

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 532 752**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

**C07D 239/22** (2006.01)

**C07D 473/02** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.09.2008** **E 08806340 (9)**

97 Fecha y número de publicación de la concesión europea: **17.12.2014** **EP 2195459**

54 Título: **Caracterización de nucleobases**

30 Prioridad:

**19.09.2007 GB 0718255**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**31.03.2015**

73 Titular/es:

**UNIVERSITY COURT OF THE UNIVERSITY OF  
EDINBURGH (100.0%)  
OLD COLLEGE, SOUTH BRIDGE  
EDINBURGH EH8 9YL, GB**

72 Inventor/es:

**BRADLEY, MARK y  
DÍAZ-MOCHÓN, JUAN J.**

74 Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

**ES 2 532 752 T3**

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

## DESCRIPCIÓN

Caracterización de nucleobases

5 **Campo técnico de la invención**

La presente invención proporciona compuestos de nucleobases modificadas, compuestos miméticos de ácidos nucleicos modificados y varios usos de estos compuestos. Más específicamente, la invención proporciona métodos para la caracterización de nucleobases, la caracterización de SNP y la secuenciación de ácidos nucleicos.

10

**Antecedentes de la invención**

Aunque ya se han publicado dos métodos originales para la secuenciación de ADN –el método químico de Maxim y Gilbert (A.M. Maxam and W. Gilbert, PNAS, 1977, 74, 560-564) y los métodos enzimáticos de Sanger (F. Sanger et al., PNAS, 1976, 5463-5467)- el método dominante que se usa actualmente se basa en la aproximación de Sanger y la llamada terminación o dideoxi-secuenciación. Esta metodología fundamental se basa en la terminación parcial de una hebra creciente de ADN, generándose una escalera de fragmentos marcados, que requieren una separación por tamaño para poder llevarse a cabo el análisis de la secuencia de ADN. Obviamente, se han introducido varias mejoras en el método desde su concepción inicial; así, los marcadores han evolucionado, sustituyéndose los nucleótidos radiactivos convencionales por compuestos fluorescentes, y ahora se usan capilares rellenos de polímero en vez de electroforesis en geles planos, para aplicaciones de secuenciación a gran escala. En consecuencia hoy en día se tiende al uso sistemático de aproximaciones masivas, tales como el uso simultáneo de 96 capilares o el uso de cuatro fluorocromos en un único canal, a la vez que las mejoras en la sensibilidad en la detección de fluorescencia y el uso de polimerasas más eficientes han ido permitiendo realizar secuenciaciones más largas. No obstante, el hecho de que este método tenga limitaciones inherentes se hace patente por el esfuerzo colosal que ha requerido la secuenciación del genoma humano. En los últimos años se han descrito varias aproximaciones nuevas que se pueden clasificar en tres categorías: (i) la secuenciación por la adición repetitiva de bases individuales; (ii) la pirosecuenciación y (iii) la ruptura mediada por enzimas de restricción o la ligación mediada por quinasas con desconvolución/descodificación.

30

(i) *Secuenciación por adición repetitiva de bases individuales.*- Se han publicado varios trabajos sobre este tema (Z.M. Li et al, PNAS, 2003, 100, 414-419; T.S. Seo et al, PNAS, 5 2005, 102, 5926-5931; L.R. Bi et al, J. Am. Chem. Soc., 2006, 128, 2542). Este enfoque se basa en la adición mediada por un enzima de una única base a una hebra cebada de ADN. La adición de la base se controla mediante una modificación del grupo trifosfato que imposibilita la adición de más de una base al mismo tiempo. Por ejemplo, mediante un bloqueante físico en el nucleótido o mediante un bloqueante químico (por ejemplo, un grupo éster en el grupo 3'OH). Este enfoque se basa en el uso de bloques marcados fluorescentemente, de manera que normalmente al eliminarse el bloqueante también se libera el fluoróforo, permitiéndose el comienzo de otro ciclo de reacción. Este enfoque tiene varios problemas; por ejemplo, este método necesita enzimas y complejos grupos que contienen trifosfato.

40

(ii) *Pirosecuenciación* (P. Nyren et.al, Anal. Biochem., 1996, 242, 84-89), En este enfoque una hebra creciente, cebadora de ADN se trata con una enzima y uno de los cuatro trifosfatos. Si la base se incorpora, el pirofosfato se libera; si no hay incorporación, entonces no se genera pirofosfato. El pirofosfato reacciona con una sulfurilasa, que lo transforma en ATP en presencia de APS (5'-fosfosulfato de adenosina). Este compuesto se trata entonces con otra enzima (luciferasa) que produce luz. Esta luz se usa para determinar cuánta y si hay adición de una base específica a la hebra creciente de ADN. Si dos o más bases del mismo tipo se añaden en una sola vez, entonces se genera más luz, cuya intensidad puede cuantificarse. El proceso se repite con el siguiente tipo de trifosfato, generándose así secuencias. Hay un número de problemas con este enfoque, que incluyen el hecho de que la cuantificación de la emisión de luz no siempre es posible, de manera que es imposible leer secuencias largas que contengan la misma base (es decir, en la práctica es imposible distinguir entre 14 y 15 bases del mismo tipo debido a la variabilidad en la emisión de luz). Este es el método usado en un artículo reciente describiendo la secuenciación por el extremo 5' usando millones de microesferas organizadas en micromatrices (M. Margulies et al., Nature, 2005, 437, 376-380).

50

(iii) *Ruptura mediada por enzima de restricción o ligación con desconvolución/descodificación* (S. Brenner et al., Nature Biotech., 2000, 18, 630-634).- En este enfoque, las secuencias se cortan usando una enzima de restricción para producir una secuencia colgante. Estas secuencias se decodifican usando una serie de 16 adaptadores codificados. Los adaptadores son a su vez cortados, lo que expone el siguiente bloque de bases a otro ciclo de decodificación. Un resultado semejante puede conseguirse mediante ligación. De nuevo, hay una serie de problemas: se necesitan varios pasos por cada desconvolución; se sigue necesitando sondas marcadas y diversas enzimas; hay cortes incompletos o no deseados, etc. Este fue el método usado por Brenner (S. Brenner et al, Nature Biotech., 2000, 18, 630-634) y por Shendure y Church en su método de secuenciación masiva en paralelo usando chips (microesferas atrapadas en un gel de poliacrilamida, J. Shendure et al., Science, 2005, 308, 1728-1732 and R. D. Mitra et al., PNAS, 2003, 100, 5926-5931).

60

*Polimorfismos de nucleótido simple:* Otro área, pero relacionada con la secuenciación es la de los polimorfismos de nucleótido único (A-C. Syvanen et al, Nature Genetics, 2005, 37, S5-S10). En realidad el análisis de SNP puede concebirse como si fuera la secuenciación de una única base. En humanos, existe por término medio un SNP por cada 300-1.000 bases, y representan hasta un 90 % de la variabilidad genética entre individuos. Un SNP puede constituir un factor de riesgo (o también una ventaja) respecto a enfermedades determinadas y también un determinante de características físicas. Hay muchos y variados métodos para analizar los SNP, pero generalmente consisten en reacciones de extensión de cebadores usando polimerasas y trifosfatos marcados fluorescentemente, aunque los métodos de captura y de análisis varían considerablemente. El análisis de los SNP es una forma sencilla de secuenciación de ADN en algunos aspectos, en el sentido de que determinar la identidad de una única base es el objetivo principal (aunque su contexto es, por supuesto, crucial).

*Ligaciones y reacciones dirigidas por ADN.* El ADN y los ácidos peptidonucleicos (APN) se han usado para síntesis de compuestos en diversas aproximaciones químicas basadas en ligación (véase sobre todo el trabajo de D. R. Liu and O. Seitz -X. Li y D.R. Liu, Angew. Chem. Int. Ed., 2004, 43, 4848-4870; S. Ficht et al, ChemBioChem, 2005, 6, 2098-2103). La ligación no enzimática también se ha usado en el sentido ADN-ADN por Kool y Richert (N. Griesang et al, Angew. Chem. Int. Ed., 2006, 45, 6144-6148 y referencias incluidas, por ejemplo P. Hagenbuch et al, Angew. Chem. Int. Ed. 2005, 44, 6588-6592), quienes usaron química clásica de adición de nucleófilos para ligar hebras de ADN (reaccionando el 3'-fosfotioato con una 5'-yodotimidina) o monómeros (reaccionando, por ejemplo, el 3'-aminonucleótido con un fosfato activado). El primer enfoque se pudo utilizar para la detección mediante color de mutaciones puntuales en ARN y ADN, aunque ello requiere cebadores largos tanto en el nucleófilo como en el electrófilo. El método de Richerts, aunque se basa en monómeros, requería los llamados cebadores de ayuda de forma que la incorporación directa requería dos cebadores que abarcasen la base. Liu usó química dinámica para componer polímeros de APN utilizando un molde de ADN (D.R. Liu et al. J. Am. Chem. Soc. 2003, 125, 13924-13925).

Los APN se han usado previamente como sondas genéticas (véase la revisión de P. Pauisova y F. Pellestor Ann. Genetique, 2004, 47, 349-358) debido a su capacidad de reconocer con mucha precisión secuencias complementarias de ADN o de ARN; no obstante, como no pueden ser reconocidos por polimerasas su uso como herramienta de análisis genético ha sido muy limitado.

*Química dinámica:* Durante la pasada década ha habido una intensa actividad en el área de las bibliotecas dinámicas (combinatorias) (P. T. Corbett et al. Chem. Rev., 2006, 106, 3652-3711; J.M. Lehn, Chem. Eur. J, 1999, 5, 2455-2463, O. Ramstrom and J.M. Lehn, Nat. Rev. Drug Discov., 2002 1, 26-36). Una "biblioteca dinámica" se puede preparar mezclando en solución dos compuestos complementarios, como una selección de aldehídos y una amina, o dioles y ácidos borónicos, o tioles y disulfuros en presencia de un molde. Debido al equilibrio dinámico que se organiza en el sistema (amina/aldehído/imina), el ligando que se une más fuertemente predomina y por tanto, en esencia, el molde "construye" y "concentra" su propia pareja. Recientemente, Dawson et al., (J. Am. Chem. Soc, 2006, 128, 15602-15603), mostraron que las cinéticas de equilibrio de los procesos dinámicos pueden ser también acelerados usando catalizadores tales como la anilina.

Como se desprende de lo anterior, la aplicación práctica de todos los nuevos métodos de análisis de ADN (y también los viejos) conlleva una serie de problemas, especialmente la necesidad de usar enzimas y trifosfatos, que con frecuencia resultan muy costosos.

El objetivo de la presente invención es obviar o mitigar al menos uno de los problemas antes citados.

### Sumario de la invención

La invención descrita en el presente documento proporciona bases modificadas, nucleobases modificadas y compuestos miméticos de ADN que pueden usarse en diversos métodos de secuenciación y/o caracterización de SNP. La invención tiene claras ventajas respecto a los métodos existentes más avanzados, puesto que cada uno de los métodos que se describen en el presente documento usan reactivos químicos y no requieren del uso de enzimas.

La invención proporciona métodos, bases modificadas, aplicaciones y kits de acuerdo a las reivindicaciones 1 a 11 que se describen en el presente documento.

Por consiguiente, en un primer aspecto, la presente invención proporciona una base modificada que comprende una parte capaz de llevar a cabo reacciones covalentes reversibles seleccionadas de los grupos funcionales (i) un aldehído y (ii) una cetona, y un marcador detectable.

Un experto en la materia apreciará que las bases (que también se conocen o se refieren aquí como "nucleobases") incluyen purinas y pirimidinas, las cuales incluyen, por ejemplo, las bases específicas adenina, guanina, timina, citosina y uracilo. De esta manera, una realización de la invención se refiere a formas modificadas de las bases adenina, guanina, timina, citosina y uracilo. Además, la presente invención abarca variantes tales como, por ejemplo, xantina, hipoxantina, isoguanina y ácido úrico.

Ha de entenderse que el término “base modificada” puede emplearse para abarcar bases/nucleobases que contienen una cadena alquílica que contiene grupos funcionales capaces de llevar a cabo reacciones covalentes reversibles. Preferentemente, la parte heterocíclica de las bases puede modificarse para contener la cadena alquílica y los grupos funcionales. Más preferentemente un heteroátomo o átomo de carbono del heterociclo puede modificarse para contener la cadena alquílica y los grupos funcionales.

Ha de entenderse que los grupos funcionales capaces de llevar a cabo “reacciones covalentes reversibles” son grupos que contienen aldehídos y/o cetonas y una forma de realización se refiere a reacciones covalentes reversibles entre los grupos aldehído/cetona de la base modificada y aminas, hidrazida e hidrazidas (A. Dirksen, et al., J. Am. Chem. Soc, 2006, 128, 15602-15603), alcoxiamina (.A. Polyakov et al., J. Phys. Org. Chem. 1999, 12, 357-363) o alcoholes, dioles y/o ácidos borónicos (O. Abed et al. Chem. Mater., 2006, 18, 1247 -1260). En una realización, el grupo capaz de una reacción covalente reversible no es un alcohol.

El término “marcador detectable” puede emplearse para abarcar marcadores o indicadores que por ejemplo se puedan distinguir entre sí ópticamente o de otro modo. Muchos de estos marcadores o indicadores son conocidos por los expertos en la materia pero, a modo de ejemplo, algunos marcadores apropiados para su uso en la presente invención pueden incluir, por ejemplo, compuestos fluorescentes o marcadores con distintas masas. Más específicamente, y en otra forma de realización, las bases/nucleobases modificadas pueden incluir uno o más marcadores detectables (tal como, por ejemplo, un fluoróforo) seleccionados de entre un grupo de marcadores con pigmentos detectables ópticamente abarcando, por ejemplo, desde el espectro azul al rojo lejano. Los ejemplos de marcadores que pueden ser apropiados incluyen, por ejemplo, dansilo, fluoresceína, rodamina, rojo Texas, IAEDANS, pigmentos de cianina (Cy3, Cy3.5, Cy5, Cy5.5, Cy7), pigmentos Bodipy (Invitrogen) y/o pigmentos Alexa Fluor (Invitrogen). En una realización, el marcador detectable no es ferroceno.

Los “marcadores de masas” apropiados pueden incluir, por ejemplo, marcadores que contienen bromuro u otros compuestos, moléculas o grupos capaces de producir un claro patrón isotópico distinguible en técnicas de espectrometría de masas tales como, por ejemplo, MALDI-TOF.

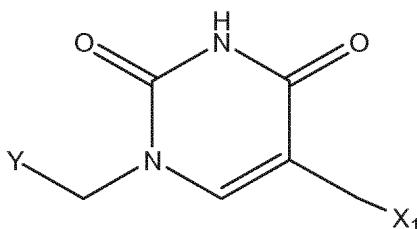
Por consiguiente, un experto en la materia apreciará que cualquiera de las nucleobases descritas en el presente documento puede detectarse, por ejemplo, mediante microscopía de fluorescencia o mediante técnicas de espectrometría de masas tales como MALDI-TOF o similares.

Ventajosamente, el heterociclo de cada una de las bases/nucleobases modificadas descritas en el presente documento puede contener un marcador detectable unido, por ejemplo, a algún número de posiciones a través de un heteroátomo o átomo de carbono. En una forma de realización de la invención, el heteroátomo puede estar modificado de forma que pueda contener espaciadores de carbono apropiados, tales como, por ejemplo, un grupo alquino, alquenileno o alquilileno que pueda sustituirse, independientemente, con uno o más de los marcadores detectables indicados anteriormente. A modo de ejemplo, el heteroátomo y/o heteroátomo modificado del heterociclo pueden contener uno o varios fluoróforos (T.S. Seo et al, PNAS, 2004, 101, 5488-5493; Z. Li et al, PNAS, 2003, 100, 414-419; L. Thoresen et al, Chem. Eur. J. 2003, 9, 4603-4610) y/o marcadores de masa, por ejemplo, bromuro, cloruro (C. Portal et al, J. Comb. Chem., 2005, 5 7, 554-560). En una realización, los heterociclos de purina y/o pirimidina pueden modificarse mediante, por ejemplo, reacciones de acoplamiento cruzado usando catalizadores de paladio (L. Thoresen et al, Chem. Eur. J. 2003, 9, 4603-4610; N.K. Garg et al. Chem. Commun., 2005,4551-4553).

Ventajosamente, cada una de las bases/nucleobases modificadas puede contener un marcador detectable diferente. De esta manera, el marcador detectable puede permitir, por ejemplo, que una nucleobase de adenina modificada se distinga de cualquier otra nucleobase modificada.

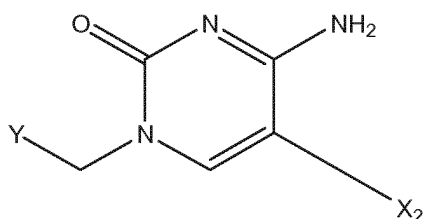
En el presente documento se describen bases modificadas seleccionadas del grupo que consiste en:

(i)



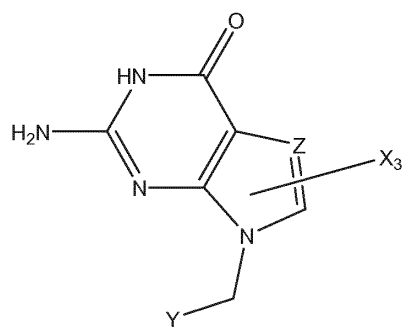
Fórmula 1

(ii)



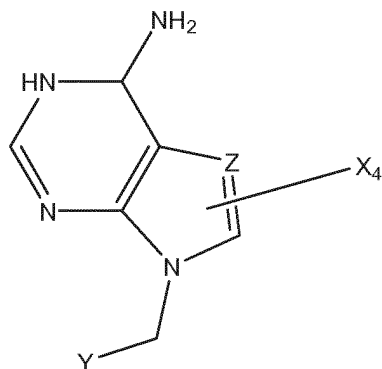
Fórmula 2

(iii)



Fórmula 3

(iv)



Fórmula 4

5 en el que Y incluye un grupo funcional capaz de llevar a cabo reacciones covalentes reversibles. Los grupos funcionales apropiados pueden ser, por ejemplo, aldehídos, cetonas y/o dioles.

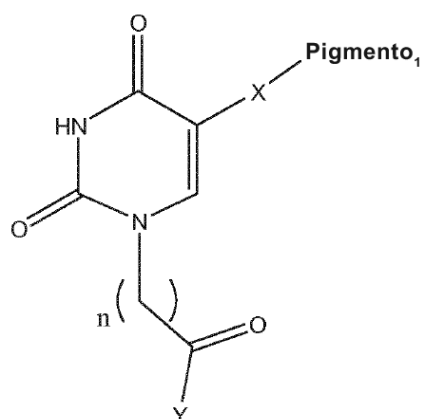
X<sub>1</sub>-X<sub>4</sub> pueden ser marcadores detectables diferentes o combinaciones de espaciadores/marcadores o hidrógeno.

10 Z puede ser carbono, nitrógeno, oxígeno o azufre.

En los casos (iii) y (iv) anteriores, X puede unirse a un heterociclo a través de Z, cuando Z es carbono, o a través del resto de carbono en posición 8.

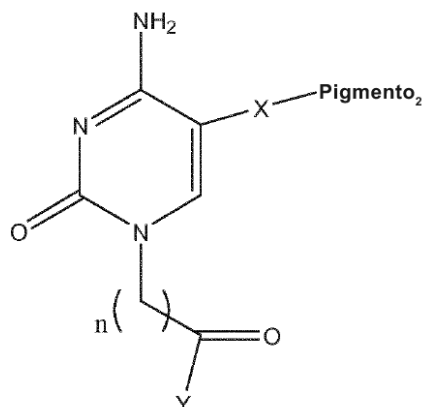
15 En el presente documento se describen bases modificadas seleccionadas del grupo que consiste en:

(i)



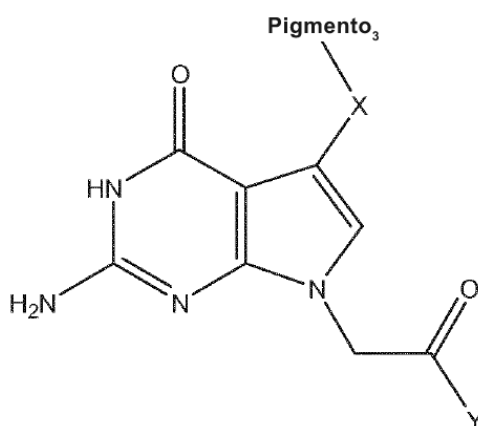
Fórmula 5

(ii)



Fórmula 6

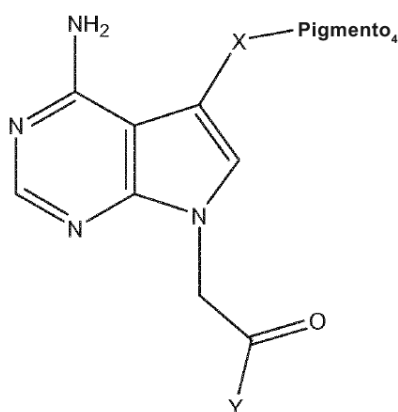
(iii)



Fórmula 7

y

5 (iv)

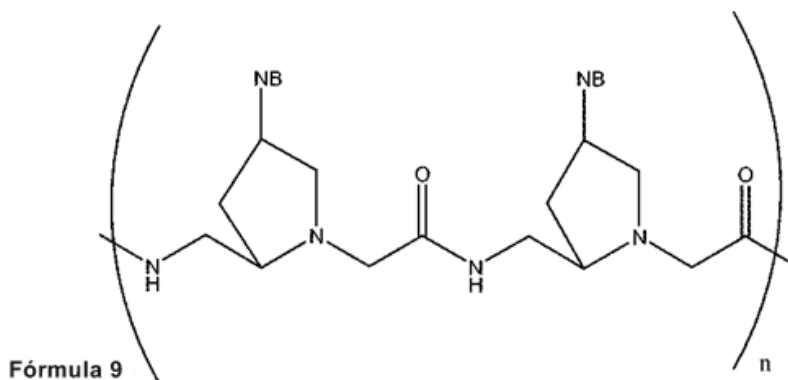


Fórmula 8

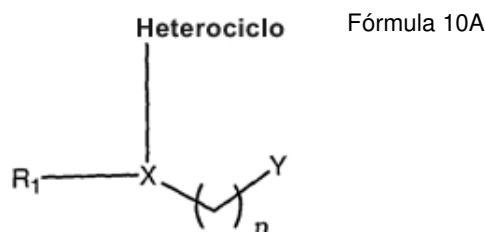
en el que X e Y pueden ser hidrógeno o una cadena hidrocarbonada, n es igual a 1, 2 o 3 y Pigmento<sub>1</sub>-Pigmento<sub>4</sub> pueden ser uno o más de los marcadores detectables listados anteriormente. Preferentemente, cada uno de los Pigmento<sub>1</sub>-Pigmento<sub>4</sub> representa un marcador detectable diferente de tal modo que, por ejemplo, la nucleobase modificada en la Formula 5 puede distinguirse de la que se muestra en la Fórmula 6-8.

Los ácidos peptidonucleicos (APN) son similares a los ácidos nucleicos existentes en la naturaleza –ácido desoxirribonucleico (ADN) y ácido ribonucleico (ARN). Mientras que el ADN y el ARN contienen estructuras de azúcares de desoxirribosa o de ribosa, respectivamente, la estructura de los APN contiene unidades repetidas de N-(2-aminoetil)-glicina que están unidas por enlaces peptídicos. Las diversas bases de pirimidina y/o purina (o nucleobases) de APN están unidas a la estructura peptídica mediante la formación de enlaces amida. Un experto en la materia entenderá que una nucleobase individual unida a través de un enlace amida a una sola unidad de N-(2-aminoetil)-glicina puede describirse como un monómero de APN, pero otros APN incluyen, por ejemplo, aquéllos que contienen estructuras de aminoetil-glicina modificadas, tales como, por ejemplo, las basadas en pirrolidina (R. J. Worthington et al. Org. Biomol. Chem., 2007, 5, 249-259) y las basadas en indol (Fórmula 9). Un

experto en la materia reconocerá otros oligómeros apropiados.



