

## MÉTODOS DE SECUENCIACIÓN DE ÁCIDOS NUCLEICOS: PRIMERA GENERACIÓN

por JOSÉ MIGUEL VALDERRAMA MARTÍN\*, FRANCISCO ORTIGOSA\* Y RAFAEL A. CAÑAS<sup>+</sup>

\*ESTUDIANTE DEL PROGRAMA DE DOCTORADO EN BIOLOGÍA CELULAR Y MOLECULAR, 4<sup>A</sup> PLANTA EDIFICIO I+D, FACULTAD DE CIENCIAS, UNIVERSIDAD DE MÁLAGA, 29071.

<sup>+</sup>PROFESOR TITULAR DEL DEPARTAMENTO DE BIOLOGÍA MOLECULAR Y BIOQUÍMICA, 4<sup>A</sup> PLANTA EDIFICIO I+D, FACULTAD DE CIENCIAS, UNIVERSIDAD DE MÁLAGA, 29071.

JMVALDERRAMA@UMA.ES, FORTIGOSA@UMA.ES, RCANAS@UMA.ES

La secuenciación de genomas, desde el descubrimiento del ADN como molécula portadora de la información genética heredable, ha ido ganando importancia debido a la gran cantidad de información que aporta y a la utilidad de esta en diferentes campos de la ciencia. Con este artículo se inicia una serie en la que se revisan los métodos de secuenciación de ácidos nucleicos. Los primeros métodos de secuenciación permitieron el avance fundamental de la Biología Molecular, abrieron las puertas para el desarrollo de nuevas tecnologías de secuenciación e iniciaron el desarrollo de nuevas disciplinas científicas que hoy tienen gran repercusión como son la Genómica y la Transcriptómica.

*Genome sequencing, since the discovery of DNA as a molecule that carries inheritable genetic information, has been gaining importance due to the large amount of information that provides and its usefulness in different fields of science. This article begins a series that reviews nucleic acid sequencing methods. The first sequencing methods allowed the fundamental advancement of Molecular Biology, opened the doors for the development of new sequencing technologies and started the development of new scientific disciplines that today have great repercussions such as Genomics and Transcriptomics.*

*Palabras clave:* secuenciación, primera generación, Maxam y Gilbert, Sanger.

Enviado: 03/09/2019

*Keywords:* sequencing, first generation, Maxam and Gilbert, Sanger.

Aceptado: 27/04/2020

Uno de los grandes intereses de los científicos y científicas siempre ha sido entender dónde y cómo se codifica toda la información que determina el desarrollo de los seres vivos y muchas han sido las suposiciones acerca de esto desde edades tempranas de la Ciencia. Pero no fue hasta 1952 que los estudios realizados por Alfred Hershey y Martha Chase demostraron cuál era realmente la macromolécula sobre la que recaía el papel de transmitir la información genética en los organismos<sup>[1]</sup>. La macromolécula en cuestión se trataba de los ácidos nucleicos (fundamentalmente el ácido desoxirribonucleico (ADN) y el ácido ribonucleico (ARN) en algunos tipos de virus), no las proteínas como habían propuesto otros investigadores. Desde entonces, innumerables han sido los estudios que se han centrado en esta macromolécula y que han llevado hasta la secuenciación de los genomas (conjunto de información genética contenido en un ser vivo, incluyendo los virus). Sólo un año después, en 1953, y siguiendo las reglas de Erwin Chargaff sobre las proporciones de las bases nitrogenadas en el ADN<sup>[2,3]</sup>, se propuso la estructura en doble hélice del ADN por parte de James Watson y Francis Crick<sup>[4]</sup>, usando los datos cristalográficos obtenidos por Rosalind Franklin y Maurice Wilkins. Además, en 1958

Matthew Meselson y Franklin Stahl demuestran que la replicación del ADN es semiconservativa<sup>[5]</sup>, lo que pone de relieve la importancia de la estructura en doble hélice del ADN. Estos y otros antecedentes, tanto científicos como técnicos, establecieron una base firme que condujo al desarrollo de los primeros métodos de secuenciación de ácidos nucleicos. Estos aparecieron en los años 60 del siglo XX aunque eran muy limitados y fueron ampliamente superados por los diseñados al final de los años 70 del siglo XX. Estos últimos, han tenido mucho más recorrido y han sido empleados de forma rutinaria durante casi medio siglo.

