

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 615 377**

51 Int. Cl.:

**G06F 19/24** (2011.01)

**G06F 19/20** (2011.01)

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **20.09.2006 PCT/US2006/036605**

87 Fecha y número de publicación internacional: **29.03.2007 WO07035806**

96 Fecha de presentación y número de la solicitud europea: **20.09.2006 E 06803889 (2)**

97 Fecha y número de publicación de la concesión europea: **14.12.2016 EP 1927064**

54 Título: **Análisis de curva de fusión con sustracción del fondo exponencial**

30 Prioridad:

**20.09.2005 US 719250 P**

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

**06.06.2017**

73 Titular/es:

**UNIVERSITY OF UTAH RESEARCH FOUNDATION  
(100.0%)  
615 Arapeen Drive, Suite 310  
Salt Lake City, UT 84108, US**

72 Inventor/es:

**PALAIS, ROBERT, ANDREW y  
WITTWER, CARL, THOMAS**

74 Agente/Representante:

**DURÁN MOYA, Luis Alfonso**

**ES 2 615 377 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**

Análisis de curva de fusión con sustracción del fondo exponencial

5 **SECTOR DE LA INVENCION**

En general, la presente invención se refiere al análisis de la curva de fusión de ácidos nucleicos. Más específicamente, varias realizaciones de la presente invención se refieren a procedimientos y sistemas para analizar los perfiles de fusión de ácidos nucleicos de doble cadena mediante la eliminación de las señales de fluorescencia de fondo.

10 **ANTECEDENTES DE LA INVENCION**

Los procedimientos para el análisis de la variación de secuencia de ADN se pueden dividir en dos categorías generales: 1) determinación del genotipo de variantes de secuencia conocida y 2) exploración de variantes desconocidas. Existen muchos procedimientos para la determinación del genotipo de variantes de secuencia conocida y están disponibles procedimientos de tubo cerrado, homogéneos, de una sola etapa, que utilizan sondas fluorescentes (Lay M. J. y otros, Clin. Chem 1997; 43: 2262-7). En cambio, la mayoría de las técnicas de exploración de variantes desconocidas requieren electroforesis en gel o separación en columna después de la PCR. Entre éstas se incluyen el polimorfismo de conformación de cadena única (Orita O. y otros, Proc Natl Acad Sci USA 1989; 86: 2766-70), la migración de heterodúplex (Nataraj A. J. y otros, Electrophoresis 1999; 20: 1177-1185), electroforesis en gel en gradiente desnaturalizante (Abrams E. S. y otros, Genomics 1990; 7: 463-75), electroforesis en gradiente de temperatura (Wartell R. M. y otros, J Chromatogr A 1998; 806: 169-85), procedimientos de escisión química o enzimática (Taylor G. R. y otros, Genet Anal 1999; 14: 181-6), así como la secuenciación de ADN. La identificación de nuevas mutaciones por secuenciación también requiere múltiples etapas después de la PCR, a saber, la secuenciación en ciclos y la electroforesis en gel. La cromatografía líquida de alto rendimiento desnaturalizante (Xiao W. y otros, Hum Mutat 2001; 17: 439-74) implica inyectar el producto de PCR en una columna.

Los polimorfismos de nucleótido único (SNP) son con diferencia las variaciones genéticas más comunes observadas en el hombre y otras especies. En estos polimorfismos, solamente una única base varía entre los individuos. La alteración puede causar un cambio de aminoácido en una proteína, alterar las tasas de transcripción, afectar al empalme y corte del ARNm o no tener ningún efecto visible sobre los procesos celulares. A veces, cuando el cambio es silencioso (por ejemplo, cuando el aminoácido para el que codifica no cambia), el genotipado de SNP aún puede ser útil si la alteración está unida (asociada) a un único fenotipo causado por otra alteración genética.

Existen muchos procedimientos para el genotipado de SNP. La mayoría utiliza PCR u otras técnicas de amplificación para amplificar la plantilla de interés. Se pueden utilizar técnicas analíticas simultáneas o sucesivas, incluyendo electroforesis en gel, espectrometría de masas y fluorescencia. Son atractivas las técnicas de fluorescencia que son homogéneas y no requieren la adición de reactivos después del comienzo de la amplificación o el muestreo físico de las reacciones para el análisis. Algunas técnicas homogéneas de ejemplo utilizan cebadores de oligonucleótidos para localizar la región de interés y marcadores o colorantes fluorescentes para la generación de señales. Varios procedimientos basados en PCR son de tubo completamente cerrado y utilizan una enzima termoestable que es estable a la temperatura de desnaturalización del ADN, de manera que después de empezar el calentamiento, no es necesaria ninguna adición.

Para el genotipado de SNP están disponibles varios procedimientos de PCR fluorescentes, homogéneos, de tubo cerrado. Entre éstos se incluyen sistemas que utilizan sondas de oligonucleótidos FRET con dos cromóforos que interactúan (sondas de hibridación adyacentes, sondas TaqMan<sup>®</sup>, balizas moleculares, Scorpions), sondas de un solo oligonucleótido con un solo fluoróforo (G-quenching probes, Crockett, A. O. y C. T. Wittwer, Anal. Biochem. 2001; 290: 89-97 y SimpleProbes<sup>®</sup>, Idaho Technology) y técnicas que utilizan un colorante de ADN de doble cadena en lugar de sondas de oligonucleótidos covalentes marcadas con fluorescencia.

Los procedimientos de PCR que controlan la fusión del ADN con colorantes fluorescentes de ADN de doble cadena se han popularizado conjuntamente con la PCR en tiempo real. Dado que la PCR produce ADN suficiente para el análisis de fusión fluorescente, tanto la amplificación como el análisis se pueden realizar en el mismo tubo, proporcionando un sistema homogéneo, de tubo cerrado, que no requiere etapas de procesamiento o separación. Los colorantes de ADN de doble cadena se utilizan habitualmente para identificar productos por su temperatura de fusión o Tf.

