

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 642 275**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **19.09.2012 PCT/GB2012/052305**

87 Fecha y número de publicación internacional: **28.03.2013 WO13041853**

96 Fecha de presentación y número de la solicitud europea: **19.09.2012 E 12762379 (1)**

97 Fecha y número de publicación de la concesión europea: **23.08.2017 EP 2758542**

54 Título: **Sonda con especificidad a región diana múltiple y de carácter tripartito**

30 Prioridad:

**19.09.2011 GB 201116131**

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

**16.11.2017**

73 Titular/es:

**EPISTEM LIMITED (100.0%)  
48 Grafton Street  
Manchester M13 9XX, GB**

72 Inventor/es:

**COBB,BEN**

74 Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

**ES 2 642 275 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**

Sonda con especificidad a región diana múltiple y de carácter tripartito

**5 Campo de la invención**

La presente invención se refiere a una sonda de ácidos nucleicos para detectar secuencias de ácidos nucleicos, y más particularmente a una sonda para detectar secuencias de ácidos nucleicos variantes de ADN genómico o amplificado. Aspectos adicionales de la invención se refieren a métodos para detectar secuencias variantes de ácidos nucleicos.

**Antecedentes de la invención**

Las sondas de ácidos nucleicos con frecuencia se usan para identificar la presencia de secuencias diana específicas en ADN genómico o amplificado. La temperatura de apareamiento o fusión de la sonda a la diana está afectada por la longitud de la región complementaria compartida por la sonda y la diana, y por la existencia de algún malemparejamiento entre los pares de bases de otra manera complementarios. Esto se puede usar para detectar la presencia de variantes, por ejemplo, SNP o repeticiones múltiples. Una sonda se puede diseñar para tener una primera temperatura de fusión ( $T_m$ ) a una secuencia de tipo natural, y el apareamiento de la sonda a la diana controlada, por ejemplo, a través del desarrollo de fluorescencia en el apareamiento. Si la  $T_m$  es diferente del valor esperado, entonces, la secuencia diana incluye una variante.

Sin embargo, este proceso puede ser difícil para implementar con regiones sonda más largas. En particular, una sonda larga no puede proporcionar variación adecuada en  $T_m$  entre una secuencia de tipo natural y una variante.

La presente invención pretende dirigirse a estas y otras desventajas con sondas convencionales.

El documento US2007/0065847 describe una colección de sondas de nucleótidos marcadas con una secuencia común para actuar como una etiqueta. Las sondas se usan para investigar genotecas de compuestos.

El documento WO2011/027966 describe sondas que tienen primeras y segundas regiones, incluyendo la primera región una molécula indicadora e incluyendo la segunda región una molécula amortiguadora de fluorescencia. Cuando la sonda se hibrida con una diana, las exonucleasas pueden escindir el indicador del amortiguador de fluorescencia, conduciendo a la detección de la señal. Cuando hay un malemparejamiento de modo que la porción indicadora no se hibrida, el indicador no se escinde y no hay detección de señal.

El documento WO2007/058499 describe sondas para virus respiratorios que tienen la primera y segunda porción de hibridación separadas por una porción de separación.

El documento US2007/0259347 describe sondas para matrices de cribado (*screening*), teniendo las sondas dos porciones separadas por un conector en las que el segmento conector se selecciona para minimizar el ruido de homología asociado con la hibridación.

El documento US2003/0170711 describe sondas de oligonucleótidos incluyendo bases universales, incluyendo 2-deoxiinosina.

El documento WO2011/087928 describe sondas de oligonucleótidos para detectar SNP KRAS y PIK3CA en un ambiente de secuencias de tipo natural, incluyendo conectores de polidesoxiinosina.

El documento WO2006/095941 describe una molécula de ácido nucleico que usa un oligonucleótido de especificidad doble que tiene una porción de especificidad a  $T_m$  alta 5', una porción de separación, y una porción de especificidad a  $T_m$  baja 3'. El segmento 3' no incluye secuencias de repetición en tándem.

El documento WO2003/054223 describe balizas moleculares tripartitas que incluyen una primera porción que tiene una estructura tallo-bucle tipo horquilla, y fluoróforos y amortiguadores de fluorescencia en la segunda y tercera porción de la baliza respectivamente. Las porciones no incluyen secuencias de repetición en tándem.

El documento WO2004/098386 describe un oligonucleótido tripartito que tiene un primer dominio capaz de hibridarse, un segundo dominio bloque de conexión, y un tercero de dominio de unión. El tercer dominio no incluye secuencias de repetición en tándem.

**Compendio de la invención**

Según un primer aspecto de la presente invención, se proporciona una sonda de ácido nucleico de acuerdo con la reivindicación 1. La invención también proporciona métodos de uso de tales sondas, como se enumeran en las reivindicaciones independientes 12 y 13. Los aspectos adicionales de la invención se describen en las

reivindicaciones dependientes.

La presencia de la región conectora permite que las sondas se dividan en elementos funcionales que tienen diferentes características de hibridación. La inclusión de estos conectores crea estructuras "burbuja", que aíslan los elementos de la sonda a partir de una perspectiva termodinámica, para proporcionar regiones con diferentes características de unión. Además, la presencia de la secuencia de ácidos nucleicos conectora permite que la sonda completa tenga las características de una molécula de polinucleótido sencilla, pero se comporta como si estuviera compuesta de sondas de ácidos nucleicos más cortas separadas. La región conectora se puede plegar para formar un bucle cuando la primera y segunda secuencia se hibridan con sus respectivas secuencias diana.

En ciertas realizaciones, además, la sonda se puede prolongar para tener regiones de hibridación múltiples separadas por conectores múltiples.

Esto se puede usar en varias diferentes maneras. Por ejemplo, la estructura sonda permite investigar regiones contiguas, en las que las sondas más largas (por ejemplo, una sonda única que abarca tanto la primera como la segunda región diana) no proporcionaría adecuada información a través del análisis de T<sub>m</sub> para diferenciar variantes.

Preferiblemente el conector es un conector de nucleósido; más preferiblemente el conector comprende polidesoxirribonucleótidos; lo más preferiblemente el conector comprende o consiste en polidesoxiinosina. La deoxiinosina tiene una temperatura de fusión baja en relación con las bases naturales debido a enlace de hidrógeno más débil. Se pueden usar otros nucleósidos.

Preferiblemente el conector es de hasta 5, 10, 15, 20, 30, 40, 50 nucleótidos de longitud.

Al menos una de la primera y segunda secuencia de ácidos nucleicos es una región indicadora. Una región indicadora incluye un resto marcado; preferiblemente un marcador fluorescente. Esto permite la detección de la sonda en el suceso de unión a una secuencia diana, y el seguimiento del apareamiento durante un intervalo de temperatura para determinar la presencia de cualquier secuencia variante diana. La sonda preferiblemente no comprende un resto amortiguador de fluorescencia, ni es el marcador previsto para usarse con un amortiguador de fluorescencia. Marcadores adecuados incluyen FAM, TET, HEX, ROX, TAMRA, Cy3, y Cy5. Los expertos en la técnica conocerán otros marcadores adecuados. Preferiblemente el marcador se incorpora a un nucleótido T, aunque se puede usar cualquier nucleótido adecuado.

La región indicadora es preferiblemente de 15 a 200 nt de longitud, más preferiblemente 15 a 150, más preferiblemente aún 15 a 100, o 20 a 100, 30 a 80, 40 a 60, o alrededor de 50 nt de longitud.

La región indicadora se diseña para tener una primera T<sub>m</sub> respecto a una secuencia diana enteramente complementaria y una T<sub>m</sub> alterada, preferiblemente inferior, respecto a una secuencia diana variante.

En esta realización, la temperatura de fusión de la primera secuencia apareada a la primera secuencia de ácidos nucleicos diana puede ser diferente a la de la segunda secuencia apareada a la segunda secuencia de ácidos nucleicos diana. Sin embargo, preferiblemente, tanto la primera como la segunda secuencia de ácidos nucleicos se seleccionan para tener T<sub>m</sub> similar a sus respectivas secuencias diana.

La región indicadora además puede comprender una región bloqueante; es decir, una porción que sirve para bloquear la prolongación de la cadena de ácidos nucleicos por la ADN polimerasa, previniendo así la prolongación de cadena durante, por ejemplo, la PCR. Un grupo bloqueante de enzima polimerasa es uno que debería tener las propiedades funcionales de bloqueo además de elongación del polímero. Un grupo bloqueante puede ser cualquier grupo químico que se pueda unir a un nucleótido que permitirá que el extremo 5' del nucleótido modificado se una a un extremo 3' del otro nucleótido en una cadena de ADN pero no permitirá la unión de un nucleótido al grupo 3' hidroxilo del nucleótido modificado. Adecuadamente, la ausencia de un grupo OH en la posición 3' prevendrá además la elongación por actividad de la polimerasa. En una realización particularmente preferida, el grupo bloqueante se selecciona de grupos acetilo, CH<sub>3</sub>, glicilo, leucilo y alanilo. En otra realización, el grupo bloqueante puede estar en forma de un di o tri péptido.