De entre todos los que aparecieron fueron dos los de mayor éxito: el método de modificación y escisión química ideado por Allan Maxam y Walter Gilbert<sup>[6]</sup> y el método de secuenciación por síntesis diseñado por Frederick Sanger<sup>[7]</sup>. Debido al extraordinario desarrollo de las técnicas de secuenciación de ácidos nucleicos, a estos métodos se los conoce actualmente como de primera generación. De entre ambos, ha sido el método de Sanger el que ha tenido una mayor aceptación y desarrollo, siendo el que ha permitido la fundación de la Genómica gracias a la primera secuenciación completa de un genoma de ADN, el

del bacteriófago  $\phi$ X174 (5.386 pb)<sup>[8]</sup> (curiosamente el primer genoma secuenciado fue el del bacteriófago MS2 que es un virus ARN y se realizó directamente sobre su molécula ARN con métodos puestos a punto previamente por Sanger en los años 60<sup>[9,10]</sup>). Además, la automatización de esta técnica fue definitiva para la consecución del Proyecto Genoma Humano. Sin embargo, el desarrollo tecnológico y bioinformático ha propiciado el surgimiento y la expansión de las técnicas de secuenciación masiva de ácidos nucleicos conocidas como secuenciación de próxima generación (NGS en inglés), provocando que las técnicas de primera generación hayan dejado de ser usadas para la secuenciación de genomas completos. No obstante, actualmente la secuenciación de Sanger sigue siendo una técnica básica utilizada o requerida por todos los laboratorios donde se realiza clonación molecular.

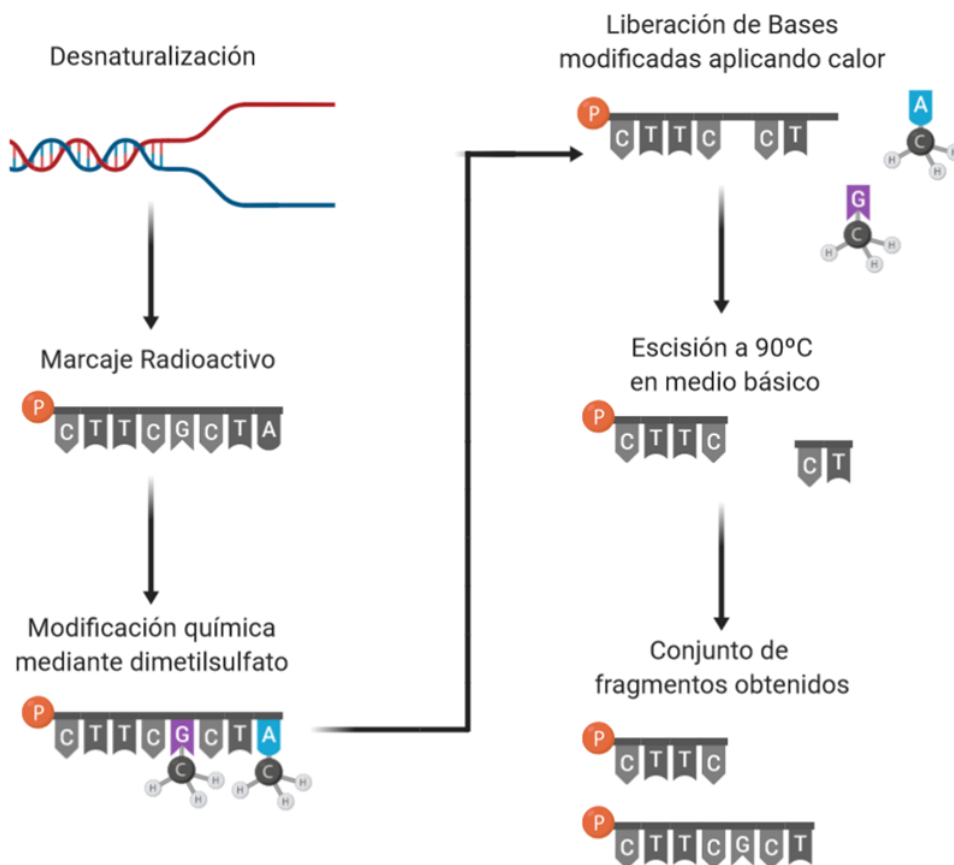
## Método de escisión química (Maxam y Gilbert)