El poder del análisis de fusión del ADN depende de su resolución. Los estudios con absorbancia UV a menudo requerían horas para recoger datos de alta resolución a una velocidad de 0,1-1,0°C/min para asegurar el equilibrio. En cambio, el análisis de fusión fluorescente se adquiere a menudo a 0,1-1,0°C/s y la resolución se limita a 2-4 puntos/C. Con los recientes avances en electrónica (por ejemplo, los convertidores A-D de 24 bits), la fusión a alta resolución se puede realizar rápidamente con 10-100 veces la densidad de datos (50-100 puntos/°C) de los instrumentos de PCR en tiempo real convencionales, tal como se ha demostrado recientemente para la fusión de la

sonda y el producto de PCR. Además, ahora están disponibles colorantes saturantes de ADN, tales como LCGreen® Plus (Idaho Technology, Salt Lake City, UT), que maximizan la detección de las dobles cadenas no emparejadas (heterodúplex) (véase, por ejemplo, las publicaciones de las patentes de Estados Unidos No. 2005/0233335 y 2006/0019253). Estos dos desarrollos aumentan de forma importante el poder de la fusión del ADN basada en fluorescencia para la identificación robusta de cambios de una sola base dentro de los productos de PCR.

El análisis de fusión de alta resolución para la exploración de genes se basa principalmente en la forma de la transición de fusión de los productos de la PCR. Un procedimiento disponible para el cribado de polimorfismos de nucleótido único (SNP) heterocigotos dentro de productos de hasta 1.000 pb tiene una sensibilidad y una especificidad del 97% y del 99%, respectivamente. En muchos casos, el análisis de alta resolución de la transición de fusión también permite el genotipado sin sondas. Se puede obtener una especificidad incluso mayor para la discriminación de variantes sobre una región más pequeña mediante la utilización de sondas no marcadas. Los genotipos específicos se deducen mediante la correlación de alteraciones de la secuencia bajo la sonda con los cambios en la Tf de la sonda. Con los recientes avances en colorantes e instrumentación, la exploración de alta resolución de genes y el genotipado con sondas no marcadas se pueden realizar opcionalmente de forma simultánea en la misma reacción. Se pueden observar las transiciones de fusión del producto de la PCR y la sonda en presencia de un colorante saturante de ADN. Además de cribar cualquier variante de secuencia entre los cebadores en el producto de la PCR, se pueden genotipar los polimorfismos y mutaciones habituales. Además, la agrupación jerárquica imparcial puede agrupar con precisión las curvas de fusión en genotipos. Se pueden utilizar una, dos o incluso más sondas no marcadas en una sola PCR.

En el genotipado y la exploración simultáneos, el análisis de fusión del producto detecta variantes de secuencia en cualquier lugar entre dos cebadores, mientras que el análisis de fusión de la sonda identifica variantes bajo una sonda. Si una variante de secuencia se encuentra entre los cebadores y bajo una sonda, se obtienen tanto la presencia de una variante como su genotipo. Si la fusión del producto indica una variante, pero la sonda no, entonces la variación probablemente se produce entre los cebadores, pero no bajo la sonda, y es necesario un análisis adicional para el genotipado. Las sondas se pueden colocar en los sitios de variación de secuencia habitual, de manera que, en la mayoría de los casos, si la exploración del producto es positiva, las sondas identificarán las variantes de secuencia, reduciendo en gran medida la necesidad de secuenciación. Con una sonda, se puede establecer el genotipo de un SNP mediante la fusión del producto de la PCR y la sonda. Con dos sondas, se puede examinar el genotipo de dos regiones separadas de la secuencia y el resto del producto de la PCR se puede explorar en busca de variantes de secuencia raras. Se pueden utilizar múltiples sondas si difieren en la temperatura de fusión y si cada alelo presenta un patrón único de fusión de la sonda y/o del producto.

En un ejemplo ilustrativo, se criba una población para detectar mutaciones de fibrosis quística. Dado que sólo el 3,8% de los caucásicos son portadores de fibrosis quística, se esperaría que el 96,2% de los individuos cribados al azar dieran negativo mediante la secuenciación completa (exón y sitio de empalme). Con 27 exones, el porcentaje de ciclos de secuenciación que se espera que sean positivos es inferior al 0,14%. Es decir, sólo aproximadamente 1 de cada 1.000 ciclos de secuenciación sería útil. Esta es la razón por la que la secuenciación no se recomienda para el cribado de la fibrosis quística. En cambio, habitualmente se realiza un panel de mutaciones seleccionadas que detecta el 83,7% de los alelos de la fibrosis quística.

De manera alternativa, se considera la exploración y el genotipado simultáneos para el cribado de la fibrosis quística mediante fusión de alta resolución. Si la longitud del amplicón se mantiene por debajo de 400 pb, la sensibilidad de la exploración de alta resolución se aproxima al 100,0%. Si se analizan las mutaciones y polimorfismos habituales con sondas no marcadas en la misma reacción, entonces también se genotiparán aproximadamente el 80% de las mutaciones. En comparación con el cribado mediante secuenciación *de novo*, la carga de la secuenciación se puede reducir en un 99,97%.

Los procedimientos de genotipado de tubo cerrado que utilizan análisis de fusión tienen la capacidad de explorar variantes inesperadas. Los procedimientos de fusión también utilizan menos sondas, y menos complejas, que los procedimientos específicos de alelo, que requieren una sonda para cada alelo analizado. La discriminación de alelos mediante la Tf o la forma de la curva es una opción interesante para el color fluorescente. Los colorantes que, en general, tiñen el ADN de doble cadena son atractivos por su simplicidad y coste. Aunque la fiabilidad del genotipado mediante la fusión del amplicón es objeto de controversia, un estudio reciente encontró que 21 de los 21 pares de heterodúplex analizados eran distinguibles mediante fusión de alta resolución de amplicones pequeños (Graham R, Liew M, Meadows C, Lyon E, Wittwer CT. Distinguishing different DNA heterozygotes by high-resolution melting. Clin Chem 2005; 51).