En una realización de la invención, tanto la primera como la segunda secuencia de ácidos nucleicos son regiones indicadoras. Pueden incluir diferentes marcadores. Tal sonda se puede usar como un indicador múltiple, permitiendo la detección de secuencias diana durante un intervalo prolongado con una única sonda.

En esta realización, la sonda se divide en múltiples regiones indicadoras discretas. Cada región indicadora tiene diferentes temperaturas de apareamiento y tiene 1 o más nucleótidos fluorescentes, preferiblemente FAM-T, o diferentes/múltiples colores. El indicador se puede usar para informar de la presencia de secuencias específicas o variantes de secuencia, SNP, inserciones, delecciones, etc. Esto permite que se detecten las secuencias múltiples durante un intervalo prolongado con una única sonda. Cada región se prepara para tener una T<sub>m</sub> similar en el caso de la secuencia diana de tipo natural, pero una T<sub>m</sub> cambiada en el caso de una mutación; esto significa que un

usuario solamente tiene que detectar la Tm cambiada para saber que la variante está presente.

Como alternativa para detectar las secuencias de diana múltiples, la sonda se puede estructurar, por ejemplo, para proporcionar una región de unión control y una región de unión de ensayo, para mejorar la fiabilidad de un ensayo.

5 En una realización alternativa de la invención, la segunda secuencia de ácidos nucleicos puede ser una región indicadora, y la primera secuencia de ácidos nucleicos es una región de anclaje. En esta realización, la temperatura de fusión de la primera secuencia apareada a la primera secuencia de ácidos nucleicos diana es mayor que la de la segunda secuencia apareada a la segunda secuencia de ácidos nucleicos diana. La secuencia conector puede formar una región bucle o se puede hibridar con cualquier secuencia de ácidos nucleicos inter-sonda. La Tm de la región de anclaje a la secuencia diana es significativamente mayor (al menos 0,5, 1, 1,5 °C, pero más generalmente hasta 5 °C a 20 °C mayor) que la Tm de la región indicadora con la secuencia diana.

15 La región de anclaje es preferiblemente de al menos 50 nt de longitud, más preferiblemente al menos 60, 70, 80, 90, 100, 125, 150 nt. La región de anclaje es preferiblemente de no más de 200 nt de longitud y no menos de 20 nt de longitud.

20 En el uso, ya que la región de anclaje tiene una temperatura de apareamiento básicamente superior sirve para sujetar la sonda a la región correcta. La región indicadora tiene una temperatura de apareamiento inferior a la región de anclaje y tiene 1 o más nucleótidos fluorescentes, preferiblemente FAM-T. El indicador se usa para informar de la presencia de una secuencia específica o variantes de secuencia.

25 En las realizaciones preferidas, la sonda se usa para detectar polimorfismos de tamaño/longitud, por ejemplo, perfilación de RCT, X Frágil, etc. La región de anclaje se usa para aparearse a la zona corriente abajo de la región de repetición, mientras que la región indicadora es homóloga a un tramo corto de las repeticiones. Puesto que el indicador se aísla de la principal región de anclaje, y puesto que los fluoróforos están solamente en la región indicadora, no se produce diferencia si el indicador se une a diferentes sitios para causar bucles en el amplicón – normalmente esto alteraría la Tm del producto, haciendo difícil la medida. En ciertas realizaciones, la región de anclaje puede estar diseñada para ser homóloga a un tramo corto de la región de repetición, así como una región flanqueadora; por ejemplo, la región de anclaje puede incluir hasta 50 nt de la secuencia de repetición y un tramo más largo, por ejemplo 150 nt, de secuencia flanqueadora.

35 En el presente documento también se describe una sonda de ácidos nucleicos que comprende una primera secuencia de ácidos nucleicos de anclaje que es complementaria a una primera secuencia de ácidos nucleicos diana; una segunda secuencia de ácidos nucleicos indicadora que es complementaria a una segunda secuencia de ácidos nucleicos diana; y una secuencia de ácidos nucleicos conectora que se une a la primera y segunda secuencia de ácidos nucleicos; en la que el conector separa la primera y segunda secuencia de modo que las temperaturas de fusión de la primera y segunda secuencia hibridadas con sus respectivas secuencias diana es discreta, y en la que la temperatura de fusión de la secuencia de anclaje apareada a la primera secuencia de ácidos nucleicos diana es mayor que la de la secuencia indicadora apareada a la segunda secuencia de ácidos nucleicos diana; y en la que cada secuencia indicadora comprende al menos un marcador detectable.

45 El indicador preferiblemente comprende una pluralidad de unidades de secuencia repetidas. Preferiblemente al menos 3, 4 o 5 unidades repetidas.

El conector preferiblemente comprende polidesoxiinosina.

50 El anclaje puede comprender una o más unidades de secuencia repetidas, pero preferiblemente comprende secuencia única. El anclaje es preferiblemente de 50 a 200 nt de longitud.

La sonda preferiblemente comprende además una secuencia bloqueante para prevenir la prolongación de la cadena durante la PCR.

55 La sonda además puede comprender uno o más secuencias conectoras e indicadoras adicionales.

60 En el presente documento también se describe una sonda de ácidos nucleicos que comprende una primera secuencia de ácidos nucleicos indicadora que es complementaria a una primera secuencia de ácidos nucleicos diana; una segunda secuencia de ácidos nucleicos indicadora que es complementara a una segunda secuencia de ácidos nucleicos diana; y una secuencia de ácidos nucleicos conectora que se une a la primera y segunda secuencia de ácidos nucleicos; en la que la temperatura de fusión de la primera secuencia indicadora apareada a la primera secuencia de ácidos nucleicos diana es similar o igual a la de la segunda secuencia indicadora apareada a la segunda secuencia de ácidos nucleicos diana; y en la que cada secuencia indicadora comprende al menos un marcador detectable.

65 La sonda preferiblemente comprende además una secuencia bloqueante para prevenir la prolongación de la cadena durante la PCR.

En el presente documento también se describe un método para detectar uno o más polimorfismos en una secuencia de ácidos nucleicos diana, comprendiendo el método:

- 5 proporcionar una sonda según un aspecto de la presente invención;  
 poner en contacto la sonda con una secuencia de ácidos nucleicos de ensayo, teniendo la secuencia de ensayo una primera secuencia de ácidos nucleicos diana y una segunda secuencia de ácidos nucleicos diana;  
 permitir que la primera secuencia de ácidos nucleicos de la sonda se hibride con la primera secuencia de ácidos nucleicos diana; y la segunda secuencia de ácidos nucleicos de la sonda se hibride con la segunda secuencia de ácidos nucleicos diana; y  
 10 determinar la T<sub>m</sub> de la primera secuencia sonda hibridada con la primera secuencia diana; y determinar la T<sub>m</sub> de la segunda secuencia sonda hibridada con la segunda secuencia diana;  
 en la que la T<sub>m</sub> es indicativa de la presencia de una secuencia de tipo natural, o de una secuencia variante.

- 15 Este método también se puede usar para mejorar la utilidad de la temperatura de fusión por lo que la T<sub>m</sub> de una primera región indicadora actúa como referencia para medir la T<sub>m</sub> de la región de secuencia variante. Generalmente la posición de la temperatura de fusión registrada que es útil en la determinación de la secuencia subyacente puede estar afectada por la concentración de las sales en la mezcla de reacción. Comparando la posición de fusión de la primera región indicadora como secuencia conocida, es posible determinar más exactamente la secuencia bajo la segunda secuencia indicadora puesto que ambas regiones estarán igualmente afectadas por sustancias  
 20 interferentes tales como sales externas.

En el presente documento también se describe un método para detectar uno o más polimorfismos en una secuencia de ácidos nucleicos diana, comprendiendo el método:

- 25 proporcionar una sonda según un aspecto de la presente invención;  
 poner en contacto la sonda con una secuencia de ácidos nucleicos de ensayo, teniendo la secuencia de ensayo una primera secuencia de ácidos nucleicos diana y una segunda secuencia de ácidos nucleicos diana;  
 permitir que la primera secuencia de ácidos nucleicos de la sonda se hibride con la primera secuencia de ácidos nucleicos diana; y la segunda secuencia de ácidos nucleicos de la sonda se hibride con la segunda secuencia de ácidos nucleicos diana; y  
 30 determinar la T<sub>m</sub> de la primera secuencia sonda hibridada con la primera secuencia diana; y determinar la T<sub>m</sub> de la segunda secuencia sonda hibridada con la segunda secuencia diana;  
 comparar la T<sub>m</sub> determinada de la primera secuencia sonda con la T<sub>m</sub> esperada de la primera secuencia sonda;  
 normalizar la T<sub>m</sub> determinada de la segunda secuencia sonda basada en la diferencia de la T<sub>m</sub> determinada y  
 35 esperada de la primera secuencia sonda;  
 en la que la T<sub>m</sub> normalizada de la segunda secuencia sonda es indicativa de la presencia de una secuencia de tipo natural, o de una secuencia variante.