El método de secuenciación químico diseñado por Allan Maxam y Walter Gilbert fue publicado en 1977<sup>[6]</sup> y mejorado en 1980<sup>[11]</sup>. Esta técnica se basa en la escisión específica de bases nitrogenadas en la cadena de ADN y la posterior ruptura de la cadena de ADN en el sitio de la escisión de la base nitrogenada debido a su mayor fragilidad (Figura 1). En la técnica original se realizan cuatro reacciones de escisión diferentes. En la primera reacción se metilan las purinas, es decir, guaninas (posición N7) y adeninas (posición N3), usando dimetilsulfato. Estas modificaciones hacen que el enlace glucosídico se pueda romper fácilmente mediante calor. Finalmente, la escisión de estas bases provoca que el enlace entre la desoxirribosa y el fosfato sea más inestable y se pueda romper en medio básico a 90° C. La siguiente reacción permite la escisión de la adenina preferentemente sobre la guanina. Puesto que su enlace glucosídico es más débil que el de la guanina cuando ambas están metiladas, se pueden usar ácidos diluidos para la escisión preferencial de la adenina. Para la ruptura de la cadena del ADN se usa el mismo procedimiento que el descrito anteriormente. En el caso de las bases pirimidínicas las reacciones químicas son mucho más complejas usándose reacciones secuenciales con hidracina y piperidina. Primero la hidracina elimina las citosinas y timinas llegando a abrir el anillo de la desoxirribosa. Posteriormente la piperidina realiza la ruptura de la cadena de ADN rompiendo el enlace con el fosfato. La última de las reacciones permite la escisión de la citosina y es la misma descrita para

ambas pirimidinas salvo que se añade 2 M de NaCl lo que inhibe la reacción de escisión de las timinas. En posteriores versiones del método se emplearon diferentes reacciones químicas. En la de 1980, cambió el método de modificación y escisión de ambas purinas, en este caso se escinden mediante el uso de un ácido. Además, se emplea la modificación por dimetilsulfato junto a piperidina para la escisión de las guaninas, y no preferentemente de las adeninas como se hacía en la versión original donde esta reacción en ocasiones podía originar ambigüedad en la interpretación.

Estas reacciones químicas no son totalmente eficaces por lo que se genera una serie de fragmentos de distinto tamaño por cada reacción de modificación de bases. Para la identificación de los fragmentos y la determinación de la secuencia de ADN se requiere el marcaje radioactivo con <sup>32</sup>P del extremo 5' del ADN a secuenciar mediante el uso de una quinasa. Los productos las reacciones de escisión son resueltos en un gel de poliacrilamida, el cual es revelado mediante una autorradiografía. Los productos de cada reacción específica de modificación se cargan en calles separadas del gel lo que permite identificar la secuencia del de la molécula de ADN en cuestión por los patrones de bandas que se obtienen en las diferentes calles determinando cada una de ellas el sitio de escisión para un nucleótido o dos en concreto (Figura 2).

El rango de separación (tamaño) de fragmentos de los geles de poliacrilamida determina el tamaño del producto de la secuenciación llegando a unas 250 bases<sup>[11]</sup>. Una de las ventajas de este método reside en que no se necesita realizar la clonación o amplificación del fragmento a secuenciar como ocurre en el de Sanger. Esta fue una de las razones que hicieron que a pesar de publicarse dos años después que el método de Sanger en un principio su difusión fuese mucho mayor que este último. La clonación molecular no se encontraba muy extendida en aquellos momentos (la primera clonación del ADN copia completo de un gen en la Universidad de Málaga no fue hasta el año 1993<sup>[12]</sup>), por ello era una gran ventaja usar un método que no necesitaba de ella, bastaba con la purificación del ADN y su fragmentación con enzimas de restricción. Sin embargo, este método requiere de ADN monocatenario, el uso de compuestos químicos peligrosos y una mayor manipulación que el método de Sanger. Finalmente, estos problemas junto con la diseminación de la clonación molecular por todo el mundo, la mejora de la síntesis de oligonucleótidos, y un mayor desarrollo instrumental y metodológico de la técnica de Sanger hicieron que su uso terminase abandonándose.



**Figura 1.** Representación esquemática de la modificación química y escisión de las purinas llevadas a cabo en el método de Maxam y Gilbert de 1977<sup>[6]</sup>. Las reacciones químicas no son totalmente eficaces y se dan lugar diferentes productos tras las escisiones. Imagen diseñada en Biorender.