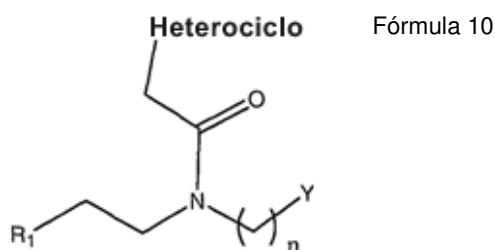
5 En el presente documento se describen monómeros de APN modificados que tienen la siguiente fórmula general:



10 en los que "heterociclo" es una base modificada (tal como, por ejemplo, citosina, adenina, guanina o timina/uracilo) que contiene un marcador detectable, n es igual a 1, 2 o 3 y además en el que X representa una forma de unión entre el heterociclo y la estructura que comprende R<sub>1</sub> e Y. R<sub>1</sub> representa un grupo capaz de producir reacciones covalentes reversibles. A modo de ejemplo, R<sub>1</sub> puede contener grupos tales como una amina, una hidrazida, una alcoxiamina, un ácido borónico, un diol y/o un tiol.

15 Y puede ser un grupo funcional capaz de producir reacciones covalentes reversibles, tal como, por ejemplo, un aldehído, una cetona, un diol, un ácido borónico o un tiol.

También se describen en el presente documento monómeros de APN modificados que tienen la siguiente fórmula general:



20 en los que "heterociclo" es una base modificada (como, por ejemplo, citosina, adenina, guanina o timina/uracilo) que puede contener un marcador detectable como se describe anteriormente, n es igual a 1, 2 o 3 y además en el que R<sub>1</sub> representa un grupo capaz de producir reacciones reversibles covalentes. A modo de ejemplo, R<sub>1</sub> puede contener grupos tales como una amina, una hidrazida, una alcoxiamina, un ácido borónico, un diol y/o un tiol.

25 Y puede ser un grupo funcional capaz de producir reacciones covalentes reversibles, tal como, por ejemplo, un aldehído, una cetona, un diol, un ácido borónico y un tiol.

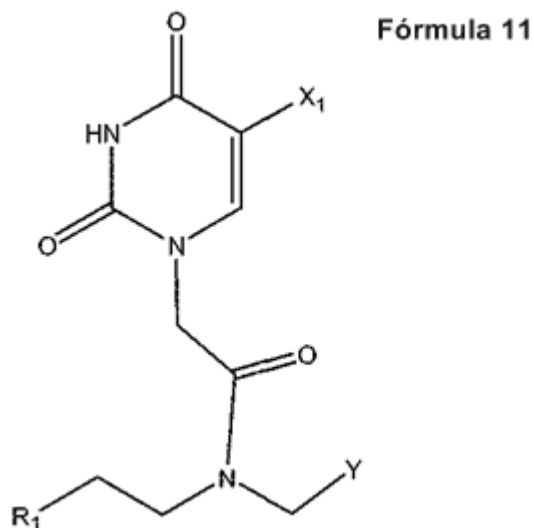
30 Ha de entenderse que R<sub>1</sub> puede derivatizarse para contener un grupo protector u opcionalmente un grupo protector que contiene (por ejemplo, unido covalentemente a) uno o más de los marcadores detectables descritos anteriormente.

35 Los grupos protectores adecuados para su uso en la derivatización adicional de R<sub>1</sub> pueden incluir, por ejemplo, grupos protectores tales como acetilo, N-[1-(4,4-dimetil-2,6-dioxociclohex-1-ilideno)etilo] (Dde), fluorenilmetoxicarbonilo (Fmoc), grupos tritilo, grupos protectores fotosensibles (basados en nitroveratilo) de disulfuro

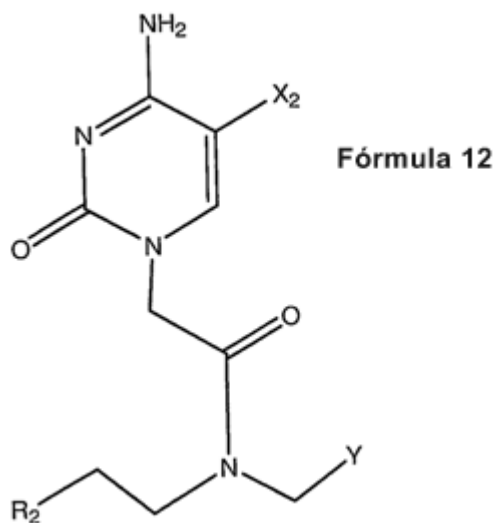
(Ardec (arilditioetiloicarbónico)), butiloxicarbónico (Boc), benciloxicarbónico (Cbz), trifluoroacetilo (Tfa), ftalimido, bencilo, aliloxicarbónico (Alloc), toluensulfonilo (Ts), éter de metoximetilo (MOM), éter de tetrahidropiránil (THP), éter de alilo, éter de butilo, acetal de bencilideno (Green, Wiley-Interscience, Nueva York, 1999).

5 En una realización, los monómeros de APN modificados pueden seleccionarse del grupo que consiste en:

(i)

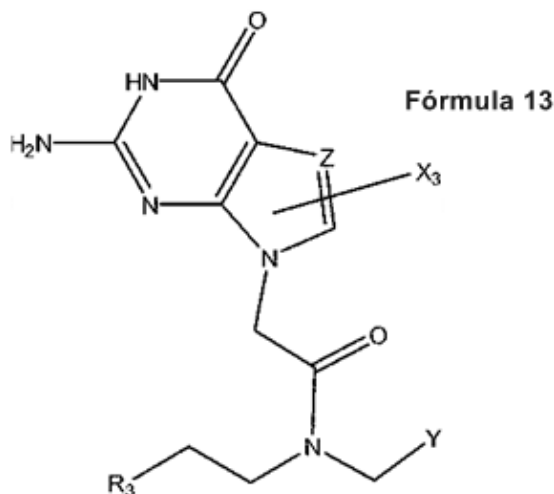


(ii)





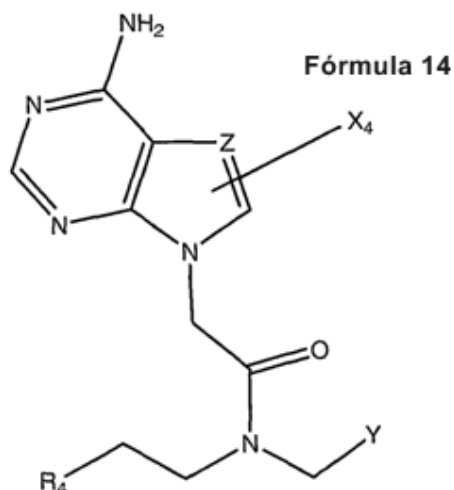
(iii)



y

5

(iv)



10 en los que R<sub>1</sub>-R<sub>4</sub> pueden incluir un grupo capaz de producir reacciones covalentes reversibles (véase arriba) opcionalmente protegido con un grupo protector (tal como se describe anteriormente). Además, o como alternativa, R<sub>1</sub>-R<sub>4</sub> pueden comprender un grupo protector que a su vez comprende (por ejemplo se une covalentemente a) uno o más de los marcadores detectables descritos en el presente documento.

15 X<sub>1</sub>-X<sub>4</sub> pueden ser uno o más de los marcadores detectables descritos en el presente documento y pueden estar unidos al heterociclo en cualquier número de posiciones a través de un heteroátomo o de un átomo de carbono. En una realización, el heteroátomo puede estar modificado de forma que pueda comprender además grupos espaciadores/espaciadores de carbono adecuados tales como, por ejemplo, un grupo alquino, alquenileno o alquilileno, que pueden ser sustituirse independientemente con uno o más de los marcadores detectables indicados anteriormente.

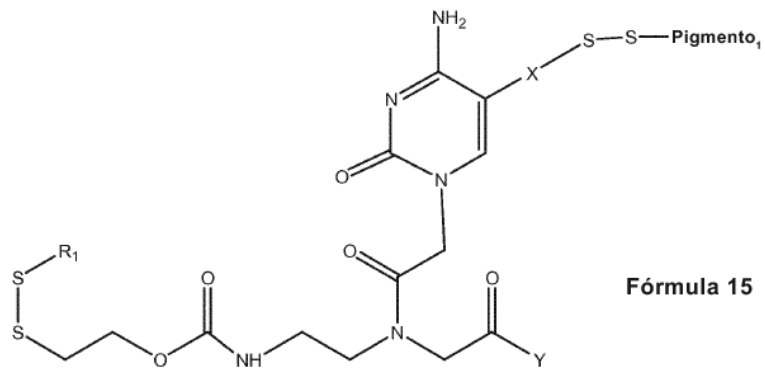
20 X<sub>1</sub>-X<sub>4</sub> es un marcador detectable unido a los heterociclos por un grupo enlazador escindible o un hidrógeno. Z puede ser carbono, nitrógeno, oxígeno y azufre y en los casos (iii) y (iv) anteriores, X puede estar unido al heterociclo a través de Z, cuando Z es un carbono, o a través del carbono en posición 8.

25 Y puede ser un grupo funcional capaz de producir reacciones covalentes reversibles tal como, por ejemplo, un aldehído, una cetona, un diol, un ácido borónico o un tiol.

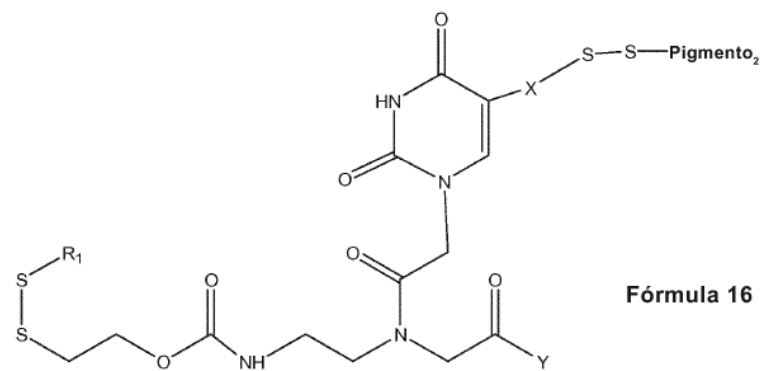
Además, se describen en el presente documento monómeros modificados de APN seleccionados del grupo que consiste en:

30

(i)

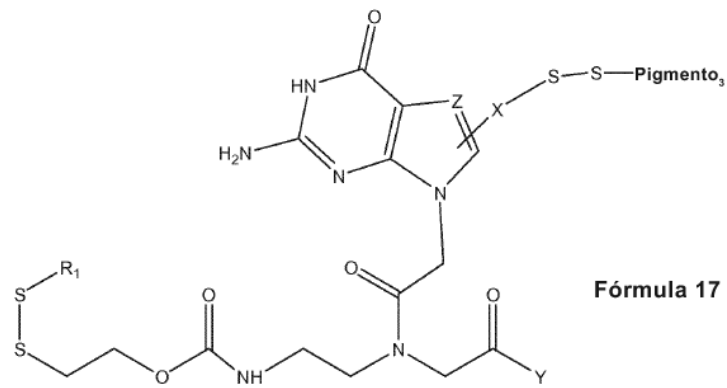


(ii)

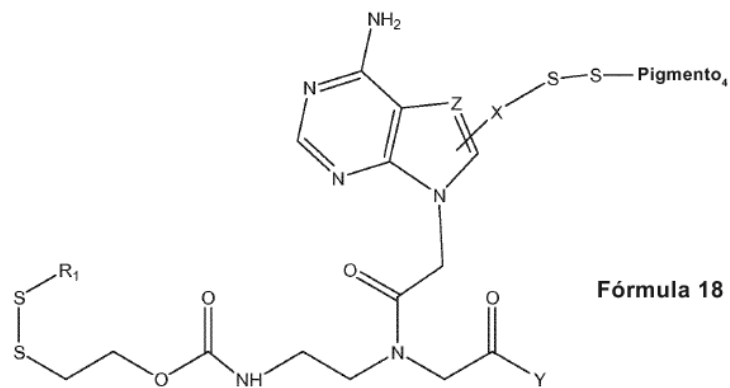


5

(iii)



y  
(iv)

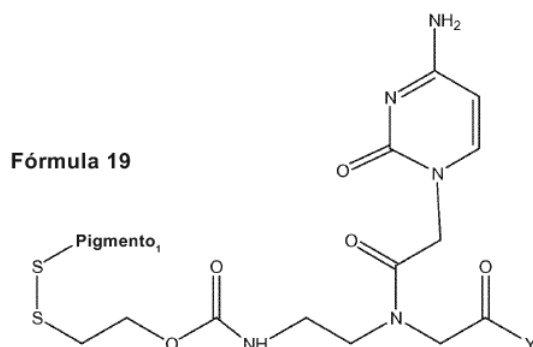


10

en el que R<sub>1</sub> puede ser una cadena hidrocarbonada, y un anillo de arilo, X puede ser una cadena hidrocarbonada e Y puede ser una cadena hidrocarbonada o hidrógeno.

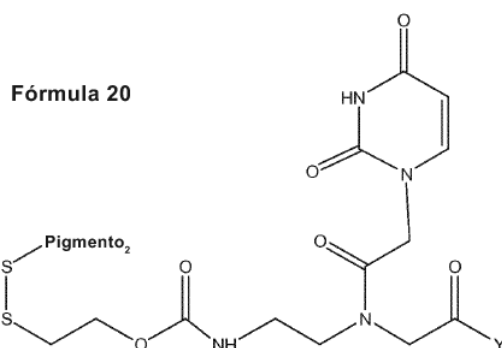
En una realización adicional, las bases modificadas pueden ser seleccionadas del grupo que consiste en:

(i)

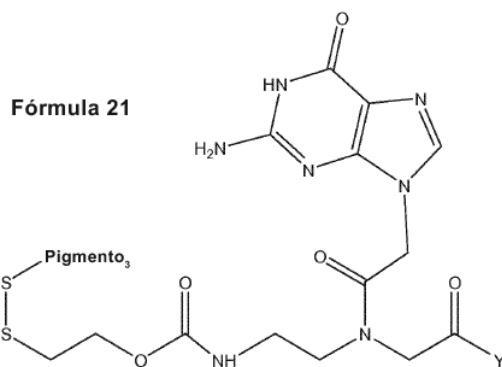


5

(ii)

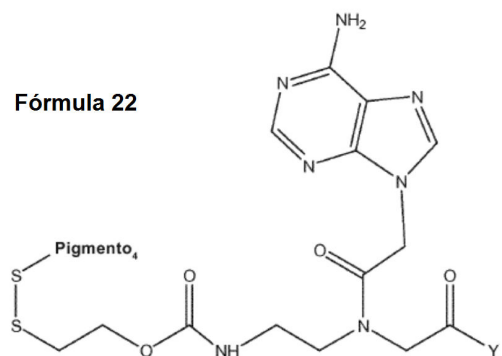


(iii)



10

y  
(iv)



en el que Y puede ser un hidrógeno o una cadena hidrocarbonada.

15

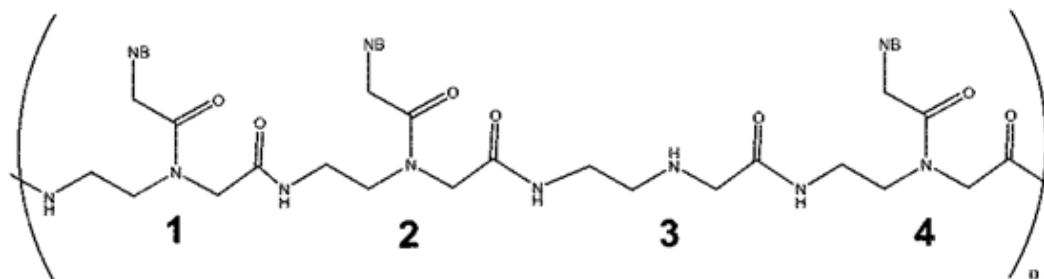
Un experto la materia apreciará que además de monómeros de APN, la presente divulgación proporciona dímeros o trímeros de APN. El término "dímero de APN" de acuerdo con esta invención debe entenderse en relación a dos (o tres, en el caso de "trímeros" de APN) monómeros de APN que están unidos covalentemente. En una realización, un dímero de APN puede incluir por lo menos una nucleobase modificada para incluir cualquiera de los marcadores detectables descritos en el presente documento. En otras realizaciones, los dímeros de APN pueden incluir en sus extremos N- o C-terminales cualquiera de los marcadores detectables descritos en el presente documento.

En situaciones en las que el extremo N- o C-terminal del dímero de APN se modifica para incluir un marcador detectable, los otros extremos N- o C-terminales pueden incluir un grupo capaz de producir reacciones covalentes reversibles. Como tales, en una realización, se describen en el presente documento dímeros de APN (o trímeros) que incluyen en el extremo N-terminal un marcador detectable y en el extremo C-terminal, un grupo capaz de producir reacciones covalentes reversibles. Además, en una realización, al menos una de las nucleobases de los monómeros de APN puede incluir además un marcador detectable.

Los ejemplos de métodos para producir dímeros de APN (o trímeros) y los ejemplos de formas específicas de dímeros de APN se describen posteriormente en más detalle.

También se describen en el presente documento oligómeros de APN y un experto en la materia entenderá con facilidad que el término "oligómero" puede usarse para referirse a una molécula que incluya al menos dos monómeros de APN unidos, por ejemplo, por un enlace peptídico. También se describen en el presente documento otros compuestos miméticos de ADN, como se indica anteriormente, en los que NB es una nucleobase (por ejemplo, una nucleobase modificada de acuerdo a la presente invención) y n es por lo menos 2.

A la vista de lo anterior, un experto en el tema apreciará que un oligómero de APN incluye normalmente una estructura peptídica continua donde cada amina secundaria de la estructura peptídica está adicionalmente derivatizada para incluir una nucleobase (tal como las nucleobases modificadas descritas anteriormente). En una realización adicional, se aporta un oligómero de APN en el que algunas de las aminas secundarias de la estructura peptídica continua no están derivatizadas para incluir una nucleobase y por tanto no quedan emparejadas. Estos oligómeros pueden citarse como oligómeros que comprenden "posiciones en blanco". La Fórmula 24 a continuación proporciona un ejemplo de un monómero de APN que incluye una posición en blanco (esto es, la amina secundaria número 3 no está derivatizada para incluir una nucleobase (NB)).

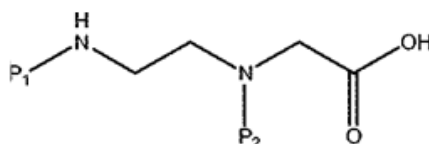


Fórmula 24

Los oligómeros de APN divulgados en el presente documento pueden incluir adicionalmente, en las posiciones N- o C-terminal, un grupo capaz de producir reacciones covalentes reversibles. Por ejemplo, la posición N-terminal puede incluir un grupo amino libre, hidrazida de aldehído/cetona, hidrazina, alcoxiamina, alcoholes, dioles y/o ácidos borónicos y la posición C-terminal puede incluir un grupo capaz de producir reacciones covalentes reversibles con el grupo en posición N-terminal. En una realización, cualquiera de las posiciones N-terminal y/o C-terminal pueden derivatizarse de forma que incluyan un grupo protector (tal como se describe anteriormente).

Los oligómeros de APN para su uso en los métodos descritos en el presente documento pueden sintetizarse usando unidades de N-2-aminoetil-glicina protegidas con grupos protectores ortogonales. Tales unidades pueden tener la siguiente fórmula general:

Fórmula 25



en la que P<sub>1</sub> y P<sub>2</sub> pueden ser disulfuros (Ardec (arilditioetiloxicarbonilo)), grupos protectores fotosensibles (basados en nitroveratilo), butiloxicarbonilo (Boc), benciloxicarbonilo (Cbz), trifluoroacetilo (Tfa), ftalimida, bencilo, aliloxicarbonilo (Alloc), N-[1-(4,4-dimetil-2,6-dioxociclohex-1-iliden)etilo] (Dde), fluorenilmetoxicarbonilo (Fmoc), t-butoxicarbonilo (Boc) y grupos tritilo (Green, Wiley-Interscience, New York, 1999).

Los oligómeros de APN descritos anteriormente pueden encontrar una aplicación particular en métodos de análisis genético.

Como tales, en un segundo aspecto, la presente invención proporciona un uso para una o más de las bases/nucleobases proporcionadas por el primer aspecto de la presente invención y/o los monómeros/oligómeros de APN descritos en el presente documento, en métodos de análisis genético. Se ha de entender que el "análisis genético" puede incluir, por ejemplo, la caracterización, identificación y/o secuenciación de nucleobases de ácidos nucleicos. En una realización, los métodos pueden usarse para caracterizar polimorfismos de un único nucleótido y/o para secuenciar ácidos nucleicos.

Un experto en la materia estará familiarizado con el término "polimorfismo de nucleótido simple" o "SNP". En resumen, un "SNP" representa una forma de variación en un genoma en el que un nucleótido particular del genoma varía entre miembros de una población. A modo de ejemplo, un SNP puede comprender dos alelos (esto es, uno de los dos posibles nucleótidos de un determinado locus) y, en dichos casos, algunos de los individuos de una población pueden portar un alelo de SNP en un determinado locus mientras otros pueden portar el otro alelo en el mismo locus.

En consecuencia, la expresión "caracterización de una nucleobase" puede usarse para abarcar el acto de identificar o determinar una nucleobase particular de una secuencia de ácido nucleico – en otras palabras, identificar qué nucleobase contiene un nucleótido particular. En los casos donde los métodos se usan para caracterizar un SNP, el término "caracterizar" puede usarse para abarcar el acto de determinar qué alelo de SNP concreto (o nucleobase) está presente en la secuencia de un ácido nucleico particular.

Por lo tanto, en un tercer aspecto, la presente invención proporciona un método para caracterizar un nucleótido en una secuencia de ácido nucleico, comprendiendo dicho método las etapas de:

- (a) poner en contacto un ácido nucleico con un oligómero de APN capaz de hibridar con una región del ácido nucleico y que carece de una base complementaria a la del nucleótido a caracterizar, para formar un dúplex de ácido nucleico/APN, y
- (b) poner en contacto el dúplex de ácido nucleico/APN con bases modificadas de acuerdo con el primer aspecto de la invención;

en el que el oligómero de APN contiene un grupo capaz de reaccionar reversiblemente con el grupo funcional Y, y donde la base modificada que se integra en el dúplex de ácido nucleico/APN es complementaria a la del nucleótido a caracterizar, caracterizándose el nucleótido mediante espectrometría de masas o a través del marcador detectable de la base modificada.

Un experto de la materia apreciará que la parte del oligómero de APN que carece de una base complementaria a una nucleobase de la secuencia de ácido nucleico, puede presentar un resto, por ejemplo una amina secundaria, capaz de reaccionar reversiblemente con un resto de las bases modificadas descritas anteriormente. Como tal, tras la puesta en contacto con el dúplex de ácido nucleico/APN, una base modificada que es complementaria (o está emparejada con) una nucleobase del ácido nucleico, puede incorporarse en el dúplex de ácido nucleico/APN mediante la formación de, por ejemplo, (i) una especie de iminio reversible entre la amina secundaria del oligómero de APN y el resto reactivo (grupo aldehído) de la base modificada y (ii) la formación de enlaces de hidrógeno entre la nucleobase modificada y la nucleobase del ácido nucleico.