- 40 En el presente documento también se describe un método para detectar repeticiones en tándem en una secuencia de ácidos nucleicos diana, comprendiendo el método:

- proporcionar una sonda según un aspecto de la presente invención, en la que la secuencia indicadora comprende varias unidades de repetición en tándem;  
 poner en contacto la sonda con una secuencia de ácidos nucleicos de ensayo, teniendo la secuencia de ensayo una primera secuencia de ácidos nucleicos diana y una segunda secuencia de ácidos nucleicos diana, en la que  
 45 la primera secuencia diana es complementaria a la secuencia de anclaje sonda, y la segunda secuencia diana comprende varias repeticiones en tándem complementarias a las unidades de repetición en tándem de la secuencia indicadora;  
 permitir que la secuencia de ácidos nucleicos de anclaje de la sonda se hibride con la primera secuencia de ácidos nucleicos diana; y la segunda secuencia de ácidos nucleicos de la sonda se hibride con la segunda  
 50 secuencia de ácidos nucleicos diana; y  
 detectar la unión o no de la secuencia indicadora a la secuencia de ácidos nucleicos de ensayo.

- La unión se puede detectar determinando la T<sub>m</sub> de la secuencia indicadora a la secuencia de ácidos nucleicos de ensayo, y/o detectando el marcador (por ejemplo, fluorescencia) a partir de la secuencia indicadora. La unión indica que el número de repeticiones en la secuencia diana es mayor que o igual al número de repeticiones en la secuencia indicadora; si hay más repeticiones en la secuencia indicadora, entonces hay no unión o unión reducida. En ciertas realizaciones, se puede considerar que la T<sub>m</sub> generalmente es proporcional al número de repeticiones en la secuencia diana, y la T<sub>m</sub> se puede determinar para determinar el número de repeticiones. Alternativamente, o  
 60 además, la secuencia indicadora de la sonda puede incluir uno o más restos marcados en cada unidad de repetición; de esta manera, el marcador detectado se puede tomar como que es proporcional al número de repeticiones en la secuencia diana.

#### Breve descripción de los dibujos

- 65 Estos y otros aspectos de la presente invención se describirán, a continuación, solamente a modo de ejemplo en

referencia a los dibujos acompañantes, en los cuales:

- La Figura 1 muestra una primera arquitectura de sonda;
- 5 las Figuras 2 a 5 muestran estructuras bicatenarias alternativas formadas por la unión de la sonda de la figura 1 a una secuencia diana;
- la Figura 6 muestra una segunda arquitectura de sonda;
- la Figura 7 muestra codones mutantes comunes en el gen *rpoB* de *M. tuberculosis*;
- la Figura 8 muestra la secuencia de una porción del gen *rpoB* de *M. tuberculosis*, junto con las secuencias de ciertos genes mutantes y la secuencia de una sonda para detectar la presencia de tales mutantes;
- 10 la Figura 9 muestra estructuras bicatenarias alternativas formadas por la unión de la sonda de la Figura 8 a las secuencias diana;
- la Figura 10 muestra curvas de fusión ejemplo obtenidas de las cadenas dobles de la Figura 9;
- la Figura 11 muestra un alineamiento de secuencia de una porción del genoma del virus vaccinia con la de otros pox virus;
- 15 la Figura 12 muestra una sonda para la detección de los pox virus;
- la Figura 13 compara la sensibilidad de la sonda conectora de la Figura 12 con la de una sonda de del virus vaccinia convencional;
- la Figura 14 muestra el análisis de la curva de fusión de la sonda conectora de la Figura 12 y una sonda convencional con diversas secuencias genómicas de pox virus amplificadas;
- 20 la Figura 15 muestra secuencias de cebador y condiciones de amplificación para amplificar el fragmento de pox virus usado en la Figura 14; y
- la Figura 16 muestra las secuencias de cebador y sonda, y el análisis de la curva de fusión, para una repetición corta en tándem en el gen de la tirosina hidroxilasa humana.

25 **Descripción detallada de los dibujos**

En el presente documento se describe una sonda de ácidos nucleicos que se divide en dos o más elementos funcionales, separados por un conector de ácido nucleico. El conector preferiblemente es deoxiinosina, la cual tiene una Tm baja en relación con las bases de ADN de origen natural. Se pueden usar otros conectores adecuados.

30 La Figura 1 muestra esquemáticamente una primera realización de una sonda. La sonda incluye una región de anclaje, de alrededor 100 a 200 nt de longitud, separada por un conector de deoxiinosina de 5 nt de longitud de una región indicadora, de desde 15 a 200 nt de longitud. La región de anclaje está diseñada para ser complementaria a una primera región diana que está corriente abajo a partir de una región de secuencia repetida en un ácido nucleico diana. La región indicadora es complementaria a la región de secuencia repetida, e incluye un número conocido de unidades repetidas. La región indicadora también incluye uno o más nucleótidos marcados, preferiblemente FAM-T. En ciertas realizaciones, cada unidad de repetición incluye un nucleótido marcado.

40 Las secuencias sonda se diseñan de modo que la Tm de la región de anclaje a la diana es significativamente mayor que la Tm del indicador a la diana, que de una en una es significativamente mayor que la Tm del conector a cualquiera de las secuencias en el ácido nucleico diana.

45 Las Figuras 2 a 5 muestran diversos escenarios diferentes para la unión de la sonda a una diana. En estos escenarios, la sonda está siendo usada para determinar los polimorfismos de tamaño/longitud, por ejemplo, perfilado de RCT, X Frágil, etc. La región de anclaje se usa para aparear la sonda a la parte corriente abajo de la región de repetición, mientras que la región indicadora es homóloga a un tramo corto de las repeticiones. Puesto que el indicador se aísla de la principal región de anclaje, y puesto que los fluoróforos están solamente sobre la región indicadora, no produce diferencia si el indicador se une a diferentes sitios para causar bucles en el amplicón – normalmente esto alteraría la Tm del producto, haciendo difícil la medición.

50 Cuando la longitud de la región de secuencia repetida diana es menor que la de la región indicadora (Figura 2), entonces, la región indicadora no se une a la diana, y no hay fluorescencia a partir del marcador, y ni curva de fusión obtenida por la medición de Tm del indicador. Sin embargo, ya que la región de anclaje se ha unido a la diana, esto se puede medir (por ejemplo, determinando la Tm), para confirmar que la sonda se une y que la región de secuencia repetida está ausente.

60 Cuando la longitud de la región de secuencia repetida diana es igual a o mayor que la longitud de la región indicadora, entonces, se pueden dar cualquiera de los escenarios en las Figuras 3 a 5. En la Figura 3, el indicador se mantiene pequeño a 1, 2, 3, 4 o 5 longitudes de repetición que limita el impacto de cualquiera de los efectos del bucle sobre la Tm de la doble cadena indicador:amplicón. Ejemplos de uso son para estructuras de repetición largas tales como X Frágil en las que es importante “dividir en bandas” el número de repeticiones; por ejemplo, <50, >100, >200. Alternativamente, se puede dar un pequeño bucle (Figura 4); de nuevo, puesto que el indicador es pequeño, hay efecto mínimo sobre la Tm de la doble cadena indicador:amplicón, y se puede determinar la presencia de al menos el número de repeticiones en la sonda.

65 Si la longitud del indicador se incrementa para acomodar repeticiones múltiples (Figura 5), entonces la Tm de la

doble cadena indicador:amplicón será proporcional al número de repeticiones. Porciones no unidas del indicador no emitirán fluorescencia debido a la auto-amortiguación del marcador. Se marcan uno o dos nucleótidos de cada repetición en el indicador. Esta metodología puede ser importante en el análisis de RCT en el que se está interesado en el número de repeticiones de 3, 4 o 5 pb hasta 15x o 20x.

