21 códigos diferentes. Por ejemplo, si hubiera dos bases por codón, entonces sólo podrían codificarse 16 aminoácidos ( $4^2=16$ ). Y dado que al menos se necesitan 21 códigos,  $4^3$  da 64 codones posibles, indicando que debe haber degeneración.

Esta propiedad del código genético lo hacen más tolerante a los fallos en mutaciones puntuales. Por ejemplo, en teoría, los codones cuatro veces degenerados pueden tolerar cualquier mutación puntual en la tercera posición, aunque el codón de uso sesgado restringe esto en la práctica en muchos organismos; los dos veces degenerados pueden tolerar una de las tres posibles mutaciones puntuales en la tercera posición. Debido a que las mutaciones de transición (purina a purina o pirimidina a pirimidina) son más probables que las de transversión (purina a pirimidina o viceversa), la equivalencia de purinas o de pirimidinas en los lugares dobles degenerados añade una tolerancia a los fallos complementaria.

Agrupamiento de codones por residuos aminoácidos, volumen molar e hidropatía  
Una consecuencia práctica de la redundancia es que algunos errores del código genético sólo causen una mutación silenciosa o un error que no afectará a la proteína porque la hidrofiliidad o hidrofobidad se mantiene por una sustitución

equivalente de aminoácidos; por ejemplo, un codón de NUN (N = cualquier nucleótido) tiende a codificar un aminoácido hidrofóbico. NCN codifica residuos aminoácidos que son pequeños en cuanto a tamaño y moderados en cuanto a hidropatía; NAN codifica un tamaño promedio de residuos hidrofílicos; UNN codifica residuos que no son hidrofílicos. Estas tendencias pueden ser resultado de una relación de las aminoacil ARNt sintetetasas con los codones heredada un ancestro común de los seres vivos conocidos.

La relación entre el ARNm y el ARNt a nivel de la tercera base se puede producir por bases modificadas en la primera base del anticodón del ARNt, y los pares de bases formados se llaman "pares de bases wobble" (tambaleantes). Las bases modificadas incluyen inosina y los pares de bases que no son del tipo Watson-Crick U-G.

### El origen del código genético

A pesar de las variaciones que existen, los códigos genéticos utilizados por todas las formas conocidas de vida son muy similares. Esto sugiere que el código genético se estableció muy temprano en la historia de la vida y que tiene un origen común en las formas de vida actuales. Análisis filogenético sugiere que las moléculas ARNt evolucionaron antes que el actual conjunto de aminoacil-ARNt sintetetasas.

El código genético no es una asignación aleatoria de los codones a aminoácidos. Por ejemplo, los aminoácidos que comparten la misma vía biosintética tienden a tener la primera base igual en sus codones y aminoácidos con propiedades físicas similares tienden a tener similares a codones.[

Experimentos recientes demuestran que algunos aminoácidos tienen afinidad química selectiva por sus codones. Esto sugiere que el complejo mecanismo actual de traducción del ARNm que implica la acción ARNt y enzimas asociadas, puede ser un desarrollo posterior y que, en un principio, las proteínas se sintetizaran directamente sobre la secuencia de ARN, actuando éste como ribozima y catalizando la formación de enlaces peptídicos (tal como ocurre con el ARNr 23S del ribosoma).

Se ha planteado la hipótesis de que el código genético estándar actual surgiera por expansión biosintética de un código simple anterior. La vida primordial pudo adicionar nuevos aminoácidos (por ejemplo, subproductos del metabolismo), algunos de los cuales se incorporaron más tarde a la maquinaria de codificación genética. Se tienen pruebas, aunque circunstanciales, de que formas de vida primitivas empleaban un menor número de aminoácidos diferentes,[12] aunque no se sabe con exactitud que aminoácidos y en que orden entraron en el código genético.

Otro factor interesante a tener en cuenta es que la selección natural ha favorecido la degeneración del código para minimizar los efectos de las mutaciones[13] . Esto ha llevado a pensar que el código genético primitivo podría haber constado de codones de dos nucleótidos, lo que resulta bastante coherente con la hipótesis del balanceo del ARNt durante su acoplamiento (la tercera base no establece puentes de hidrógeno de Watson y Crick).

DNA to Amino Acid Conversion

DNA Sequence -> Protein Sequence converter

DNA to protein translation (6 frames/17 genetic codes)

Tablas del código genético

The Genetic Codes & Genetic Code Tables