En una realización, el método puede incluir la etapa adicional de atrapar la base integrada en el dúplex de ácido nucleico/APN y complementaria (o sea, emparejada) al nucleótido del ácido nucleico. Por ejemplo, puede detenerse la reacción reversible entre la amina secundaria del oligómero de APN y el grupo capaz de reacciones covalentes reversibles de la base modificada. Por ejemplo, pueden reducirse las especies de iminio para dar lugar a una amina terciaria estable usando agentes reductores como cianoborohidruro de sodio.

Se ha de entender que la expresión "capaz de hibridar con una región de la secuencia de ácido nucleico" debe interpretarse en el sentido de que el oligómero de APN es complementario a, o comparte cierto grado de homología con, una porción de la secuencia de ácido nucleico.

El experto entenderá fácilmente que mediante el marcador detectable presente en cada una de las nucleobases modificadas en contacto con el dúplex de ácido nucleico/APN, puede ser posible detectar qué base modificada que se ha incorporado al dúplex de ácido nucleico/APN. De este modo se puede conseguir fácilmente la caracterización del nucleótido del ácido nucleico. Por ejemplo, si la nucleobase modificada que se ha encontrado integrada con el dúplex de ácido nucleico/APN contiene un marcador que indica que se trata de una nucleobase

de timina, de acuerdo con el emparejamiento convencional de bases complementarias, el nucleótido del ácido nucleico debe contener una nucleobase de adenina.

5 Sin desear quedar ligados a teoría alguna, la integración de una nucleobase modificada que es complementaria a una nucleobase del ácido nucleico puede representar un ejemplo de un proceso de selección dinámica que está relacionado con las diversas fuerzas de interacción de las nucleobases complementarias (emparejadas) y no emparejadas, así como con la concentración relativa de las cuatro nucleobases modificadas y puede controlarse por cambios en la concentración de los tampones, el pH, la temperatura y también por el uso de diferentes catalizadores. Los expertos en este campo conocen bien los procesos de selección dinámica e incluyen los sistemas en los que un número de compuestos complementarios se mezclan en presencia de un molde (J.M. Lehn, *Nat. Rev. Drug Disc.*, 2002 1, 26-36). Debido al equilibrio dinámico que aparece en un sistema de este tipo, el ligando que se une más fuertemente predominará y por tanto el molde “construye” o selecciona de entre los diversos componentes añadidos su propio “ligando” o “pareja”.

15 Ventajosamente, el dúplex de ácido nucleico/APN se pone en contacto con cada una de las nucleobases modificadas descritas anteriormente. Normalmente, el dúplex de ácido nucleico/APN contactará con nucleobases modificadas que contienen nucleobases complementarias a cada una de las nucleobases probablemente presentes en la muestra de ácido nucleico. Por ejemplo, y en una realización, el dúplex de ácido nucleico/APN puede ponerse en contacto con la forma modificada de las bases adenosina, guanina, citosina y timina descritas anteriormente. Además, al asegurarse de que cada tipo de base modificada usada en los métodos descritos en el presente documento contiene un marcador que permite distinguirla de las otras nucleobases modificadas, puede lograrse fácilmente la caracterización de un nucleótido en una secuencia de ácido nucleico.

25 Aunque los métodos descritos en el presente documento pueden llevarse a cabo en solución, puede ser ventajoso inmovilizar o como alternativa unir el oligómero de ácido nucleico o APN a alguna forma de sustrato de soporte, preferentemente un sustrato sólido. En una realización, el sustrato de soporte puede incluir vidrio, nitrocelulosa, celulosa, plástico, agarosa, microesferas, un metal (por ejemplo oro) o similares. En el caso de microesferas, se prefieren tamaños de aproximadamente 1 nm a aproximadamente 2 mm.

30 También se describe en el presente documento un método alternativo para caracterizar un nucleótido en un ácido nucleico. Tal como se ha indicado, el término “caracterizar” engloba el acto de identificar una nucleobase particular de un nucleótido. Cabe destacar que todas las definiciones descritas anteriormente también son de aplicación en este aspecto de la invención. El método comprende las etapas de:

- 35 (a) hibridar una secuencia de ácido nucleico con un oligómero de APN complementario a una región de la secuencia de ácido nucleico cadena arriba del nucleótido a caracterizar y que comprende además un grupo funcional capaz de producir reacciones covalentes reversibles, para formar un dúplex de ácido nucleico/APN; y  
(b) poner en contacto el dúplex de ácido nucleico/APN con monómeros de APN modificados de acuerdo al segundo aspecto de esta invención.

40 en el que el monómero de APN modificado que se integra con el dúplex de ácido nucleico/APN es complementario a la nucleobase del nucleótido, caracterizándose dicho nucleótido gracias al marcador detectable del monómero de APN.

45 Preferentemente, el oligómero de APN hibrida con, o es complementario a, una secuencia de ácido nucleico que se sitúa inmediatamente cadena arriba en relación al nucleótido a caracterizar. En otras palabras, el oligómero de APN puede hibridar con o unirse a una secuencia de ácido nucleico en una posición 3' a la del nucleótido del ácido nucleico tal que el resto terminal (o N-terminal) del oligómero de APN se sitúa inmediatamente adyacente al nucleótido a caracterizar.

50 Ventajosamente, el dúplex de ácido nucleico/APN se pone en contacto con monómeros modificados de APN (tales como aquéllos proporcionados por el segundo aspecto de la invención) que contienen nucleobases complementarias a cada una de las nucleobases probablemente presentes en la muestra de ácido nucleico. Por ejemplo, y en una realización, el dúplex de ácido nucleico/APN puede ponerse en contacto con monómeros de APN modificados que contienen las bases adenosina, guanina, citosina y timina descritas anteriormente.

55 Un experto en la materia entenderá que cuando los monómeros de APN se ponen en contacto con el dúplex de ácido nucleico/APN, el monómero de APN que contiene las nucleobases modificadas complementarias a la nucleobase del ácido nucleico, debido a la selección dinámica (como se describe anteriormente), se integrará en el dúplex de ácido nucleico/APN.

60 Los métodos descritos en el presente documento pueden utilizar los dímeros de APN proporcionados por esta invención. En dichos casos, en vez de usar los monómeros de APN tal como se describen, por ejemplo, en la etapa (b) descrita anteriormente, el método puede utilizar los dímeros (o trímeros) de APN descritos en el presente documento. Además, el oligómero de APN hibridado a la secuencia de ácido nucleico que contiene al nucleótido a caracterizar, puede hibridarse de forma tal que, mientras esté cadena arriba en relación al nucleótido a

caracterizar, el resto terminal (o N-terminal) del dímero de APN se sitúa adyacente a un nucleótido que a su vez está situado inmediatamente adyacente al nucleótido a caracterizar. De esta manera, para hibridar correctamente con la hebra de ácido nucleico que contiene al nucleótido a caracterizar, el dímero de APN tiene que contener dos nucleobases complementarias, una complementaria al nucleótido a caracterizar y la otra complementaria al nucleótido inmediatamente cadena arriba del mismo.

Un experto apreciará que cuando se usan trímeros de APN en vez de monómeros o dímeros, el oligómero de APN puede hibridarse a la secuencia de ácido nucleico que comprende al nucleótido a caracterizar, de tal forma que hay dos nucleótidos de las secuencias de ácido nucleico entre el extremo N-terminal del oligómero de APN y el nucleótido a secuenciar. De esta manera, un trímero de APN que se integre correctamente debe poseer tres nucleobases complementarias; dos complementarias a los nucleótidos inmediatamente cadena arriba respecto al nucleótido a caracterizar y una complementaria al nucleótido a caracterizar.

El método puede comprender además la etapa de atrapar el monómero de APN modificado (o dímero/trímero de APN) que comprende la nucleobase complementaria a la nucleobase de ácido nucleico, por ejemplo, deteniendo la reacción reversible. Por ejemplo, las especies de imina pueden reducirse en un proceso conocido como aminación reductora usando agentes reductores tales como cianoborohidruro de sodio.

Ventajosamente, puesto que cada uno de los monómeros de APN modificados (o dímeros o trímeros de APN) está marcado con al menos un marcador detectable que es distinguible de los marcadores detectables en otros tipos de monómeros de APN (dímeros o trímeros), puede lograrse fácilmente la detección del monómero específico (dímero o trímero) que se ha integrado.

Un experto entenderá fácilmente que en los casos en los que se usan dímeros de APN, hay 16 posibles combinaciones de los cuatro nucleótidos estándar (A, G, T y C) que deben tenerse en consideración. Así, cuando se usan dímeros de APN, los métodos descritos en el presente documento pueden requerir la adición de todos los 16 dímeros posibles de APN. De forma similar, cuando se usan trímeros de APN, hay 64 combinaciones posibles de los cuatro nucleótidos estándar, así, cuando se usan trímeros de APN, los métodos descritos en el presente documento pueden requerir la adición de todos los 64 trímeros de APN.

Como se indica, se ha de entender que mientras que cada uno de los métodos descritos anteriormente se ha descrito en referencia a la caracterización de una nucleobase/nucleótido particular de una secuencia de ácido nucleico, los métodos pueden también permitir al usuario caracterizar un SNP presente en una muestra de ácido nucleico. Por ejemplo, si se sabe que hay una SNP en un locus determinado en un gen, puede ser posible, mediante el diseño de oligómeros de APN que hibridan con cualquiera de los lados del locus del SNP o inmediatamente cadena arriba del locus del SNP (como se describe anteriormente) puede ser posible caracterizar la SNP (es decir, identificar qué alelo de SNP particular está presente en el locus). Dichos métodos pueden ser particularmente útiles para detectar mutaciones asociadas con trastornos genéticos particulares.

También se describe en el presente documento un método para secuenciar un ácido nucleico, comprendiendo dicho método las etapas de:

- (a) hibridar una secuencia de ácido nucleico con un oligómero de APN capaz de hibridar con una región de dicha secuencia de ácido nucleico y que tiene en su posición N-terminal un grupo funcional capaz de interactuar con un monómero de APN de acuerdo con el segundo aspecto de la invención, en condiciones que permitan la formación de un dúplex de ácido nucleico/APN; y
- (b) poner en contacto el dúplex de ácido nucleico/APN con monómeros de APN de acuerdo con el segundo aspecto de la invención;

en el que los monómeros de APN que se integran en el dúplex de ácido nucleico/APN son complementarios a una nucleobase de la secuencia de ácido nucleico que puede identificarse posteriormente gracias al marcador detectable del monómero de APN.

Puesto que cada uno de los monómeros de APN puede estar marcado con un marcador detectable que es distinguible de los marcadores detectables de los monómeros de APN que comprenden otras nucleobases, detectando la etiqueta del monómero de APN que se ha integrado en o con el dúplex de ácido nucleico/APN, puede ser posible secuenciar un ácido nucleico.

Además, y tal como se ha descrito anteriormente, el método divulgado anteriormente puede, en vez de usar los monómeros de APN en la etapa (b), usar los dímeros y/o trímeros de APN proporcionados por esta invención.

Preferentemente, cada uno de los monómeros de APN (o dímeros o trímeros) que se ponen en contacto con el dúplex de ácido nucleico/APN puede comprender en su posición N-terminal un grupo bloqueante (como se describe anteriormente). Dichos monómeros de APN (o dímeros o trímeros) se citan en lo sucesivo como "monómeros (dímeros/trímeros) de APN bloqueados". Los métodos que usan monómeros (dímeros/trímeros) de APN bloqueados son particularmente ventajosos porque solo un monómero de APN puede integrarse con un dúplex de ácido

nucleico/APN cada vez. Para poder integrar posteriormente monómeros (dímeros o trímeros) adicionales de APN, el grupo bloqueante del monómero (dímero o trímero) de APN integrado debe eliminarse previamente (opcionalmente junto con cualquier marcador detectable). De esta manera antes de la adición de un monómero (dímero o trímero: bloqueado o no) de APN adicional, el marcador detectable de la nucleobase modificada integrada puede identificarse y determinarse la correspondiente nucleobase del ácido nucleico.

Las técnicas que pueden usarse para eliminar el grupo protector (junto a cualquier marcador presente) son conocidas por cualquier experto en la materia y pueden incluir, por ejemplo, escisión basada en bases, escisión basada en ácidos, reducción de disulfuros, reacciones catalíticas basadas en metales, reacciones de escisión basadas en luz (Green, Wiley-Interscience, Nueva York, 1999).

La eliminación del grupo protector y de cualquier marcador presente en la nucleobase modificada integrada puede exponer o producir un resto (tal como un grupo de amina libre, aldehído/cetona, hidrazida, hidrazina, alcoxiamina, alcoholes, dioles y/o ácidos borónicos) capaz de reaccionar reversiblemente con otro monómero de APN. Además, el método puede comprender la etapa adicional de atrapar el monómero integrado de APN para evitar reacciones reversibles adicionales. Por ejemplo, reduciendo las especies de imina con cianoborohidruro de sodio y posterior captura de la amina secundaria generada mediante, por ejemplo, una etapa de amidación usando cloruro de acetilo. Puesto que cada nucleobase modificada a integrar con el dúplex de ácido nucleico/APN se une a una nucleobase complementaria del ácido nucleico, los métodos descritos en el presente documento hacen posible determinar secuencialmente la secuencia del ácido nucleico.

Cada uno de los métodos descritos en el presente documento ofrece muchas ventajas sobre la técnica anterior. En particular, no se necesitan enzimas y no son necesarios trifosfatos marcados, sólo se usan monómeros dímeros o trímeros de APN marcados. Además, no hay necesidad de usar las estrategias actuales de unión de fluoróforos dado que las estrategias existentes, como por ejemplo el uso de alquinos, sólo se usan debido a requerimientos de enzimas (Q. Meng *et al*, *J. Org. Chem.*, 2006, 71, 3248-3252).

Un experto en la materia apreciará que cualquiera de los métodos divulgados en el presente documento y proporcionados por esta invención, puede requerir el uso de la tecnología de micromatrices. Por ejemplo, el ácido nucleico que comprende al nucleótido a caracterizar puede inmovilizarse sobre algún tipo de sustrato adecuado usando, por ejemplo, un sistema de microimpresión o similar. De esta manera, puede inmovilizarse un gran número de ácidos nucleicos diferentes sobre sustratos en zonas concretas.

En otras realizaciones, el ácido nucleico que comprende los nucleótidos a caracterizar puede mantenerse en solución de manera que los otros componentes (por ejemplo, los oligómeros de APN, las nucleobases modificadas, etc.) también se añadirían en solución.

En realizaciones adicionales, los ácidos nucleicos que comprenden los nucleótidos a caracterizar pueden inmovilizarse sobre sustratos tales como, por ejemplo, superficies de oro adecuadas para el análisis por espectrometría de masas.

En el presente documento se divulga un kit que comprende los reactivos y componentes necesarios para los métodos divulgados en el presente documento y los métodos de la invención. En una realización, el kit puede proporcionar reactivos y componentes útiles en métodos para la caracterización de un nucleótido de un ácido nucleico y/o para secuenciar un ácido nucleico, comprendiendo dicho kit componentes seleccionados del grupo que consiste en:

- (a) un oligómero de ácido peptidonucleico (APN) capaz de hibridar con una región de un ácido nucleico y que carece de una nucleobase complementaria a la del nucleótido a caracterizar;
- (b) nucleobases modificadas de acuerdo con la presente invención;
- (c) monómeros, dímeros, trímeros y/u oligómeros de APN tales como los descritos en el presente documento; y
- (d) un oligómero de APN complementario a una porción de la secuencia del ácido nucleico cadena arriba en relación al nucleótido a caracterizar y que además comprende un grupo funcional capaz de reacciones covalentes reversibles.

Un experto en la materia apreciará que, mientras que la presente invención se ha descrito en relación a los APN miméticos de ADN, también podrían usarse otros compuestos miméticos de ADN siempre que permitan una incorporación dinámica de nucleobases. Dichos compuestos miméticos de ADN alternativos incluyen aquéllos divulgados por R.J. Worthington *et al*. (R. J. Worthington *et al*, *Org. Biomol. Chem.*, 2007, 5, 249-259). Además, puede ser posible usar dímeros, trímeros y/u oligómeros de ADN modificados en los métodos descritos en el presente documento. Más específicamente, aquellas etapas que requieren del uso de un oligómero de APN capaz de hibridar con una secuencia de ácido nucleico que se quiere secuenciar o que comprende un nucleótido a caracterizar pueden, en realizaciones alternativas, usar dímeros, trímeros y/u oligómeros de ADN modificados con objeto de incluir los grupos funcionales requeridos capaces de producir reacciones reversibles y/o posiciones en blanco correspondientes a los nucleótidos a caracterizar.



## Descripción detallada de la invención

La presente invención se describirá a continuación en detalle y haciendo referencia a las siguientes figuras, que muestran:

5

Figura 1: Estructuras de APN y ADN mostrando hibridación APN-ADN.

Figura 2: Análisis de SNP basado en dinámica. Una secuencia de APN complementaria que carece de una base enfrentada al sitio donde se localiza un SNP se hibrida con una secuencia de ácido nucleico que comprende un SNP, para formar un dúplex de ácido nucleico/APN. Mediante acoplamiento dinámico, la base complementaria al nucleótido de la SNP se integra con el dúplex de ácido nucleico/APN. Cada base modificada puede marcarse con marcador específico que puede ser un fluoróforo (véase la Figura 1).

10

Figura 3: Ilustración del análisis de SNP basado en dinámica mostrado en la Figura 2.

Figura 4: Método alternativo para el análisis de SNP basado en dinámica - los oligómeros de ADN se hibridan con oligómeros de APN complementarios que comprenden grupos amino libres en el extremo N-terminal que permiten el acoplamiento dinámico de la base complementaria a la nucleobase del SNP.

15

Figura 5: Ilustración del análisis de SNP basado en dinámica mostrado en la Figura 4.

Figura 6: Ilustración de la secuenciación de ADN basada en dinámica. (a) Oligómeros de APN, que pueden acoplarse a superficies o en solución, que contienen un grupo amino libre en el extremo N-terminal. (b) Los moldes de ADN hibridan con sus correspondientes "cebadores de APN". (c) Adición de los aldehídos de los cuatro monómeros de APN protegidos en N- (d) acoplamiento dinámico de la nucleobase correspondiente (e) eliminación de ambos grupos protectores y marcadores "fijando" la hebra creciente tal vez con un proceso de reducción y (f) repetición.

20

Figura 7: Ilustración de la secuenciación de ADN basada en dinámica. (a) Oligómeros de ADN, que pueden acoplarse a superficies o en solución. (b) "Cebadores de APN" que contienen un grupo amino libre en posición N-terminal hibridan con su correspondiente molde de ADN. (c) Adición de los cuatro aldehídos de monómeros de APN. (d) Acoplamiento dinámico de la correspondiente nucleobase. (e) Eliminación de ambos grupos protectores y marcadores y "fijado" de la hebra creciente mediante un proceso de reducción y (f) repetición.

25

Figura 8: Secuenciación de ADN basada en dinámica de las Figuras 6 y 7.

Figura 9: Síntesis de aldehídos de las bases (i) Alquilación en N de una nucleobase usando bromoalquil acetal (ii) marcaje de las nucleobases mediante la reacción de Sonogashira (iii) desprotección del grupo protector acetal.

30

Figura 10: Síntesis de aldehídos de APN protegidos/marcados: (i) protección de los primarios (ii)  $C^{Mmt}CH_2COOH$ ,  $A^{Mmt}CH_2COOH$  o  $G^{Mmt}CH_2COOH$  o  $T-CH_2COOH$ , DCC, HOBt; NB = Nucleobases.

Figura 11: Representación esquemática de la disposición usada para imprimir 8 oligómeros de ADN (Tabla 1) y un indicador marcado con fluorescencia.

35

Figura 12: Imagen en el canal FITC de una lámina que contiene 8 oligómeros de ADN (Tabla 1) y un marcador fluorescente de ADN hibridado con APN **13** que contiene una posición en blanco. Solo se detecta el marcador fluorescente de ADN.

40

Figura 13: Imagen en el canal Cy5 de una lámina que contiene 8 oligómeros de ADN (Tabla 1) y un marcador fluorescente de ADN hibridado con APN **13** que contiene una posición en blanco, solo los dúplex de APN-ADN con una orientación antiparalela fueron capaces de hibridar.

40

Figura 14: (A) Muestra la disposición de amino oligonucleótidos modificados (Tabla 2) impresos sobre láminas de alhído usando un robot Microdrop equipado con una pipeta piezoeléctrica (patrón 3x8). (B)

Representación esquemática de dúplex antiparalelos formados por APN **13** y oligonucleótidos de ADN mostrados en la Tabla 2. (C) Imagen fluorescente (canal Cy5) que muestra APN **13** hibridado con oligo 1 y 2 (Tabla 2). (D) Imagen en el canal FITC que muestra la incorporación dinámica del aldehído de citosina **10** marcado con fluoresceína después de la incubación de matrices que contienen dúplex de ADN-APN **13** con aldehídos **9** y **10**. La señal fluorescente se detectó solo donde se imprimió el oligo 2 G-antipar (Tabla 2, Figura 14A y 14B). (E) Imagen adicional en el canal FITC que muestra el resultado de una aproximación in situ y la

45

incorporación dinámica de aldehído de citosina marcado con fluoresceína. La señal de fluoresceína se detectó solo cuando se imprimió el oligo 2 G-antipar (Tabla 2, Figura 14A y 14B).

50

Figura 15: (A) Síntesis de dímeros de aldehído para secuenciación dinámica donde la segunda nucleobase se define por un pigmento. (B) Un dímero de APN en el que la primera nucleobase se identifica por un pigmento.

50

Figura 16: (A) Método para la caracterización de nucleótido usando dímeros de APN en los que la segunda base del dímero se marca con un marcador detectable. (B) Método alternativo para caracterización de nucleótido utilizando dímeros de APN en los que la primera nucleobase se marca con un marcador detectable.

55

Figura 17: Diagrama esquemático que muestra un método para el análisis de SNP mediante espectrometría de masas.

55

Figura 18: Muestra la estructura general de nucleobases modificadas que se usan en el análisis de SNP mediante espectrometría de masas

60

Figura 19: Muestra la estructura general de nucleobases modificadas que pueden usarse en métodos de secuenciación de ácidos nucleicos basados en espectrometría de masas.

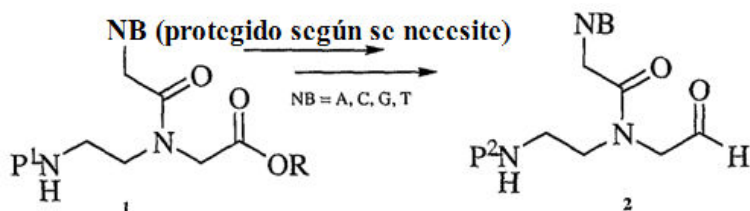
## Métodos

### Síntesis de las bases marcadas, bloques de construcción y cebadores marcados.

5 (i). Monómeros de APN-aldehído y bases de aldehído. Se prepararon como se muestra en las Figuras 9 y 10. Este método es aplicable a muchos grupos protectores, incluyendo el grupo Dde, el grupo Fmoc, el grupo protector disociable por tiol (Ardec (arilditioetiloxicarbonilo)), los grupos protectores fotodisociables (basados en nitroveratilo) así como fluoróforos.