A pesar de que las variantes de secuencia habituales normalmente se pueden genotipar con una o dos sondas no marcadas en la misma reacción, también se pueden utilizar más de dos sondas y/o reacciones secuenciales. Por ejemplo, múltiples sondas solapadas pueden localizar variantes raras inesperadas en la región cubierta por una sonda. Se pueden diseñar sondas adicionales para identificar la posición exacta y la secuencia de la variación. Sin embargo, la secuenciación de ADN es una estrategia más directa para la identificación de nuevas variantes, previamente desconocidas, en particular cuando la región amplificada es muy variable. No obstante, en la gran mayoría de los análisis genéticos, la secuencia de tipo natural amplificada es conocida y las posibles variantes

habituales son limitadas. En estos casos, la exploración y el genotipado se pueden realizar en una sola etapa mediante la fusión del ADN con oligonucleótidos simples. No se requieren sondas fluorescentes ni separaciones, y la amplificación (15 min) y el análisis de fusión (1-2 min) pueden ser rápidos.

5 Tal como se ha descrito anteriormente, el genotipado y la exploración simultáneos, así como otras técnicas de genotipado que utilizan el análisis de fusión, han sido áreas de investigación prometedoras. Sin embargo, el análisis de la curva de fusión antes de disponer de capacidad de alta resolución proporcionaba falta de especificidad y precisión. Con la llegada del análisis de la curva de fusión de alta resolución, el ruido de la fluorescencia de fondo puede interferir con la utilización de curvas de fusión para genotipar de manera precisa los SNP, detectar variaciones de secuencia y detectar mutaciones. Dependiendo del amplicón, las técnicas anteriores de eliminación de la fluorescencia de fondo han dado lugar a la obtención de datos erróneos. Por ejemplo, la técnica de la línea basal utiliza la extrapolación lineal como procedimiento para normalizar las curvas de fusión y eliminar la fluorescencia de fondo. Esta técnica funciona bien con sondas marcadas. Sin embargo, ésta y otras técnicas de normalización anteriores no han funcionado tan bien con sondas no marcadas (Zhou L, Myers AN, Vandersteen JG, Wang L, Wittwer CT. Closed-Tube Genotyping with Unlabeled Oligonucleotide Probes and a Saturating DNA Dye. Clin Chem. 2004; 50: 1328-1335), fusión de amplicones múltiplex pequeños (Liew M, Nelson L, Margraf R, S Mitchell, Erali M, Mao R, Lyon E, Wittwer CT. Genotyping of human platelet antigens 1-6 and 15 by high-resolution amplicon melting and conventional hybridization probes. J Mol Diag, 2006; 8: 97-104) y la fusión combinada de amplicones y sondas no marcadas (Zhou L, Wang L, Palais R, Pryor R, Wittwer CT. High-resolution analysis for simultaneous mutation scanning and genotyping in solution. Clin Chem 2005; 51: 1770-7), ni funcionan tan bien para amplicones pequeños. Esto es debido, como mínimo, en parte, a que los procedimientos de fusión de sondas no marcadas y amplicones pequeños a menudo requieren la sustracción del fondo a temperaturas más bajas (40-80°C) a las habituales para la fusión de amplicones estándar a 80-95°C. A estas temperaturas más bajas, la línea basal de baja temperatura no es lineal, sino una curva con un rápido aumento de la fluorescencia a bajas temperaturas. Cuando se utiliza la extrapolación lineal, las líneas se pueden cruzar antes de completarse la transición de fusión y, cuando esto ocurre, las técnicas anteriores no proporcionan el medio más preciso para el análisis de la curva de fusión, en parte debido a su dependencia matemática de la fluorescencia absoluta.

30 Sería ventajoso un sistema y un procedimiento para genotipar SNP, detectar variaciones de secuencia y/o detectar mutaciones con una alta precisión en ácidos nucleicos de doble cadena a través de la utilización de técnicas de perfil de fusión de alta resolución. Además, sería adicionalmente ventajoso que la fluorescencia de fondo se pudiera separar de manera automática y precisa de un perfil de fusión de una muestra de ácido nucleico de doble cadena. Sería adicionalmente ventajoso que el sistema y el procedimiento realizaran un análisis preciso de la curva de fusión para amplicones pequeños y grandes, así como con sondas no marcadas.

### 35 CARACTERÍSTICAS DE LA INVENCION

En un aspecto de la presente invención, se da a conocer un procedimiento para generar una curva de fusión corregida, tal como se define en la reivindicación 1.

40 En otro aspecto de la presente invención, se da a conocer un sistema para generar una curva de fusión, tal como se define en la reivindicación 19.

45 Otras características de la presente invención serán evidentes para los expertos en la técnica tras la consideración de la siguiente descripción detallada de realizaciones preferentes, que ejemplifican el mejor modo de llevar a cabo la presente invención, tal como se considera actualmente.

### DESCRIPCIÓN BREVE DE LOS DIBUJOS

50 Las **figuras 1A-B** son curvas de fusión de alta resolución de un gen del factor V Leiden diana examinado con una sonda no marcada. La **figura 1A** muestra las curvas de fusión originales. La **figura 1B** muestra el genotipado después de la sustracción del fondo exponencial.

55 La **figura 2A** muestra las curvas de fusión originales del gen del factor V Leiden diana (parte superior) y las curvas de fusión después de la sustracción del fondo exponencial y la función mejorada con la superposición de curvas (parte inferior). La **figura 2B** muestra el intento fallido de una función previa con superposición de curvas en la curva de fusión original de la **figura 2A**.