5 Una arquitectura de sonda alternativa se muestra en la Figura 6. Esta sonda es un indicador múltiple, e incluye dos regiones indicadoras separadas unidas por un conector. Cada región indicadora tiene temperaturas de apareamiento diferentes y tiene 1 o más nucleótidos fluorescentes, preferiblemente FAM-T, o diferentes/múltiples colores. El indicador se usa para informar de la presencia de una secuencia específica o variantes de secuencia (por ejemplo, SNP, inserciones, deleciones, etc.). Esto permite que se detecten las secuencias múltiples durante un intervalo prolongado con una sonda única. Cada región se prepara para tener una T<sub>m</sub> similar, pero no idéntica, en el caso de la secuencia de tipo natural, pero una T<sub>m</sub> cambiada en el caso de una mutación para que un usuario solamente tenga que detectar la T<sub>m</sub> cambiada para saber si la variante está presente. Por "similar" se quiere decir que la T<sub>m</sub> difiere como máximo 2, 1,5, 1, 0,5 grados C.

15 La sonda también se podría estructurar, por ejemplo, para proporcionar una región de unión control y región de unión de ensayo.

20 A continuación, se da un ejemplo de uso de una sonda indicadora múltiple para detectar variantes en el gen *rpoB* de *Mycobacterium tuberculosis*. La resistencia a fármaco múltiple en *M. tuberculosis* es compleja. Rifampin es una medicación de *M. tuberculosis* de primera línea y es la diana principal para identificar en el campo antes del tratamiento. *M. tuberculosis* resistentes a Rifampin tienen mutaciones en la región central de 81 pb del gen *rpoB*, la cual codifica la subunidad β de la ARN polimerasa. El 96 % de los aislados clínicos resistentes a Rifampin de *M. tuberculosis* tienen mutaciones en este gen. Las mutaciones en los codones 516, 526 o 531 dan como resultado alto nivel de resistencia a Rifampin. Sin embargo, detectar mutaciones a través de una región génica de 81 pb requerirá generalmente sondas convencionales múltiples, varias de las cuales necesitarían superponerse, requiriendo así etapas de detección múltiple.

30 La Figura 7 muestra codones de tipo natural y variantes en una sección del gen *rpoB*; indicar que se muestran 9 codones a partir de una región de 21 codones (63 nt). Los presentes inventores han diseñado una sonda (mostrada en la Figura 8, SONDA marcada) para detectar e informar de la presencia de cualquiera de los genotipos potenciales de *M. tuberculosis* con mutaciones en los principales codones de resistencia 516, 526 o 531.

35 La Figura 8 también muestra la secuencia de tipo natural (WT) del gen *rpoB*, junto con tres secuencias mutantes (MUT516, MUT526 y MUT531). Las regiones recuadradas de la secuencia WT son aquellas a las que la sonda está diseñada para hibridarse. La sonda por sí misma incluye una primera secuencia indicadora que abarca los codones 531 a 525, incluyendo los potenciales codones mutantes 526 y 531, un conector que comprende cinco residuos de polidesoxiinosina (indicados como "I"), y una segunda secuencia indicadora que abarca los codones 518 a 510, incluyendo el potencial codón mutante 516. Cada una de las secuencias indicadoras incluye dos restos T marcados (resaltados). La segunda secuencia indicadora también es adyacente a un resto bloqueante de PCR.

La unión de sonda a la secuencia WT en ambas posiciones da la misma T<sub>m</sub> de 60 °C para cada secuencia indicadora. La presencia de secuencias mutantes cambia ésta a 52 °C a 56 °C.

45 La Figura 9 muestra los posibles patrones de unión de la sonda a la secuencia diana. Cuando ambas secuencias indicadoras se unen al tipo natural (WT:WT), la T<sub>m</sub> del primer indicador (T<sub>m</sub>1) iguala la T<sub>m</sub> del segundo indicador (T<sub>m</sub>2). Un mutante único en el primer indicador (MUT:WT) baja la T<sub>m</sub> del primer indicador a T<sub>m</sub>3, la cual es inferior a T<sub>m</sub>2. Un mutante único en el segundo indicador (WT:MUT) baja la T<sub>m</sub> del segundo indicador a T<sub>m</sub>4, la cual es inferior a T<sub>m</sub>1. Un mutante único en cada indicador (MUT:MUT) cambia las respectivas T<sub>m</sub> a T<sub>m</sub>3 y T<sub>m</sub>4. Por último, la presencia de un segundo mutante en la primera diana indicadora (MUT:WT) baja la T<sub>m</sub> a T<sub>m</sub>5, la cual es inferior a T<sub>m</sub>3.

55 Cada una de estas T<sub>m</sub> es distinguible una de otra, y así la sonda sencilla se puede usar para determinar la presencia de tres mutaciones distintas en el gen *rpoB*. La Figura 10 da curvas de fusión representativas a partir de tales experimentos. La presencia de secuencias de tipo natural y mutantes se puede distinguir claramente. Por tanto, esta arquitectura de sonda permite la detección múltiple sencilla de mutaciones múltiples a través de un tramo relativamente largo de genoma.

60 Un uso como ejemplo adicional de tal sonda se muestra en las Figuras 11 a 15, las cuales ilustran el uso de una sonda para la detección del virus vaccinia, y que distingue el virus de los pox virus relacionados.

La Figura 11 compara una porción de 30 nucleótidos de la secuencia genómica del virus vaccinia con la de los virus de la viruela del mono, viruela bovina, ectromelia y viruela del camello. Se verá que hay ocho bases que pueden diferir a través de esta secuencia.

65 La Figura 12 ilustra una sonda para la detección de esta región del genoma del virus y distinguir los diversos virus

uno de otros. La sonda incluye una primera región indicadora separada de una segunda región indicadora por una secuencia de cinco residuos de inosina. La primera región indicadora incluye una caperuza de trimetoxiestilbeno. Cada región indicadora incluye dos restos T fluorescentemente marcados.

5 La secuencia sonda es:

5'-GAG T\*AT TTG T\*CA TTT IIIII TAT ATT T\*GT TGG CT\*G-3': SEQ ID NO: 1 Con caperuza de un 5' trimetoxiestilbeno, y un 3' fosfato. I = inosina. T\* = T marcado con fluoresceína.

10 Las dos regiones indicadoras abarcan la secuencia de 30 nucleótidos mostrada en la Figura 11, y ambas cubren varios sitios de variación potenciales.

La Figura 13 compara la sensibilidad de la sonda con la de una sonda de virus vaccinia patrón que abarca la misma secuencia de 30 nt. La sonda conectora es significativamente más sensible en el mismo número de copia. Para comparar la sensibilidad, una porción del genoma se amplifica usando los cebadores y las condiciones indicadas en la Figura 15 y, a continuación, un ensayo de la curva de fusión conducido usando la sonda relevante en el amplicón.

15

La Figura 14 muestra curvas de fusión para la sonda del virus vaccinia ("sonda conectora") y una sonda de virus vaccinia convencional ("sonda patrón") cuando se hibrida con diferentes secuencias diana de pox virus. El intervalo de Tm abarcado por la sonda conectora es mucho más pequeño, pero la sonda conectora proporciona discriminación mejorada entre los diversos pox virus, y permite identificar y distinguir claramente los cinco virus usando una única sonda. El uso de una sonda conectora comparada con una sonda convencional baja la Tm del producto, y baja el intervalo de Tm para permitir la multiplexación mayor en el intervalo. Las sondas de la invención permiten que se identifiquen hasta cinco apareamientos de bases sobre una corta región de 30 bases, permitiendo así la identificación de los pox virus.

20

La Figura 16 muestra secuencias cebador y sonda para analizar una repetición corta en tándem (RCT) en el gen de la tirosina hidroxilasa humana.

30 Los cebadores directos (*forward*) e inversos (*reverse*) son:

TH01-FWD

ATT CAA AGG GTA TCT GGG CTC TGG: SEQ ID NO: 2

35

TH01-REV

GTG GGC TGA AAA GCT CCC GAT TAT: SEQ ID NO: 3

40 El gen de TH incluye una secuencia de repetición corta en tándem (AATG)<sub>n</sub>, la cual se conoce por ser polimórfica. Para confirmar que la presente invención es capaz de distinguir entre diferentes longitudes de RCT, se compararon dos sondas para sensibilidad y precisión. Ambas sondas eran de la misma secuencia, se preparó la sonda 1 sin un conector, mientras que la sonda 2 incluye una secuencia de cinco bases de inosina que actúan como un conector entre las dos porciones de la sonda. Las secuencias sonda eran:

45

TH01-Sonda1

M ACAGACTCCAFG GFGAATGAAFGAATGAGGGAAATAAGGG AGGA P: SEQ ID NO: 4

50 TH01-Sonda2

M ACAGACTCCAFG GFGAATGAAFGAATGA\*\*\*\*\* GGGAAATAAGGG AGGA P: SEQ ID NO: 5

(M = trimetoxiestilbeno; P = fosfato; \* = inosina; F = T marcado con fluoresceína).

