

# 28

## CAPÍTULO

### Código genético

Uno de los problemas centrales de la genética molecular consiste en conocer cómo la información contenida en el ADN y transcrita en un ARNm puede dirigir la síntesis de una cadena polipeptídica específica, o sea, cuál es la relación entre la secuencia de bases del ARNm y la de los aminoácidos de una proteína; éste es el contenido del problema del código genético.

El código genético surge como una necesidad natural, debido a las diferencias existentes entre la estructura de las moléculas que conservan la información (ADN) y aquellas que la expresan en funciones específicas (proteínas). Se acostumbra decir que esta información está en 2 lenguajes diferentes: el primero, en forma de una secuencia de bases y el segundo, de aminoácidos. Al pasar la información del ADN al ARNm se mantiene en el mismo lenguaje, por lo que el proceso se denomina transcripción. Pero al pasar del ARNm a las proteínas hay un cambio de lenguaje y por tanto es una traducción. Para traducir es necesario poseer la relación de equivalencia que existe entre los signos utilizados en un lenguaje y los empleados por el otro. En esto consiste la necesidad del código genético.

En nuestros días la solución de este problema pudiera parecer algo muy sencillo, pues bastaría con determinar la secuencia de bases de un gen y la de aminoácidos de la proteína por él codificada y correlacionarlas, pero esto no fue posible en los primeros años de la década de los 60 cuando se enfrentó la solución de este problema. Los métodos de determinación de la estructura primaria de las proteínas estaban ya establecidos desde el trabajo de Sanger con la insulina (1956) pero los procedimientos que permitieron hacer lo mismo con el ADN, con un elevado grado de fidelidad, no aparecen hasta finales de la década de los 70 cuando el código ya había sido descifrado.

Este estudio comenzará por un esbozo de los trabajos fundamentales para descifrar el código genético y después se realizará un análisis y se destacarán sus características principales. Desde el punto de vista bioquímico el código se define como la relación de equivalencia entre la secuencia de bases del ARNm y la secuencia de aminoácidos de la proteína.

### Primeros pasos

El problema del código genético fue formulado por primera vez por *George Gamow* en 1953, y en su solución participaron numerosos investigadores, empleando fundamentalmente métodos químicos y biológicos.

**Gamow** supuso que la cadena polipeptídica se forma directamente sobre la doble hélice del ADN, estando cada aminoácido situado en un hueco entre 4 nucleótidos; este hueco tendría aproximadamente la forma de un rombo. Dos nucleótidos pertenecerían a una banda y 2 a la otra. El «código de rombos rojos» de **Gamow** asegura precisamente las 20 letras, pues como utiliza 4 bases da origen a ( $4^4$ ) 256 combinaciones; no obstante, esta forma de codificación directa imponía algunas limitaciones a la secuencia de aminoácidos de la proteína, esto quiere decir que una vez fijado un aminoácido, el siguiente no podía ser cualquiera de los 20, sino un número muy reducido de ellos. Sin embargo, las investigaciones demostraron que tal limitación no existía en las proteínas, cuyas secuencias de aminoácidos eran conocidas.

El descubrimiento de que la síntesis de proteínas se realizaba en una estructura citoplasmática (los ribosomas) y la existencia de los ARNm rechazaron definitivamente la hipótesis de **Gamow**.

Por razonamientos teóricos se estableció la relación de codificación -que se comprobó después experimentalmente-, o sea, cuántas bases son necesarias para codificar un aminoácido. Como el lenguaje del ARNm sólo tiene 4 letras (A, U, G, C) y el de las proteínas 20, la relación no podía ser 1:1, pues las proteínas constarían de 4 aminoácidos; si fueran 2 las bases alcanzaría para ( $4^2=16$ ) 16 aminoácidos; pero con 3 bases ( $4^3=64$ ), el número de combinaciones era mayor que el número de aminoácidos, por lo que quedó establecido que a cada aminoácido en la proteína le correspondían 3 bases en el ARNm, y la relación de codificación es 3:1. Este triplete de bases, por ser la unidad de codificación, recibió el nombre de codon.

Teniendo en cuenta que son posibles 64 combinaciones de bases y que sólo 20 aminoácidos diferentes se incorporan en la síntesis de proteínas, se deduce que al menos para algunos aminoácidos existe más de un codon, o sea, el código es degenerado. Los codones que se traducen en el mismo aminoácido reciben el nombre de sinónimos.

Otra consideración importante se refiere a la forma en que el ARNm es traducido, si el código es o no superpuesto. Si el código no es superpuesto cada base nitrogenada formará parte de un solo codon, mientras que si fuera superpuesto formaría parte de más de uno:

1. UAG CCU UAG CCG GAA UUA GCU	No superpuesto
2. UAGCCUAGCCGGAAUUAGCU	ARNm
3. UAG AGC GCC CCU CUU UUA UAG AGC GCC	Superpuesto

Esta alternativa pudo resolverse teniendo en cuenta que se conocía la secuencia de aminoácidos de algunas proteínas. Si el código fuera superpuesto, un cambio en una base ocasionaría el cambio de más de un aminoácido continuo en la proteína. Sin embargo eran conocidas algunas proteínas en las cuales sólo se producía el cambio de un aminoácido por otro, lo que implica que el código no es superpuesto.