60 Las **figuras 3A-C** muestran las curvas de fusión analizadas del gen de la lipasa hepática. La **figura 3A** muestra una curva de fusión original (panel superior) y un intento fallido de genotipado de la muestra mediante la función previa de genotipado (panel inferior). La **figura 3B** muestra una curva de fusión original (panel superior) y un resultado exitoso del genotipado de la muestra mediante la nueva función de agrupamiento (panel inferior). La **figura 3C** proporciona la curva de fusión de la **figura 3B** (panel inferior) a lo largo de la placa de reacción de 96 pocillos.

65 Las **figuras 4A-F** muestran una sonda no marcada y el genotipado del amplicón completo del gen de hemocromatosis diana. La **figura 4A** es una curva de fusión original de alta resolución. La **figura 4B** es un gráfico

de la derivada negativa de la curva de fusión de la **figura 4A**. La **figura 4C** es el resultado de la sustracción del fondo exponencial de la curva de fusión de la **figura 4B**. La **figura 4D** muestra la curva de fusión original y la **figura 4E** muestra una normalización lineal fallida de línea basal de los datos de la **figura 4D**. La **figura 4F** muestra el genotipado del gen de hemocromatosis después de la sustracción del fondo exponencial.

Las **figuras 5A-D** muestran las curvas de fusión de alta resolución del locus del factor V Leiden y la agrupación jerárquica imparcial después de varias funciones. La **figura 5A** muestra los datos originales de la curva de fusión. La **figura 5B** muestra la derivada negativa de los datos de fusión originales. La **figura 5C** muestra una sustracción de la línea basal lineal realizada sobre los datos de la **figura 5B**. La **figura 5D** muestra la función de sustracción del fondo exponencial realizada sobre los datos de la **figura 5A**, seguida de la normalización y el trazado como la derivada negativa.

Las **figuras 6A-D** muestran las curvas de fusión de alta resolución del locus del factor V Leiden con genotipado con sonda no marcada. La **figura 6A** muestra los datos originales de la curva de fusión. La **figura 6B** muestra el resultado de un gráfico de la derivada negativa de los datos de la **figura 6A**. La **figura 6C** muestra un gráfico de la derivada negativa de la región de la sonda después de la sustracción del fondo exponencial utilizando pendientes de las regiones indicadas en los datos de la **figura 6B**. La **figura 6D** muestra la agrupación de 3 genotipos realizada mediante la función de agrupación de los datos de la **figura 6C**.

Las **figuras 7A-D** muestran los datos de exploración y genotipado del exón 11 del gen regulador de la transconductancia de la fibrosis quística (CFTR). La **figura 7A** muestra las secuencias variantes analizadas con las sondas no marcadas. La **figura 7B** muestra las curvas de fusión normalizadas después de la sustracción del fondo exponencial. La **figura 7C** muestra el gráfico de la derivada negativa de la región de sonda. La **figura 7D** muestra un gráfico de la diferencia de la transición de fusión del producto de la PCR.

Las **figuras 8A-E** muestran las curvas de fusión de alta resolución del exón 10 del gen CFTR. La **figura 8A** muestra las secuencias variantes con las sondas. La **figura 8B** muestra una curva de fusión normalizada después de la sustracción del fondo exponencial. La **figura 8C** muestra el gráfico de la derivada negativa realizada sobre los datos de la **figura 8B**. La **figura 8D** muestra un gráfico de la diferencia de la transición de fusión del producto de la PCR.

La **figura 9A** muestra las curvas de fusión originales para un amplicón de  $\beta$ -globina que incluye los locus SNP HbS y HbC. La **figura 9B** muestra los datos de las curvas de fusión después de la sustracción del fondo exponencial. La **figura 10C** muestra los genotipos agrupados de las curvas de fusión normalizadas de la **figura 9B**.

Las **figuras 10A-C** muestran las curvas de fusión de alta resolución de 60 tipos naturales de muestra en el locus del factor V Leiden. La **figura 10A** es una curva de fusión normalizada después de la sustracción del fondo exponencial. La **figura 10B** muestra el gráfico de diferencia después de la técnica de representación de la diferencia vertical previa. La **figura 10C** muestra el gráfico de diferencia después de la técnica de representación de la diferencia ortogonal.

La **figura 11** es un diagrama de bloques de un ejemplo ilustrativo del sistema de análisis nucleico.

#### DESCRIPCIÓN DETALLADA DE LAS REALIZACIONES PREFERENTES

En referencia a la figura 1, se muestra un perfil de curva de fusión antes (figura 1A) y después (figura 1B) de realizar la normalización con la sustracción del fondo exponencial (SFE). La SFE es un procedimiento para normalizar los datos de la curva de fusión original y proporciona un conjunto de datos mejor para el análisis y el genotipado. La realización de la SFE sobre una curva de fusión derivada u original da lugar a una curva de fusión corregida más adecuada para un análisis detallado. Las curvas de fusión originales (figura 1A) a menudo representan valores de fluorescencia en función de la temperatura.

En un ejemplo, la sustracción del fondo exponencial se calcula mediante el ajuste de la pendiente de la curva de fusión original a dos temperaturas,  $T_L$  y  $T_R$ . La curva de fusión original está representada por el conjunto de ecuaciones (1), a continuación, en las que  $F(T)$  representa la curva de fusión original,  $M(T)$  representa la señal de la muestra de ácido nucleico, y en las que  $B(T)$  representa la señal de fondo.  $T_L$  y  $T_R$  se obtienen a partir de puntos lejanos a las temperaturas de transición de fusión de la señal de la muestra, donde la transición de fusión no afecta significativamente a la pendiente, por lo tanto, la pendiente ( $M'(T)$ ) de la señal ( $M(T)$ ) es esencialmente cero y desaparece de manera eficaz de manera exponencial. Esto se refleja en el conjunto de ecuaciones (2).

