

**Cómo el ARN se convirtió en la clave
del origen y el futuro de la vida**

**El
Catalizador**

Thomas R. Cech

Premio Nobel de Química

Ariel

Thomas R. Cech

El catalizador

Cómo el ARN se convirtió en la clave
del origen y el futuro de la vida

Traducción de Pedro Pacheco González

Ariel

Título original: *The Catalyst. RNA and the Quest to Unlock Life's Deepest Secrets*

Primera edición: marzo de 2025

© Thomas R. Cech, 2024

Publicado originalmente en inglés por W. W. Norton & Company

© Pedro Pacheco González, por la traducción, 2025

Derechos exclusivos de edición en español:

© Editorial Planeta, S. A., 2025

Avda. Diagonal, 662-664, 08034 Barcelona

Editorial Ariel es un sello editorial de Planeta, S. A.

www.ariel.es

www.planetadelibros.com

ISBN: 978-84-344-3845-3

Depósito legal: B. 3.606-2025

Impreso en España

La lectura abre horizontes, iguala oportunidades y construye una sociedad mejor. La propiedad intelectual es clave en la creación de contenidos culturales porque sostiene el ecosistema de quienes escriben y de nuestras librerías. Al comprar este libro estarás contribuyendo a mantener dicho ecosistema vivo y en crecimiento.

En **Grupo Planeta** agradecemos que nos ayudes a apoyar así la autonomía creativa de autoras y autores para que puedan seguir desempeñando su labor.

Dirígete a CEDRO (Centro Español de Derechos Reprográficos) si necesitas fotocopiar o escanear algún fragmento de esta obra. Puedes contactar con CEDRO a través de la web www.conlicencia.com o por teléfono en el 91 702 19 70 / 93 272 04 47.

Queda expresamente prohibida la utilización o reproducción de este libro o de cualquiera de sus partes con el propósito de entrenar o alimentar sistemas o tecnologías de inteligencia artificial.



Índice

<i>Introducción: La era del ARN</i>	13
---	----

PRIMERA PARTE LA BÚSQUEDA

1. El mensajero	23
2. Ajuste vital	47
3. Ir por libre	67
4. Formas cambiantes	93
5. La nave nodriza	119
6. Orígenes	139

SEGUNDA PARTE LA CURA

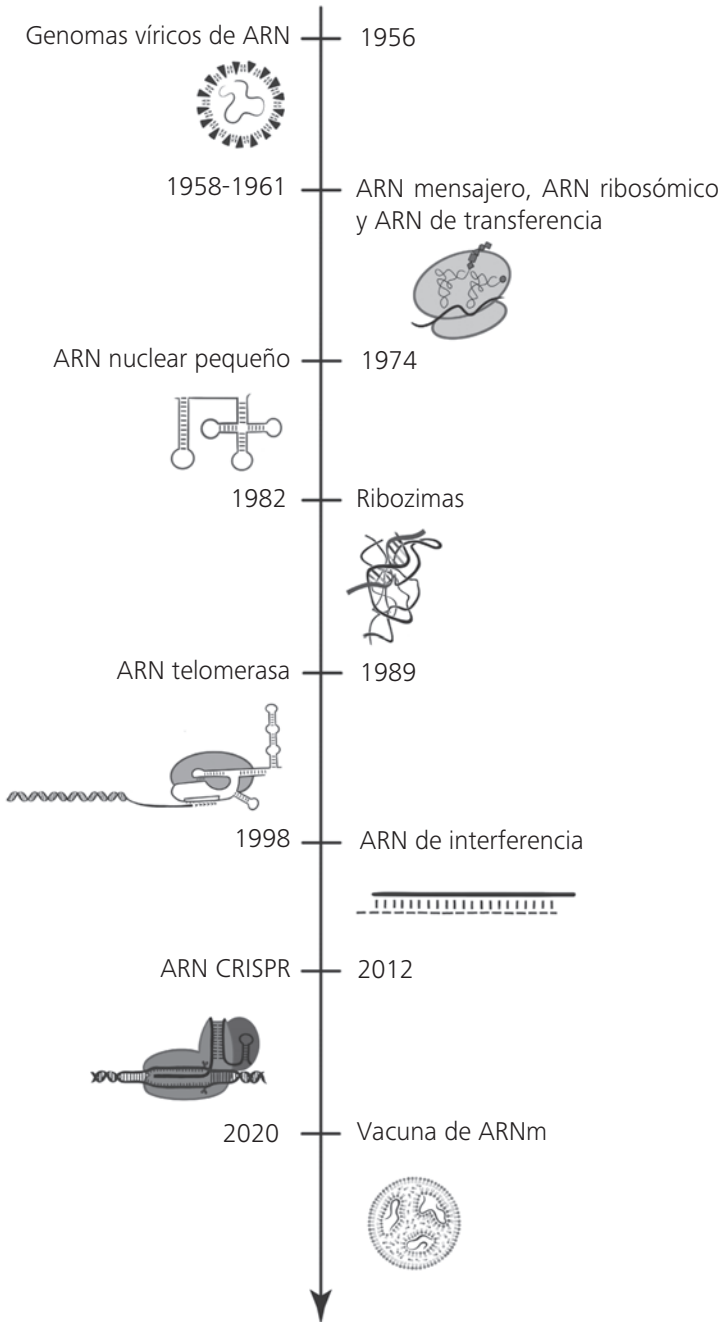
7. ¿Es la fuente de la juventud una trampa mortal?	161
8. Gusanos que se retuercen	189
9. Parásitos precisos, copias chapuceras	209
10. ARN contra ARN	225
11. Utilizando las tijeras	253
<i>Epílogo: El futuro del ARN</i>	283

<i>Agradecimientos</i>	293
<i>Notas</i>	297
<i>Glosario</i>	337
<i>Índice analítico</i>	349