55

Después de amplificar la secuencia diana con los cebadores, se realizó el análisis de la curva de fusión con el amplicón y o bien la sonda 1 o la sonda 2. Las curvas se muestran en la Figura 16. Aunque ambas sondas son capaces de distinguir entre 3, 6, 9 y 12 copias de la repetición, la sonda 2, con el conector, proporciona discriminación entre diferentes números de repeticiones durante un intervalo de temperatura mucho más estrecho.

60



**REIVINDICACIONES**

1. Una sonda de ácidos nucleicos que comprende una primera secuencia de ácidos nucleicos de anclaje que es complementaria a una primera secuencia de ácidos nucleicos diana, una segunda secuencia de ácidos nucleicos indicadora que es complementaria a una segunda secuencia de ácidos nucleicos diana y una secuencia de ácidos nucleicos conectora que se une a la primera y la segunda secuencias de ácidos nucleicos, en donde el conector separa las dos secuencias primera y segunda de modo que la temperatura de fusión de la secuencia de anclaje apareada a la primera secuencia de ácidos nucleicos diana y de la secuencia indicadora apareada a la segunda secuencia de ácidos nucleicos diana son distintas y en donde la temperatura de fusión de la secuencia de anclaje apareada a la primera secuencia de ácidos nucleicos diana es mayor que la de la segunda secuencia apareada a la segunda secuencia de ácidos nucleicos diana y la secuencia indicadora comprende secuencias de repetición en tándem, en donde la repetición en tándem comprende al menos 3 copias idénticas de una secuencia en tándem y en donde la sonda no comprende un resto amortiguador.
2. La sonda de la reivindicación 1 en la que el conector comprende un polidesoxirribonucleótido y/o en la que el conector comprende o consiste en polidesoxiinosina.
3. La sonda de cualquier reivindicación anterior en la que el conector es de hasta 5, 10, 15, 20, 30 40, 50 nucleótidos de longitud.
4. La sonda de cualquier reivindicación anterior en la que al menos una de la primera y la segunda secuencias de ácidos nucleicos es una región indicadora que incluye un resto marcado, preferiblemente un marcador fluorescente, preferiblemente en donde la región indicadora es de 15 a 200 nt de longitud, más preferiblemente 15 a 150, más preferiblemente aún 15 a 100 o 20 a 100, 30 a 80, 40 a 60 o alrededor de 50 nt de longitud y/o en donde la región indicadora está diseñada para tener una primera Tm respecto a una secuencia diana enteramente complementaria y una Tm alterada, preferiblemente inferior, respecto a una secuencia diana variante.
5. La sonda de cualquier reivindicación anterior que comprende además una región bloqueante que sirve para bloquear la prolongación de las cadenas de ácidos nucleicos mediante ADN polimerasa.
6. La sonda de cualquier reivindicación anterior en la que tanto la primera como la segunda secuencias de ácidos nucleicos son regiones indicadoras, preferiblemente en la que ambas regiones indicadoras se seleccionan para tener una Tm similar en el caso de unión a la secuencia diana tipo natural y Tm reducidas en el caso de unión de una secuencia diana mutante; preferiblemente la Tm reducida para la primera región indicadora difiere de la de la segunda región indicadora.
7. La sonda de cualquier reivindicación 1 a 5 en la que la Tm de la doble cadena región de anclaje:secuencia diana es significativamente mayor (al menos 0,5, 1, 1,5, pero más generalmente hasta 5 a 20 grados C mayor) que la Tm de la región indicadora a la secuencia diana y/o en donde la región de anclaje es al menos de 50 nt de longitud, más preferiblemente al menos 60, 70, 80, 90, 100, 125, 150 nt.
8. La sonda de la reivindicación 7 en la que la región de anclaje es no más de 200 nt de longitud y opcionalmente en la que la región de anclaje comprende secuencias de repetición en tándem y secuencias únicas.
9. La sonda de cualquiera de las reivindicaciones 1 a 6 en la que la primera y la segunda secuencias de ácidos nucleicos diana se encuentran en el genoma de *M. tuberculosis* o en la que la primera y la segunda secuencias de ácidos nucleicos diana se encuentran en el genoma del pox virus vaccinia.
10. La sonda de cualquier reivindicación anterior en la que la primera y la segunda secuencias diana están dentro de 200, 150, 100, 50, 40, 30, 20, 15, 10, 5, 4, 3, 2, 1 nucleótidos de uno al otro en una secuencia genómica.
11. Una sonda de ácidos nucleicos según cualquier reivindicación anterior que comprende además una o más secuencias conectoras adicionales y una o más secuencias de ácidos nucleicos adicionales para la hibridación con una o más secuencias diana adicionales.
12. Un método para detectar repeticiones en tándem en una secuencia de ácidos nucleicos diana, comprendiendo el método:
  - proporcionar una sonda según cualquier reivindicación anterior, comprendiendo la sonda una secuencia indicadora que tiene varias unidades de repetición en tándem;
  - poner en contacto la sonda con una secuencia de ácidos nucleicos de ensayo, teniendo la secuencia de ensayo una primera secuencia de ácidos nucleicos diana y una segunda secuencia de ácidos nucleicos diana, en donde la primera secuencia diana es complementaria a la secuencia de anclaje de la sonda y la segunda secuencia diana comprende varias repeticiones en tándem complementarias a las unidades de repetición en tándem de la secuencia indicadora;
  - permitir que la secuencia de ácidos nucleicos de anclaje de la sonda se hibride con la primera secuencia de

ácidos nucleicos diana y la segunda secuencia de ácidos nucleicos de la sonda se hibride con la segunda secuencia de ácidos nucleicos diana; y  
detectar la unión o no de la secuencia indicadora a la secuencia de ácidos nucleicos de ensayo.

- 5 13. Un método para detectar uno o más polimorfismos en una secuencia de ácidos nucleicos diana, comprendiendo el método:

10 proporcionar una sonda según una cualquiera de las reivindicaciones 1 a 11;  
poner en contacto la sonda con una secuencia de ácidos nucleicos de ensayo, teniendo la secuencia de ensayo una primera secuencia de ácidos nucleicos diana y una segunda secuencia de ácidos nucleicos diana;  
15 permitir que la primera secuencia de ácidos nucleicos de la sonda se hibride con la primera secuencia de ácidos nucleicos diana y la segunda secuencia de ácidos nucleicos de la sonda se hibride con la segunda secuencia de ácidos nucleicos diana; y  
determinar la  $T_m$  de la primera secuencia sonda hibridada con la primera secuencia diana; y  
15 determinar la  $T_m$  de la segunda secuencia sonda hibridada con la segunda secuencia diana;  
comparar la  $T_m$  determinada de la primera secuencia sonda con la  $T_m$  esperada de la primera secuencia sonda;  
normalizar la  $T_m$  determinada de la segunda secuencia sonda basada en la diferencia de la  $T_m$  determinada y esperada de la primera secuencia sonda;  
20 en donde la  $T_m$  normalizada de la segunda secuencia sonda es indicativa de la presencia de una secuencia de tipo natural o de una secuencia variante.