Conjunto de ecuaciones (1)  $F(T) = M(T) + B(T)$

Conjunto de ecuaciones (2)  $F'(T_L) = B'(T_L)$  y  $F'(T_R) = B'(T_R)$

Para la ecuación 3 se ajusta un modelo exponencial, en la que la forma exponencial se desplaza hasta  $T_L$  para la estabilidad numérica a estos dos valores:  $B'(T) = aCe^{a(T-T_L)}$  a  $T = T_R$ ,  $T_L$ , a  $T = T_L$ , esto da  $aC = B'(T_L)$  y a  $T_R$  esto da  $aCe^{a(T_R-T_L)} = B'(T_R)$ .

Conjunto de ecuaciones (3)

$$B(T) = Ce^{a(T-T_L)}$$

$$B'(T) = aCe^{a(T-T_L)} \text{ a } T = T_R, T_L$$

$$aC = B'(T_L) \text{ a } T = T_L$$

$$aCe^{a(T_R-T_L)} = B'(T_R) \text{ a } T = T_R$$

Se entiende que  $T_L$  y  $T_R$  se han medido al generar la curva de fusión original y, por lo tanto, se utilizan estos valores para obtener los parámetros (a) y (C), tal como se muestra en el conjunto de ecuaciones (4).

Conjunto de ecuaciones (4)

$$e^{a(T_R-T_L)} = B'(T_R)/B'(T_L), \text{ de manera que}$$

$$\ln(B'(T_R)/B'(T_L))$$

$$a = \frac{\ln(B'(T_R)/B'(T_L))}{(T_R - T_L)}$$

$$C = B'(T_L)/a$$

$$M(T) = F(T) - Ce^{a(T-T_L)} \text{ (fondo eliminado)}$$

Dado que las pendientes de  $T_L$  y  $T_R$  se utilizan para determinar el fondo exponencial en lugar de los valores de fluorescencia, la sustracción del fondo se puede calcular sin hacer referencia a la cantidad de señal presente, la cual puede variar debido a la cantidad de materiales presentes o simplemente la variación de muestra a muestra.

La señal  $M(T)$ , opcionalmente, se puede normalizar, de manera ilustrativa al intervalo 0-100 mediante la aplicación del desplazamiento y reescalado lineal según el conjunto de ecuaciones (5) en el intervalo de interés.

Conjunto de ecuaciones (5)

$$M(T) = 100(M(T) - m)/(M - m),$$

en el que  $m = \min\{M(T)\}$  y  $M = \max\{M(T)\}$

De manera alternativa, en otro ejemplo, se puede ajustar un exponencial a la fluorescencia de fondo de una curva de fusión derivada. El fondo se elimina mediante el ajuste de la altura de una curva derivada numéricamente por ordenador con un exponencial, a continuación, se sustrae el fondo de la curva de fusión original. Dado que la derivada de un exponencial es un exponencial con el mismo grado de disminución, la sustracción del fondo exponencial se aplica en la presente realización de la curva derivada mediante la sustracción de un ajuste exponencial de los valores a las temperaturas de interés. En este procedimiento, se puede utilizar el valor (altura) de la curva derivada colectiva a dos temperaturas  $T_L$  y  $T_R$  y ajustar estos valores al modelo correspondiente para  $B'(t) = De^{a(T-T_L)}$ , en la que D corresponde a  $aC$  de la derivada anterior. En esta situación, los parámetros D y a se resuelven de la siguiente manera. A  $T = T_L$ , esto da  $D = B'(T_L)$ , así que no hay necesidad de resolver para el parámetro D; aparece directamente como una medición. A  $T_R$  esto da  $De^{a(T_R-T_L)} = B'(T_R)$ . Dividiendo la segunda ecuación entre la primera se obtiene  $e^{a(T_R-T_L)} = B'(T_R)/B'(T_L)$ , de manera que  $a = \ln(B'(T_R)/B'(T_L))/(T_R - T_L)$  es coherente con el procedimiento anterior (los grados de disminución exponencial de un exponencial y su derivada son los mismos), aunque ahora los valores de B' se determinan a partir de la altura de la derivada numérica de los datos medidos en lugar del ajuste de la pendiente de los datos medidos. Finalmente, se obtiene la derivada de la señal con la derivada del fondo eliminado mediante sustracción:  $M'(T) = F'(T) - De^{a(T-T_L)}$ , con los parámetros D y a determinados como antes. Al igual que antes, la señal derivada  $M'(T)$ , opcionalmente, puede normalizarse, de manera ilustrativa hasta el intervalo 0-100 mediante la aplicación del desplazamiento y reescalado lineal  $M'(T) = 100(M'(T) - m)/(M' - m)$ , en la que  $m = \min\{M'(T)\}$  y  $M' = \max\{M'(T)\}$  en el intervalo de interés, respectivamente.

En la figura 1, se muestran curvas de fusión ilustrativas para varios genotipos en un sistema modelo del gen del factor V Leiden utilizando una sonda no marcada. Las transiciones para la fusión de la sonda no marcada y del amplicón son visibles. En este ejemplo ilustrativo, los pares de líneas 1 (líneas 1, 2) y 2 (líneas 3, 4), tal como se muestran en la figura 1A, representan los pares de cursores respectivos. Cada par de cursores proporciona una región fuera de la transición de fusión para la extracción de los intervalos de temperatura de los datos de la curva de fusión original, que se utilizan para la determinación de  $F'(T_L)$  y  $F'(T_R)$  en el análisis de la SFE. Cada una de las dos regiones se selecciona de manera que sea lo suficientemente pequeña para que la pendiente no cambie significativamente en la región, pero sea lo suficientemente amplia para proporcionar una muestra precisa de la pendiente local. Sin embargo, dado que el fondo es un exponencial, la pendiente no será constante si la región se ensancha demasiado. En un ejemplo, se pueden disponer algoritmos de posicionamiento inicial automáticos en el software. Por ejemplo, se pueden identificar regiones de fondo en las que se satisfice la ecuación diferencial exponencial,  $y' = Cy$ , en la que C es una constante. Si se desea, el software puede permitir al usuario ajustar los cursores para tratar de mejorar los resultados. De manera alternativa, el software puede permitir al usuario fijar los cursores a áreas específicas, si se conocen las regiones de transición de fusión. Son posibles otros procedimientos de fijación de las regiones.