Las 3 primeras características del código fueron: que estaba formado por tripletes (codones), éstos no eran superpuestos y que el código tenía carácter degenerado.

## Descifrado del código

Los primeros experimentos encaminados a descifrar el código fueron realizados por **Marshall Nirenberg** y **Heinrich Mathei** en 1961. Ellos tuvieron un éxito inicial al conseguir un extracto celular estable que sintetizaba proteínas activamente, y lograron dirigir la síntesis de un polipéptido por traducción de un polirribonucleótido sintético que actuaba como ARNm (Fig. 28.1).

Este sistema fue útil no sólo en la dilucidación del código genético, sino también para esclarecer muchos aspectos relacionados con el mecanismo de síntesis de proteínas.

Los polinucleótidos se sintetizaron utilizando una enzima - polinucleótido fosforilasa-, descubierta en 1955 por **Marianne Grumberg-Manago** y **Severo Ochoa**. Esta enzima se diferencia notablemente de la ARN polimerasa, pues utiliza como sustratos los nucleósidos difosfatados; como no libera pirofosfato su acción no progresa acoplada a la hidrólisis de éste, y lo más importante es que esta enzima no es dirigida por un molde de ADN, por lo que la secuencia del polinucleótido es al azar dictada por la proporción relativa de los diferentes nucleósidos difosfatos presentes en la mezcla de reacción.

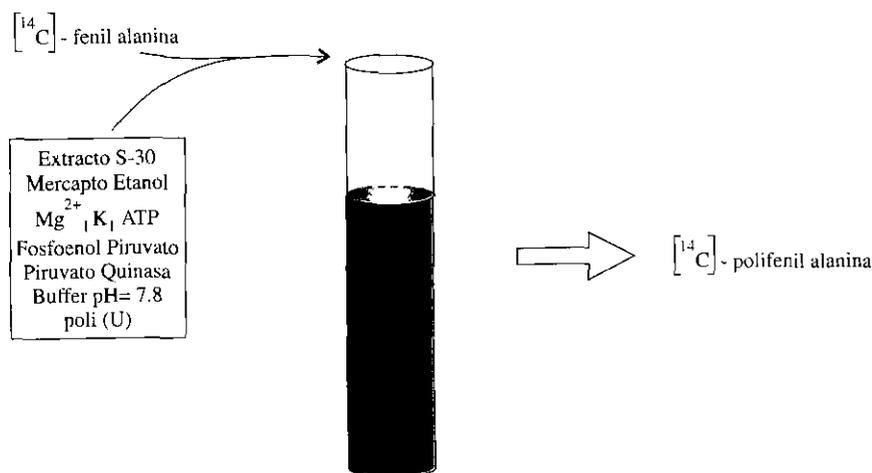


Fig. 28.1. Obtención de extractos celulares sintetizadores de proteínas. Las bacterias son trituradas con alumina como abrasivo y se le añade desoxirribonucleasa que, al hidrolizar el ADN, evita que en el extracto pueda aparecer ARNm endógeno. Una primera centrifugación a baja velocidad elimina la alumina y los restos celulares que no están bien homogeneizados. Una segunda centrifugación a 30 000 g sirve para eliminar membranas y paredes celulares, lo que deja un sobrenadante el cual se denominó S-30 (sobrenadante de 30 000 g) con el que se realizaron los primeros experimentos. Una tercera centrifugación a 100 000 g permite separar los ribosomas, y a partir del sobrenadante S-100 se pueden obtener los ARNt y las enzimas activadoras. El preparado S-100 se utilizó en experimentos posteriores.

El procedimiento consistía en preparar un conjunto de tubos de ensayo, de composición similar, en cada uno de los cuales se añadía un aminoácido diferente marcado con  $^{14}\text{C}$ .

Cuando al sistema se le añadía poliuridina se obtenía la formación de polifenilalanina. Se logró de esta manera descifrar la primera palabra del código: el codon UUU significa fenilalanina (Fen).

Por un procedimiento similar se pudieron asignar otras 2 palabras: el CCC para la prolina (Pro) y el AAA para la lisina (Lis). Los polímeros de poliguanosina no pudieron emplearse porque forman una compleja estructura tricaténaria que no sirve de matriz para la traducción.

La etapa siguiente consistió en determinar la composición -no la secuencia- de los demás tripletes. Si la enzima polinucleótido fosforilasa se incubaba con UDP, el resultado por supuesto es el poli U. Si se incubaba con UDP y GDP se obtiene un copolímero que contiene U y G. La secuencia de bases es azarosa, pero controlando la proporción de nucleótidos se puede calcular la frecuencia con que deben producirse cada uno de los codones. La tabla 28.1. muestra los resultados obtenidos en un polinucleótido preparado a partir de una mezcla que contenía 76 % de U y 24 % de G.