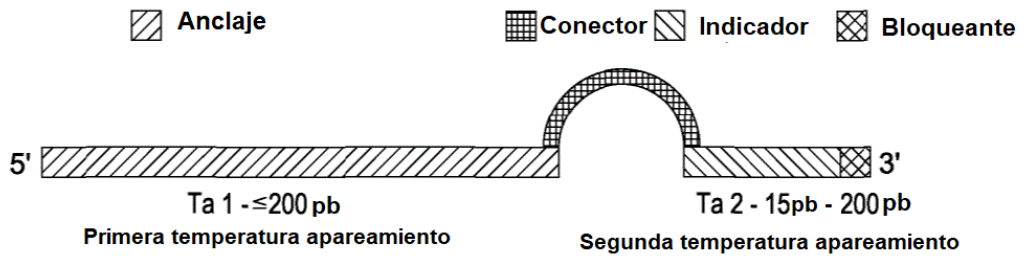


Fig.1

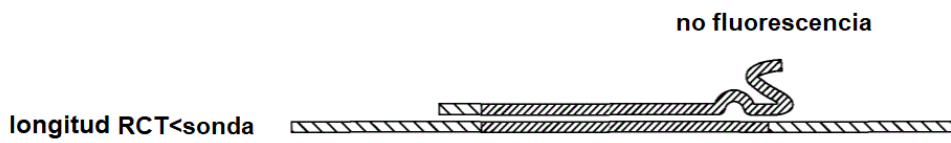


Fig.2

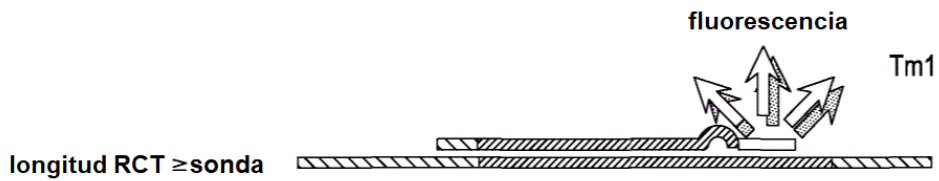


Fig.3

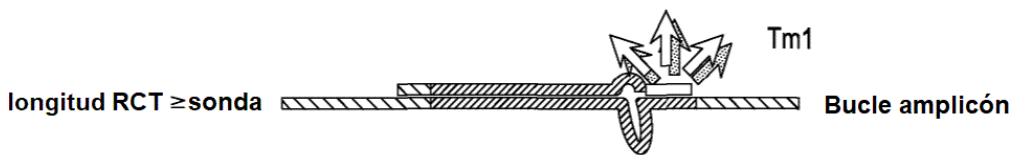


Fig.4

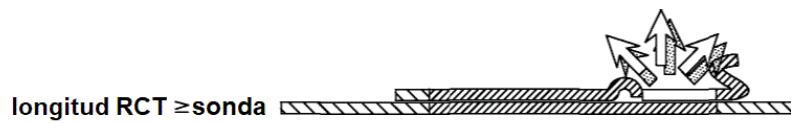


Fig.5

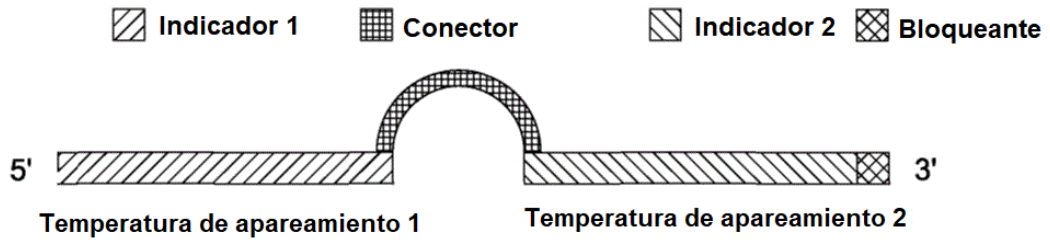


Fig.6

							5	3		3	1	
	Wt	511	512	513	514	515	516	522	526	527	531	533
		L	S	Q	F	M	D	S	H	K	S	L
SEQ ID No: 6		CTG	AGC	CAA	TTC	ATG	GAC	TCG	CAC	AAG	TCG	CTG
SEQ ID No: 7												
SEQ ID No: 8	Mut	CCG	ACC	CTA	TTC	ATT	GTC	CAG	GAC	CAG	TTG	CCG
		2P	3T	IL	1Dup	2I	1IV	1Q	3D	1Q	87L	3P
				AAA			TAC	TGG	CTC		TGG	
				1Q			2Y	1W	1L		2W	
				GAA					TAC		CAG	
				1E			1Del		10Y		1Q	
									CAA			
									3Q			
									AAC			
									2N			
									CCC			
									1P			
									CGC			
									1R			

Fig.7

Secuencia para detectar

SEQ ID No: 9 WT 511 513 515 517 519 521 523 525 527 529 531  
 GCCAGCTGAGCCAAATTCATGGACCAGAACAAACCCGCTGTCGGGGTTGACCCACAAAGCCCGGACTGTGGGGCTG  
 512 514 516 518 520 522 524 526 528 530

SEQ ID No: 10 MUT 516 GCCAGCTGAGCCAAATTCATGGTCAGAACAAACCCGCTGTCGGGGTTGACCCACAAAGCCCGGACTGTGGGGCTG  
 SEQ ID No: 11 MUT 526 GCCAGCTGAGCCAAATTCATGGACCAGAACAAACCCGCTGTCGGGGTTGACCCACAAAGCCCGGACTGTGGGGCTG  
 SEQ ID No: 12 MUT 531 GCCAGCTGAGCCAAATTCATGGACCAGAACAAACCCGCTGTCGGGGTTGACCCACAAAGCCCGGACTGTGGGGCTG

Secuencia sonda

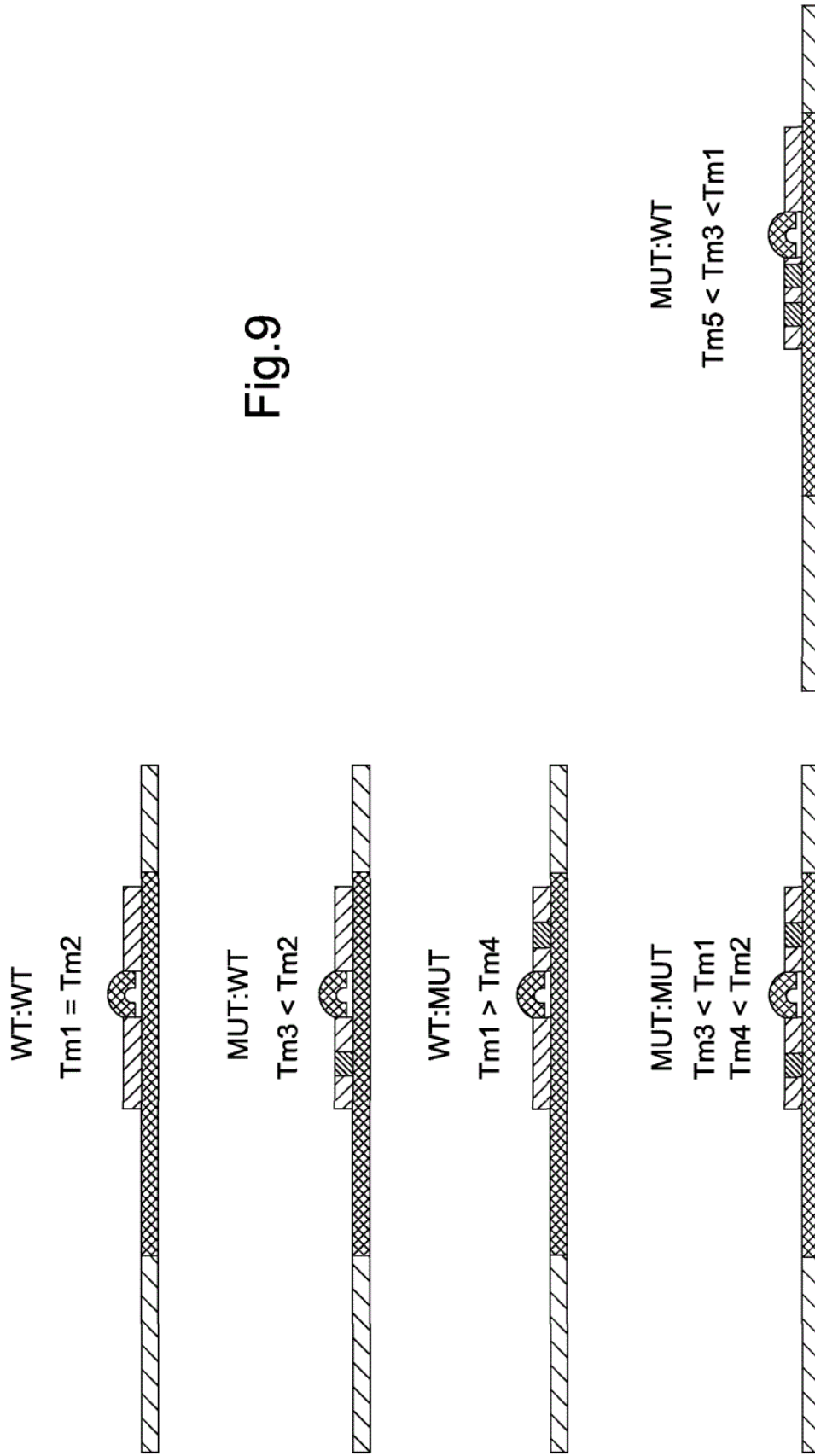
SEQ ID No: 13 SONDA 5' CGACAGCGGGGCTTCGGGT IIIIII TCTGGCCCATGAATGGGCTCAGCTG (BLOQUE) 3'

(IIIIII) = 5x (cinco) restos de polidesoxiinosina

Datos región sonda

Región	Sal de fusión ajustada	Longitud	GC%	Peso	RlnK	deltaG	deltaH	deltaS
rpoB-Z1	67,3	21	67	6494,3	33,404	32,6	190,1	491,7
rpoB-Z2	67,4	25	52	7660,0	33,404	33,8	201,1	523,3