En el presente ejemplo, se generaron curvas de fusión a partir de los datos extraídos del instrumento de fusión HR-1. Se genera un valor de pendiente  $T_L$  a partir del par de cursores 1 y se genera  $T_R$  a partir del par de cursores 2. En este ejemplo, los pares de cursores están situados, respectivamente, a aproximadamente 76°C y 85°C, pero se entiende que la colocación de los pares de cursores variará dependiendo de las transiciones de fusión del ácido o

ácidos nucleicos presentes en la muestra. Los pares de cursores 1 y 2 agrupan la transición de fusión de la sonda del ácido nucleico de muestra, pero el par de cursores 2 también se encuentra antes de la región de transición de fusión del amplicón. Después de realizar el ajuste exponencial con las pendientes generadas a partir de los pares de cursores, se realiza la sustracción del fondo exponencial. La figura 1B representa la región de fusión de la sonda después de haber sustraído el fondo, utilizando las ecuaciones de la SFE y la normalización descritas anteriormente.

La sustracción del fondo exponencial para la región de fusión del amplicón utilizando los mismos o diferentes pares de cursores se puede llevar a cabo (no se muestra) con el mismo exponencial encontrado para la región de fusión de la sonda.

Aunque ilustrativamente los pares de cursores se colocan a cada lado, como mínimo, de una transición de fusión, se entiende que la posición de los pares de cursores puede variar en la práctica de la presente invención. Por ejemplo, en una realización alternativa, los valores de las pendientes ( $T_L$  y  $T_R$ ) se obtienen ambos a partir de los puntos antes de la región de fusión de la sonda. En aún otra realización alternativa, los valores de las pendientes ( $T_L$  y  $T_R$ ) se obtienen ambos a partir de puntos posteriores a la región de fusión de la sonda, pero antes de la región de fusión del amplicón. En aún otra realización alternativa, los valores de las pendientes ( $T_L$  y  $T_R$ ) se obtienen ambos a partir de puntos posteriores a la región de fusión del amplicón. En aún otra realización alternativa, los valores de las pendientes ( $T_L$  y  $T_R$ ) se obtienen a partir de puntos anteriores a la región de fusión de la sonda y posteriores a la región de fusión del amplicón. En otra realización alternativa, los valores de las pendientes ( $T_L$  y  $T_R$ ) se obtienen en dos puntos cualesquiera de la curva de fusión original, en los que ni la sonda ni el amplicón se funden. Aunque los pares de cursores ilustrativos están separados el uno del otro, es posible realizar la sustracción del fondo exponencial con dos valores de pendiente ( $T_L$  y  $T_R$ ) que están muy juntos. Esto puede ser útil para las curvas de fusión muy llenas con regiones sin fusión limitadas.

La eliminación del fondo exponencial identifica la señal de la curva de fusión de la muestra independiente de la elección del usuario de dónde el usuario decide ajustar el fondo, siempre y cuando  $T_L$  y  $T_R$  estén fuera del intervalo de fusión. En general, el ruido de fondo incluye señales de fluorescencia de fondo y señales de fusión alternativas no correspondientes a ácidos nucleicos, las cuales interfieren con el análisis de los datos de la muestra. Por ejemplo, cuando se utilizan sondas no marcadas y amplicones múltiplex a temperaturas más bajas (menos de 80°C), el ruido de fondo a baja temperatura ha impedido previamente, de manera ocasional, la identificación de curvas de señales de la muestra.

Los procedimientos de la técnica anterior para la sustracción del fondo pueden fallar a medida que las señales de fondo y las señales de muestreo se aproximan a una amplitud igual. Incluso cuando la señal de la muestra es significativamente más alta que el fondo, se ha encontrado que la sustracción del fondo exponencial es más coherente y precisa.

Como ejemplo, la sustracción del fondo exponencial proporciona una mayor precisión y especificidad cuando se distinguen las curvas de fusión de las muestras. Un estudio reciente encontró el 100% de precisión en la distinción entre una muestra de tipo natural normal y una muestra mutante homocigótica cuando los amplicones tenían aproximadamente 40 pares de bases de longitud. Antes de la utilización de la sustracción del fondo exponencial, se entendía teóricamente que una fracción de amplicones pequeños de una muestra de tipo natural normal y su muestra mutante homóloga tendría curvas idénticas, en particular si el contenido de GC se mantenía igual entre los dos amplicones. Por lo tanto, la utilización de una fusión con mayor resolución y la eliminación del fondo exponencial se atribuye a una precisión muy elevada del genotipado y la exploración de mutaciones de las muestras de ácido nucleico de doble cadena.