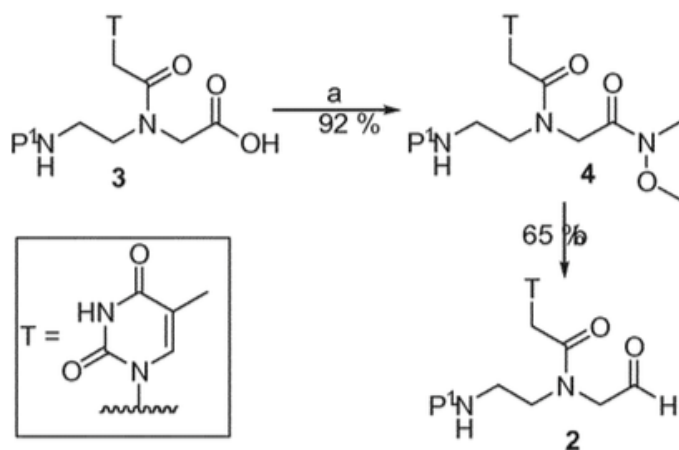
10 Los métodos detallados anteriormente pueden adaptarse para producir un material marcado con fluorescencia mediante el acoplamiento clásico de Sonogashira de derivados de bromo- y yodo-pirimidina y purina a propargilaminas seguido de marcaje fluorescente con varios fluoróforos (véase esquema 3). Este enfoque puede aplicarse también a la síntesis de las "bases de aldehído" marcadas con fluorescencia (Figura 6). La elección del fluoróforo estará dictada por la necesidad de permitir la detección individual de las bases.

15 **Síntesis de aldehídos de APN protegidos y etiquetados:** Los aldehídos de APN **2** se prepararon a partir de ácidos carboxílicos de APN, ésteres de APN o alcoholes de APN siguiendo química convencional (Esquema 1). **1** se preparó de acuerdo a métodos publicados (L. Bialy *et al. Tetrahedron* 2005, *61*, 8295-8305).



20 **Esquema 1.** Síntesis de la diana.

A modo de ejemplo, **3** se transformó en una amida de Weinreb **4** (Esquema 2), después se redujo para obtener el aldehído diana **2**. Para evitar una reducción excesiva de **4**, se utilizó el agente reductor más suave hidruro de tri-*tert*-butoxialuminio y litio (LiAlH(O-*t*-Bu)<sub>3</sub>) en lugar de LiAlH<sub>4</sub> (M. Paris *et al. Tetrahedron Lett.* 1998, *39*, 1341-1344).



30 **Esquema 2.** Síntesis del aldehído diana mediante reducción de una amida de Weinreb: (a) MeONHMe.HCl, EDC.HCl, HOBT.H<sub>2</sub>O, Et<sub>3</sub>N, DMF, temperatura ambiente; (b) LiAlH(O-*t*-Bu)<sub>3</sub>, THF, temperatura ambiente. 65 % de rendimiento (HPLC).

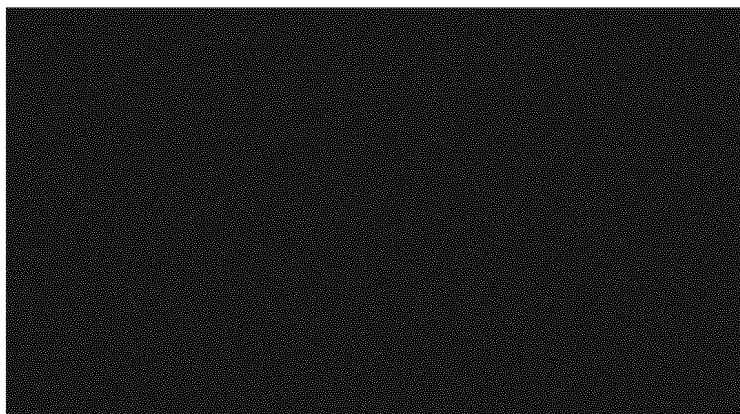
35 Los aldehídos se purificaron siguiendo la estrategia de captura y liberación sobre una resina de treonilo (D.R. Liu *et al. J. Am. Chem. Soc.* 2003, *125*, 13924-13925).

40 **(a) Preparación de la resina:** El NovaGel HL de aminometilo se hinchó con DMF durante aproximadamente 10 min. Mientras tanto, se añadió DIPEA a una solución de Fmoc-Thr(*t*-Bu)-OH y TBTU en DMF y la mezcla de reacción se agitó durante 5 min. La resina hinchada se filtró entonces y se añadió la solución de treonina activada y protegida. La suspensión resultante se agitó a temperatura ambiente durante 2 h. La resina se

separó entonces de la mezcla de reacción mediante filtración y se lavó con DMF (5 x), THF (5 x) y DCM (5 x), y se secó a vacío a 40 °C durante la noche. La resina se hinchó en DMF durante aproximadamente 10 min y se filtró, entonces se agitó con piperidina al 20 % (v/v) en DMF x2. La resina se agitó entonces con TFA (ácido trifluoroacético) al 80 % (v/v) en DCM, se filtró, se lavó con DCM (1 x) y se agitó de nuevo con TFA al 80 % (v/v) en DCM. La resina se filtró, se lavó con DCM (5 x) y se secó a vacío a 40 °C.

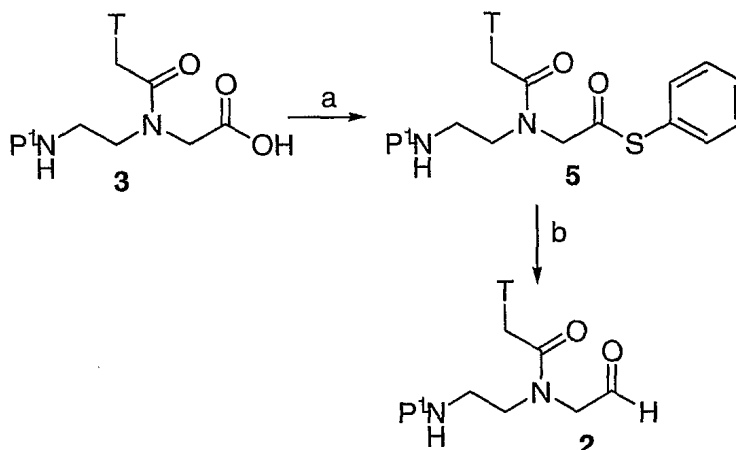
**(b) Purificación de aldehído mediante captura y liberación:**

**Captura:** Se añadió una solución de aldehído crudo a la resina de captura de treonilo desprotegida. La mezcla se agitó a temperatura ambiente durante 1 h, entonces la resina se filtró y lavó. **Liberación:** La resina se agitó y lavó con una mezcla de AcOH/H<sub>2</sub>O/DCM/MeOH (10/5/5/80, 2 ml x 5) durante 20 min y los lavados se concentraron a vacío para obtener el aldehído.



**Esquema 3.** La condensación del aldehído impuro con una resina modificada de treonina produce una oxazolidina unida al soporte. Después de lavar, la ruptura genera el aldehído puro.

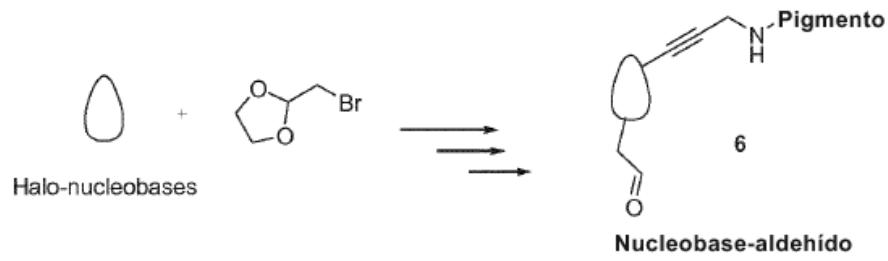
1. Una ruta alternativa para obtener el aldehído **2** es a través del S-bencil tioéster **5** como se muestra en el Esquema 4 (P.T. Ho, *et al J. Org. Chem.* 1993, 58, 2313-2316)



**Esquema 4.** Síntesis de aldehídos diana a través del S-bencil tioéster: (a) bencil mercaptano, DMAP, DCC, DMF temperatura ambiente, (b) trietilsilano, Pd/C, THF, temperatura ambiente, 20 h, 95 % de rendimiento (HPLC).

Rutas alternativas incluyen la reducción del metil éster al correspondiente alcohol primario y la subsecuente oxidación o síntesis directa del alcohol de APN y oxidación.

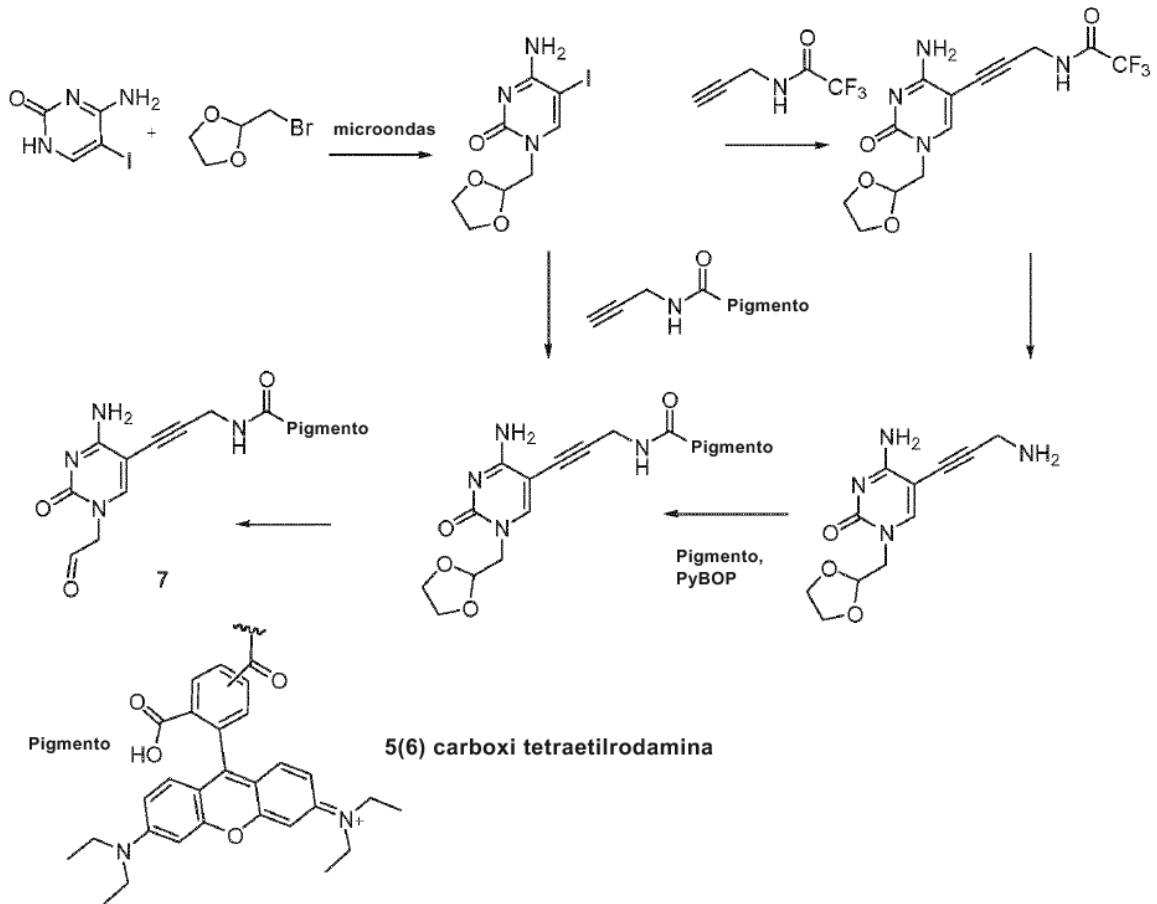
**Síntesis de aldehídos de nucleobases etiquetados para análisis de SNP (Método 1):** Los aldehídos de nucleobases **6** se prepararon a partir de halo-nucleobases disponibles comercialmente mediante alquilación con 2-(bromometil)-1,3-dioxolano (Esquema 5) bajo irradiación de microondas seguida por la reacción de Sonogashira con propargilamina protegida con Tfa, desprotección del grupo protector Tfa y acoplamiento con un pigmento derivatizado con ácido carboxílico.



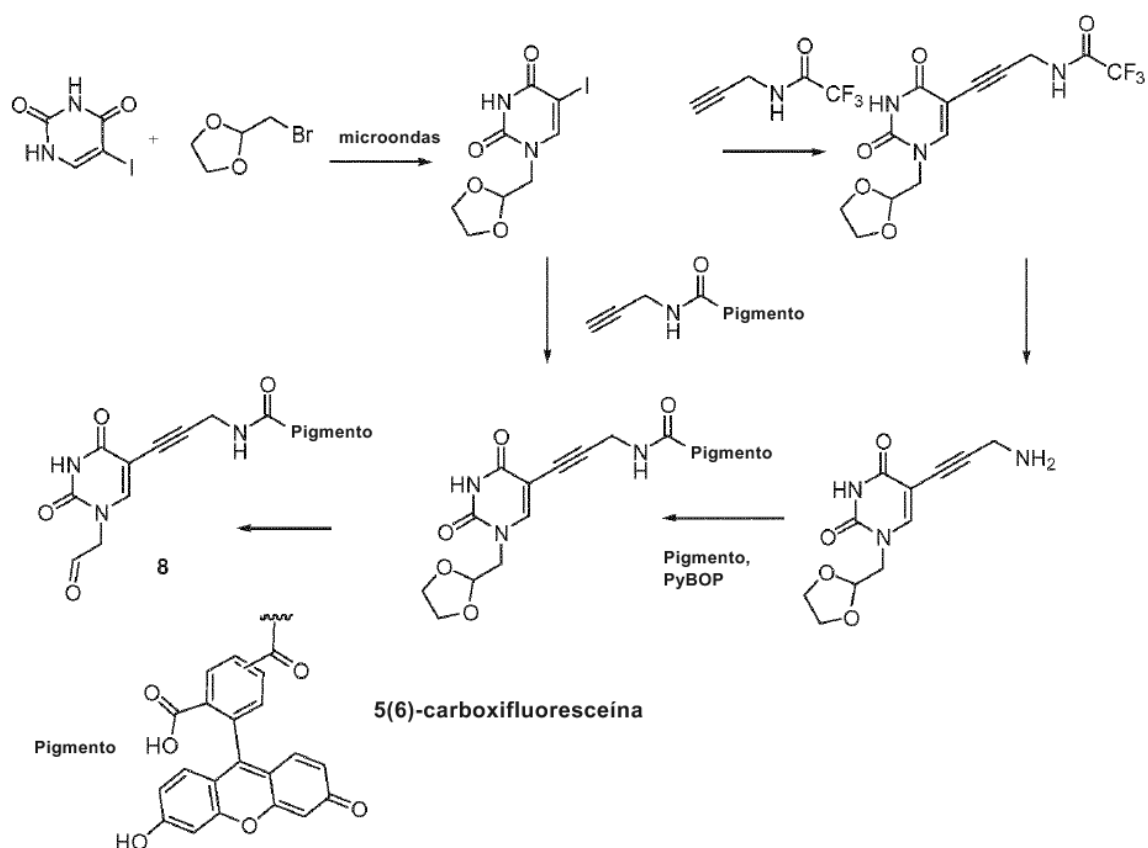
**Esquema 5.** Estrategia general para sintetizar un aldehído de nucleobase.

A modo de ejemplo, los derivados de timina y citosina se sintetizaron como se describe en los Esquemas 6 y 7.

5



**Esquema 6.** Síntesis del aldehído de citosina marcado con rodamina 7



**Esquema 7.** Síntesis del aldehído de timina marcado con fluoresceína **8**

#### Alquilación de nucleobases con acetales

5

Este proceso se optimizó usando radiación de microondas a 100 °C durante 30 min en THF. El uso de 1,2 equiv. de 2-(bromometil)-1,3-dioxolano con  $\text{NEt}_3$  generó el producto mono-alkilado en una proporción 4:1.

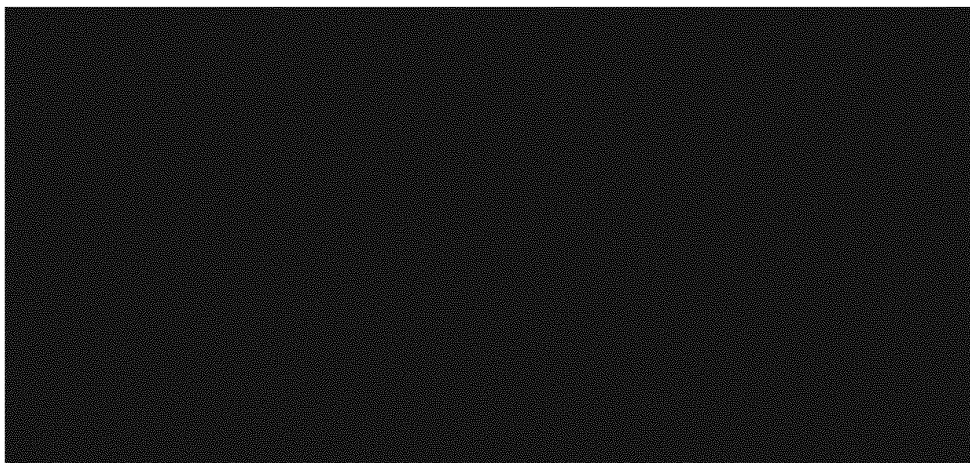
10

El marcaje de las nucleobases se consiguió a través de la reacción de acoplamiento cruzado de Sonogashira utilizando aminometilacetileno.

(a) Reacción de Sonogashira con aminometilacetileno protegido con Tfa antes de desproteger el grupo Tfa con amoniaco. PyBOP como reactivo de acoplamiento.

15

(b) La segunda ruta explorada consistió en llevar a cabo la reacción de Sonogashira con el grupo acetileno portando ya el pigmento. Esa reacción se llevó a cabo haciendo reaccionar aminometil acetileno con pigmentos activados unidos al soporte usando una resina de hidroxinitrobenzoico (Esquema 8).



**Esquema 8.** Síntesis de acetileno marcado usando pigmentos activados unidos a un soporte.

5 Finalmente, los acetales se desprotegeron usando HCl 2N en H<sub>2</sub>O para dar lugar a los productos **7** y **8**.

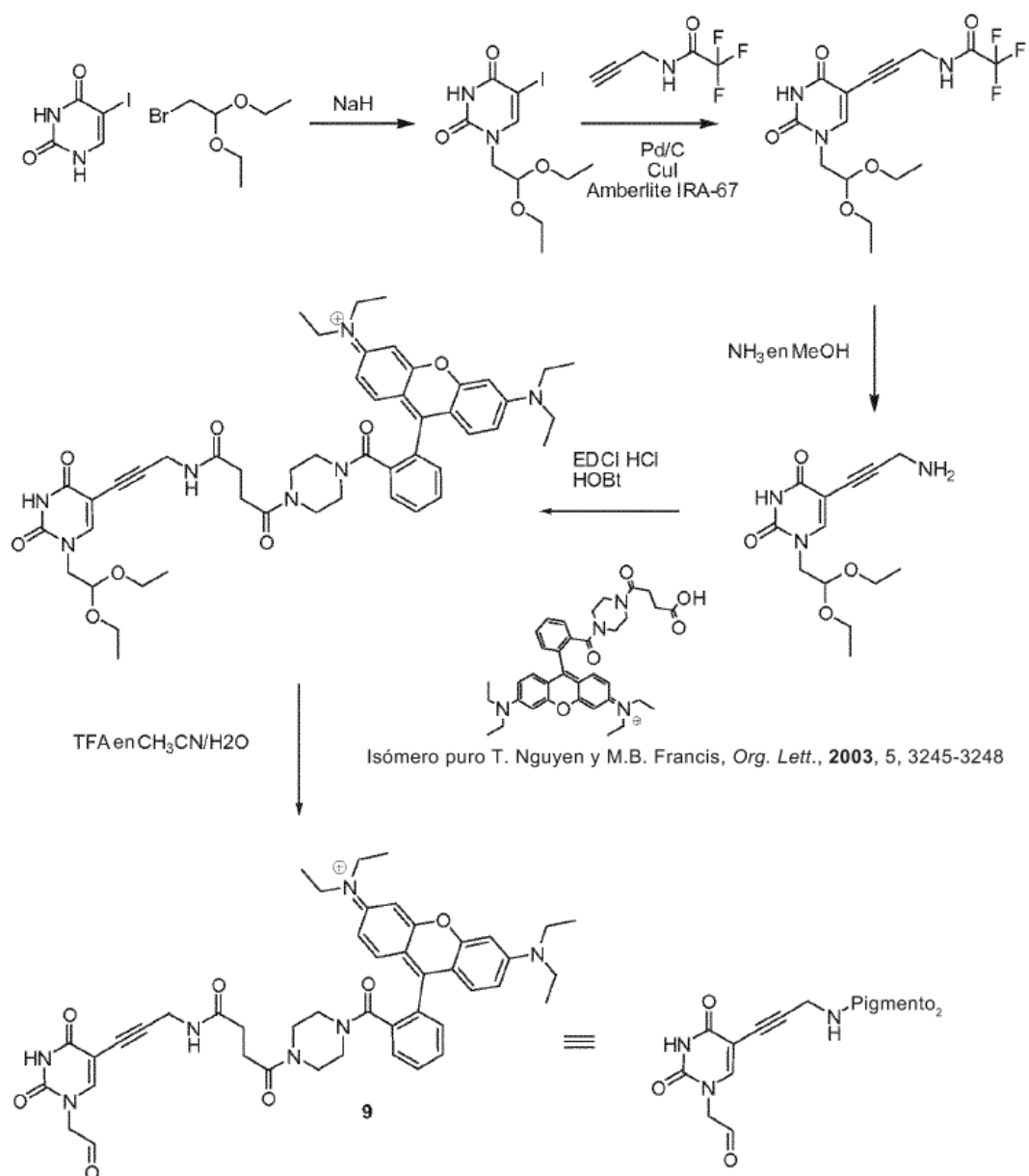
**Síntesis de aldehídos de bases marcadas con fluorescencia para análisis de SNP usando el método de "clamp"**

10 La alquilación de las nucleobases se consiguió usando un procedimiento modificado descrito por L. Christensen et al., *Nucl. Acids Res.*, 1998, 26, 2735-2739. Un equivalente de nucleobase halogenada se disolvió en DMF con NaH (1.2 equivalentes) y se agitó durante 30 min a temperatura ambiente. Entonces, se añadieron 1,12 equivalentes de dietilacetal de bromoacetaldehído y la solución se agitó bajo radiación de microondas durante 30 min a 130 °C.