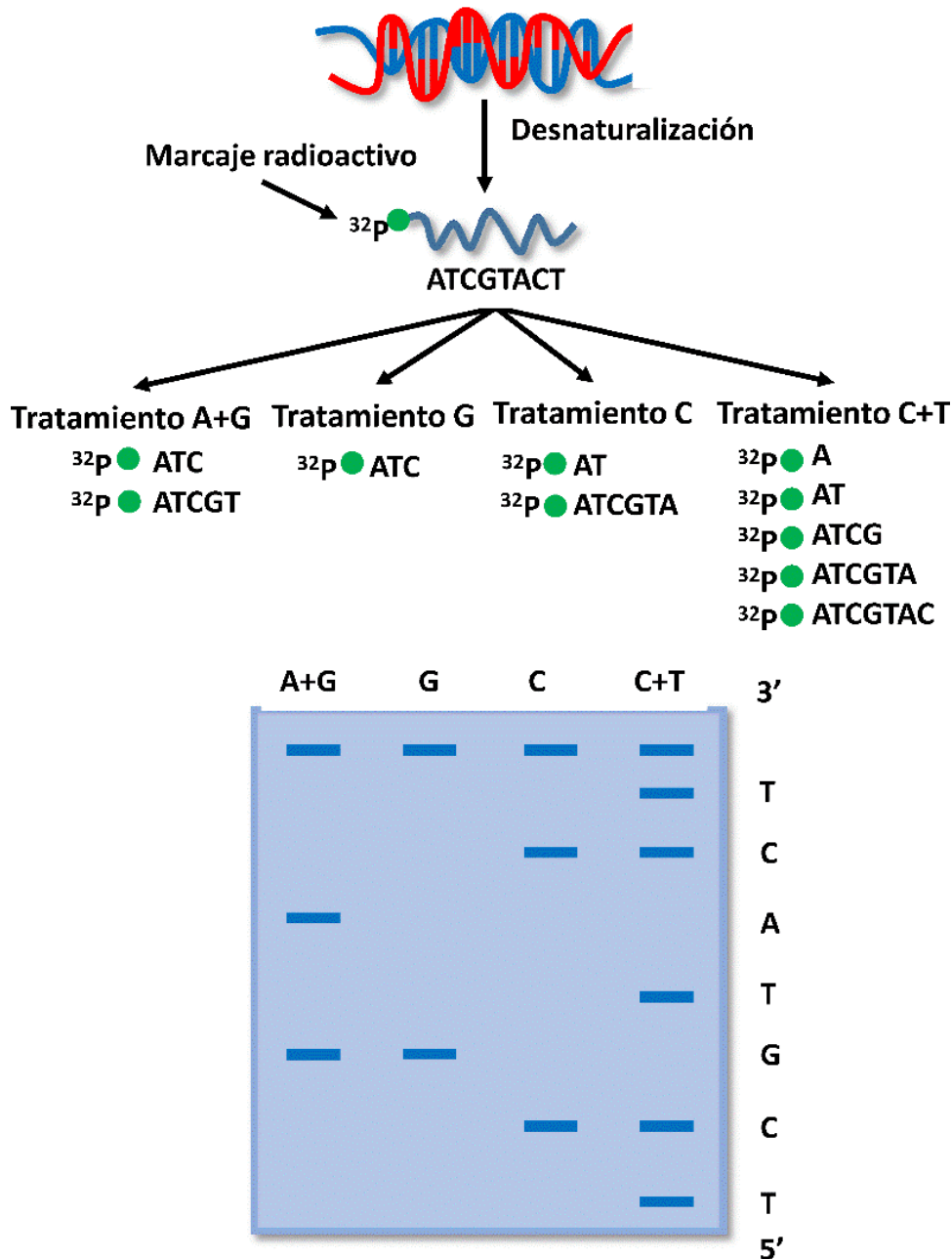
## Método de terminación de la cadena (Sanger)

La primera versión del método de Sanger se publicó inicialmente en 1975<sup>7</sup> y fue completado en 1977 con el uso de didesoxinucleótidos terminadores<sup>[13]</sup>. Se trata de un método de secuenciación por síntesis, es decir, basado en la replicación de una molécula de ADN por la ADN polimerasa. A partir de un oligonucleótido cebador la ADN polimerasa origina un conjunto de oligonucleótidos de diferentes tamaños y complementarios a la hebra molde (Figura 3A). La actividad enzimática de todas las moléculas de la ADN polimerasa no es sincrónica por lo que se pueden generar productos (oligonucleótidos) de tamaños aleatorios. Además, en esta reacción uno de los cuatro desoxinucleótidos trifosfato (dNTP) se encuentra marcado radioactivamente (<sup>32</sup>P). Posteriormente, tras purificar los productos de la primera reacción se debían realizar dos tipos de reacciones con la ADN polimerasa llamadas *minus* y *plus*, y basadas en el uso diferencial de los desoxinucleótidos y que por combinatoria dan lugar a ocho reacciones diferentes. En las reacciones *minus* se incorporan tres de los desoxinucleótidos (ej. dATP, dTTP y dGTP) pero no el cuarto (ej. dCTP) (Figura 3B). En las reacciones *plus*