PRIMERA PARTE

LA BÚSQUEDA

Distintas formas y funciones del ARN



El mensajero

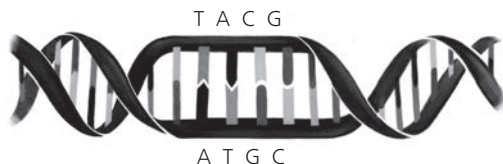
Antes de dedicarse a descifrar el código de la vida, George Gamow había resuelto muy diversos y relevantes problemas científicos. Nacido en 1904 en Odesa, ciudad portuaria del mar Negro, Gamow empezó a contemplar el universo a los seis años, cuando vio el cometa Halley desde la azotea del edificio de apartamentos donde residía con su familia. Cuatro décadas después, se convertiría en el principal defensor mundial de la teoría que aseguraba que el universo comenzó con una «gran explosión» o *Big Bang*.¹ Los colegas científicos de Gamow lo consideraban un genio, «otro Heisenberg»,² dijo Niels Bohr, comparándolo con el premio Nobel pionero de la mecánica cuántica, pero también un bicho raro, «un diablillo gigante, que salta de los átomos a los genes y de ahí a los viajes espaciales», afirmó Jim Watson.³

Gamow destacaba por su 1,90 m de estatura y su particular sentido del humor, del que hacía gala incluso en los ambientes académicos más serios. Cuando publicó la teoría del origen cosmológico de los elementos químicos que había desarrollado con su alumno Ralph Alpher, Gamow incluyó el nombre de su colega Hans Bethe, simplemente para crear una lista de autores Alpher-Bethe-Gamow acorde con el alfabeto griego.

Tras huir de la Unión Soviética en 1933, Gamow llegó a Estados Unidos un año después. Fue profesor de Física durante veinte años en la Universidad George Washington de Washington D. C., y, finalmente, llegó a mi universidad,

la de Boulder, donde el edificio más alto del campus (que incluye el Departamento de Física) se llama torre Gamow en su honor. Aunque se había curtido en física nuclear y cosmología, a principios de la década de 1950 Gamow tenía claro que la pregunta científica más apasionante que faltaba por responder no tenía nada que ver con el origen del universo ni con el comportamiento de las partículas subatómicas. De hecho, no tenía nada que ver con la física.

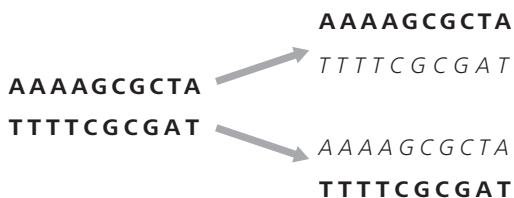
En junio de 1953, Gamow leyó en la revista *Nature* el trascendental artículo de Jim Watson y Francis Crick en el que anunciaban que la estructura del ADN era una doble hélice. Esa estructura ofrecía una solución clara a lo que había sido un gran misterio: cómo se duplica la información genética para que pueda transmitirse de una generación a la siguiente. Las cuatro unidades químicas, o *bases*,* dispuestas a lo largo de una cadena de ADN (adenina [A], timina [T], guanina [G] y citosina [C]) forman pares de bases con sus complementarias situadas en la otra cadena de la doble hélice: A siempre se empareja con T, mientras que G siempre se empareja con C.



Los pares de bases mantienen unidas las dos hebras de la doble hélice del ADN. La sección central se muestra desenrollada para que se puedan apreciar los pares de bases A-T y G-C.

* Las mismas abreviaturas (A, T, G y C) se utilizan tanto para las bases como para los nucleótidos: un nucleótido es una base unida a un azúcar desoxirribosa y a un grupo fosfato. Los nombres químicos completos de los nucleótidos son adenosina, timidina, guanosina y citidina. Aunque distinguir entre nucleótidos y bases tiene importancia bioquímica, su contenido informativo es idéntico.

La doble hélice parece una cremallera retorcida. Si se abriera la cremallera, cada lado tendría toda la información necesaria para dirigir la construcción del lado opuesto, ya que los dos lados son siempre perfectamente complementarios. Watson y Crick llegaron a la conclusión de que así era como se debía replicar la información genética.



Las letras en negrita representan el ADN parental y las letras en cursiva el ADN hijo recién sintetizado.

En una carta fechada el 8 de julio de 1953, Gamow felicitaba a Watson y Crick por convertir la biología en una de las «ciencias exactas».⁴ Y tuvo la audacia de proponer la colaboración entre varias disciplinas científicas para abordar la siguiente gran cuestión que todo el mundo tenía en mente: cómo se leía realmente la información codificada en esas cadenas de A, T, G y C para crear, en última instancia, una mano, un corazón, un hígado o un cerebro; o los guisantes arrugados y lisos en el jardín de Gregor Mendel, el monje agustino que propuso por primera vez que esos rasgos se transmitían de generación en generación en forma de una unidad elemental que ahora llamamos gen. Gamow sugirió que podía ayudar a Watson y Crick a utilizar las matemáticas y la física para descifrar este código genético.

Watson y Crick se sintieron halagados de que un físico famoso se interesara por su trabajo, pero la carta manuscrita de Gamow «era tan extravagante —recordaba Watson—, que no sabíamos lo serio que podía ser».⁵ Gamow iba muy en serio. Durante los meses siguientes se dedicó a descifrar el código genético. En aquella época era asesor de la Marina estadounidense, por lo que no solo buscó la ayuda de quími-

cos y físicos, sino también de criptógrafos militares. Sin embargo, para resolver el secreto del ADN se necesitaba el ácido ribonucleico, o ARN, hijo del ADN.