Tabla 28.1. Frecuencia relativa de tripletes de bases obtenidas al controlar la proporción de cada una en la mezcla de reacción

Triplete	Probabilidad	Frecuencia
UUU	$0,76 \times 0,76 \times 0,76 = 0,439$	100
UUG	$0,76 \times 0,76 \times 0,24 = 0,139$	31,6
UGU	$0,76 \times 0,24 \times 0,76 = 0,139$	31,6
GUU	$0,24 \times 0,76 \times 0,76 = 0,139$	31,6
UGG	$0,76 \times 0,24 \times 0,24 = 0,043$	10,0
GUG	$0,24 \times 0,76 \times 0,24 = 0,043$	10,0
GGU	$0,24 \times 0,24 \times 0,76 = 0,043$	10,0
GGG	$0,24 \times 0,24 \times 0,24 = 0,012$	3,1

Nota: Aparecen 4 clases de frecuencias para los 8 codones posibles.

Si se utiliza este polímero en un sistema como el descrito anteriormente, con 20 tubos de ensayo que contengan los 20 aminoácidos, pero sólo uno de ellos marcado con  $^{14}\text{C}$ , se puede medir la frecuencia con que cada aminoácido se incorpora al polipéptido que se sintetiza (tabla 28.2).

**Tabla 28.2.** Composición de los codones según la frecuencia de incorporación de los aminoácidos

Aminoácido	Cantidad incorporada	Composición del codon
Fen	100	3 U
Val	37	2 U + 1 G
Leu	36	2 U + 1 G
Cys	35	2 U + 1 G
Trp	14	2 G + 1 U
Gli	12	2 G + 1 U

La comparación de los 2 resultados sugiere que los codones correspondientes a la valina (Val), cisteína (Cis) y leucina (Leu) contienen 2 U y 1 G, mientras que los del triptófano (Trp) y glicina (Gli) tienen 2 G y 1 U.

Por variaciones en la composición del polímero se lograron otras combinaciones hasta determinar todos los codones que contenían U, después se procedió de igual forma para los que contenían C y por último, los que tenían A. Los de G no se determinaron por las razones expuestas.

Mediante este procedimiento se pudo conocer la composición (no la secuencia) de aproximadamente 50 codones.

Otros experimentos de gran relevancia fueron realizados para asignar los codones que faltaban y para comprobar los existentes.

Una nueva línea experimental fue asumida por *Nirenberg* y otros. En 1964 se conocía que algunos polinucleótidos sintéticos estimulaban la unión de los ARNt con sus aminoácidos correspondientes y los ribosomas, este complejo de 3 componentes podía ser identificado por el método de ultracentrifugación, en un gradiente de densidad de sacarosa. Desafortunadamente este método de análisis era demasiado laborioso; *Nirenberg* y otros diseñaron una forma ingeniosa y rápida de aislar el aminoacil-ARNt unido al ribosoma y al ARNm, utilizando un filtro de nitrocelulosa. Los ribosomas quedaban adheridos al filtro, sin embargo, los aminoacil-ARNt pasaban por el filtro fácilmente, excepto que estuvieran unidos a los ribosomas. En particular *Nirenberg* y *Phillip Leder* (1964) fueron capaces de inducir la unión de Fen-ARNt al ribosoma, utilizando poli U sin otro homopolímero, el complejo formado fue retenido en el filtro de nitrocelulosa. La utilización de copolímeros como ARNm sintéticos, traía por resultado la adsorción de aminoacil-ARNt-ribosoma, cuyo aminoácido coincidía con los resultados de los experimentos realizados en los sistemas libres de células. Estos experimentos también demostraron la existencia de un complejo aminoácido-ARNt-ARNm-ribosoma como intermedio durante la síntesis de proteínas.

Los trinucleótidos estimulaban la formación del complejo, dando una confirmación directa del carácter de tripletes del código genético. *Nirenberg* se dio cuenta que era más fácil sintetizar trinucleótidos de secuencia conocida que un ARNm; su trabajo que comenzó con el triplete GUU correspondiente a Val, progresó rápido, y alrededor de 50 de los 61 codones fueron asignados directa y convincentemente por este método.

Otra línea experimental diferente siguió *H. Ghobind Khorana* (1966), quien combinando procedimientos enzimáticos y de síntesis orgánica logró la síntesis de polinucleótidos de secuencia conocida, a partir de la polimerización de dinucleótidos; uno de ellos fue el poli (AC) que produce ACACACACAC, el cual puede ser leído en grupos de 3 bases como ACA|CAC|ACA|CAC de manera que el polipéptido sintetizado, tomando este polímero como ARNm, constaría de 2 aminoácidos alternantes, que fue el siguiente:

**Tre-His-Tre-His-Tre-His**

Por los experimentos descritos anteriormente, se sabía que la composición del codon de la treonina (Tre) era de 2 A y 1 C y el de His de 2 C y 1 A, por lo tanto ACA codifica para Tre y CAC para His.

La utilización del poli(UG) producía:

**Cis-Val-Cis-Val-Cis-Val**

con lo que se comprobaba que UGU correspondía a la Cis, y se asignó el GUG a la Val.