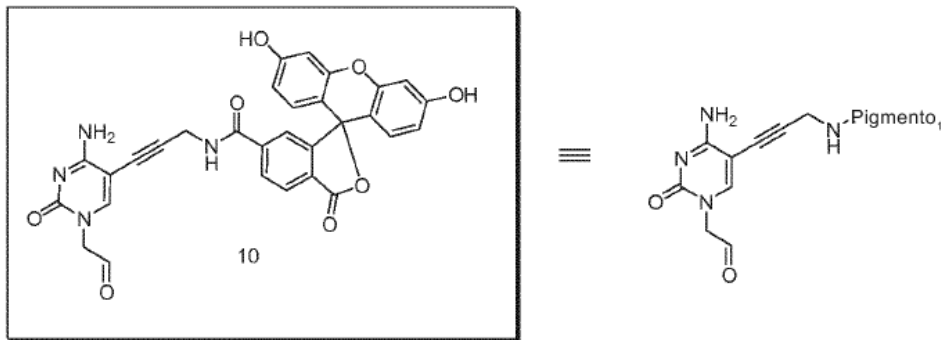
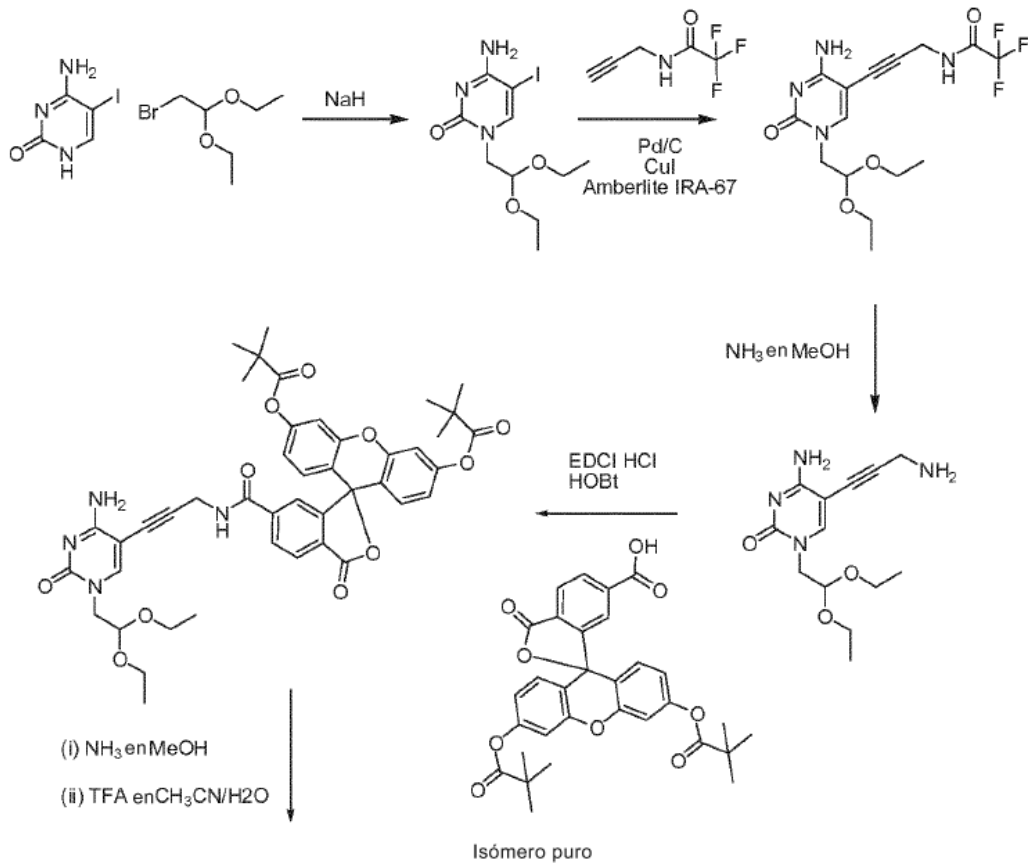
15 Las halobases alquiladas se sometieron a acoplamiento cruzado de Sonogashira siguiendo el procedimiento descrito en N.K. Garg *et al. Chem. Commun.*, 2005, 4551-4553 y usando aminometilacetileno protegido con Tfa. La desprotección del grupo Tfa con amoníaco en MeOH dio lugar a una amina primaria libre que se usó para acoplar los pigmentos conteniendo grupos carboxílicos. El acoplamiento de la amida se llevó a cabo usando los agentes acoplantes HOBt/EDCI HCl.

20 La desprotección del acetal se llevó a cabo tratando con TFA/12,5 %H<sub>2</sub>O/12,5 %CH<sub>3</sub>CN al 75 % durante 24 horas a temperatura ambiente. Como alternativa, calentando a 60 °C bajo radiación de microondas durante 2 horas. Los acetales se purificaron mediante RP-HPLC. La cromatografía RP-HPLC se llevó a cabo en un sistema HP 1100 equipado con una columna de fase reversa Phenomenex Prodigy C18 (250 mm x 10 mm x 5 mm) con un flujo de 2,5 ml/min y eluyendo con (A) TFA al 0,1 % en H<sub>2</sub>O y (B) TFA al 0,042 % en acetonitrilo, con un periodo isocrático inicial de 4 min a 0 % (B) seguido de un gradiente de 0-50 % (B) durante 25 min y 50-100 % (B) durante 10 min, manteniéndose a 100 % (B) durante 5 min. Los análisis ESI-MS se llevaron a cabo en un espectrómetro de masas de cuadrupolar (QMS) modelo LC/MSD Series 1100 de Agilent Technologies usando el modo de ionización mediante electropulverización (ESI). Los aldehídos finales se identificaron por NMR y LC-MS (ESI).

25 A modo de ejemplos, el aldehído de timina marcado con rodamina (Esquema 9) y el aldehído de citosina marcado con fluoresceína (Esquema 10) se prepararon como se describe arriba. Las nucleobases modificadas de esta manera pueden usarse en cualquiera de los métodos descritos en el presente documento y, en particular, en los métodos para caracterización y/o análisis de SNP.



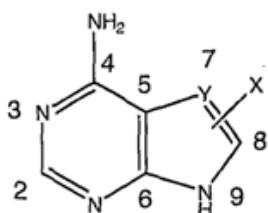
**Esquema 9** - Síntesis del aldehído de timina marcado con rodamina **9** para análisis de SNP usando el método de "clamp"



**Esquema 10** - Síntesis del **aldehído de citosina marcado con fluoresceína 10** para el análisis de SNP usando el método de "clamp"

- 5 Un experto en este campo apreciará que los derivados de adenina y guanina usando pigmentos BODIPY se prepararán de forma similar. En esos casos las halonucleobases son las siguientes:

**Adenina**

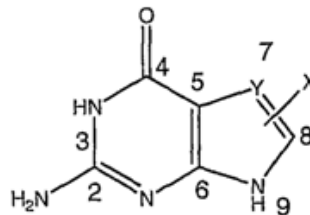


1

X = Br en posición 7 u 8

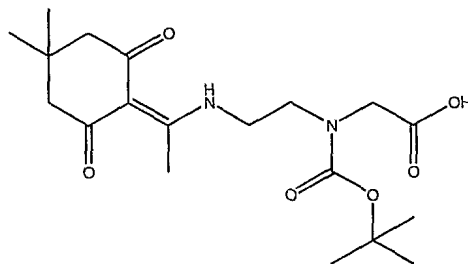
Y = C cuando X en posición 7; N cuando X en posición 8

**Guanina**



1



**Síntesis de N-2(Dde-amino)etil-N-boc-glicina 11**

5 A una solución de éster de metil N-2(Dde-amino)etil-glicina (1 mmol) (L. Bialy *et al.*, *Tetrahedron*, **2005**, *61*, 8295-8305) en THF seco (10 mL, 0,1 M) se añadió Boc<sub>2</sub>O (1,1 mmol) y trietilmaina (1,1 mmol) y la reacción se agitó durante 5 horas, monitorizándose por TLC. Después de eliminar el disolvente el producto en bruto se disolvió en DCM y se lavó con NaHCO<sub>3</sub>, KHSO<sub>4</sub> y salmuera. La fase orgánica se secó sobre NaSO<sub>4</sub> anhidro y se concentró para dar lugar a un sólido amarillo. Sin ninguna purificación adicional el producto crudo se disolvió en MeOH y se añadió una solución de Cs<sub>2</sub>CO<sub>3</sub> 2 M en agua. Después de 1,5 h la reacción se acidificó a pH 3 con KHSO<sub>4</sub>. El ácido se precipitó, se filtró y se secó para dar lugar al ácido **11** en forma de sólido blanco.

**Síntesis de los oligómeros de APN 12 y 13**

15 Los oligómeros de APN **12** (H<sub>2</sub>N-TACTACATC\_CTTCC-CONH<sub>2</sub>) y **13** (Cy5COHN-TACTACATC\_CTTCC-CONH<sub>2</sub>) = monómero blanco desprotegido de boc **9** se sintetizaron usando monómeros protegidos con Dde (Bradley *et al*, *Tetrahedron*, 2005, *61*, 8295-8305) en fase sólida (J.J. Diaz-Mochon *et al*, *Org. Lett* 2004, *6*, 1127-1129). Para insertar el monómero blanco se usó N-2(Dde-amino)etil-N-boc-glicina **11**.  
HPLC y MALDI-TOF  
20 APN **12** (MALDI-TOF; masa calculada: 3780, masa observada: 3781 (M+1).  
APN **13** (MALDI-TOF; masa calculada: 4244, masa observada: 4246 (M+1).

**Ensayo basado en matrices en superficies**

25 8 oligonucleótidos modificados con amino (Tabla 1) se imprimieron por contacto en láminas Code-link (Amersham) para el análisis de SNP. Esos oligos se diseñaron para tener una orientación paralela (extremo C-terminal del APN encarando el extremo 3'-terminal del ADN) cuando se hibridaron con APN **13** o antiparalela (extremo N-terminal del APN encarando el extremo 3'-terminal del ADN).

**Tabla 1.** Oligómeros de ADN para el análisis dinámico de SNP (método 1)

A	A-antipar	TTT TTT GGA AGA GAT GTA GTA
B	G-antipar	TTT TTT GGA AGG GAT GTA GTA
C	C-antipar	TTT TTT GGA AGC GAT GTA GTA
D	T-antipar	TTT TTT GGA AGT GAT GTA GTA
E	A-par	TTT TTT ATG ATG TAG AGA AGG
F	C-par	TTT TTT ATG ATG TAG CGA AGG
G	G-par	TTT TTT ATG ATG TAG GGA AGG
H	T-par	TTT TTT ATG ATG TAG TGA AGG

35 Las superficies se imprimieron usando un Genetix Qmini Arrayer y puntas sólidas. Se utilizó un oligo de ADN marcado con FITC como marcador y se usó el siguiente diseño, como se muestra en la Figura 11.

40 Una solución 2 μM de APN 13 en tampón HybGen (Genetix) se hibridó sobre las láminas usando una estación de hibridación Hyb4 (de 65 °C a 40 °C en 6 h y entonces a 40 °C durante 2 h). Después de lavarla, la lámina se escaneó usando un escáner Lavisión Biotech con los filtros ajustados para la detección de FITC y de Cy5 (Figuras 12 y 13). La Figura 12 muestra el canal de FITC y sólo se detecta el marcador; en la Figura 13 (canal de Cy5), sólo los oligos con una orientación antiparalela fueron capaces de hibridar el APN **13** modificado.

**Análisis de SNP**

45 Los oligonucleótidos modificados con aminos (Tabla 2) fueron impresos mediante inyección de tinta sobre láminas de aldehído (Genetix) para el análisis de SNP. Esos oligos se diseñaron para hibridar siguiendo una orientación antiparalela (el extremo N-terminal del APN encarando el extremo 3'-terminal del ADN) cuando hibrida con APN **13**.

**Tabla 2.** Oligómeros de ADN para el análisis dinámico de SNP

A	A-antipar Oligo 1	TTT TTT GGA AGA GAT GTA GTA
B	G-antipar Oligo 2	TTT TTT GGA AGG GAT GTA GTA

Las superficies se imprimieron usando un robot Microdrop y una pipeta piezoeléctrica. La Figura 14A muestra el diseño empleado.

5

Una solución 2  $\mu\text{M}$  de APN **13** en tampón HybGen (Genetix) se hibridó sobre las láminas usando una estación de hibridación Hyb4 de 55  $^{\circ}\text{C}$  a 30  $^{\circ}\text{C}$  en 12 h. Después de lavar, la lámina se escaneó usando un escáner Lavidon Biotech. La Figura 14B muestra el dúplex que se forma. La Figura 14C (canal Cy5) muestra que los oligos con orientación antiparalela fueron capaces de hibridar al APN modificado **13**.

10

Una vez hibridadas las superficies con matrices, los aldehídos de las bases **9** y **10** se incubaron en las superficies con matrices. La incorporación dinámica se observó cuando las matrices se incubaron con 5  $\mu\text{M}$  de cada aldehído junto con 1 mM de NaBCNH<sub>3</sub> a temperatura ambiente durante 16 h (véase Figura 14D) tanto a pH 6 (NH<sub>4</sub>OAc 0,1M) como a pH 8,5 (NaHCO<sub>3</sub> 0,2M, NaCl 0,3 M). Las imágenes obtenidas usando el canal de fluoresceína (canal de FITC) detectan oligo ADN **2** (antiparalelo) (Figura 14D). Esta señal viene del aldehído de la base portando un pigmento de fluoresceína, en este caso aldehído de citosina **10**, correspondiente al apareamiento perfecto para G.

15

También se usó un segundo enfoque: a una solución 2  $\mu\text{M}$  de APN **13** en tampón HybGen (Genetix) se añadieron los aldehídos **9** y **10** a una concentración 2  $\mu\text{M}$  junto con NaBCNH<sub>3</sub>. Esta solución se usó para hibridar una lámina conteniendo los oligos de ADN indicados en la Tabla 2. La hibridación tuvo lugar de 55  $^{\circ}\text{C}$  a 30  $^{\circ}\text{C}$  en 12 h. En estas condiciones, las imágenes obtenidas usando el canal de fluoresceína (canal FITC) detectan el oligo de ADN **2** (antiparalelo G: véase Figura 14E). Esta señal viene del aldehído de la base portando un pigmento de fluoresceína, en este caso el aldehído de citosina **10**, que corresponde al apareamiento perfecto para G.

20

## 25 **Ensayo en solución.**

### **Síntesis de los oligómeros de APN 12.**

El oligómero de APN **14** (NH<sub>2</sub>-CATTCTTCCTCT-CONH<sub>2</sub>) se sintetizó usando monómeros protegidos con Dde (L. Bialy *et al*, *Tetrahedron*, 2005, 61, 8295-8305) en fase sólida (JJ. Diaz-Mochon *et al*, *Org. Lett.* 2004, 6, 1127-1129). APN **14** (MALDI-TOF; masa calculada: 3345; masa observada: 3348 (M+1) Se usaron 2 oligonucleótidos modificados con amino complementarios al APN **14** para el análisis de ADN en solución usando espectrometría de masas y análisis en fase sólida.

30

35 **Tabla 3.** Oligómeros de ADN usados para la secuenciación dinámica y análisis de SNP (método 2)

I	C-extensión	TTTTTTAGAGGAAGAATGGGTAA
J	T- extensión	TTTTTTAGAGGAAGAATGAAGTT

A una solución de APN **14** a una concentración 1  $\mu\text{M}$  en tampón HybGen se añadió una solución de oligómero I de ADN a 1,2  $\mu\text{M}$  en tampón TE (Tabla 3). La mezcla se calentó a 65  $^{\circ}\text{C}$  y se enfrió lentamente a 40  $^{\circ}\text{C}$ . En esta etapa se realizaron diferentes modificaciones del pH antes de la adición del aldehído del monómero de APN **2**. La reacción de extensión se monitorizó mediante HPLC y espectrometría de masas usando una columna de fase reversa y tampones de amoniaco.

40

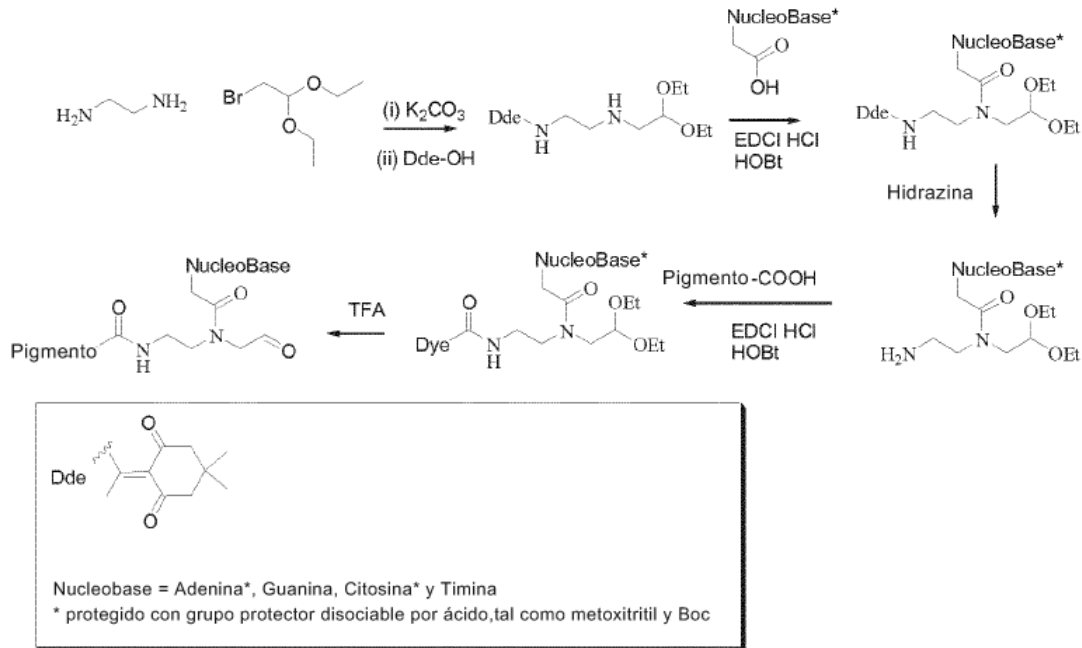
### **Síntesis de aldehídos de monómeros de APN marcados fluorescentemente para la extensión dinámica.**

45 Esos compuestos se sintetizaron siguiendo protocolos modificados desarrollados por L. Bialy *et al.*, (L. Bialy *et al.* *Tetrahedron*, 2006, 61, 8295-8305) para la síntesis de monómeros de APN. La diferencia principal es la alquilación inicial de etilendiamina con bromoacetaldehído dietil acetal usando irradiación de microondas (véase el Esquema 11).

La desprotección del grupo Dde en solución se llevó a cabo usando hidrazina y agua a temperatura ambiente durante 16 h. El acoplamiento del pigmento se realizó usando EDCI y HOBt. La desprotección final se llevó a cabo usando TFA en acetonitrilo durante 30 min. La purificación y los análisis se llevaron a cabo como se menciona anteriormente.

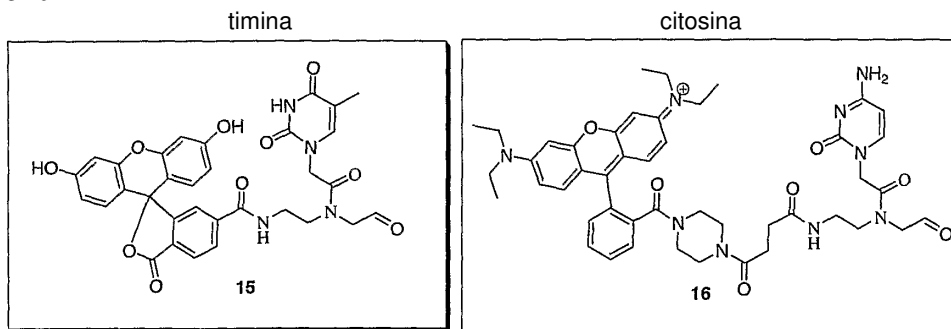
50

**Esquema 11**



5 Como ejemplos, se prepararon aldehído de APN de timina marcado con fluoresceína y aldehído de APN de citosina marcado con rodamina - véase esquema 12.

**Esquema 12**



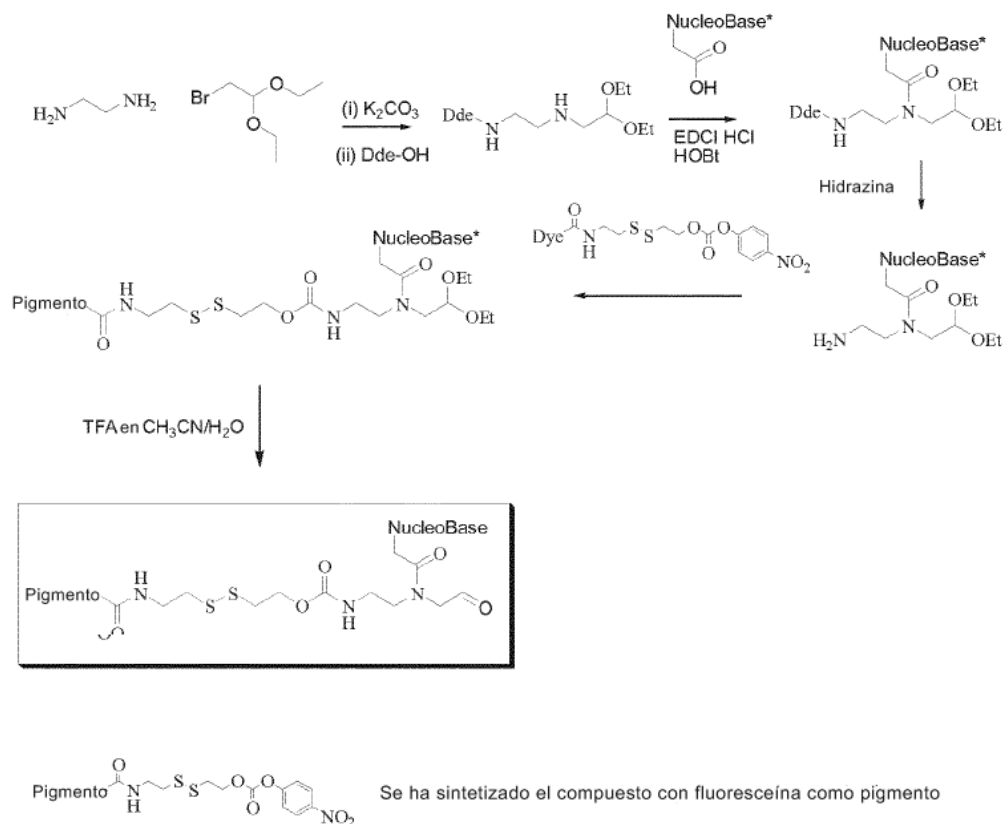
10 Un experto apreciará que los derivados de adenina y guanina usando pigmentos de BODIPY se prepararán de forma similar

**Uso de dímeros para secuenciación**

15 Los aldehídos se preparan mediante acoplamiento a un bloque de construcción de APN adicional. Ello necesita la preparación de una mezcla de 16 compuestos llevada a cabo mediante métodos de fase sólida usando una estrategia de separación y mezcla. En este caso los 4 aldehídos protegidos en N- terminal (A, T, C y G) se inmovilizan sobre un conector de hidrazina (véase A. Lee *et. al*, *J. Am. Chem. Soc.*, 1999, 121, 9907-9914) o una resina de captura de treonilo ( ). Los grupos protectores se liberan entonces y las cuatro resinas se mezclan y vuelven a separar en cuatro partes para acoplar los monómeros protegidos de APN convencionales. Después de la

20 desprotección y el marcaje usando disulfuro activado (Esquema 13) conteniendo un pigmento específico de acuerdo a la última nucleobase, la mezcla global de la resina y la muestra liberada produce 16 dímeros de APN (Figura 15A).

Esquema 13



5 La Figura 16 (A y B) detalla métodos de caracterización de SNP/nucleótidos y/o de secuenciación que utilizan los dímeros/trímeros de APN proporcionados por esta invención. Cuatro oligómeros de ADN acoplados a la lámina variando sólo en una posición (por simplicidad solo se muestran 6 bases). Una secuencia complementaria de APN simple se hibrida con la matriz (de nuevo nótese que el APN tendrá 12 bases - sólo se muestran 2 por claridad). Los 16 aldehídos de los dímeros de APN se añaden permitiendo un acoplamiento dinámico del dímero correspondiente. En este caso la segunda base del dímero se identificará gracias a un marcador detectable.

10 En una realización alternativa, los dímeros se crean de manera que la primera nucleobase se identifica mediante un marcador detectable y la segunda es aleatoria. En estos casos la primera nucleobase tiene un pigmento en el anillo mientras que el grupo protector no porta ningún pigmento. La Figura 15B muestra un ejemplo de esta forma de dímero.

15 **Esquema para la detección mediante espectrometría de masas en matrices de oro**

20 Las superficies de oro con oligos de ADN acoplados a través de monocapas autoensambladas (SAM: *self-assembled monolayers*) de oro-tiol se pueden usar para el análisis de material genético. Por ejemplo, después de la formación de SAM en superficies de oro usando oligos de ADN modificados con tioles y la hibridación de oligómeros de APN, la incorporación dinámica usando nucleobases modificadas con aldehído se puede usar para caracterizar SNPs/nucleótidos y/o para secuenciar ácidos nucleicos. Como se describe anteriormente, la base incorporada en la hebra de APN se puede detectar por MALDI-TOF (para el uso de superficies de oro y la detección de la hibridación APN-ADN véase Brandt et al., (Brandt et al, *Nucleic Acid Research*, 2003, 31, el 19).