UN ACERTIJO Y UNA CORBATA

En términos biológicos, «descifrar el código genético» significa entender cómo el ADN codifica las proteínas. La vida se construye a partir de proteínas, que son los principales motores de todos los organismos de nuestra biosfera. En los seres humanos, algunas proteínas forman estructuras como las fibras musculares, la piel y el pelo. Otras actúan como enzimas, descomponiendo los alimentos que ingerimos en sus componentes constituyentes y reciclando estas piezas para construir nuevas máquinas celulares. Otras hacen agujeros en las envolturas que recubren nuestras células, lo que permite la entrada selectiva de algunos nutrientes y sales y la expulsión de otros. Las hay que actúan como moléculas señalizadoras, es decir, reciben información del mundo exterior y activan procesos internos en consecuencia. Otras son anticuerpos, que nos protegen de invasores extraños como los virus. En resumen, las proteínas presentan una enorme diversidad.

Desde el punto de vista meramente químico, las proteínas son polímeros, cadenas de cien o incluso mil aminoácidos. Hay veinte tipos diferentes de aminoácidos con nombres como lisina, valina y fenilalanina. Cada proteína tiene un orden específico, o secuencia, de aminoácidos a lo largo de su cadena.⁶ Esta secuencia determina cómo se pliega la proteína creando una entidad tridimensional que tiene una función particular, como digerir los alimentos en el estómago o transmitir señales neuronales en el cerebro. Así, aunque oigamos decir que necesitamos comer pescado, tofu o una hamburguesa «para ingerir suficiente proteína», no existe una proteína, sino miles diferentes. Y cada una está codificada por su propio gen, que está hecho de ADN.

Mientras que las proteínas están formadas por veinte aminoácidos, el ADN está formado por cuatro bases. Por tanto, por un lado tenemos los genes del ADN compuestos a partir de la secuencia de A, G, T y C y, por otro, las proteínas correspondientes compuestas por veinte aminoácidos diferentes. Gamow trataba de averiguar cómo las diferentes disposiciones de las cuatro bases del ADN especificaban, o codificaban, cada uno de los veinte aminoácidos. En 1954 ya había reclutado a Watson, Crick y un ecléctico grupo de científicos de renombre (veinte en total, uno por cada aminoácido) para que se ocuparan del problema de la codificación.

Gamow mandó tejer corbatas de lana, cada una bordada con una hebra de ARN, para los miembros de esta pequeña fraternidad de intelectuales descifradores de códigos. Bautizó al grupo como «el Club de la Corbata de ARN». ¿Por qué no el «Club de la Corbata del ADN»? Fue Jim Watson quien convenció a Gamow de que tenían que centrar su atención en el ARN debido a la estructura básica de las células.⁷ En los organismos superiores el ADN se encuentra dentro del núcleo celular, pero las proteínas se fabrican fuera del núcleo,⁸ en la región conocida como citoplasma. Esta separación espacial obligaba a que algún tipo de mensajero transportara la información desde el ADN hasta el lugar de producción de las proteínas. Algunos científicos con visión de futuro pensaron que el ARN podría ser ese mensajero, ya que se encontraba en grandes cantidades tanto en el núcleo como en el citoplasma.

Desde principios del siglo xx, los químicos han analizado el ARN y el ADN y han descubierto que están cortados por el mismo patrón bioquímico. Sus nombres completos (ácido desoxirribonucleico y ácido ribonucleico) revelan lo estrechamente relacionados que están. «Desoxi» indica que el ADN tiene un átomo de oxígeno menos que el ARN en cada una de sus unidades; el átomo de oxígeno extra hace que el ARN sea químicamente mucho menos estable que el ADN.

Durante muchas décadas, se creía que el ADN y el ARN eran compuestos químicos sin función alguna.⁹ Tenían mucho menos interés para los científicos que las proteínas, moléculas como la insulina, cuyas funciones como enzimas y como moléculas de señalización resultaban apasionantes. El ADN se convirtió en la pieza central de la biología cuando, en 1944, Oswald Avery y sus colegas del Instituto Rockefeller descubrieron que el ADN era la molécula responsable de los cambios hereditarios que se producían en las bacterias.¹⁰ Poco después, en 1953, Watson y Crick descubrieron la estructura doble helicoidal del ADN.¹¹ Como los científicos ya habían teorizado en 1947 que el ARN era una copia del ADN, supusieron que también debía desempeñar un papel importante en la química de la vida.¹²

Aunque el ARN suele ser monocatenario y el ADN bicatenario, ambos hablan el mismo idioma. Al igual que el ADN, el ARN se escribe con un alfabeto de cuatro letras. Las tres primeras letras (A, G y C) son las mismas que las del ADN. La cuarta letra, U (uracilo), ocupa el mismo lugar en el alfabeto del ARN que el que ocupa la T en el alfabeto del ADN. Así que cuando el ARN se copia del ADN, porta la misma información.

Dedicado como siempre a las grandes teorías, Gamow trató primero de encontrar una solución al problema de la codificación usando tan solo lápiz, papel y su cerebro. En 1953, el mismo año en que se descubrió la doble hélice, planteó que era posible que tres bases codificaran un aminoácido.¹³ Su razonamiento era matemático. Si se tomaban las cuatro letras del alfabeto del ADN o ARN y se intentaban crear tantas combinaciones de dos letras como fuera posible, solo se obtendrían dieciséis combinaciones, demasiado pocas para especificar los veinte aminoácidos que componen las proteínas. Sin embargo, si se intentaran ordenar las cuatro letras del alfabeto del ADN o ARN en tantas combinaciones de tres letras como fuera posible, se obtendrían sesenta y cuatro de estos *codones* o *tripletes*, suficientes para

especificar los veinte aminoácidos. Grupos de bases más largos podrían hacer el trabajo, pero con tres había suficiente; era el formato más económico para el código genético.

Pero ¿qué grupos de tres bases codificaban qué aminoácidos? Aunque Gamow había reunido a algunos de los mayores genios del siglo xx, al final los distinguidos miembros del Club de la Corbata de ARN llegaron a un callejón sin salida. A finales de la década de 1950, Gamow había renunciado a descifrar el código, convencido de que el problema «no tenía solución» que pudiera encontrarse «sobre la base de la teoría pura».¹⁴

Lo que Gamow realmente necesitaba era el equivalente biológico de la piedra de Rosetta. En lugar de un texto escrito tanto en jeroglífico egipcio como en griego, debería mostrar la *secuencia de aminoácidos* de una proteína, así como la *secuencia de ARN* que la codifica. Pero no existía tal piedra; los científicos tendrían que inventar la suya propia. Y, para ello, habría que determinar si ese hipotético «mensajero» existía en realidad, es decir, si el ARN era el eslabón perdido entre nuestros genes y los componentes proteínicos de la vida.