*Khorana* logró también la polimerización de trinucleótidos, por ejemplo, el poli (UAC) que puede ser leído de 3 formas diferentes:

1. UAC|UAC|UAC|UAC|UAC|UAC|UAC que origina politirosina
2. ACU|CU|ACU|ACU|ACU|ACU|ACU que codifica politreonina
3. CUA|CUA|CUA|CUA|CUA|CUA|CUA que forma polileucina

Experimentos como estos consiguieron no sólo determinar la secuencia de muchos de los codones de composición conocida, sino además, corroborar los que habían sido asignados por otros métodos.

**Codones de terminación**

Los codones que indican el final de la cadena polipeptídica y que no codifican ningún aminoácido también fueron asignados por la utilización de polímeros de secuencia conocida.

Si se utilizaba el poli(UAUC) la traducción podía comenzar en 4 sitios diferentes,

UAU|CUA|UCU|AUC|UAU|CUA|UCU|AUC|  
 AUC|UAU|CUA|UCU|AUC|UAU|CUA|UCU|  
 UCU|AUC|UAU|CUA|UCU|AUC|UAU|CUA|  
 CUA|UCU|AUC|UAU|CUA|UCU|AUC|UAU|

originando sólo 4 codones diferentes (UAU, AUC, UCU y CUA) que produce la secuencia alternante Tir-Leu-Ser-Ile que puede comenzar con cualquiera de ellos. Esta situación se repetía con muchos polinucleótidos del tipo poli(MNPQ).

Cuando se ensayó el poli(GUAA) se esperaba encontrar un resultado similar, pues éste debe producir también 4 codones y por tanto un tetrapéptido repetido:

GUA|AGU|AAG|UAA|GUA|AGU|AAG|UAA|  
 UAA|GUA|AGU|AAG|UAA|GUA|AGU|AAG|  
 AAG|UAA|GUA|AGU|AAG|UAA|GUA|AGU|  
 AGU|AAG|UAA|GUA|AGU|AAG|UAA|GUA|

Sin embargo, se encontró que no se producían largos polipéptidos, sino que se formaba sólo 2 productos: el tripéptido Val-Ser-Lis y el dipéptido Ser-Lis. A partir de

los codones ya asignados (Val=GUA, Ser=AGU y Lis=AGG) se pueden alinear los péptidos con los distintos tipos de secuencia y se obtienen los resultados siguientes:

1. GUA AGU AAG UAA GUA AGU AAG UAA  
Val - Ser - Lis - ?
2. UAA GUA AGU AAG UAA GUA AGU AAG  
Val - Ser - Lis - ?
3. AAG UAA GUA AGU AAG UAA GUA AGU  
Lis ?
4. AGU AAG UAA GUA AGU AAG UAA GUA  
Ser - Lis - ?

Lo que muestra que la síntesis se detiene en el codon anterior al UAA y esto indicaba que éste debía una señal de terminación.

Experimentos similares mostraron que el poli(AUAG) producía sólo el tripéptido Ile-Asp-Arg y el dipéptido Asp-Arg, probando que UAG era otro codon de terminación. Por último, con poli(UGA) sólo se formaba polimetionina y poliaspártico, lo que llevó a la conclusión que UGA era también un codon de terminación.

Como resultado de una broma de laboratorio a estos 3 codones se le dieron nombres propios -ámbar para el UAG, ocre al UAA y ópalo al UGA- y éstos se utilizan corrientemente en la literatura científica.

### Codon de iniciación

En los sistemas *in vitro* la síntesis de proteínas comienza en cualquiera de las bases del polinucleótido sintético utilizado, pero *in vivo* esto no es así, se requiere de un codon de iniciación. Por experimentos *in vitro* se ha podido determinar que el más utilizado para ello es el AUG, aunque en muy pocos casos también el GUG; esto fue corroborado más tarde en numerosos estudios de secuencias de ADN. Estos codones tienen además la función de codificar aminoácidos específicos -el AUG metionina y el GUG valina- para incorporarlos dentro de la cadena. En el capítulo 30 se explicará cómo la célula puede diferenciar una de otra.

Como resultado de todos estos trabajos a finales de 1966 el código genético había sido descifrado completamente. El total de codones y su significado aparece en la figura 28.2.

Aunque el código fue descifrado por experimentos *in vitro*, la exactitud de la asignación de cada codon ha sido confirmada por análisis genéticos y bioquímicos. La

Fig. 28.2. El código genético casiuniversal. Estructura actual del código genético. Nótese que en la mitad izquierda predominan los aminoácidos apolares y en la derecha los polares; estas 2 mitades difieren en que la segunda base sea pirimidínica a la izquierda, o purínica a la derecha. Ocho de las 16 casillas están ocupadas por el mismo aminoácido, lo que significa que la tercera base es redundante. Sólo 2 aminoácidos están codificados por un codon cada uno. INI se refiere al codon de iniciación y TER a los de terminación.