Fig.8



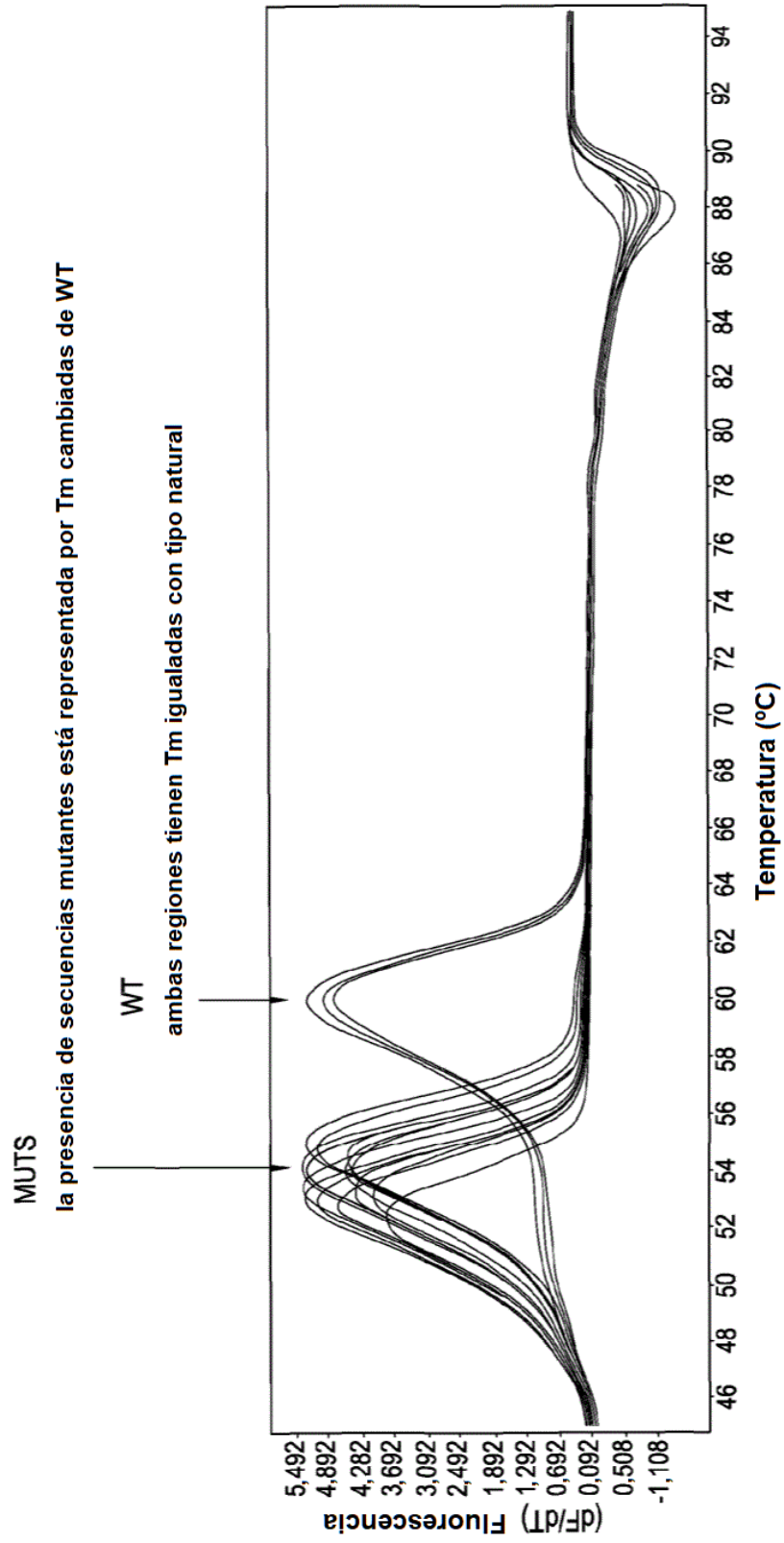


Fig.10

SEQ ID No: 14 virus Vaccinia GAG TAT TTG TCA TTT TAT ATT TGT TGG CTG  
 SEQ ID No: 15 virus de viruela del mono GAG TTT TTG TAA TTT TAT ATT TGT TGG CTG  
 SEQ ID No: 16 virus de viruela bovina GAG TAT TTA TCA TTT TAT ATC TGT TGG GTG  
 SEQ ID No: 17 virus ectromelia GAA CAT TTA TCA TTT TAT ATT TAT TGG GTG  
 SEQ ID No: 18 virus de viruela del camello GAG CAT TTA TCA TTT TAT ATT TGT TGG CTG

Fig.11

VAC\_SONDA2  
 5' MGAGFATTTGFCATTT \*\*\*\*\*TATATTTGGTGGCFGP 3' : SEQ ID NO: 1

M = TRIMETOXIESTILBENO  
 P = 3' FOSFATO  
 F = Fluoresceína dT  
 \*\*\*\*\* = 5x polyI