¿RECIBISTE MI MENSAJE?

En la década de 1950, muchos científicos, entre ellos Jim Watson y Francis Crick, creían firmemente que el ARN podría transportar la información del ADN, desde el núcleo de la célula al citoplasma, donde se sintetizan las proteínas. Pero cuando los científicos trataron de encontrar pruebas empíricas de que el ARN es el mensajero de la vida, sus resultados iniciales fueron desalentadores. La primera decepción llegó cuando descubrieron que la mayor parte del ARN del citoplasma celular tenía la misma proporción de bases A, G, C y U, con independencia de las proteínas que se sintetizaran. Esto no tenía sentido: sería como descubrir que la Novena Sinfonía de Beethoven tiene exactamente la misma proporción de cada nota musical que «Bad Romance» de Lady Gaga.

Cabría esperar que estas piezas musicales tan diferentes tuvieran distribuciones distintas de fa sostenido y si bemol, etc. Del mismo modo, cabría esperar que diferentes proteínas, con sus diferentes composiciones de aminoácidos, fueran codificadas por diferentes proporciones de A, G, C y U en los ARN mensajeros, o ARNm, que las especifican.

La segunda decepción fue que la mayor parte del ARN del citoplasma es muy estable: desde que se crea, tiene una vida muy larga. Pero los científicos habían observado casos en los que los tipos de proteínas que se fabricaban en una célula cambiaban rápidamente, en pocos minutos, pasando de producirse un conjunto de proteínas a otro muy distinto. Por ejemplo, si se cambiaba la fuente de alimento de las bacterias, estas dejaban de fabricar las enzimas que utilizaban para digerir el alimento anterior y empezaban a fabricar inmediatamente enzimas adecuadas para el nuevo alimento. Del mismo modo, si una bacteria se infecta con un virus (las bacterias están plagadas de sus propios virus, llamados *bacteriófagos* o *fagos*), pasa de producir proteínas bacterianas a producir proteínas de fagos. Por consiguiente, un ARNm verdadero tendría que ser inestable para permitir cambios rápidos en la producción de proteínas. La naturaleza extremadamente estable del ARN celular parecía descalificarlo para ser el mensajero que buscaban los científicos.

Entre los científicos que seguían creyendo que debía existir un ARNm oculto estaban François Jacob y Jacques Monod, del Instituto Pasteur de París. Tras sus importantes descubrimientos sobre la activación y desactivación de los genes en las bacterias, pasaron a centrarse en el ARNm. En 1960, durante una visita a Cambridge (Inglaterra),¹⁵ Jacob se reunió con sus amigos Francis Crick y Sydney Brenner, entre otros, en las habitaciones de este último en el King's College. Jacob describió sus últimos experimentos sobre la forma en que se regulaban los genes en las bacterias, y la conversación pronto derivó hacia especulaciones sobre el papel del ARN mensajero como vínculo entre el gen y la proteína.

De repente, y casi de manera simultánea, Crick y Brenner dieron un respingo. Recordaron los recientes experimentos de dos científicos del Laboratorio Nacional Oak Ridge de Tennessee, Ken Volkin y Larry Astrachan. En estudios sobre *Escherichia coli* infectada por un fago llamado T2, Volkin y Astrachan habían observado que se formaban a gran velocidad nuevos ARN más pequeños que el ARN estable de la célula (ahora sabemos que es el ARN incrustado en *los ribosomas*, las fábricas de proteínas de la célula).¹⁶ Por una serie de complejas razones, Volkin y Astrachan habían interpretado que sus datos indicaban que el ARN se estaba metamorfoseando en ADN. Pero ¿y si, por el contrario, hubieran vislumbrado el escurridizo ARNm?

Era una posibilidad apasionante, pero para confirmarla sería necesaria una prueba más directa de su hipótesis del ARNm. Sydney Brenner y François Jacob fueron a visitar al genetista Matt Meselson en Caltech unas semanas más tarde, y se decidió que los tres realizarían los experimentos necesarios. Meselson y su colega Frank Stahl habían utilizado recientemente una nueva técnica, consistente en usar una ultracentrifugadora, cuyo rotor gira a gran velocidad, para comprobar la hipótesis de Watson y Crick de que las hebras de la doble hélice se separaban durante *la replicación* del ADN. Los resultados corroboraron la teoría, ya que cada doble hélice hija conservaba una hebra «antigua» heredada de su progenitora, y se emparejaba con una hebra recién sintetizada.

Brenner, Jacob y Meselson utilizarían esta misma ultracentrifugadora para intentar encontrar las agujas de ARNm en el pajar de ARN ribosómico. El plan consistía en infectar la bacteria *E. coli* con un fago que haría que las bacterias cambiaran el tipo de proteína que producían. Al mismo tiempo que añadían el fago, los científicos añadían una versión radiactiva del uracilo (marcada con carbono 14) que se incorporaría solo al ARNm del fago recién producido, no a los ARN bacterianos preexistentes. Si la hipótesis del ARNm era correcta, el nuevo ARNm del fago aparecería justo en ese

momento, lo que les permitiría encontrar el eslabón perdido entre el ADN y la proteína.

Y así fue. Brenner, Jacob y Meselson detectaron el ARNm radiactivo del fago en la ultracentrifugadora, ya que era claramente más pequeño que el ARN ribosómico. Como esperaban, este ARN era efímero y se unía a los ribosomas preexistentes de las bacterias infectadas para producir los nuevos tipos de proteínas que necesitaba el fago para hacer su trabajo sucio.¹⁷ Por fin habían localizado ARNm, a pesar de lo escurridizo que era.