sólo se incorpora un desoxinucleótido (ej. dCTP) así los productos de esta reacción deben tener terminar en el mismo nucleótido (en este caso C) y presentar un nucleótido más que el producto equivalente de la reacción *minus* (Figura 3C). Los productos de las reacciones marcados radioactivamente (<sup>32</sup>P) posteriormente deben resolverse en un gel desnaturante de poliacrilamida y revelarse en una autorradiografía. El problema de esto es que requiere la realización 8 reacciones diferentes (4 reacciones *plus* por cada uno de los diferentes dNTPs y sus reacciones *minus* correspondientes) y los resultados eran difíciles de interpretar, además que la longitud de secuenciación no era muy grande debido al sistema empleado. En 1977 se resolvieron los principales problemas al usarse nucleótidos terminadores<sup>[13]</sup>. La ADN polimerasa usa una hebra de ADN como molde y va incorporando nucleótidos complementarios hasta introducir un terminador de la secuenciación. Estos terminadores son didesoxinucleótidos trifosfato (ddNTPs) que carecen del grupo hidroxilo (-OH) en el carbono 3' de la ribosa, de manera que no se puede introducir el siguiente desoxinucleótido (dNTPs). De este modo, en reacciones independientes sobre el fragmento que queremos secuenciar, se añaden los cuatro dNTPs (ej. dATP, dCTP, dGTP y dTTP) así como una fracción

de un ddNTPs (ej. ddATP). Esto hace que en vez de ocho reacciones se realicen cuatro y que las reacciones puedan desarrollarse a partir de un único cebador que puede ser de secuencia conocida, por ello la longitud de secuenciación puede ser mayor limitándola únicamente el método de separación de los fragmentos de ADN obtenidos. Estadísticamente hablando, se deberían producir fragmentos de todos los tamaños posibles por la introducción de esos ddNTPs en dife-

rentes momentos. Gracias a esto, y a la separación del producto de cada reacción en una electroforesis en un gel desnaturante de poliacrilamida, se puede obtener la secuencia de ADN de ese fragmento concreto atendiendo a la posición y orden de cada una de las bandas respecto a las cuatro diferentes calles que corresponden al mismo número de reacciones de secuenciación por cada ddNTP (Figura 4A).



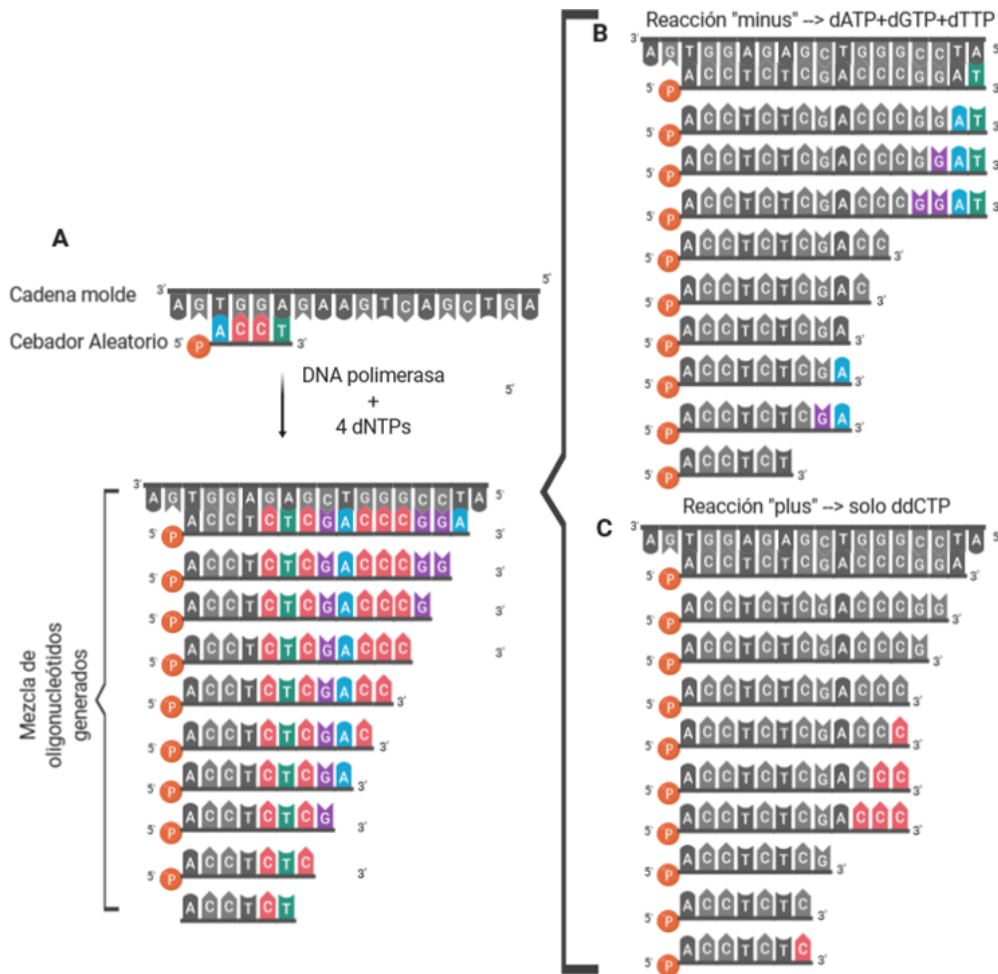
**Figura 2.** Ilustración esquemática del método de secuenciación por escisión química de Maxam y Gilbert en su versión de 1980<sup>[11]</sup>. Las diferentes reacciones, llevadas a cabo de manera independiente, se separan en un gel de poliacrilamida y en función de las bandas se puede resolver la secuencia del fragmento analizado. El revelado de las bandas es posible gracias al marcaje con isótopos radioactivos.