		Segunda letra				
		U	C	A	G	
U	Primera letra (5')	Fen	Ser	Tir	Cys	U
		Fen	Ser	Tir	Cys	C
		Leu	Ser	Ter	Ter	A
		Leu	Ser	Ter	Tir	G
C	Leu	Pro	His	Arg	U	
	Leu	Pro	His	Arg	C	
	Leu	Pro	GIN	Arg	A	
	Leu	Pro	GIN	Arg	G	
A	Ile	Tre	AsN	Ser	U	
	Ile	Tre	AsN	Ser	C	
	Ile	Tre	Lis	Arg	A	
	Met/Ini	Tre	Lis	Arg	G	
G	Val	Ala	Asp	Gli	U	
	Val	Ala	Asp	Gli	C	
	Val	Ala	Glu	Gli	A	
	Val	Ala	Glu	Gli	G	

reciente tecnología de secuenciación del ADN ha dado la confirmación definitiva a este problema, que parecía un sueño casi imposible, cuando fue postulado en 1953 y que poco más de 10 años después quedaba totalmente resuelto.

## Universalidad del código

Tanto los experimentos de *Niremberg* como los de *Khorana* fueron realizados utilizando los componentes del sistema biosintético de proteínas de la *E. coli*, pero después fueron realizados repetidamente utilizando bacterias, virus, hongos, plantas y animales, incluyendo a los mamíferos. Los resultados en todos los casos coincidieron y llevaron a postular la universalidad del código.

Más tarde se ha encontrado que el sistema informativo de las mitocondrias presenta desviaciones con respecto al código universal. Esas variaciones ya fueron estudiadas en el capítulo 37 (tabla 37.4). Para otros organismos el código mitocondrial ofrece otras variaciones.

Estos descubrimientos limitaron el carácter universal del código y por ello hoy se dice que es casi universal.

## Estructura del código

Mucho se ha escrito y especulado sobre la forma de organización del código, tratando de descubrir cuáles son las leyes que subyacen en su organización. Aunque este objetivo no parece haberse alcanzado totalmente, se han podido señalar algunas de sus regularidades más sobresalientes.

El punto central de la atención ha sido el carácter degenerado del código. Para 32, o sea, la mitad de los codones, se tiene una degeneración completa de la tercera base, el aminoácido queda codificado totalmente por las 2 primeras bases (por ejemplo, UC codifica Ser) para los demás (con la excepción de las parejas UGA/UGG y AUA/AUG) la degeneración es casi completa, pues sólo requiere el carácter purínico o pirimidínico de la tercera base. Esto se resume diciendo que los codones del tipo XYU y XYC son siempre sinónimos al igual que los del tipo XYA y XYG, con las excepciones mencionadas (tabla 28.3).

Tabla 28.3. Grupo I de codones, donde las 2 primeras bases son suficientes para la codificación del aminoácido

X	Y	N	Aminoácido
G	G	6	Gli
G	C	6	Ala
C	G	6	Arg
C	C	6	Pro
C	U	5	Leu
G	U	5	Val
A	C	5	Tre
U	C	5	Ser

Nota: Grupo I. Para el codon XYZ. Z =(A,G,C,U)

Teniendo esto presente se pueden agrupar los codones sólo por las 2 primeras bases, en 2 grupos de 8 cada uno: en el primer grupo los pares XY, que codifican los aminoácidos con independencia total de Z y, en el segundo, los que varían según Z sea purínica (Pu) o pirimidínica (Pi).

La columna n indica el número de puentes de hidrógeno que puede formar la pareja XY al unirse a un par complementario (X'Y'). El valor n pudiera llamarse grado de complementariedad, que en el primer grupo es 6 y 5 (promedio 5,5) y el segundo 5 y 4 (promedio 4,5). Se puede pensar que mientras mayor sea n, menor valor tendrá la interacción Z-Z', ya que la unión XY-X'Y' es suficientemente estable. El valor de la tercera base será discutido nuevamente en el siguiente acápite. Estas características han hecho suponer que el código primitivo era de sólo 2 letras y que ha evolucionado hacia su estado actual de 3 letras.

**Tabla 28.4. Grupo II de codones, donde la tercera base tiene una función determinante**

X	Y	N	Aminoácido	
			Z=Pu	Z=Pi
G	A	5	Glu	Asp
A	G	5	Arg	Ser
U	G	5	Trp/Ter	CyS
C	A	5	GIN	His
A	A	4	Lis	AsN
A	U	4	Ile/Met	Ile
U	A	4	Ter	Tr
U	U	4	Leu	Fen

Nota: Grupo II. Para el codon XYZ.

La secuencia de bases codifica sólo la secuencia de aminoácidos de la proteína, pero su estructura tridimensional depende de su secuencia. Alteraciones en la secuencia pueden ocasionar alteraciones en la conformación y, por lo tanto, en la función. Como se vio en la sección de biomoléculas, en esto influye notablemente la polaridad de la cadena R de los aminoácidos, los de carácter hidrofóbicos son los determinantes en el plegamiento de la cadena. La sustitución de un aminoácido apolar por uno polar puede causar cambios drásticos en la conformación de la proteína; alteraciones de este tipo sólo ocurren cuando existe el cambio de una purina por una pirimidina que esté ocupando la segunda posición del codon. Los cambios en la primera base también influyen menos en la naturaleza polar o apolar del aminoácido codificado. Sin embargo debido al carácter degenerado del código es muy poco probable que esto suceda.