Una forma de concebir este proceso biológico es imaginar un tocadiscos. El ribosoma es el tocadiscos, el ARNm es el disco de vinilo y la proteína es la música que se escucha al bajar la aguja. Puede que cambies el disco en función de tu estado de ánimo, pero el resto de la configuración sigue siendo la misma. Al igual que un tocadiscos puede reproducir cualquier disco de vinilo, los ribosomas pueden trabajar con cualquier ARNm que aparezca. El ARNm es lo que determina la proteína específica producida, ya sea una proteína de fago, una proteína de *E. coli* u otra; igual que el disco determina la música que estás escuchando.

La razón por la que el ARNm era tan difícil de detectar para los científicos es que solo alrededor del 5 por ciento del ARN de una bacteria *E. coli* es ARN mensajero. El otro 95 por ciento es, sobre todo, ARN ribosómico. Además, *E. coli* tiene 4.000 genes distintos, cada uno de los cuales produce un ARNm de tamaño y secuencia diferentes. Por lo tanto, cualquier ARNm supone mucho menos que el 5 por ciento total. Por último, la mayoría de los ARNm de *E. coli* tienen una vida útil de apenas unos minutos, lo que dificulta su detección. Por el contrario, los ARN ribosómicos no solo son omnipresentes en el citoplasma, sino que también son muy longevos, por lo que es fácil entender por qué los primeros científicos que estudiaron este tema tardaron tanto en ver más allá de todo ese ARN ribosómico.

UN INTERLUDIO LINGÜÍSTICO

Pensar en un tocadiscos puede ayudarnos a explicar cómo los ribosomas y el ARNm se unen para fabricar proteínas. Pero cuando se trata de entender cómo el ARN codifica las proteínas, tenemos que apagar la música y abrir un libro.

Las páginas de la mayoría de los libros están llenas de letras escritas, palabras y frases. En el libro de la vida, que está escrito en el lenguaje del ADN, no tenemos un alfabeto de veintiséis letras como en inglés o veintidós letras como en hebreo, sino cuatro letras: A, G, C y T. La disposición de esas cuatro letras (su secuencia) determinará el significado de cada palabra y cada frase.

Entonces, ¿cuántos libros necesitamos para recoger el sentido de la vida, para reflejar el significado de todo el ADN de un organismo determinado? La respuesta depende, por supuesto, del tamaño del genoma, que es la totalidad del ADN de un organismo. El genoma humano, agrupado en veintitrés cromosomas,* tiene unos 3.000 millones de bases, o letras.

Suponiendo un tamaño de letra típico que permita unos 3.000 caracteres por página, se necesitarían un millón de páginas para registrar el genoma humano. Teniendo en cuenta que un libro largo puede tener quinientas páginas, la secuencia del genoma requeriría unos 2.000 libros, demasiados para la biblioteca de tu casa, pero solo una fracción de los que tienes a mano en tu biblioteca pública local. Cada cromosoma humano, compuesto por una única molécula lineal de ADN, ocuparía unos noventa libros. Un genoma bacteriano, como el de *E. coli*, es mucho más pequeño; todo el genoma reside en un único cromosoma circular.

* Los cromosomas están formados por moléculas de ADN empaquetadas con proteínas. Los espermatozoides y los óvulos humanos tienen un juego de veintitrés cromosomas, mientras que las células somáticas tienen dos juegos: uno procedente de la madre y otro del padre.

Sus 4,5 millones de bases cabrían con facilidad en tres libros de la biblioteca.

Como hemos aprendido, para fabricar una proteína necesitamos ARN mensajero que lleve al ribosoma la información escrita en el ADN. Pero los ARN naturales no son copias de páginas enteras del gran libro del ADN. Más bien tienen puntos de inicio y final específicos que es poco probable que coincidan con saltos de página. Así, en lugar de una fotocopia de una página de un libro físico, podríamos pensar en el ARNm como una sección de texto copiada de un libro electrónico a otro documento electrónico con unos pocos clics. Movemos el cursor para resaltar una parte concreta de una página y luego la pegamos en una página nueva de nuestro procesador de textos. (Y con la práctica función de buscar y reemplazar, podemos cambiar de manera simultánea todas las T del ADN por las U del ARN, simulando el proceso químico que ocurre en la naturaleza.)

Este proceso de copiar y pegar (en el que el ADN se copia en ARNm) se produce constantemente en nuestro organismo cada vez que hay que sintetizar una nueva proteína. La región del ADN que se copia en ARNm varía en función de las necesidades de cada momento y lugar. En los niños en edad de crecimiento se copian partes del genoma diferentes de las que se copian en los adultos, y en el corazón las partes que se copian del genoma son diferentes de las que se copian en el cerebro, el hígado y la piel. En otras palabras, la copia está muy regulada.

Tras copiar y pegar, podemos examinar la secuencia de nuestro ARNm. Es un conjunto de A, G, C y U, como, por ejemplo, el siguiente:

GUAGGGCAUGCCUUCGAAAAUAUUUUGUUAGCGCCUCCUUGGAGUAGAA

Digamos que sabemos cómo descodificar los grupos de tres bases (los codones o tripletes) a lo largo de este ARNm. El siguiente es nuestro diccionario. En lugar de aminoáci-

dos, hemos elegido palabras en castellano como significado de cada codón:

AUG = El

CCU = gran

UCG y UCC = gato

AAA y AGA = come

AUA = una

UUU = gorda

UGU y UAU = rata

CGC = pero

CUC = dos

CUU = y

GGA = seis

GUA y GUU = para

GAA y GAG = ti

AAU = ahora

AUU = ver

UUG = zorro

UUA y UAA = corre

GCG y GCC = fuera

UGG = juerga

AGU y AGC = sol

Mi diccionario tiene un vocabulario muy limitado: solo veinte palabras, análogas a los veinte aminoácidos naturales. Algunas de estas palabras están codificadas por un solo codón: *diversión*, por ejemplo, está codificado por UGG. Otras palabras (*correr, fuera, sol*, etc.) están codificadas por más de un codón, igual que en el código genético. Cabe preguntarse por qué la naturaleza ha desarrollado un sistema que utiliza tantos codones (sesenta y cuatro en total) para codificar solo veinte palabras (aminoácidos). ¿No hay una forma más elegante de estructurar el código?