El análisis de fusión de alta resolución de fusión es útil para obtener resultados viables para identificar la eficacia de la eliminación del fondo y el genotipado. El análisis de fusión se puede realizar en una variedad de instrumentos de fusión, incluyendo los instrumentos de fusión de alta resolución HR-1<sup>®</sup> (un horno de fusión basado en capilares) y LightScanner<sup>®</sup> (un horno de fusión basado en placas) (ambos de Idaho Technology, Salt Lake City, Utah). Sin embargo, se entiende que el análisis de las curvas de fusión se puede realizar en ausencia de la amplificación, en particular sobre muestras de ácido nucleico muy uniformes. En un protocolo de fusión ilustrativo que utiliza el HR-1, en primer lugar las muestras se amplificaron en el LightCycler<sup>®</sup> (Roche Diagnostics Corp., Indianapolis, IN) (o en el RapidCycler<sup>®</sup> (Idaho Technology, Salt Lake City, Utah)), a continuación, se calentaron momentáneamente en el LightCycler a 94°C y rápidamente se enfriaron (ajuste del programa de -20°C/s) hasta 40°C. Los capilares del LightCycler se pueden transferir de uno en uno al instrumento de alta resolución y calentarse, de manera ilustrativa a 0,3°C/s. El HR-1 es un instrumento de una muestra que rodea un capilar de LightCycler con un cilindro de aluminio. El sistema se calienta mediante calentamiento Joule a través de una bobina enrollada alrededor del exterior del cilindro. Se pueden adquirir aproximadamente 50 puntos de datos por cada °C. El LightScanner es un sistema basado en placas que proporciona una fusión de alta resolución sobre placas de microtitulación de 96 ó 384 pocillos. La PCR se puede realizar en cualquier ciclador térmico basado en placas compatibles.

En algunos casos, es ventajoso no desnaturalizar el producto después de la PCR antes de la obtención de la curva de fusión. Por ejemplo, cuando el objetivo es examinar el número de secuencias de repetición (por ejemplo, STR, VNTR), la amplificación se puede detener en la etapa de extensión durante la fase exponencial de la reacción antes de la fase meseta y, a continuación, se realiza el análisis de fusión. De esta manera, se pueden analizar productos

de extensión homodúplex. En el examen de las repeticiones, los productos homodúplex pueden ser más informativos que los productos heterodúplex, en especial porque muchos productos diferentes heterodúplex pueden formarse a partir de diferentes alineaciones de las repeticiones. En algunos casos, puede ser útil obtener una curva de fusión de productos homodúplex (sin desnaturalización previa) y una curva de fusión de productos heterodúplex (con la desnaturalización y la formación de todas las combinaciones posibles de dobles cadenas). La diferencia entre estas dos curvas de fusión proporciona una medida de la extensión de los heterodúplex que se pueden formar utilizando la misma muestra como "control homodúplex".

Las técnicas de sustracción del fondo anteriores a menudo incluían la diferenciación numérica como elemento. La diferenciación numérica de los datos originales implica un ajuste y suavizado artificial que pueden afectar a la eficacia de los datos. Entre las técnicas de diferenciación numérica habituales se incluyen curvas derivadas negativas y curvas derivadas integradas. La presente realización no requiere que los datos originales se diferencien numéricamente. Esto es una ventaja sobre las técnicas de sustracción del fondo anteriores. La conversión de los datos originales en curvas derivadas a menudo implica la amplificación del ruido de fondo y el suavizado artificial de características significativas de los datos de la fusión. La presente realización es capaz de distinguir diferencias sutiles, pero molecularmente significativas, en los datos de fusión, lo cual es una ventaja sobre las técnicas anteriores, que implicaban el análisis de la curva derivada.

Aunque el procedimiento de sustracción del fondo exponencial anterior se utiliza en referencia a curvas de fusión de ácidos nucleicos, se entiende que este procedimiento se puede aplicar a una variedad de conjuntos de datos, incluyendo otros conjuntos de datos biológicos, que tienen ruido de fondo exponencial y es particularmente adecuado cuando es importante la sustracción del fondo sin utilizar el valor de la señal para el cálculo.

**Función de superposición de curvas**

Un ejemplo alternativo incluye una función de superposición de curvas para utilizar en el análisis de curvas de fusión. Los procedimientos anteriores de superposición de curvas, o de desplazamiento de temperatura, incluyen las etapas de seleccionar un intervalo de fluorescencia, habitualmente a baja fluorescencia (por ejemplo, del 5 al 15%) de la curva de fusión normalizada, ajustar un polinomio de segundo grado a todos los puntos dentro del intervalo para cada curva y, a continuación, desplazar cada curva para superponer mejor las representaciones en este intervalo. La superposición de curvas corrige cualquier variación menor de temperatura entre pruebas y aumenta la capacidad de distinguir heterocigotos de homocigotos. Sin embargo, el procedimiento anterior no era satisfactorio cuando había menos de tres (3) puntos en el intervalo. En ausencia de los tres puntos de datos, el procedimiento anterior no podía proporcionar un medio para la superposición automática y precisa de curvas, ya que los procedimientos de superposición matemática conocidos no podían aplicarse con gran precisión. Entre los procedimientos de superposición matemática conocidos se incluyen el procedimiento de mínima distancia, el procedimiento de valores promedio más bajos de la distancia absoluta y el procedimiento de mínimos cuadrados.

El presente ejemplo analiza las curvas de fusión normalizadas mediante la superposición de las mismas entre un valor variable dependiente inferior ( $y_L$ ) y un valor variable dependiente superior ( $y_H$ ). Esto se realiza mediante la extracción de todos los valores numéricos ( $x$ ,  $y$ ) en el intervalo, siempre que todos los valores ( $y$ ) continúen disminuyendo, o todos los valores ( $y$ ) continúan creciendo. En lugar de ajustar ( $y$ ) en función de ( $x$ ) y buscar la mejor superposición de dichos ajustes, los pares ordenados se invierten, haciendo así que ( $x$ ) sea función de ( $y$ ). Un ejemplo del ajuste óptimo de mínimos cuadrados de una función desplazada horizontalmente se muestra en el conjunto de ecuaciones 6,  $x_1(y) + c$  a otro,  $x_2(y)$  se obtiene mediante la búsqueda de la constante  $c$  que hace que la diferencia promedio de  $(x_1(y) + c) - x_2(y)$  sea igual a cero.

$$\text{Conjunto de ecuaciones 6} \quad \min_c \int ((f(z) + c) - g(z))^2 dz = \int g(z) - f(z) dz$$

El valor  $x_2(y)$  se obtiene mediante la búsqueda de la constante ( $c$ ), haciendo así que la diferencia promedio de  $(x_1(y) + c) - x_2(y)$  sea igual a cero. El valor ( $z$ ) representa una variable de integración. El valor ( $z$ ) en el conjunto de ecuaciones 6 representa el valor ( $x$ ) o el valor ( $y$ ), lo cual es, en parte, debido a que la variable independiente es la ( $y$ ) original e igual al valor de fluorescencia. Las funciones  $f(z)$  y  $g(z)$  representan secciones entre dos valores de fluorescencia normalizados (bajo y alto) que se eligen para la superposición de dos curvas de fusión normalizadas, en las que la temperatura se representa en función de la fluorescencia. El valor ( $dz$ ) es una medida normalizada sobre un intervalo de fluorescencia normalizado.