25 Cuando se lleva a cabo el análisis de SNP, las nucleobases pueden ser las nucleobases modificadas descritas considerablemente arriba, o sea las que tienen un pigmento acoplado, nucleobases no modificadas o nucleobases modificadas para incluir una etiqueta de masa, como una etiqueta de bromuro, para producir una envoltura isotópica claramente observable (Figura 17).

30 La estructura general de las nucleobases modificadas para su uso en el análisis de SNP mediante espectrometría de masas (tal como los que usan técnicas tales como MALDI-TOF) se proporciona en la Figura 18.

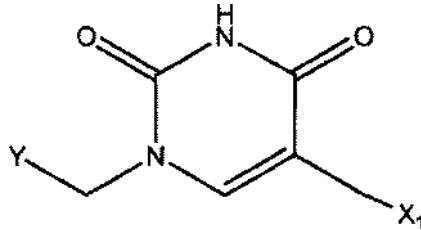
35 Un experto apreciará que un método de análisis similar "basado en la masa" puede usarse para la secuenciación y las nucleobases modificadas que pueden ser útiles en dichos métodos que se detallan en la Figura 19.

REIVINDICACIONES

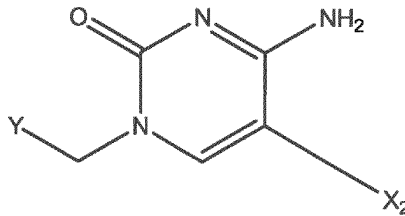
1. Un método para caracterizar un nucleótido en una secuencia de ácido nucleico, comprendiendo dicho método los siguientes pasos:

- 5 (a) poner en contacto un ácido nucleico con un oligómero de ácido peptidonucleico (APN) capaz de hibridar con una región del ácido nucleico y que carece de una base complementaria a la del nucleótido a caracterizar, para formar un dúplex de ácido nucleico/APN; y  
 10 (b) poner en contacto el dúplex de ácido nucleico/APN con una o más bases modificadas seleccionadas entre el grupo que consiste en:

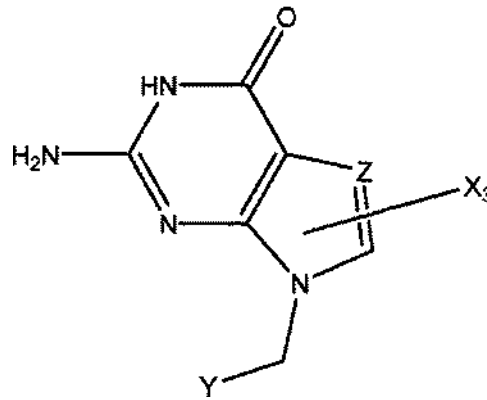
(i)



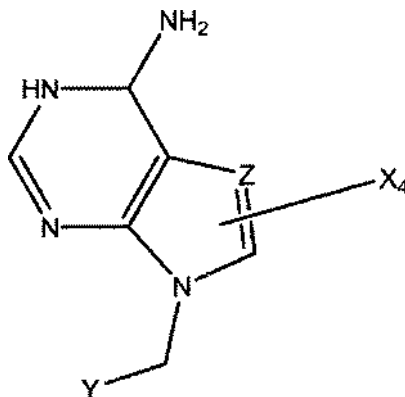
15 (ii)



(iii)



20 y (iv)

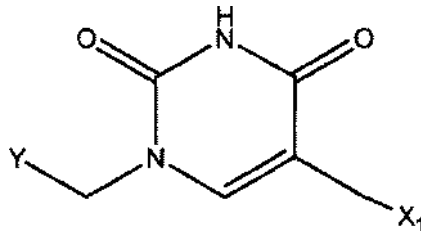


25 en donde Y es un grupo funcional capaz de llevar a cabo reacciones covalentes reversibles; X<sub>1</sub>-X<sub>4</sub> es un marcador detectable, combinaciones de espaciador y marcador o hidrógeno; y Z es carbono o nitrógeno;

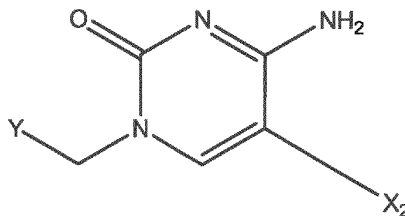
en donde el oligómero de APN incluye un grupo capaz de reaccionar reversiblemente con el grupo funcional Y, y en donde la base modificada que se integra con el dúplex de ácido nucleico/APN es complementaria a la del nucleótido a caracterizar, caracterizándose el nucleótido mediante espectrometría de masas o mediante el marcador detectable de la base modificada.

- 5 2. El método de la reivindicación 1, en el que Y se selecciona entre el grupo que consiste en aldehídos, cetonas, tioles y/o dioles.
- 10 3. El método de las reivindicaciones 1 o 2, en el que el método de caracterización de un nucleótido se usa para caracterizar polimorfismos de nucleótido simple.
4. El método de las reivindicaciones 1 o 2, en el que el método de caracterización de un nucleótido se usa para secuenciar ácidos nucleicos.
- 15 5. El método de las reivindicaciones 1-3, en el que el marcador detectable se identifica mediante espectrometría de masas o métodos de microscopía.
6. El método de las reivindicaciones 1-5, en el que el ácido nucleico y/o el APN se inmovilizan sobre o se unen de otro modo a un sustrato de soporte.
- 20 7. El método de la reivindicación 6, en el que el ácido nucleico y/o el APN se inmovilizan o se unen como una matriz o micromatriz.
8. Una base modificada seleccionada entre el grupo que consiste en:

(i)

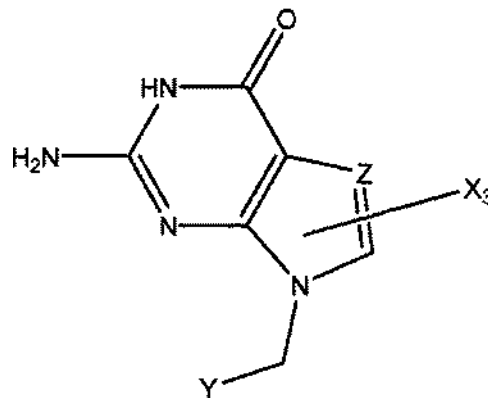


(ii)



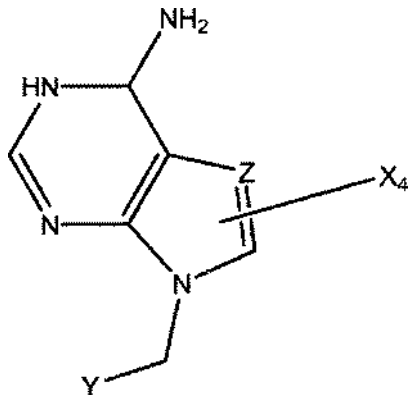
30

(iii)



y

(iv)



en las que Y es un grupo funcional capaz de producir reacciones covalentes reversibles seleccionado entre el grupo que comprende (i) un aldehído; y (ii) una cetona;

5

X<sub>1</sub>-X<sub>4</sub> es hidrógeno o comprende un marcador detectable seleccionado entre el grupo que consiste en: (i) uno o más seleccionados entre el grupo que comprende dansilo, fluoresceína, rodamina, rojo Texas, IAEDANS, pigmentos de cianina (Cy3, Cy3.5, Cy5, Cy5.5, Cy7), pigmentos Bodipy (Invitrogen), pigmentos Alexa Fluor (Invitrogen) y pigmentos SNARF; y (ii) un marcador masa;

10

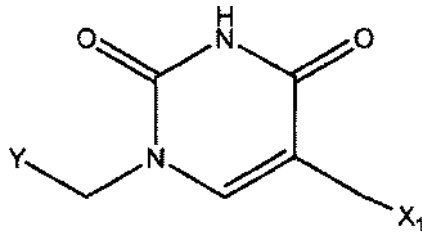
Z es carbono o nitrógeno;

con la condición de que X<sub>1</sub> no sea Br.

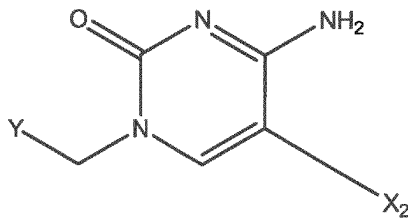
9. El uso de una o más bases/nucleobases modificadas seleccionadas entre el grupo que consiste en:

15

(i)

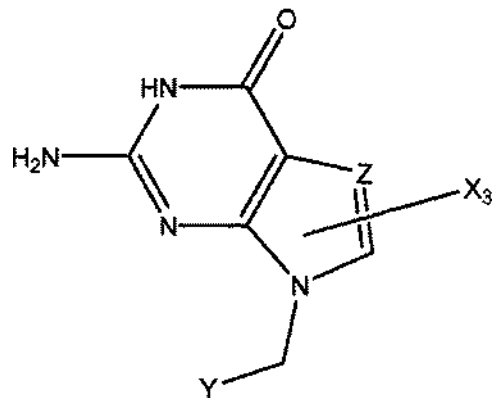


(ii)



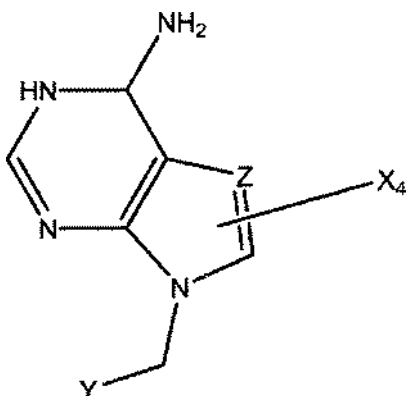
20

(iii)



y

(iv)



en las que Y es un grupo funcional capaz de llevar a cabo reacciones covalentes reversibles seleccionado entre el grupo que consiste en (i) aldehído; y (ii) cetona; X<sub>1</sub>-X<sub>4</sub> es hidrógeno o comprende un marcador detectable seleccionado entre el grupo que consiste en (i) uno o más seleccionados entre el grupo que consiste en dansilo, fluoresceína, rodamina, rojo Texas, IAEDANS, pigmentos de cianina (Cy3, Cy3.5, Cy5, Cy5.5, Cy7), pigmentos Bodipy (Invitrogen), pigmentos Alexa Fluor (Invitrogen) y pigmentos SNARF; y (ii) un marcador de masa; y Z es carbono o nitrógeno; en métodos de análisis genético.

5

10

10. El uso de la reivindicación 9, en el que los métodos de análisis genético incluyen la caracterización e identificación de nucleobases de ácidos nucleicos con el propósito de caracterizar polimorfismos de un único nucleótido y/o secuenciar ácidos nucleicos.

15

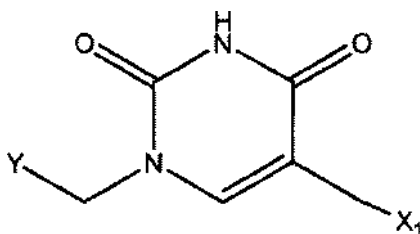
11. Un kit que proporciona reactivos y compuestos útiles en métodos para la caracterización de un nucleótido de un ácido nucleico y/o para secuenciar un ácido nucleico, comprendiendo dicho juego componentes seleccionados entre el grupo que consiste en:

20

(a) un oligómero de ácido peptidonucleico (APN) capaz de hibridar con una región de un ácido nucleico que carece de una nucleobase complementaria a aquella del nucleótido a caracterizar;

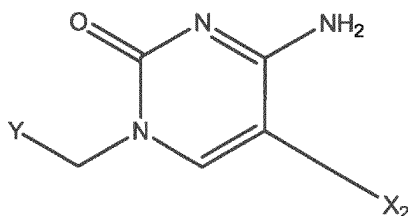
(b) nucleobases modificadas seleccionadas entre el grupo que consiste en:

(i)



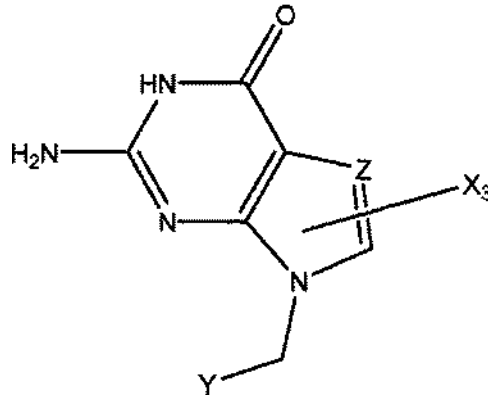
25

(ii)



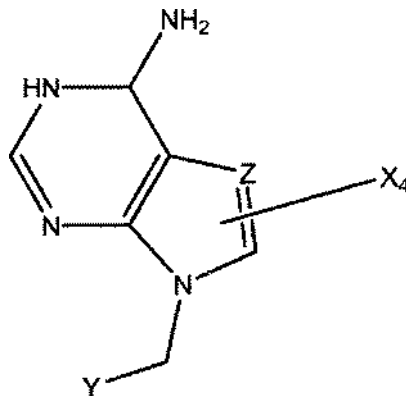


(iii)



5

y  
(iv)



10

en las que Y es un grupo funcional capaz de llevar a cabo reacciones covalentes reversibles seleccionado entre el grupo que comprende (i) un aldehído y (ii) una cetona;

X<sub>1</sub>-X<sub>4</sub> es hidrogeno o comprende un marcador detectable seleccionado entre el grupo que consiste en (i) uno o más seleccionados entre el grupo que consiste en dansilo, fluoresceína, rodamina, rojo Texas, IAEDANS, pigmentos de cianina (Cy3, Cy3.5, Cy5, Cy5.5, Cy7), pigmentos Bodipy (Invitrogen), pigmentos Alexa Fluor (Invitrogen) y pigmentos SNARF; y (ii) un marcador de masa; y

15

Z es carbono o nitrógeno; y  
(c) instrucciones de uso.