Esta aparente falta de elegancia es uno de los aspectos de la biología que confundió a físicos como Gamow. En física,

los acontecimientos son en gran medida predecibles. Si conoces las ecuaciones de Maxwell, puedes resolver tu problema de física clásica. Pero con la biología, la única regla es: lo que sea que funcione. Cuando un sistema, por envejecido que parezca, empieza a funcionar bien, queda bloqueado por la evolución y resulta muy difícil cambiarlo. El hecho de que un buen ingeniero pudiera idear un sistema más sencillo o más eficaz es irrelevante.

Ahora que sabemos cómo traducir el mensaje escrito en las cadenas de codones a palabras con sentido, necesitamos saber dónde empieza y dónde acaba el mensaje. ¿Empezamos por el extremo izquierdo, leyendo GUA como «para», y seguimos leyendo hasta el extremo derecho? Aunque eso pueda parecer razonable, la naturaleza tiene una solución diferente. Utiliza el triplete AUG para designar el inicio de una frase. Nuestra regla equivalente aquí es que las frases deben empezar por «El», que está codificado por AUG. Entonces, simplemente escaneamos desde el extremo izquierdo, buscamos el primer AUG y empezamos a leer desde ahí, con el diccionario en la mano.

GUAGGGC AUG CCU UCG AAA AUA UUU UGU UAG CGC CUC CUU GGA GUA GAA
El gran gato come una gorda rata

Todo va bien hasta que llegamos al codón UAG y no encontramos ninguna entrada para él en nuestro diccionario. Ah, sí, necesitamos otro codón triplete para marcar el final de la frase: un punto.* Por tanto, consideramos:

UAG = final de frase = punto

* Ten en cuenta que es muy posible que los nucleótidos que preceden al codón de inicio y los que siguen al codón de finalización no formen parte de otras «frases» (o genes), pero podrían tener funciones reguladoras.

Ya podemos decir que hemos descodificado el mensaje:

GUAGGGC AUG CCU UCG AAA AUA UUU UGU UAG CGC CUC CUU GGA GUA GAA

El gran gato come una gorda rata

Ahora que hemos especificado cuáles son los codones de inicio y fin (AUG y UAG) nos damos cuenta de que no es necesario separar los codones con espacios. Si nos dan la secuencia de ARNm como una cadena ininterrumpida de letras, y sabemos que la leemos de tres en tres, obtenemos la misma información:

GUAGGGCAUGCCUUCGAAAAUAUUUUGUUAGCGCCUCCUUGGAGUAGAA

El gran gato come una gorda rata

Durante la década de 1950, la hipótesis del triplete de Gamow no pasó de ser una suposición. En 1961, un ingenioso conjunto de experimentos llevados a cabo por Francis Crick y sus colegas, en los que inducían una mutación en un gen de fago mediante colorantes de acridina, demostró finalmente la existencia de los tripletes.¹⁸ Los colorantes de acridina son moléculas planas muy parecidas a las bases del ADN, por lo que se deslizan en el interior de la doble hélice y pueden confundirse con bases al replicar el ADN, provocando así la inserción de una base errónea en el ADN hijo. Imagínate una fila de libros colocados uno al lado del otro en una estantería. Si metes otro libro, todos los que le siguen deben moverse una posición para acomodar la inserción. Si metes dos, los demás se desplazan el espacio equivalente a la anchura de dos libros.

Nuestra analogía lingüística nos muestra cómo esas inserciones afectarían al mensaje y cómo las inserciones múltiples apoyan la idea de un código creado a base de tripletes. Partiendo de nuestro mensaje hipotético, he marcado los codones de inicio y fin con cursiva:

GUAGGGC AUG CCU UCG AAA AUA UUU UGU UAG CGC CUC CUU GGA GUA GAA
El gran gato come una gorda rata

Esto es lo que ocurre cuando el colorante de acridina provoca la inserción aleatoria de una U, subrayada a continuación:

GUAGGGC AUG CCU UCG UAA AAU AUU UUG UUA GCG CCU CCU UGG AGU AGA
El gran gato corre ahora ver zorro corre fuera gran gran juega sol come

Nuestra frase empieza correctamente, pero la U insertada desplaza el marco de lectura. A esto lo llamamos «mutación por desplazamiento del marco de lectura», proceso que ocurre con bastante frecuencia en el genoma humano y puede causar resultados desagradables como la fibrosis quística, la enfermedad de Crohn y la enfermedad de Tay-Sachs. Un desplazamiento del marco de lectura resulta muy nocivo porque la traducción a partir de ese punto carece de sentido. Además, el UAG que antes designaba el final de la frase ahora está fuera de marco, por lo que la frase sin sentido sigue y sigue. Si se tratara de un ARNm que codificara una proteína, esta sería inútil.

Veamos ahora qué ocurre cuando añadimos dos acridinas a nuestra secuencia, provocando la inserción de dos bases. Es tan destructivo como la inserción de una sola base:

GUAGGGC AUG CCU UCG UUA AAA UAU UUU GUU AGC GCC UCC UUG GAG UAG
El gran gato corre come rata gorda para sol fuera gato zorro tu

En este caso, el desplazamiento del marco de lectura produce un codón de parada UAG en el extremo derecho, pero la frase resultante sigue sin tener sentido. Si se tratara de ARNm, la proteína mutada resultante sería también inútil.