En un principio los ddNTPs se encontraban marcados radioactivamente, con  $^{32}\text{P}$  o  $^{35}\text{S}$  en la posición  $\alpha$  correspondiente a los grupos fosfato de uno de los cuatro desoxinucleótidos trifosfato. Sin embargo, esto

presenta dos importantes desventajas. Por un lado, la necesidad de realizar cuatro reacciones separadas, para cada nucleótido o combinación de nucleótidos, debido a que no se puede discernir entre las señales



radioactivas de las diferentes reacciones, y en segundo lugar el riesgo contra la salud del investigador o técnico que realizaba este proceso.



**Figura 3.** Representación esquemática de la primera versión del método de Sanger donde se realizaba una primera reacción con cebadores, ADN polimerasa y los 4 dNTPs para generar una mezcla de oligonucleótidos (A) que posteriormente sería usada en las reacciones *minus*, que empleaba todos los dNTPs menos uno de ellos (B), y una reacción *plus* donde se usaba el dNTP restante (C). La combinatoria da lugar a un total de 8 reacciones independientes que eran resueltas en un gel desnaturante de poliacrilamida para leer e interpretar los resultados.

Durante el proceso de secuenciación del genoma humano, estos inconvenientes fueron solventados mediante el uso de nucleótidos terminadores unidos a fluorocromos. Actualmente, se usan fluorocromos que al ser excitados a 488 nm de longitud de onda mediante un láser de Argon emiten luz a otra longitud de onda determinada. Para cada tipo de didesoxinucleótido se asocia un fluorocromo que emite luz a una longitud de onda diferente (por ejemplo, dicloro[R6G] para la adenina; dicloro[ROX] para la citosina; dicloro[R110] para la guanina; dicloro[TAMRA] para la timina)<sup>[14]</sup>. Esto permite discernir entonces entre cada base nitrogenada dentro de una única reacción de secuenciación. Los fluorocromos son derivados de la rodamina, un compuesto orgánico heterocíclico y fluorescente. El problema que tienen estos fluorocromos es que a una misma longitud de onda de excitación (488 nm) presentan diferentes intensidades en sus respuestas de emisión. Por ello se han desarrollado