Si se toma como ejemplo el codon GUG y se analizan todos los cambios posibles y sus resultados, se verá en la figura 28.3 que de 9 cambios posibles, sólo en un caso que afecta la segunda base, la val se cambiaría por un aminoácido polar (glu) y, por tanto, podría influir notablemente en la estructura de la proteína. En codones de otros aminoácidos hidrofóbicos esto es menos evidente, pero en general 2 de cada 3 cambios no producen alteraciones en el carácter hidrofóbico del residuo de aminoácido.

No obstante, lo dicho anteriormente, en ocasiones existen aminoácidos que son tan valiosos en la estructura de la proteína, que no admiten ser cambiados por ningún otro aunque éste sea muy semejante, por ejemplo, un aminoácido cuya cadena R aporte el grupo catalítico al centro activo de una enzima.

## Descodificación

Una vez conocido el código genético, o sea, la relación entre la secuencia de bases del ARNm y la de aminoácidos de las proteínas, es necesario conocer cómo se realiza el proceso de descodificación, cómo se logra la conversión de un lenguaje en otro. Al estudio detallado de los mecanismos de este proceso está dedicada el capítulo 30. Ahora sólo se estudiará brevemente el fundamento del proceso de descodificación.

Para poder descifrar el código hace falta otro tipo de ARN llamado de transferencia y cuya estructura fue estudiada en la sección de biomoléculas. Es necesario recordar que en estas moléculas existen 2 sitios bien definidos que ocupan los extremos de la estructura de la «L». En uno de ellos (el extremo 3'-OH) se encuentra el triplete 5'-CCA-3' a cuyo 3'-OH se une un aminoácido específico; en el otro extremo se encuentra el asa anticodon con su secuencia característica 5'---Pi--U--XYZ--Pu(modificada)--3', donde XYZ representa el anticodon. Es este triplete el que se une al codon del ARNm durante la traducción, por la formación de pares de bases complementarias. Por ejemplo, si el anticodon tiene la secuencia 5'--AAA--3' su codon complementario será el UUU. A este tipo de ARNt se unirá por el extremo 3'-OH el aminoácido fenilalanina. Si el anticodon fuera 5'--GGG--3' se complementarían con el codon 5'--CCC--3' y el ARNt llevaría en el extremo 3'-OH a la prolina. La reacción de unión del aminoácido al ARNt correspondiente se estudiará con más detalles en el capítulo 30.

Teniendo en cuenta que existen 61 codones que codifican aminoácidos, cabría suponer la existencia de 61 tipos de ARNt con sus anticodones correspondientes a cada uno de los codones del ARNm. Esta idea fue sustentada inicialmente al hallar que cuando se purificaba ARNt de la leucina (leu-ARNt) y se examinaba en un sistema artificial de síntesis de proteínas, había una especie (leu-ARNt<sub>I</sub>) que incorporaba el aminoácido cuando se utilizaba poli(UC), pero no al emplear poli(UG), mientras otra especie (leu-ARNt<sub>II</sub>) lo hacía con el segundo pero no con el primero, esto significa que el leu-ARNt<sub>I</sub> lee el codón CUU y el leu-ARNt<sub>II</sub>, el GUU.

Sin embargo, la idea un codon-un ARNt no prosperó mucho tiempo debido a los experimentos de Holley, quien logró establecer la secuencia completa de bases del ala-ARNt de levadura. Esta molécula lee 3 (GCU, GCC y GCA) de los 4 codones de la alanina. El examen de la secuencia de las zonas no apareadas mostró que en un solo sitio había una secuencia 5'--GC--3' por lo que esta pareja debía ser parte del anticodon (se debe recordar que por convención, las secuencias de los ácidos nucleicos se escriben del extremo 5' al 3' y el apareamiento se realiza en sentido antiparalelo, por lo que el codon 5'--AUC--3' debe tener un anticodon 5'--GAU--3').

Esto implicaba que el anticodon fuera la secuencia 5'--IGC--3', en la cual I simboliza el nucleótido raro de inosina. Como el ala-ARNt responde a los 3 codones (GCU, GCC y GCA) entonces, la inosina no forma puentes de hidrógeno o puede formarlos con U, C y A. Si fuera la primera situación también leería el codon GCC, lo cual no sucedía, por lo tanto, I puede formar puentes de hidrógeno, aunque no sea en la forma de los pares habituales U-A y G-C (Fig. 28.4).

La aparición de la inosina en el anticodon se debe a que siempre que en la primera posición aparezca A en el transcrito primario, ésta es desaminada enzimáticamente produciendo hipoxantina, que es la base que forma parte de la inosina.