Fig.12



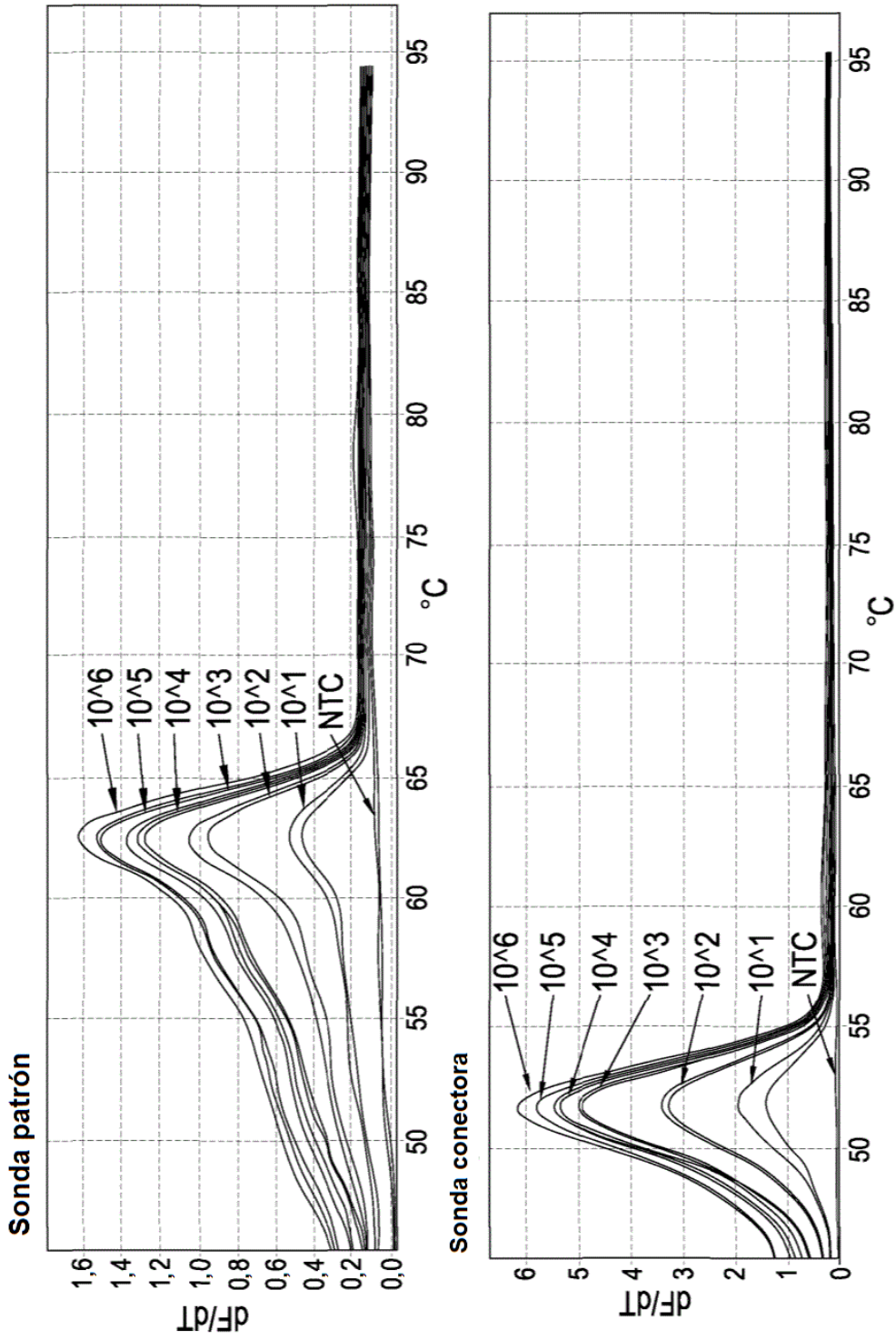


Fig.13

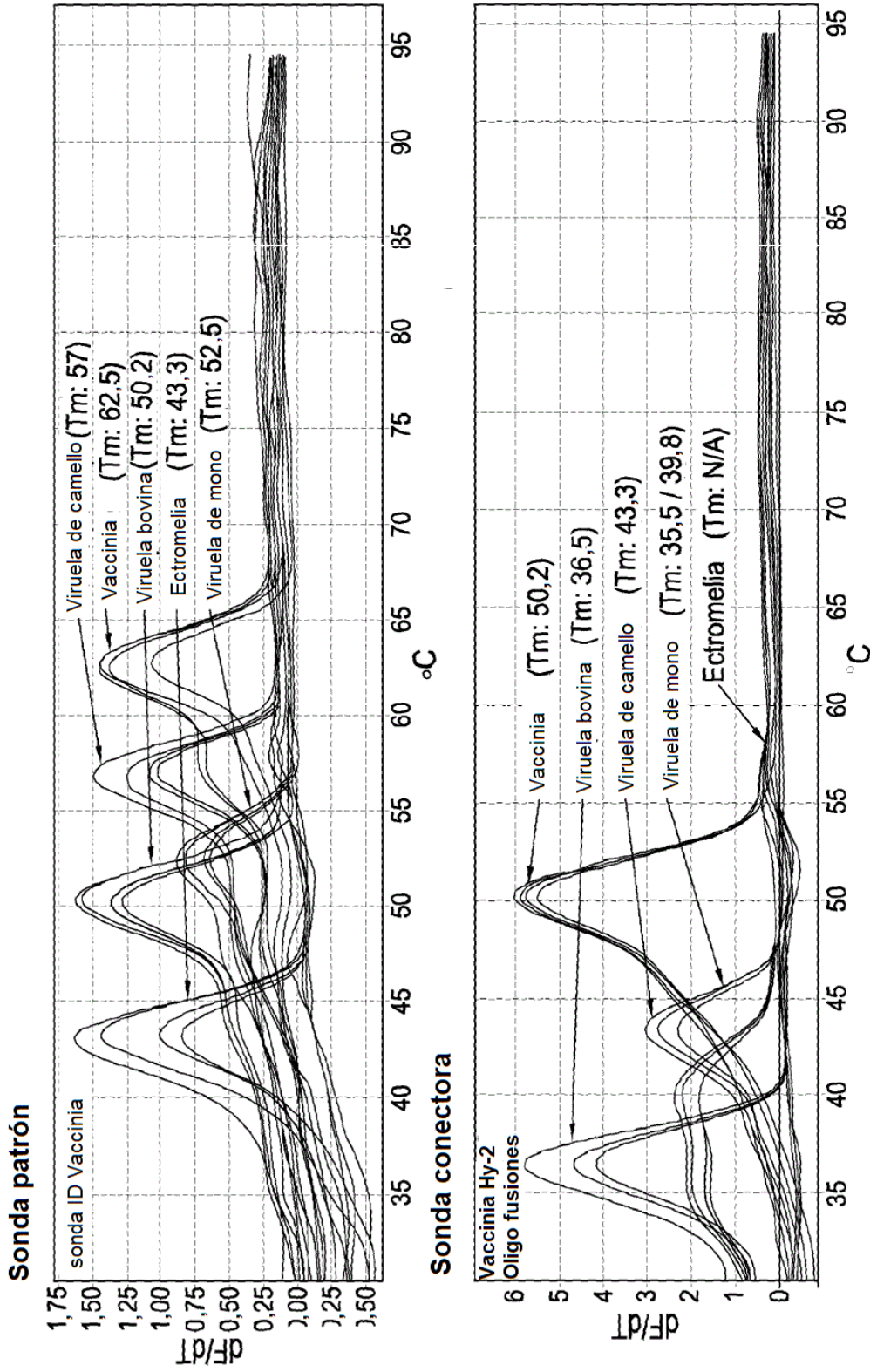


Fig.14

Detalles de amplicón			
Longitud Producto	Tm	Distancia desde 3' UTR	Estructura
142	71	176	Ninguno
		Sí	

Detalles Cebador				
Cebador sentido	Posición	Tm	GC%	Dímero
CAACTAGTGTGATTAATATGTGACACGTT	364	60,2	34,5	-4,6
				-4,6

Cebador antisentido	Posición	Tm	GC%	Dímero
CCATCAAGATCCCTACCAACCC	505	60,2	54,5	-5,1
				-2,2

Información adicional

3GAG1ATTTG1CATT\*\*\*\*\*TATATT1GTTGGC1G2

1= fluoresceína dT

2= fosfato

3= TRIMETOXISTILBENO

\*= Inosina

Fig.15

## ES 2 642 275 T3

TH01-Sonda 1

M ACAGACTCCAFG G~~F~~GAATGAA~~F~~GAATGAGGGAAATAAGGG AGGA P: SEQ ID  
NO: 4

TH01-Sonda2

M ACAGACTCCAFG G~~F~~GAATGAA~~F~~GAATGA\*\*\*\*\* GGGAAATAAGGG AGGA P: SEQ  
ID NO: 5

TH01-FWD

ATT CAA AGG GTA TCT GGG CTC TGG: SEQ ID NO: 2

TH01-REV

GTG GGC TGA AAA GCT CCC GAT TAT: SEQ ID NO: 3

### Fig.16

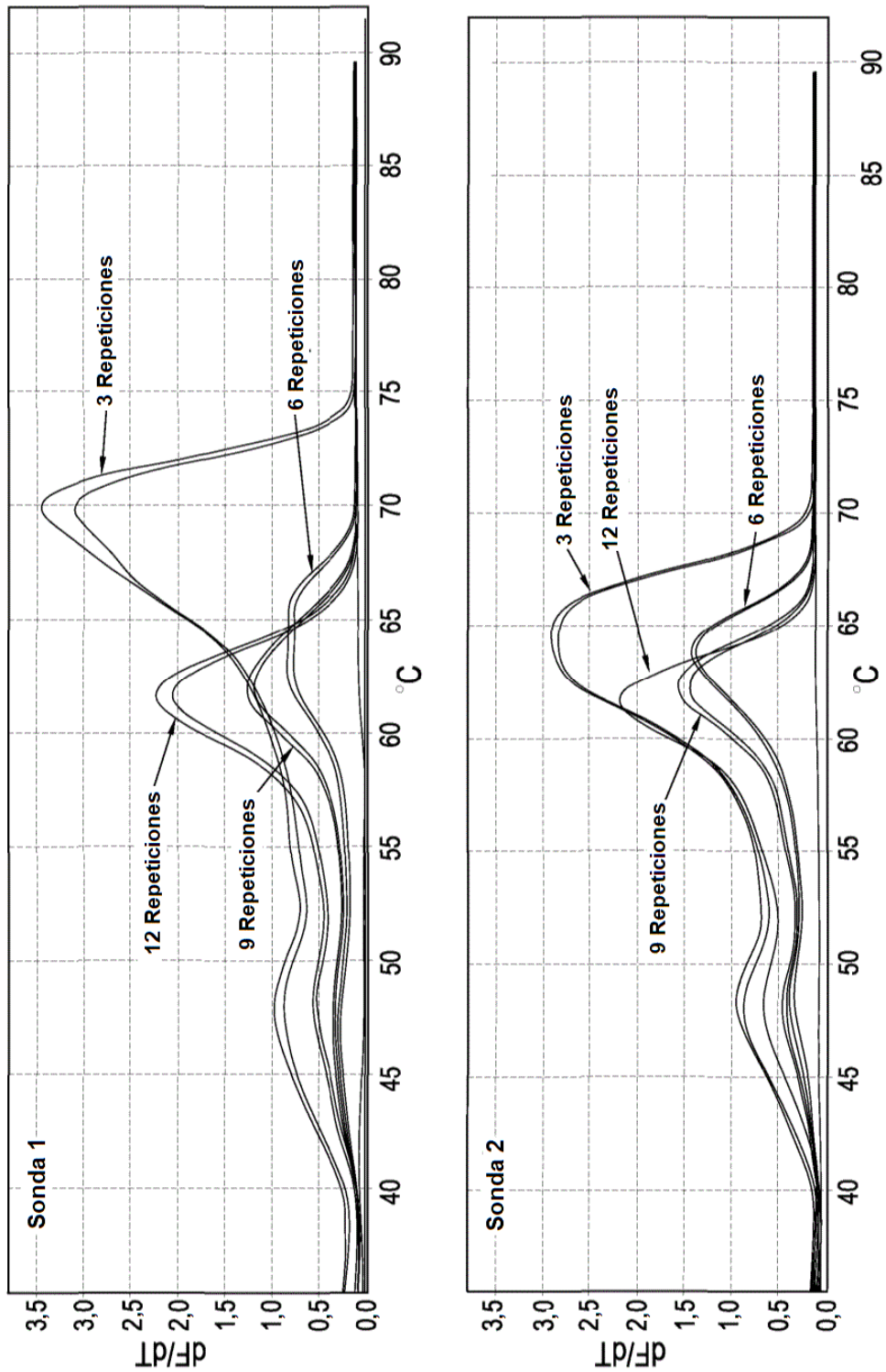


Fig.16 (continuación)