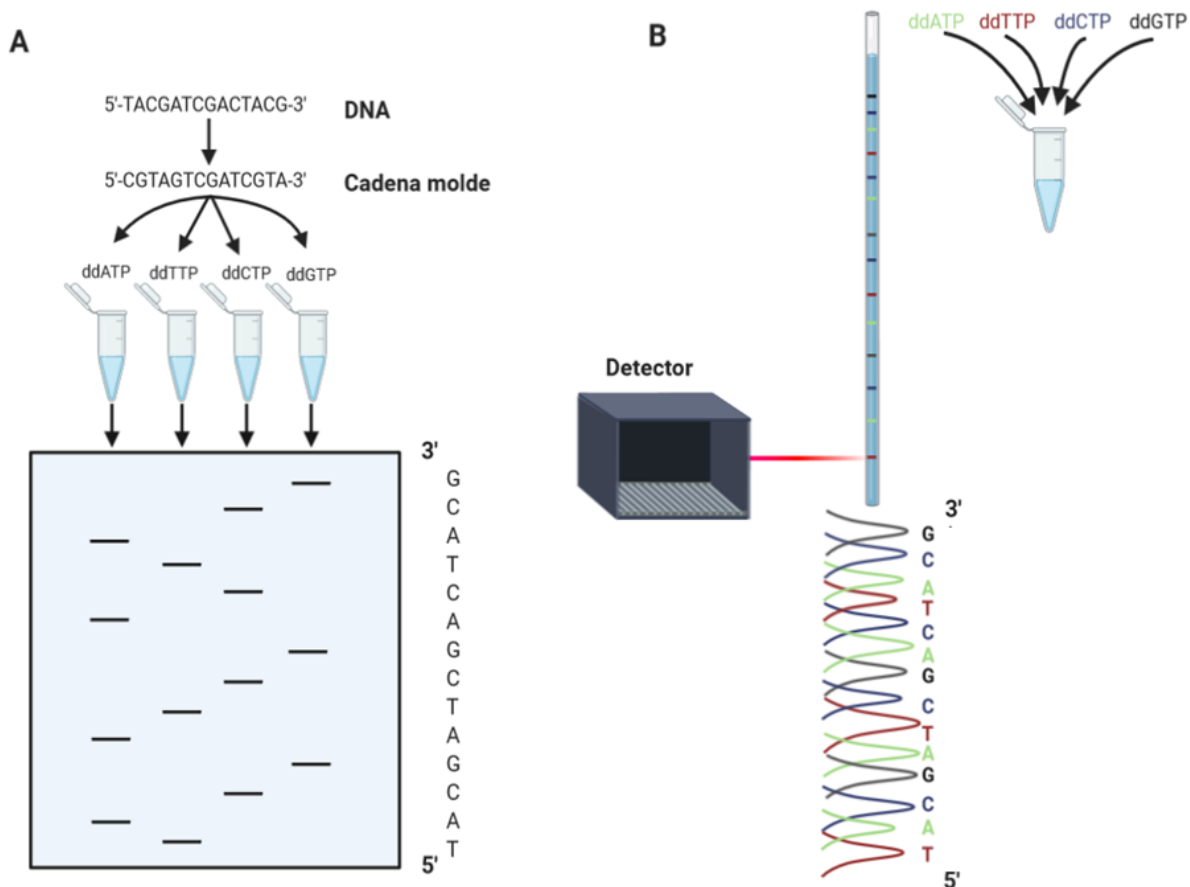
nuevos terminadores que contienen dos fluorocromos, como los *BigDye* (*Applied Biosystems*), para potenciar e igualar las señales de emisión mediante el establecimiento de un efecto de transferencia de energía entre fluorocromos («FRET», en inglés)<sup>[14]</sup>. En los equipos actuales la detección se realiza de manera automática por parte de un fluorímetro situado en el extremo de una electroforesis capilar (Figura 4B).

En un principio tanto en el método de Maxam y Gilbert como en el de Sanger, las electroforesis se realizaban en geles desnaturantes de poliacrilamida debido a que presentaban resoluciones de hasta un nucleótido entre fragmentos de ADN, la necesaria para la secuenciación. Además, el tamaño de los geles solía ser lo mayor posible para aumentar el rango de separación de fragmentos y, por tanto, el resultado (Figura 4B). El problema de esto estribaba en su difícil manipulación unido a que las muestras cargadas en estos geles emitían radioactividad. Por ello, uno

de los principales avances en el método de Sanger fue la sustitución de estas electroforesis «convencionales» por el empleo de la electroforesis capilar. Esta implementación técnica ha permitido aumentar el tamaño de los fragmentos de secuenciación (en torno a 1.000 pb), el alto voltaje que permite aplicar ha aumentado la velocidad del proceso (anteriormente podía durar hasta ocho horas en comparación con los 15 minutos actuales), ha permitido disminuir la cantidad del material de partida hasta el nivel de nanogramos (ng,  $10^{-9}$  g) y ha posibilitado la paralelización del proceso pudiendo realizarse hasta 384 muestras a la vez debido al pequeño diámetro de los capilares (50-100 mm)<sup>[15]</sup>. El interior de estos capilares se encuentra silanizado por lo que su carga es negativa y no ejerce resistencia sobre el ADN cuya carga neta también es negativa. Además se encuentran rellenos de polímeros lineales que ejercen de fase estacionaria en la electroforesis.