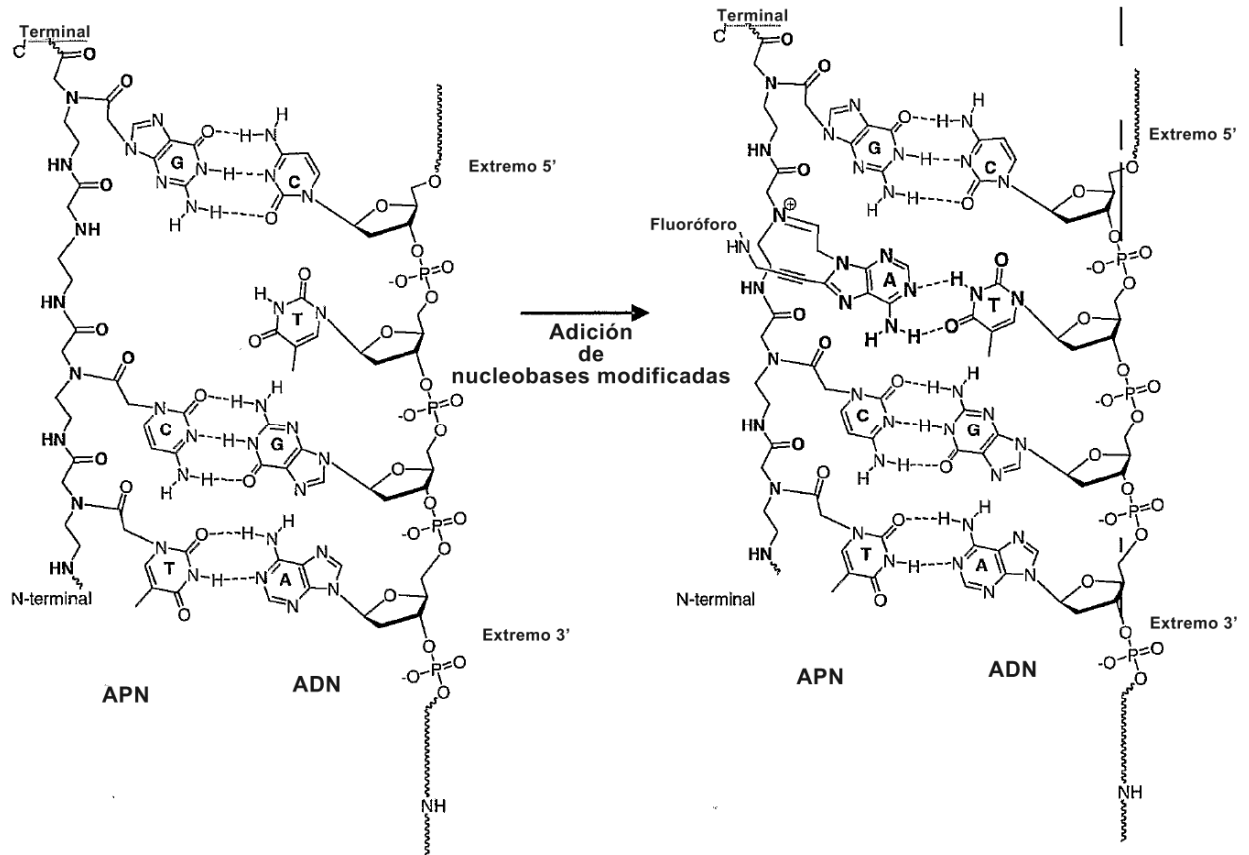


Figura 1

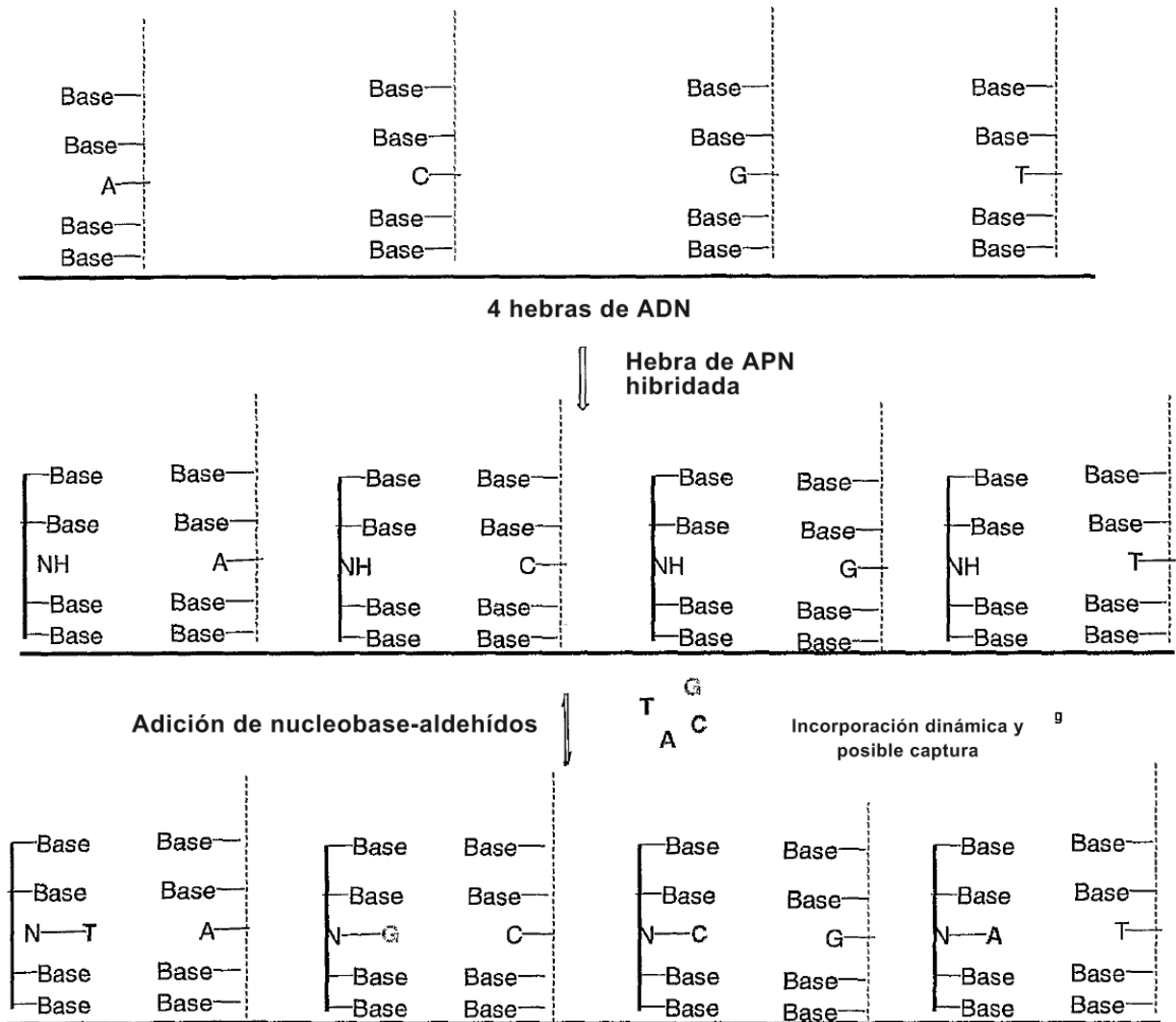


Figura 2

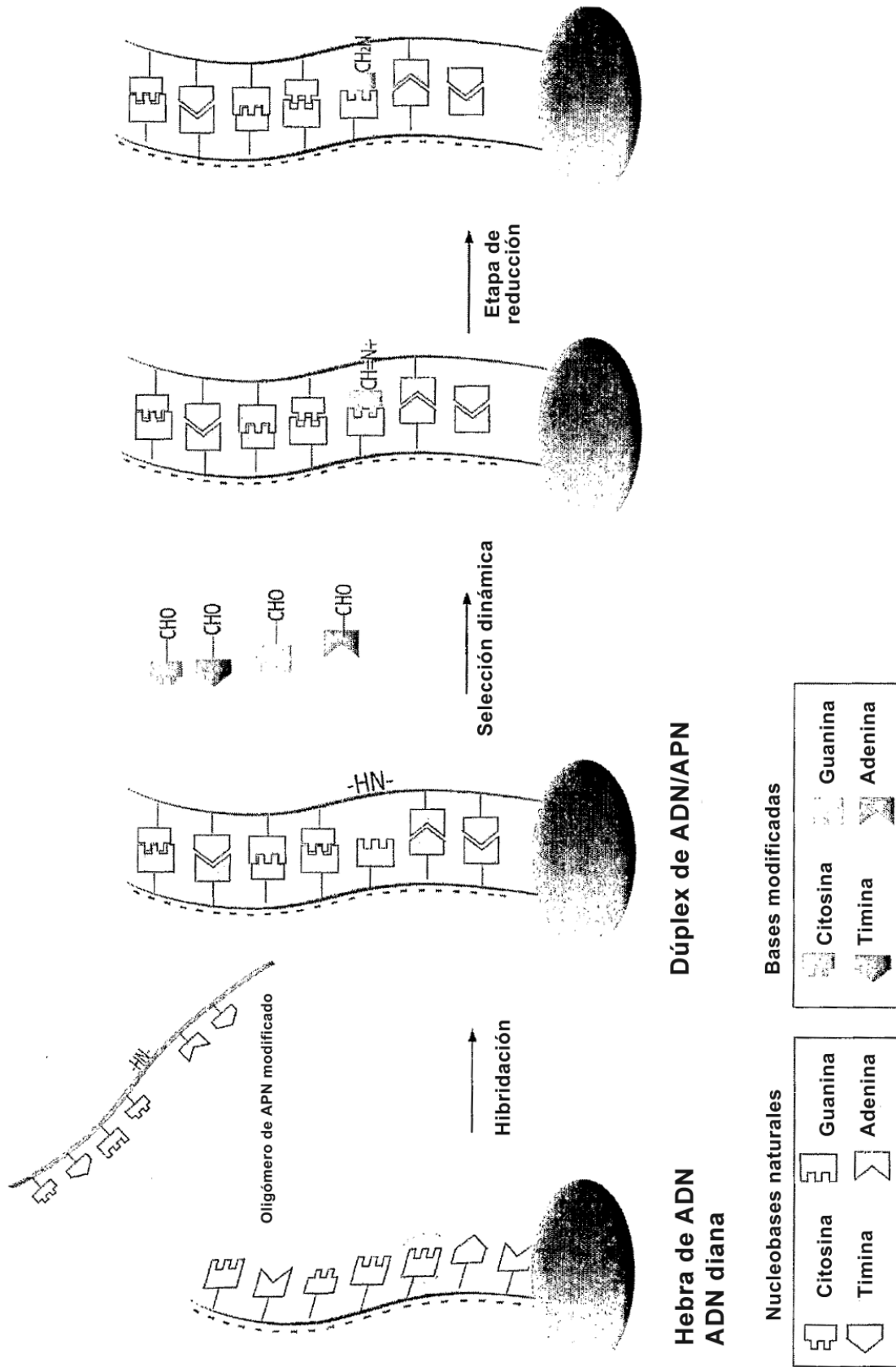


Figura 3

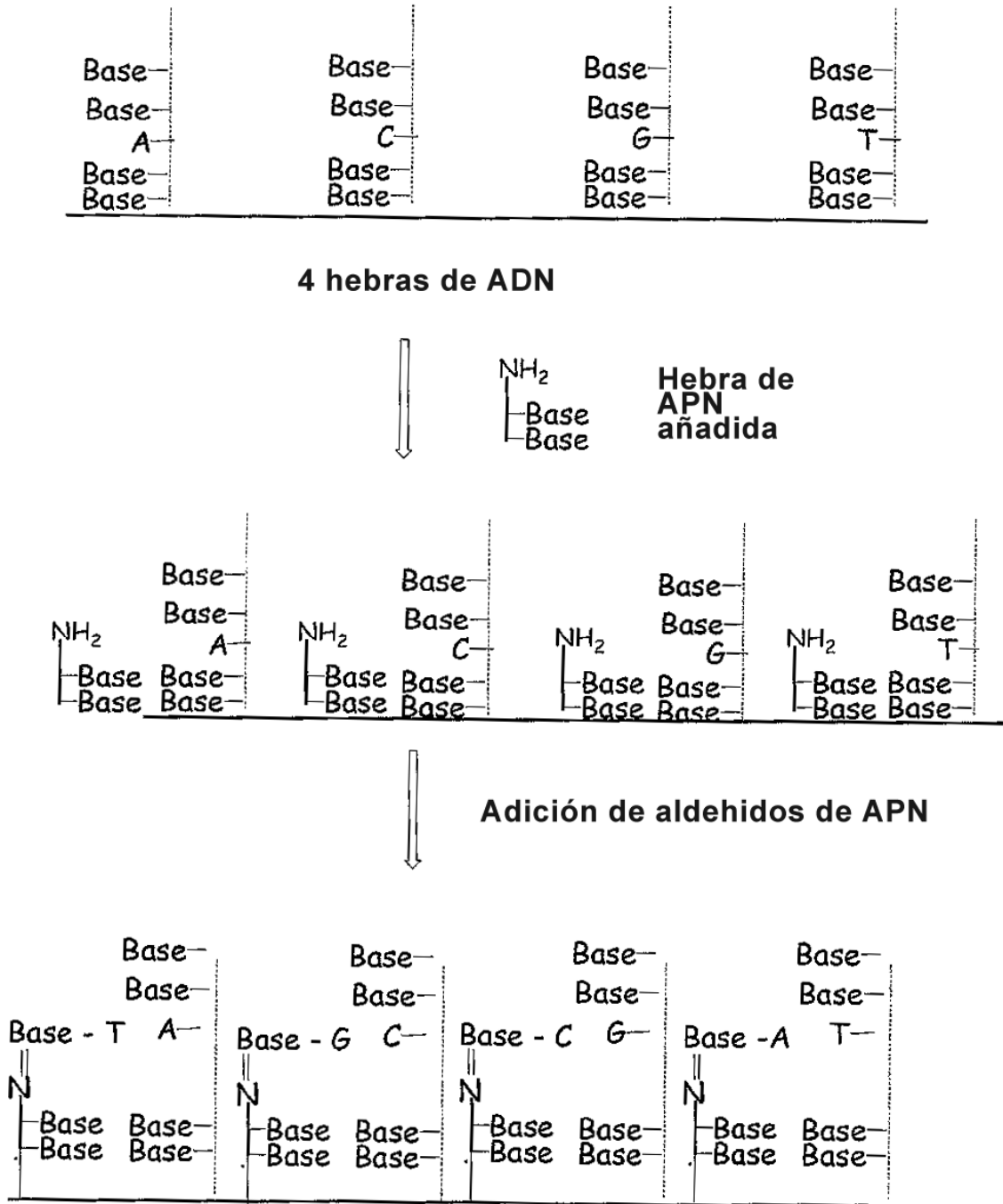


Figura 4

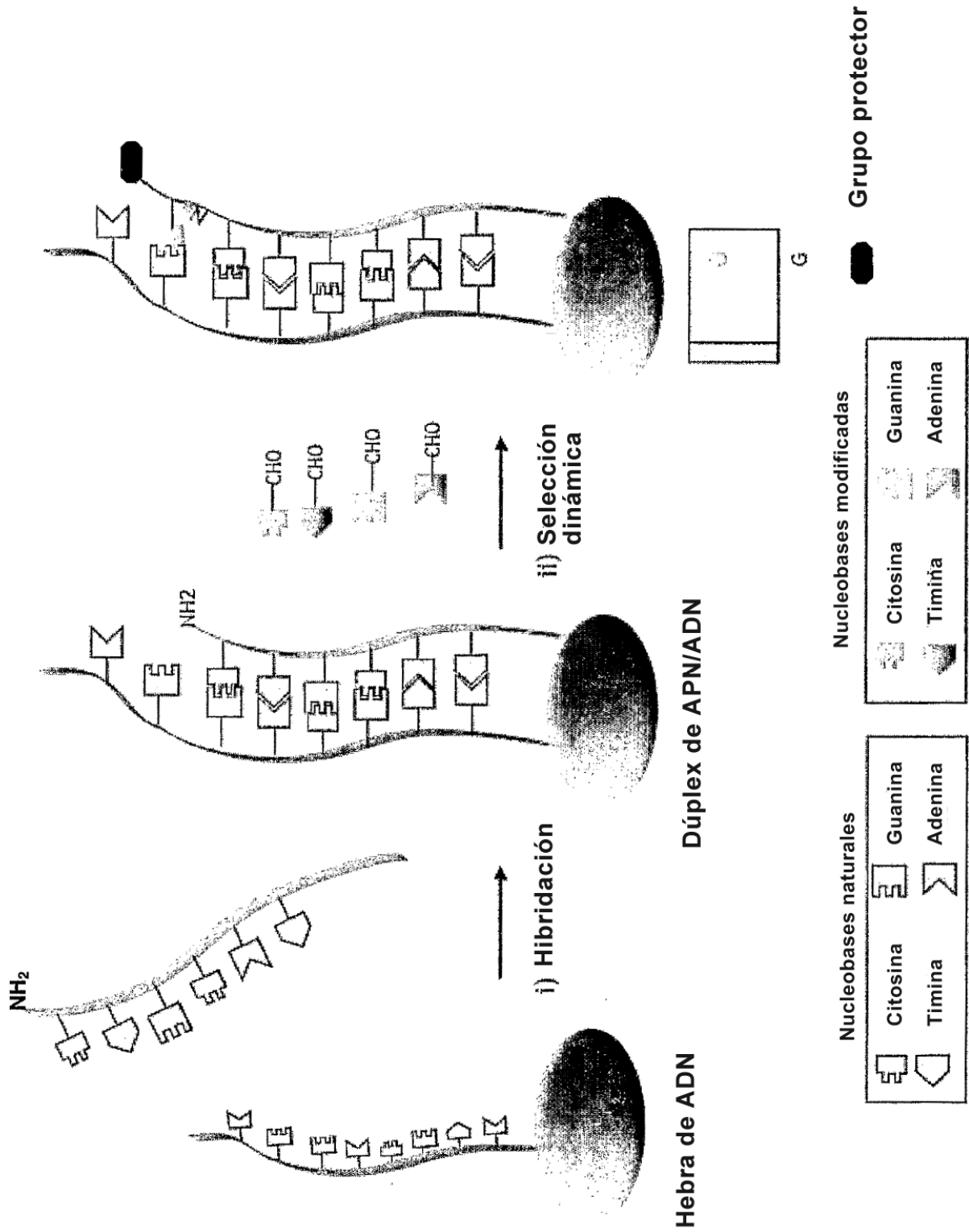


Figura 5

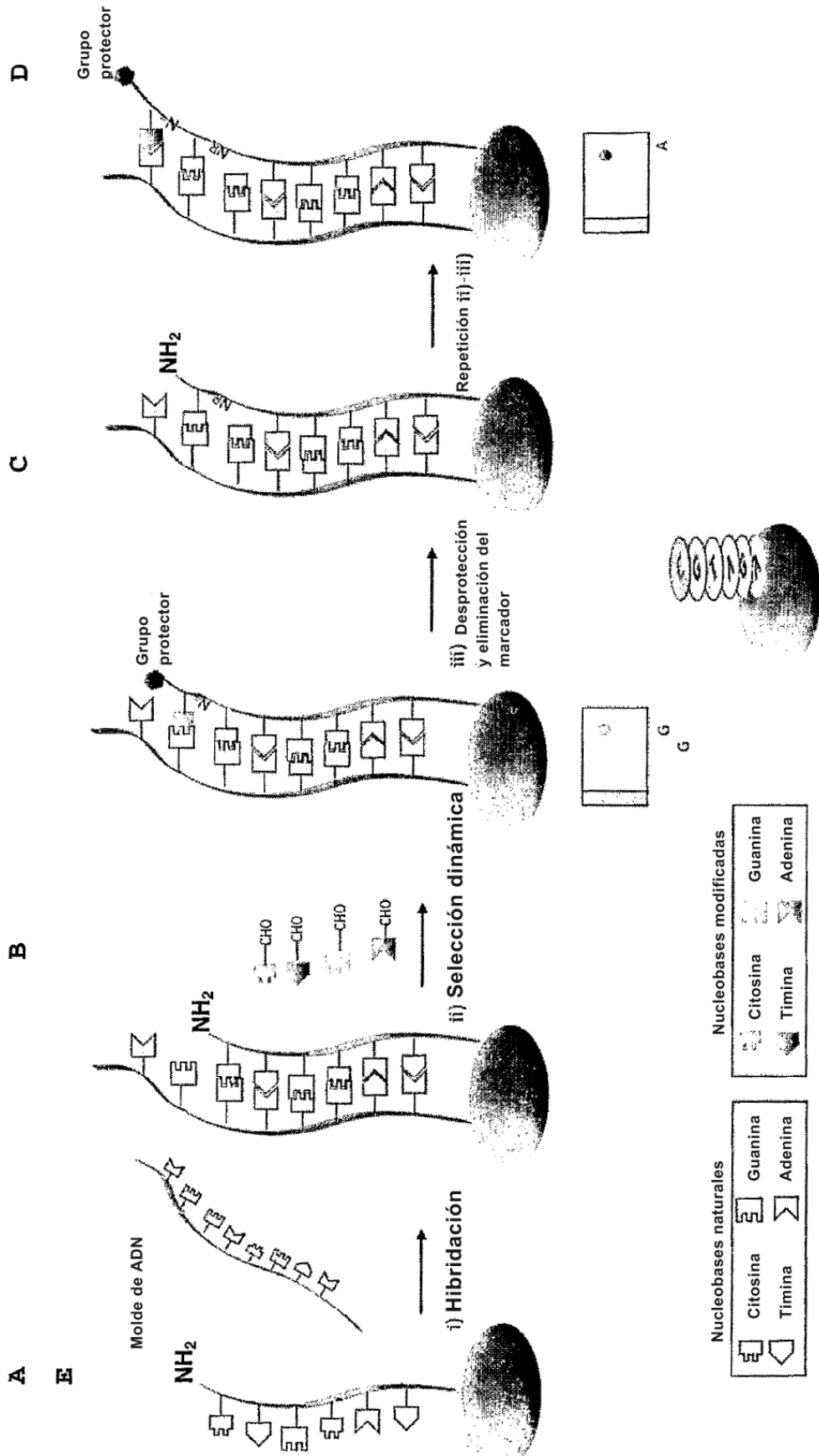


Ilustración de la secuenciación

Figura 6

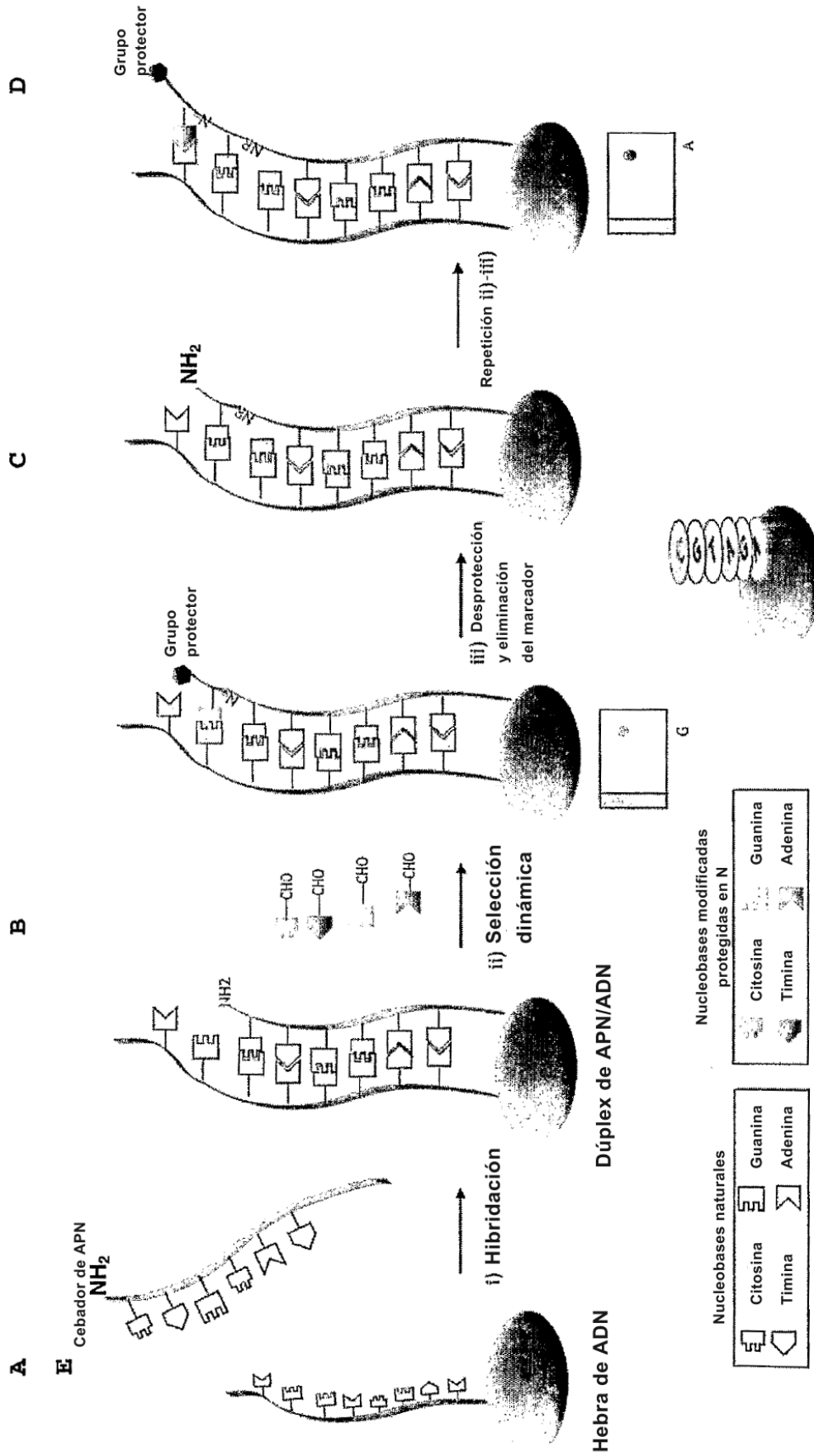


Ilustración de la secuenciación

Figura 7



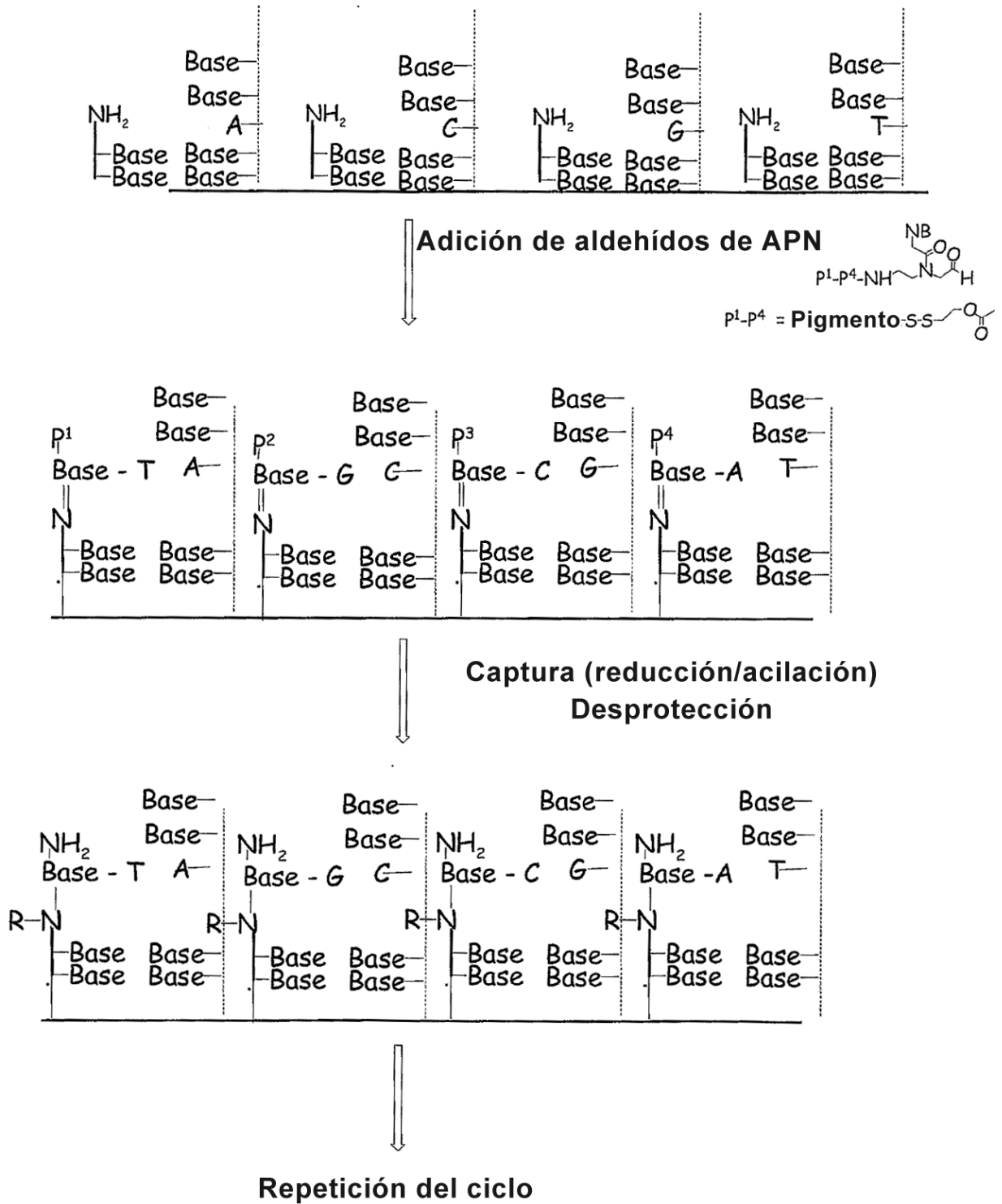


Figura 8

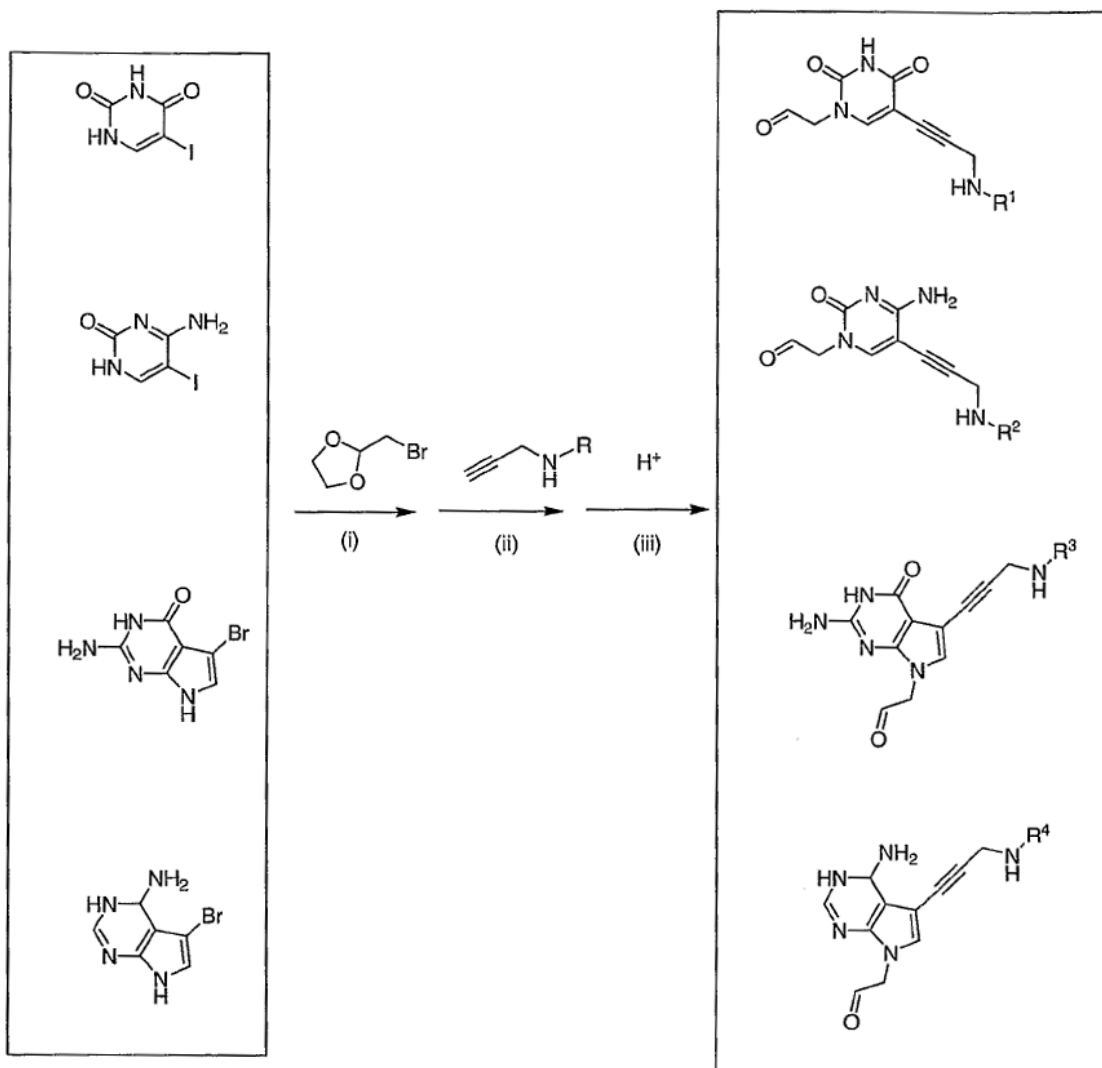


Figura 9

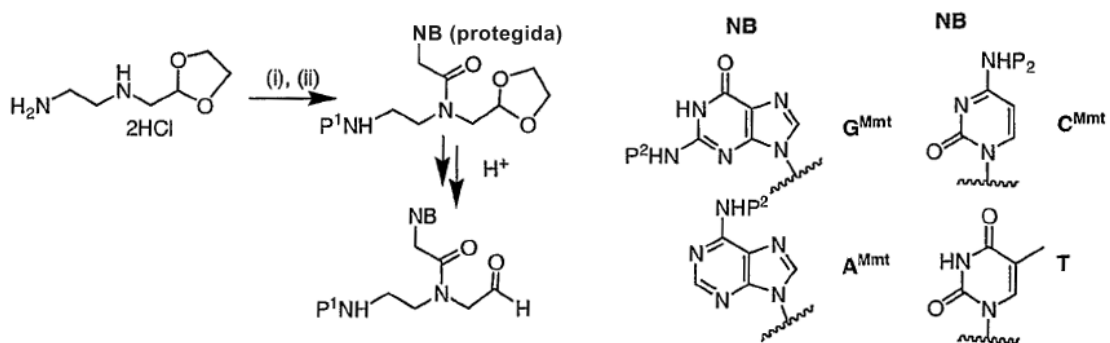


Figura 10

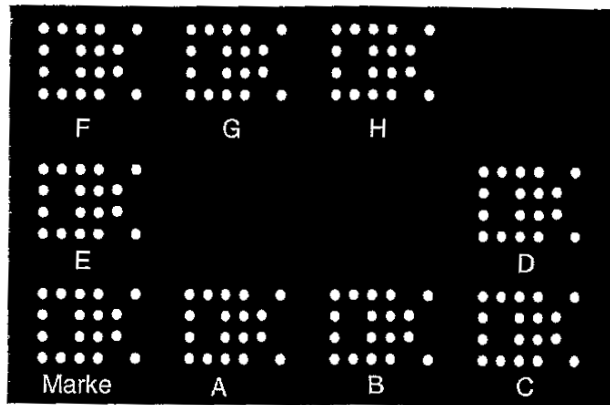


Figura 11

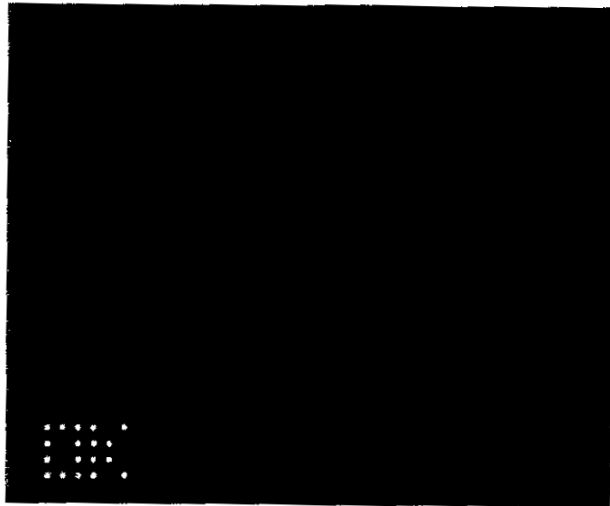


Figura 12

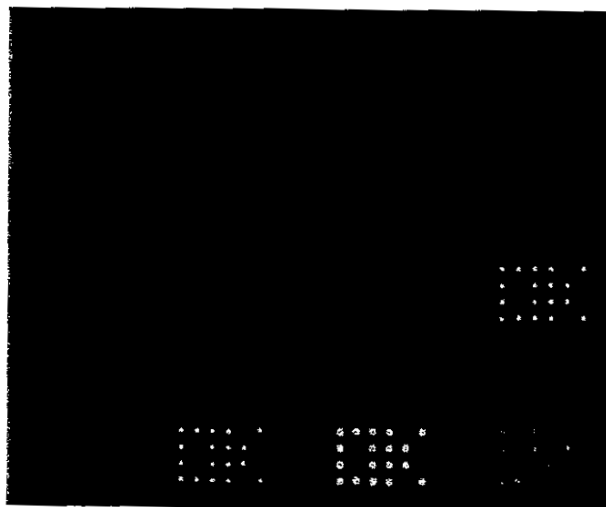


Figura 13

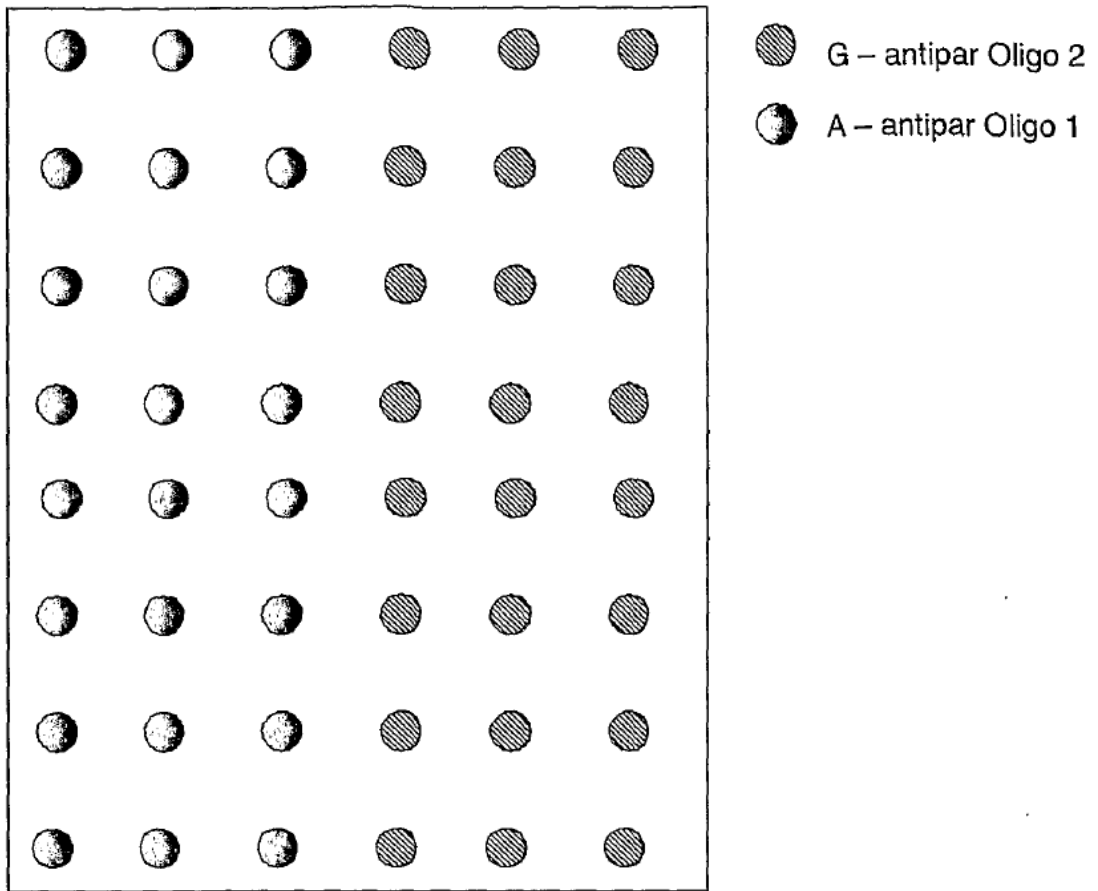


Figura 14A



Figura 14B

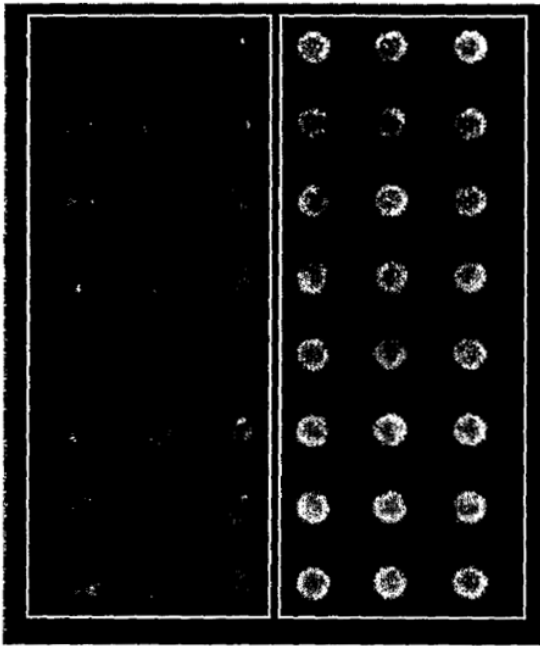


Figura 14C

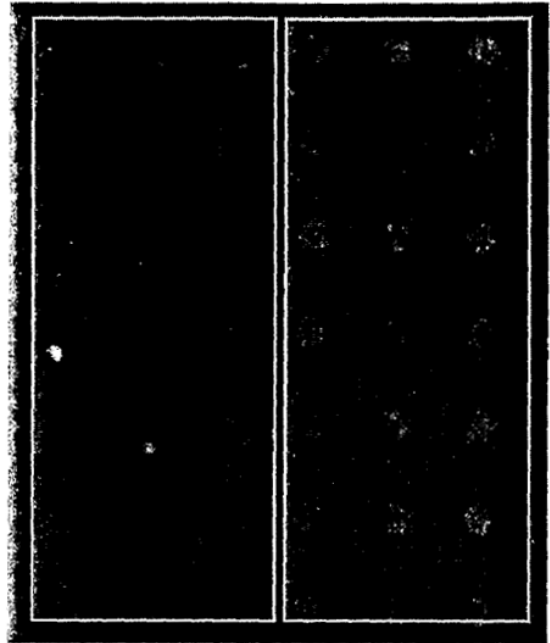


Figura 14D



Figura 14E