Por último, examinemos las consecuencias de una inserción de tres bases:

GUAGGGC AUG CCU UCG UUU AAA AUA UUU UGU UAG CGC CUC CUU GGA GUA

El gran gato gorda come una gorda rata

Si el código consta de tripletes, como se muestra en la imagen, en este caso se inserta una palabra aberrante («gorda») en la frase, pero todas las demás palabras son correctas. La frase es más o menos comprensible.

Y eso es exactamente lo que Crick y sus colegas dijeron haber descubierto, confirmando de esa forma la hipótesis original de Gamow. Una inserción era mortal. Dos inserciones también eran mortales. ¡Pero con tres inserciones la cosa seguía funcionando más o menos! Por lo tanto, los codones deben estar compuestos por grupos de tres bases.

DESCIFRAR EL CÓDIGO

Gamow y su Club de la Corbata de ARN habían descubierto que el código genético estaba escrito en grupos de tres letras y que el ARN era, probablemente, el puente entre el ADN y las proteínas. Brenner, Jacob y sus colegas habían demostrado esto último, identificando de forma concluyente el ARN mensajero, que transmite la información de nuestro ADN a la maquinaria de construcción de proteínas de la célula, los ribosomas. Pero seguíamos sin poder leer esa información. En los albores de la década de 1960, el código genético seguía siendo indescifrable, hasta que apareció un joven científico llamado Marshall Nirenberg.

En 1959, Nirenberg consiguió un puesto de investigador independiente en el Instituto Nacional de Artritis y Enfermedades Metabólicas de Bethesda (Maryland). Antes de dedicarse a la bioquímica y obtener su doctorado, ya había cursado un máster en la Universidad de Florida estudiando tricópteros, un grupo de insectos cuyas larvas son acuáticas, pero que de adultos son parecidos a polillas. (Quizá Nirenberg fuera reacio a reconocer que había co-

menzado su carrera científica trabajando con insectos, ya que omitió este hecho en su discurso de aceptación del Nobel.) En su intento por descifrar el código genético, Nirenberg dio un paso ambicioso. Tenía poca relación con los líderes de este campo de investigación y nunca había recibido una corbata de lana de ARN. Sin embargo, apartándose del enfoque de Gamow, Nirenberg decidió abordar el problema de la codificación mediante la bioquímica. Para ello sería necesario recrear la síntesis proteica (el proceso por el que el ARNm se convierte en proteína) en tubos de ensayo.

Nirenberg se basó en el trabajo pionero de sus colegas bioquímicos Elizabeth Keller y Paul Zamecnik, del Hospital General de Massachusetts (Boston), que habían desarrollado un método para sintetizar proteínas fuera de las células utilizando ingredientes obtenidos de hígados de rata.¹⁹ Posteriormente, Zamecnik y otros científicos demostraron que se podían utilizar extractos de *E. coli* para hacer lo mismo. Y este fue el sistema elegido por Nirenberg. En pocas palabras, el método consistía en romper la bacteria *E. coli* para obtener una preparación bruta de ribosomas y, a continuación, añadir diferentes secuencias de ARN a modo de moldes para la síntesis de proteínas.*

Tras múltiples pistas falsas, Nirenberg y su becario posdoctoral, Heinrich Matthaei, se toparon con el hecho de que una simple molécula de ARN compuesta íntegramente por bases U, conocida como poli(U), dirigía la síntesis de una única proteína compuesta íntegramente por el aminoácido fenilalanina. El atractivo del poli(U) radicaba en que, con independencia de dónde se iniciara la lectura, solo podía haber un tipo de codón, UUU, y solo se tra-

* Las condiciones de sus experimentos en tubos de ensayo eran tales que los ribosomas no necesitaban una señal de inicio en el ARNm (el AUG en nuestra analogía anterior), sino que iniciaban el proceso de forma aleatoria en algún punto de la cadena de ARNm.

ducía en fenilalanina. Así se descifró la primera pieza del rompecabezas del código genético: UUU codifica fenilalanina.²⁰

Ahora el camino a seguir estaba claro: introducir diferentes secuencias de ARN en los ribosomas y ver qué proteínas salían. De este modo, se descubrió que la poli(C) y la poli(A) codifican la poliprolina y la polilisina, respectivamente. Así que la tabla de asignaciones de codones creció: tres menos, faltan sesenta y uno. Naturalmente, estos tres primeros eran fáciles, porque cada letra del codón era idéntica. Los codones más complicados requerían un enfoque diferente.

Este es el momento de hablar de Gobind Khorana.* Nacido en el seno de una humilde familia hindú de Raipur, entonces en la India, pero ahora parte de Pakistán, estudió en Inglaterra y Suiza. Saltó a la fama por su trabajo de descifrado del código cuando era profesor de la Universidad de Wisconsin-Madison. Junto a su equipo ideó métodos para sintetizar químicamente el ADN, uniendo los *nucleótidos* de uno en uno. A continuación, utilizaban una enzima recién descubierta para obtener ARN copiado del ADN.²¹ Lo mejor de este método era que les permitía determinar con precisión la secuencia de bases del ADN y, por tanto, del ARN copiado a partir de él. En el laboratorio, estos ARN se empleaban como mensajeros para especificar las secuencias de aminoácidos correspondientes.²² Khorana sintetizó todas las permutaciones posibles de tres bases, las introdujo en

* Conocí a Khorana en el MIT cuando yo era becario posdoctoral de biología en el edificio 16 y él era un químico galardonado con el Premio Nobel que trabajaba en el edificio 18. Debido a la severidad de los inviernos bostonianos, muchos de los edificios del MIT están interconectados, de modo que yo podía caminar hasta el departamento de química sin abrigarme. Khorana me pareció modesto, contagiosamente entusiasmado con su investigación y nada molesto por el hecho de que un humilde investigador posdoctoral criado en Iowa se presentara sin avisar.

su sistema de síntesis de proteínas y observó qué cadena de aminoácidos obtenía. De esta forma, ayudó a completar la tabla del código genético, con lo que ganó el Premio Nobel de Fisiología o Medicina de 1968 junto a Marshall Nirenberg.²³