La explicación a este fenómeno fue propuesta por Francis Crick (1965) con la idea del acoplamiento vacilante (*wobble*). Hasta ese momento se creía que apareamientos diferentes de A-U, A-T y G-C no apareaban en los ácidos nucleicos. Esto es cierto en el

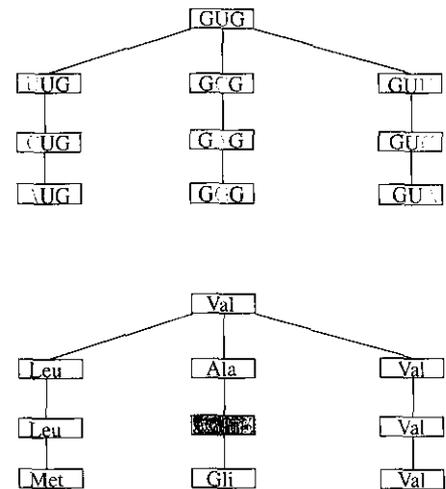


Fig. 28.3. Importancia del carácter degenerado. A partir del codon GUG se pueden obtener 9 codones, cambiando una sola base cada vez. Sin embargo en 3 ocasiones el significado del codon no cambia y de las 9 posibilidades sólo en un caso se altera el carácter polar del aminoácido codificado.

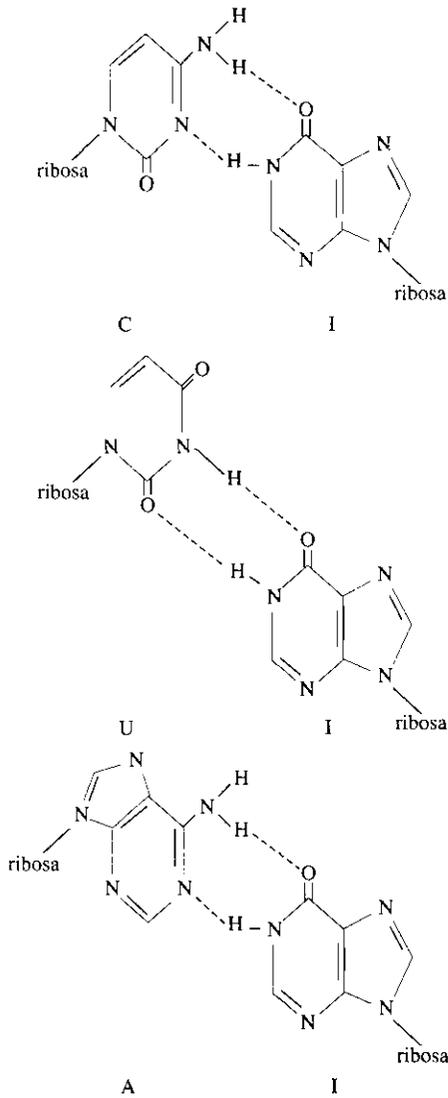


Fig. 28.4. La inosina como formadora de pares de bases. La inosina (en color) puede formar pares de bases con la citosina (arriba), uracilo (centro) y adenina (abajo), por lo cual su aparición en el anticodon del ARNt no impide el apareamiento con el codon del ARNm. La aparición de la inosina en la primera posición del anticodon permite que varios codones sean leídos por el mismo ARNt.

ADN por las características estructurales de la molécula, como se vio en la sección de biomoléculas. Sin embargo, como el apareamiento codon-anticodon se produce entre 2 moléculas de ARN no es necesario conservar la estructura regular de la doble hélice. *Crick* demostró que pueden existir otros apareamientos en la interacción codon-anticodon, pero que esto requería en primer lugar que las 2 primeras bases formaran apareamientos típicos, para que se garantizara una estabilidad máxima y, en segundo lugar, que el tercer par de bases no produjera mucha distorsión como ocurre habitualmente con los pares purina-purina, identificando como posibles los apareamientos que aparecen en la figura 28.5.

La posibilidad de formar los 4 pares adicionales de bases (A-I, U-I, C-I y G-U) explica cómo una sola molécula de ARNt puede aparearse con varios codones.

Ahora es comprensible el caso de los ala-ARNt de levadura. Uno de ellos, estudiado por *Holley*, que tiene el anticodon IGC puede formar pares de bases con 3 de los codones (GCU, GCC y GCA), pues la inosina puede aparearse con U, C y A. El otro (denominado ala-ARNt<sub>II</sub>) sólo acopla con el codón GCG, para lo cual serían posibles 2 anticodones el CGC y el UGC, pues G puede formar pares de bases con C y con U. Si fuera el UGC entonces debía leer también GCG y GCA, pero como este no es el caso, el anticodon debía ser CGC y lo es en realidad.

Aunque en solución los trinucleótidos se aparean mal, pues su tamaño no les permite estabilizarse por las interacciones de apilamiento de bases; en el proceso de traducción estas interacciones codon-anticodon se estabilizan debido a varios factores:

1. Tanto el codon como el anticodon forman parte de moléculas mayores, lo cual permite que la unión entre ellos pueda estabilizarse por apilamiento de bases.
2. Hacia el lado 5' del anticodon siempre hay una U, lo que sugiere que debe tener alguna participación en la estabilización de la unión codon-anticodon.
3. El apareamiento se produce normalmente cuando ambos (ARNt y ARNm) están unidos a los ribosomas. Es posible que esa unión haga que el codon y el anticodon se posicionen de forma que la estabilidad de la unión sea máxima.