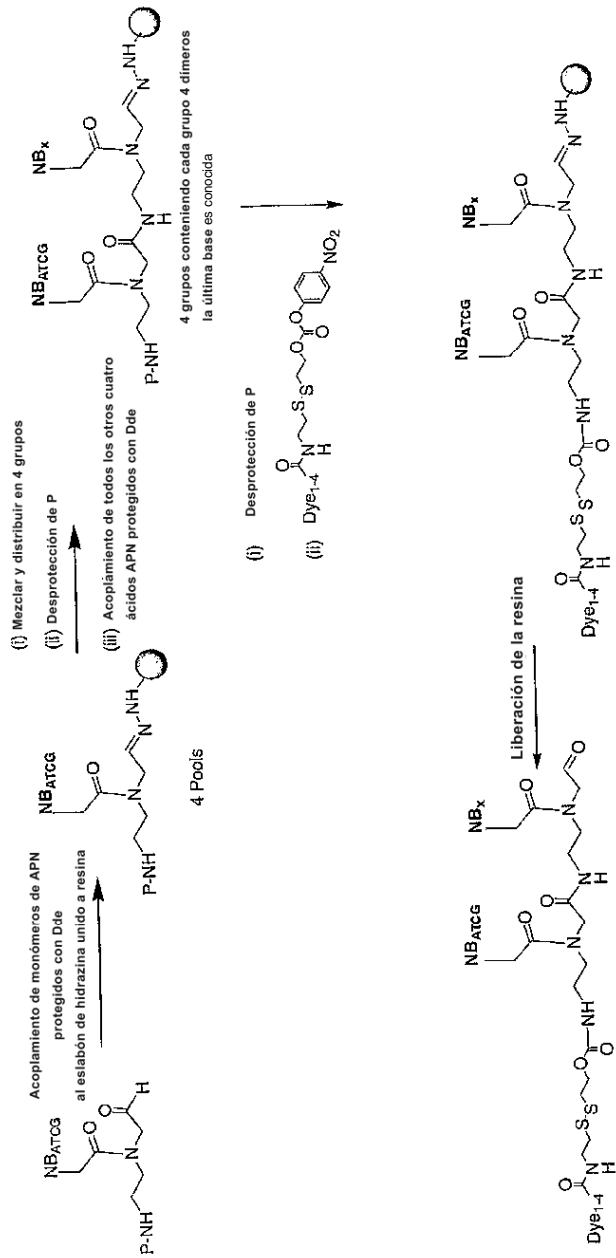


Figura 15A

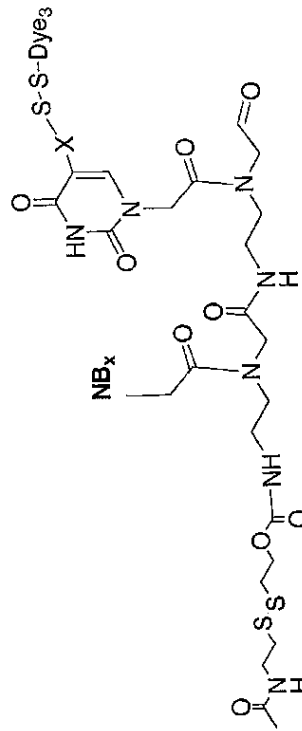
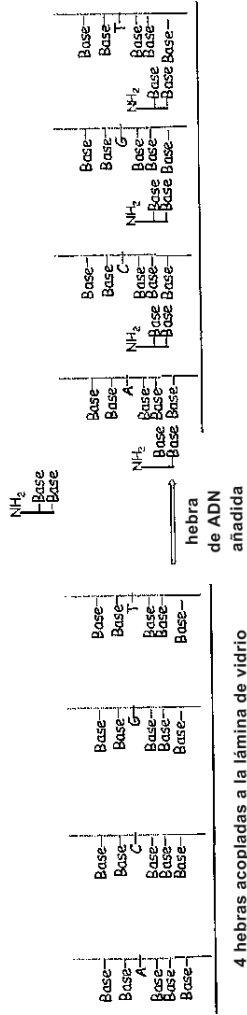


Figura 15B



(i) Añadir 16 dímeros de aldeídos de APN  
(ii) Reducción, adición de las cabezas

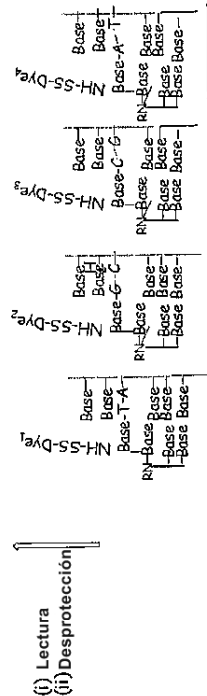
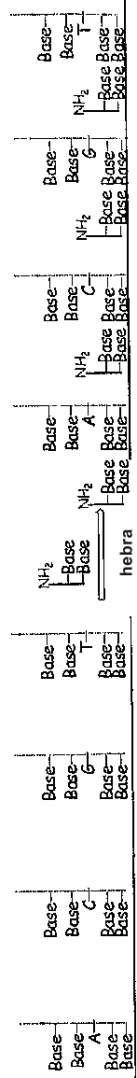
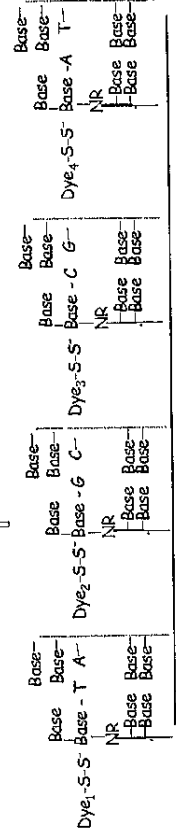


Figura 16A



(i) Añadir 16 dímeros de aldeídos de APN  
(ii) Reducción, adición de las cabezas

(i) Lectura  
(ii) Desprotección



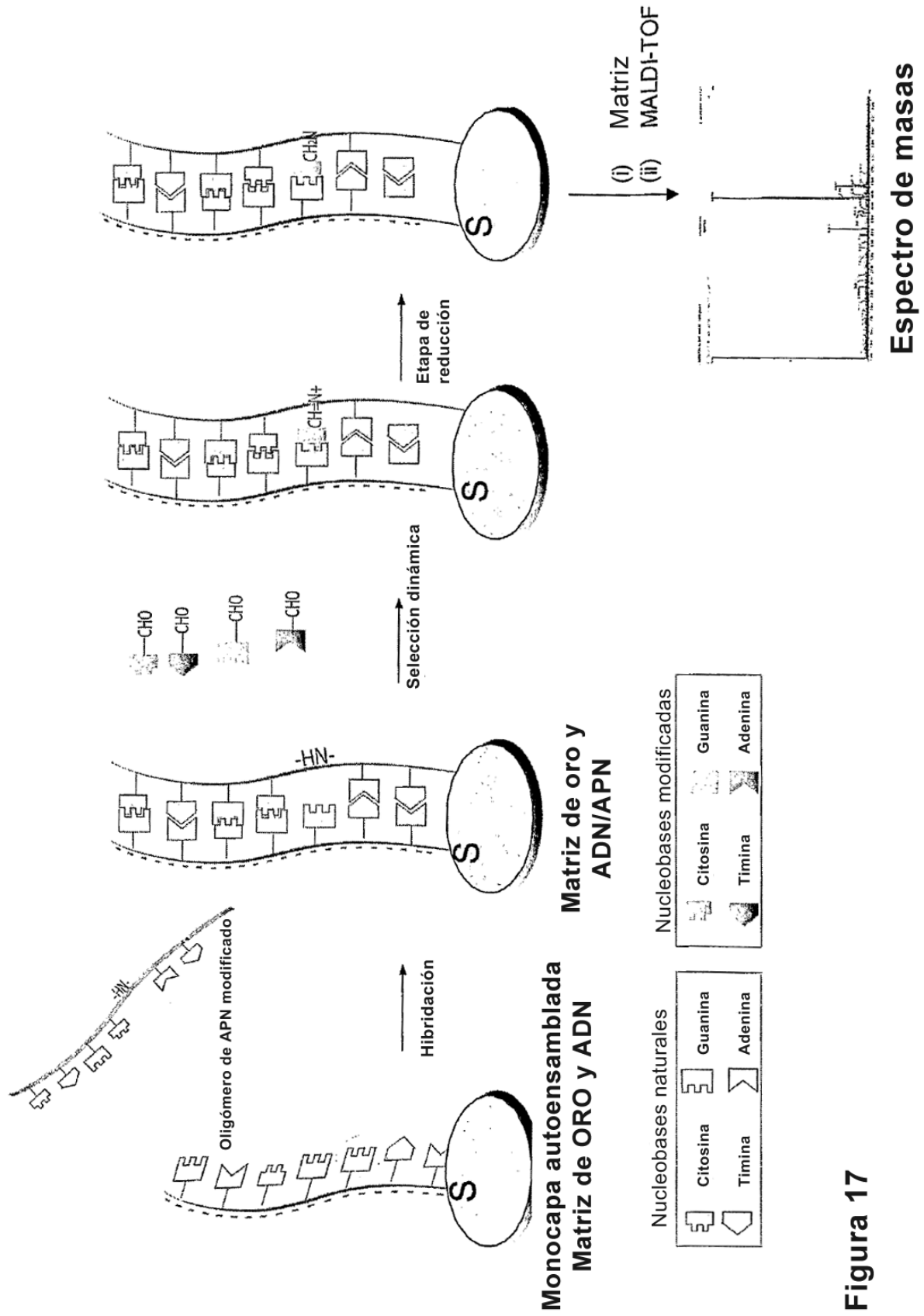
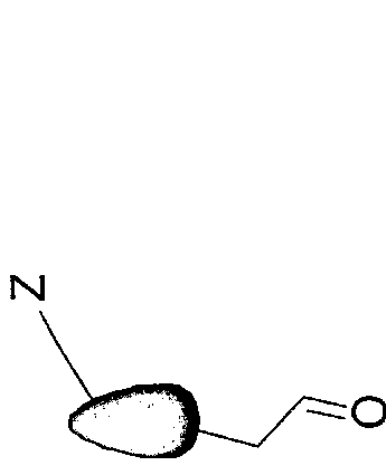


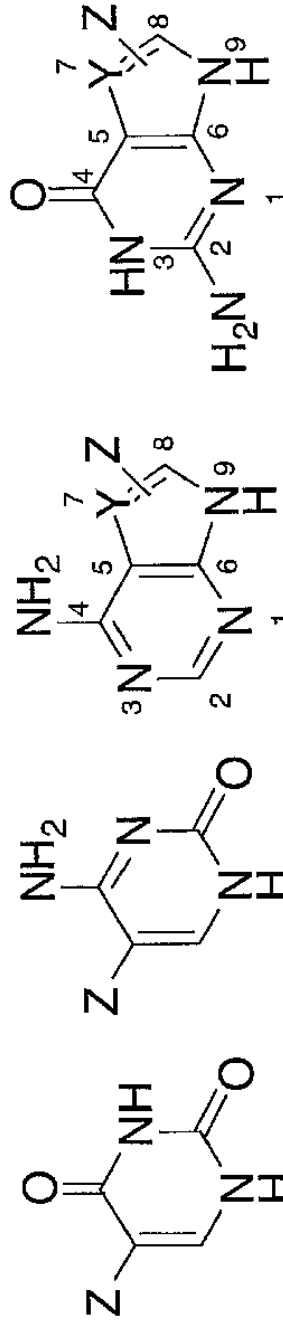
Figura 17





Nucleobases A, T, C, G

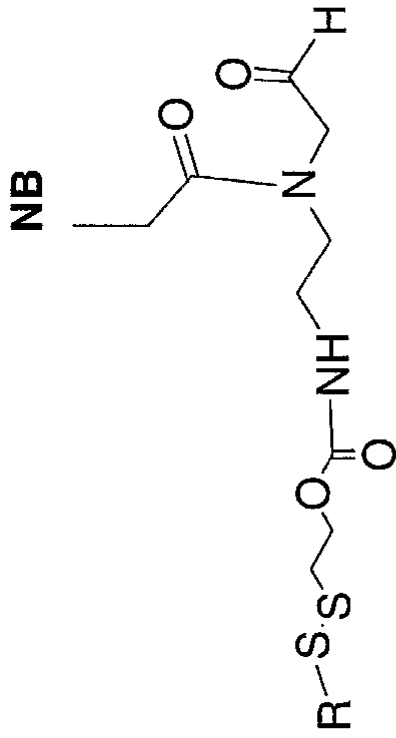
Z: H, pigmento o bromuro



Z=en 7 ó en 8

Y=C cuando Z en 7; N cuando Z en 8

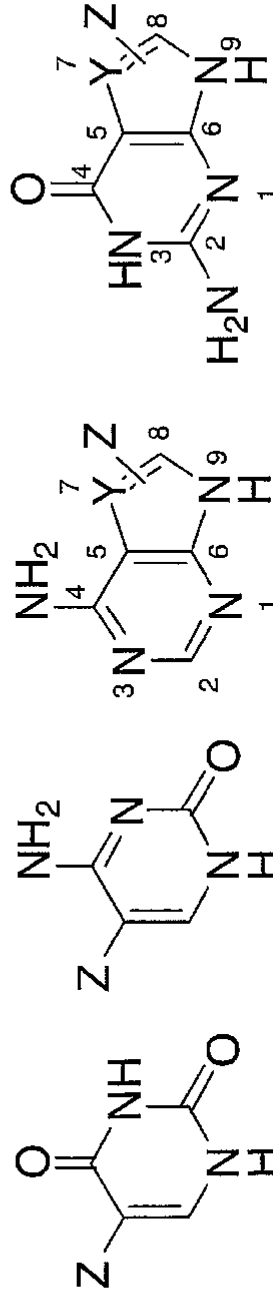
Figura 18



Nucleobases A, T, G, C

R=pigmento, alquil, acetamido etil

Z=H, pigmento o bromuro



Z=en 7 ó en 8

Y=C cuando Z en 7; N cuando Z en 8