Tras leer sobre estos descubrimientos en las noticias, Gamow se sintió satisfecho de que alguien hubiera resuelto por fin el problema de la codificación, aunque no pudo evitar criticar el enfoque experimental: «La solución parece bastante menos elegante que la simple correlación teórica que yo había visualizado en un principio —dijo—. Pero tiene la indiscutible ventaja de ser correcta.»²⁴

UN PEQUEÑO ARN CON UN TRABAJO IMPORTANTE

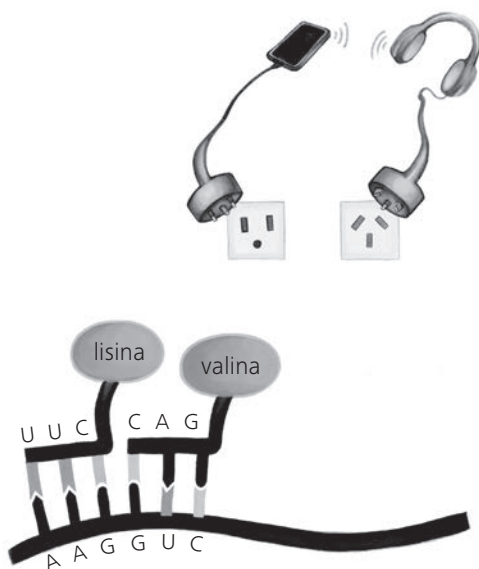
Dilucidar el código que utiliza el ARNm para especificar las proteínas fue un gran logro, pero no era la única gran pregunta que había que responder. Disponer de un código era fundamental, pero ¿qué venía después? Tiene que haber un proceso por el que el código se «lea». ¿Cómo conseguía el ARNm colocar los aminoácidos en la posición correcta para encadenarlos y construir una proteína?

Francis Crick dio una respuesta a esta pregunta en 1955, y su propuesta destaca por ser un singular triunfo teórico. Propuso que debía existir una clase de moléculas, que nunca se habían visto hasta entonces, que serían las piezas que faltaban en la gran teoría del código genético.²⁵ Imaginó un grupo de «moléculas adaptadoras», cada una con dos extremos. Un extremo estaría unido a uno de los veinte aminoácidos. El otro extremo reconocería y se uniría al codón de ARNm correspondiente.²⁶

Una forma de entender qué y cómo son estos adaptadores de aminoácidos a ARNm es imaginar que son como los adaptadores que conectan los aparatos electrónicos a las

tomas de corriente. Para cargar un par de auriculares, se necesita el adaptador adecuado para el enchufe adecuado, y ese enchufe puede ser diferente según se esté en Estados Unidos, el Reino Unido o Alemania. Lo mismo que ocurre con los auriculares, ocurre con los aminoácidos. Tienes una toma de corriente (un codón) y un enchufe de tres clavijas (llamado *anticodón*) que encaja en su interior. Como el codón y el anticodón son complementarios, se conectan entre sí por emparejamiento de bases. Y, como si de una regleta de enchufes se tratara, las moléculas adaptadoras se alinean para conectar los codones del ARNm con sus aminoácidos correspondientes formando cadenas que constituyen las proteínas. O al menos eso es lo que propuso Crick.

Pero la teoría solo nos hace avanzar hasta un punto. Alguien tenía que ir al laboratorio y ver si era posible encontrar ese adaptador. Ese alguien fue Paul Zamecnik, que había desarrollado el método de síntesis de proteínas que permitió a Nirenberg descubrir el primer codón. Empezó con aminoácidos marcados con un isótopo radiactivo del carbono para poder seguir su rastro. Sería similar a lo que hacen los bancos, que esconden un pequeño bote con pintura que explota en el interior de una bolsa de dinero para poder rastrearlo en caso de robo. Cuando Zamecnik añadió estos aminoácidos radiactivos a su sistema de hígado de rata, ocurrió algo sorprendente: pequeños ARN se volvieron radiactivos, lo que indicaba que un aminoácido se había unido a un ARN. Nunca antes se había observado una unión de este tipo, pero era exactamente lo que ocurriría si existiera un ARN cuyo trabajo consistiera en adaptar los codones del ARN a los aminoácidos.²⁷ Más adelante, estos pequeños ARN recibieron el nombre de *ARN transferentes* o *de transferencia* (ARNt), porque transfieren el aminoácido correcto a la cadena proteica en crecimiento dentro del ribosoma. Como veremos una y otra vez, los ARN de transferencia pueden ser pequeños, pero son muy poderosos.



Francis Crick propuso que cada codón de una cadena de ARNm (abajo) debía ser reconocido por una «molécula adaptadora» portadora del aminoácido correspondiente, como la lisina o la valina de este ejemplo. Esto es análogo a los cables de carga (de los teléfonos móviles o de los auriculares) que sirven de adaptadores, conectando varias tomas eléctricas de tres clavijas con los dispositivos.

A mediados de la década de 1960, a algunos científicos les pareció que la historia del ARN había terminado. Cada ARNm dirigía la síntesis de una proteína diferente basándose en un código que se había podido descifrar. En la síntesis de proteínas intervenían otros dos tipos de ARN estables: los ARN de transferencia, que conectaban cada codón de ARNm con el aminoácido correspondiente, y los ARN ribosómicos, que (como veremos con más detalle más adelante) construían la proteína.

Incluso hoy en día, mucha gente piensa que el ARN solo sirve para lo que acabo de explicar. Lo ven como un curran-te a las órdenes del ADN, que engrasa las ruedas de la maqui-

naria celular que transforma el código de la vida en la materia de la vida. Por supuesto, ese proceso biológico es esencial para la supervivencia de todos los seres vivos de la Tierra. Y si realmente fuera el final de la historia del ARN, seguiría siendo un logro colosal para nuestra minúscula molécula. Pero resulta que la mensajería es solo el primero de los muchos superpoderes del ARN.