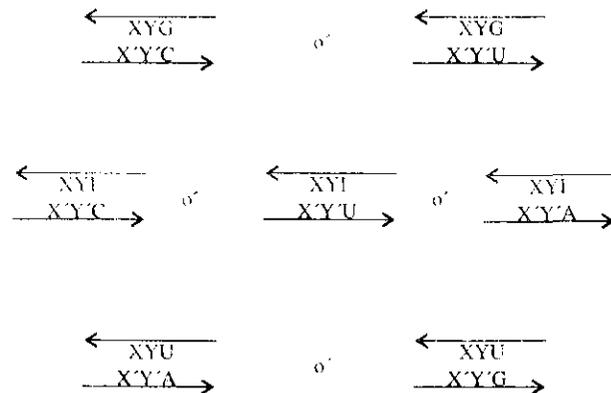


Fig. 28.5. Modelo de acoplamiento vacilante. Según la hipótesis del apareamiento vacilante en la unión del codon con el anticodon, pueden aparecer pares de bases que no se encuentran normalmente en el ADN. En cada caso el anticodon está en rojo y el codon en azul. La flecha indica la dirección 5'→3'. En todos los casos sólo se representa la tercera base. Obsérvese que los codones que aparecen en la misma horizontal son siempre sinónimos.

## Resumen

Desde el punto de vista bioquímico el código genético es la relación de equivalencia entre la secuencia de bases nitrogenadas de los ARNm y la secuencia de aminoácidos de las proteínas. El codon es la unidad de codificación y está formado por 3 bases, por lo que la relación de codificación es 3:1. Cada base nitrogenada del ARNm forma parte de un solo codon, lo que quiere decir que el código no es super-

puesto. Por lo general cada aminoácido es codificado por más de un codon, lo que significa que el código es degenerado.

El descifrado del código fue posible gracias a 2 grupos de trabajo principalmente el de Niremberg y el de Khorana. El primero, utilizando un sistema libre de células y polinucleótidos sintéticos, como inicio, y trinucleótidos de secuencia conocida, después, logró establecer la estructura de numerosos codones. El segundo, utilizando polinucleótidos de secuencia conocida confirmó las asignaciones hechas por Niremberg y estableció la secuencia de otros codones, entre ellos, los de terminación. A finales de 1966 el código estaba totalmente descifrado.

Todas las bacterias, virus, hongos, vegetales y animales utilizan el mismo código; sólo en las mitocondrias aparecen variaciones, por lo que se considera que el código tiene carácter quasiuniversal.

Los análisis de la estructura del código llevan a la idea de que su forma actual es resultado de un proceso evolutivo, en el cual se ha ido ganando en seguridad. El carácter degenerado posibilita amplias variaciones en la secuencia de bases del ADN, y por lo tanto del ARNm, sin modificar esencialmente la estructura primaria de las proteínas.

El proceso de descodificación se lleva a cabo en los ribosomas con la participación de los ARNt, que contienen el triplete anticodon, que al aparearse con el codon facilitan que cada aminoácido ocupe una posición precisa en la cadena polipeptídica. El apareamiento se produce con mayor fuerza en las 2 primeras bases que en la tercera, lo que explica que un mismo ARNt pueda leer varios codones.

## Ejercicios

1. Como se observa en la figura 28.2 existen 6 codones para la Leu, Ser y Arg ¿Cuál será el número mínimo de Leu-ARNt, Ser-ARNt y Arg-ARNt que debe tener una célula para sintetizar proteínas eficientemente si todos los codones son utilizados con la misma frecuencia?
2. El codon UAC codifica Tir y el UAG es un codon de terminación ¿Podría un tir-ARNt aparearse al UAG y provocar que la síntesis de la proteína continuara más allá de su límite normal?
3. El codon de iniciación preferencial es el AUG, pero se ha visto que también puede emplearse el GUG y en ambos casos se incorpora Met ¿Pudiera explicarse esta situación, suponiendo que un mismo ARNt pueda leer los 2 codones?
4. Los ácidos aspártico y glutámico ocupan la misma casilla del código, difieren sólo en la tercera base ¿Cuáles deben ser los anticodones del asp-ARNt y el glu-ARNt para que en la cadena polipeptídica no pueda incorporarse uno de ellos en vez del otro?
5. Un bioquímico afirmó haber establecido la secuencia de bases de un Tir-ARNt cuyo anticodon tenía la secuencia 5' - 1CA - 3', pero este hallazgo se puso en duda por todos sus colegas ¿Cuáles serían las razones que hicieron poner en duda el hallazgo anunciado?
6. Un ARNm es modificado artificialmente incluyéndole una G entre los codones 15 y 16, y se observa que el polipéptido sintetizado difiere del original en la secuencia de aminoácidos a partir del 16; después se hace una segunda modificación al ARNm y se observa que ahora la secuencia de aminoácidos del polipéptido sólo difiere del original en los aminoácidos 16, 17 y 18 ¿En qué consistió la segunda modificación? ¿Cómo se pueden explicar los 2 resultados obtenidos?