Estos desarrollos técnicos hicieron que el método

de Sanger se hiciese con el monopolio de la secuenciación de ADN durante cerca de dos décadas que coinciden con los primeros avances importantes de la Genómica: secuenciación del primer procarionta (*Haemophilus influenzae*; 1,83 Mb)<sup>16</sup>, el primer genoma eucariota (*Saccharomyces cerevisiae*; 12,1 Mb)<sup>[17]</sup>, el primer genoma de una planta (*Arabidopsis thaliana*; 157 Mb)<sup>[18]</sup> o el genoma humano (*Homo sapiens*; 3,2 Gb)<sup>[19]</sup>. A pesar de los grandes avances en los métodos de secuenciación experimentados en los últimos quince años, que han cambiado los estudios genómicos, el método inventado por Sanger en 1977 sigue siendo el mejor para la secuenciación de un fragmento individual de ADN en el contexto de la manipulación de los ácidos nucleicos en el laboratorio. Sin estos pioneros de la Biología Molecular nada de lo hecho posteriormente existiría, la secuenciación de los ácidos nucleicos nos concedió la herramienta para leer los diferentes tomos del libro de la vida.



**Figura 4.** Secuenciación enzimática de Sanger. A) Las diferentes reacciones del método de Sanger se llevan a cabo por separado. Cada reacción cuenta con un didesoxinucleótido trifosfato marcado con un isótopo radiactivo que actúa como terminador de la reacción. Las reacciones son resueltas en un gel de poliacrilamida donde, gracias a una autorradiografía, se pueden observar los diferentes fragmentos de DNA y así descifrar la secuencia. B) Esquema del método de Sanger utilizado en la actualidad utilizando ddNTPs marcados con fluorocromos y el uso de la electroforesis capilar. Imagen diseñada en Biorender.

## Referencias

- [1] Hershey A.D. y Chase M. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol* 36: 39-56, 1952.
- [2] Elson D. y Chargaff E. On the desoxyribonucleic acid content of sea urchin gametes. *Experientia*. 8: 143-145, 1952.
- [3] Chargaff E., Lipshitz R. y Green C. Composition of the desoxy-pentose nucleic acids of four genera of sea-urchin. *J Biol Chem* 195: 155-160, 1952.
- [4] Watson J.D. y Crick F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171: 737-738, 1953.
- [5] Meselson M. y Stahl F.W. The replication of DNA in *Escherichia coli*. *Proc Natl Acad Sci USA* 44: 671-682, 1958.
- [6] Maxam A.M. y Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci USA* 74: 560-564, 1977.
- [7] Sanger F. y Coulson A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 94: 441-446, 1975.
- [8] Sanger F., Air G.M., Barrell BG, Brown N.L., Coulson A.R., Fiddes C.A., Hutchison C.A., Slocombe P.M. y Smith M. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 265: 687-695, 1977.
- [9] Min Jou W., Haegeman G., Ysebaert M. y Fiers W. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature* 237: 82-88, 1972.
- [10] Fiers W., Contreras R., Duerinck F., Haegeman G., Iserentant D., Merregaert J., Min Jou W., Molemans F., Raeymaekers A., Van den Berghe A., Volckaert G. y Ysebaert M. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* 260: 500-507, 1976.
- [11] Maxam A.M. y Gilbert W. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol* 65: 499-560, 1980.
- [12] Cantón F.R., García-Gutiérrez A., Gallardo F., de Vicente A. y Cánovas F.M. Molecular characterization of a cDNA clone encoding glutamine synthetase from a gymnosperm, *Pinus sylvestris*. *Plant Mol Biol* 22: 819-828, 1993.
- [13] Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-5467, 1977.
- [14] Kumar S. y Fuller C.W. Advances in Dye-Nucleotide Conjugate Chemistry for DNA Sequencing. *Perspectives in Bioanalysis* 2: 119-149, 2007.
- [15] Miksík I. y otros. Matrices for capillary gel electrophoresis—a brief overview of uncommon gels. *Biomed. Chromatogr.* 20: 458-465, 2006.
- [16] Fleischmann R., Adams M., White O., Clayton R., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A. y Merrick J.M. et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269: 496-512, 1995.
- [17] Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dujon B., Feldmann H., Galibert F., Hoheisel J.D., Jacq C., Johnston M., Louis E.J., Mewes H.W., Murakami Y., Philippsen P., Tettelin H., Oliver S.G. Life with 6000 genes. *Science* 274: 546: 563-567, 1996.
- [18] The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815, 2000.
- [19] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-945, 2004.