A efectos de explorar heterocigotos en un producto de PCR, la forma es más importante que la temperatura absoluta o  $T_f$ . Los heterocigotos producen heterodúplex que funden a temperaturas más bajas y distorsionan la forma de una curva de fusión general. Las diferencias en la forma son más eficaces de utilizar cuando se identifican diferentes genotipos que con las  $T_f$  absolutas, ya que la variación de temperaturas puede estar causada por diferencias menores de las muestras y la variabilidad del instrumento. Entre las posibles diferencias de las muestras se incluyen, pero sin limitarse a las mismas, las variaciones en la fuerza iónica y la aparición de evaporación durante los procedimientos de procesamiento. La variabilidad del instrumento también es posible con respecto a las posiciones relativas y exactas de los pocillos sobre las placas de muestra.



Tal como se ha descrito anteriormente, los procedimientos anteriores para comparar la forma de las curvas incluían la superposición de curvas mediante su desplazamiento a lo largo del eje de temperatura, implementado mediante el ajuste de un polinomio de segundo grado a un intervalo pequeño de fluorescencia de cada curva. A continuación, se elegía una curva patrón arbitraria y las curvas restantes se desplazaban para superponerse con la curva patrón sobre esta región. Sin embargo, cuando la densidad de datos y el intervalo de fluorescencia son pequeños, el procedimiento anterior es propenso a fallar, tal como se muestra en la figura 2B. El panel superior (figura 2B) incluye curvas de fusión para un fragmento de PCR del factor V Leiden de 96 muestras de ADN genómico, habiéndose normalizado mediante el procedimiento de sustracción del fondo exponencial. En la figura 2B son visibles tres grupos de genotipos distintos. En el panel superior también se muestran marcadores del intervalo de fluorescencia F1, F2 sobre los que se intenta la superposición. El primer marcador del intervalo F1 se encuentra aproximadamente al 10% de fluorescencia y el segundo marcador del intervalo F2 se encuentra aproximadamente al 15% de fluorescencia. En el panel central se muestra un gráfico aumentado de puntos de datos reales de 80 a 82°C contenido en el intervalo de fluorescencia seleccionado. Varias curvas muestran que menos de tres puntos están incluidos dentro de este intervalo, lo que hace que un ajuste cuadrático sea imposible y, en última instancia, conduce a la imposibilidad del procedimiento de superposición de curvas anterior, tal como se muestra en el panel inferior.

El presente ejemplo utiliza el algoritmo del conjunto de ecuaciones (6) (figura 2A) con el mismo conjunto de datos. La presente realización analiza los datos con precisión y éxito, y, de hecho, para hacerlo sólo necesita un punto de datos de cada una de las respectivas curvas, proporcionando de este modo un procedimiento de análisis nuevo y mejorado. En el panel superior de la figura 2A, los datos originales se muestran junto con cursores verticales (véase la figura 1A y la descripción anterior) que definen las regiones para la estimación de la pendiente del fondo exponencial. En el panel inferior (figura 2A) se proporcionan las curvas normalizadas con el fondo sustraído que han sido desplazadas en la temperatura de manera satisfactoria, de manera que todas las curvas se superponen dentro de la región de fluorescencia del 10 al 15%.

El presente ejemplo es un procedimiento ventajoso sobre los procedimientos anteriores por diversas razones, incluyendo que sólo se requiere un (1) punto de datos para la validez y exactitud de la función de superposición de curvas. La presente realización es ventajosa adicionalmente, ya que tiene una optimalidad rigurosa para el ajuste de mínimos cuadrados. Se contempla que la presente realización de la función de superposición de curvas se puede representar en numerosas representaciones matemáticas diferentes de las descritas en el presente documento, y el conjunto de ecuaciones (6) se considera uno de dichos ejemplos.

### **Función de representación de la diferencia**

Otro ejemplo alternativo incluye una función mejorada de representación de las diferencias para el análisis de las diferencias entre las curvas de fusión de las muestras de ácido nucleico. Los procedimientos anteriores sustraían todas las curvas de una curva de referencia arbitraria o un promedio de curvas de referencia. El resultado de la sustracción era estrictamente la distancia vertical entre las curvas en cada punto de temperatura. Estas representaciones de la diferencia aumentaban visualmente la diferencia entre las curvas para que pudieran ser observadas más fácilmente. Sin embargo, la distancia vertical entre las curvas no representa con precisión la distancia más corta entre las curvas. Éste es especialmente el caso cuando el valor de fluorescencia de las curvas de fusión disminuye a una velocidad significativa o con pendiente negativa. Esta deficiencia daba lugar a un artefacto conocido como "burbuja de variación", que a menudo es claramente visible cuando la disminución de la fluorescencia es máxima.

Una nueva mejora en la presente función de representación de las diferencias es ponderar las diferencias entre las curvas según la pendiente de las curvas. Esto proporciona una mejor medida de la distancia entre las curvas, de manera que se pueden identificar de manera correcta y automática los genotipos comunes y distintos. La ponderación equilibra el efecto de la pendiente en la medida de la diferencia entre las curvas. Cuando no se realiza la ponderación, la diferenciación vertical estándar sobredimensiona la diferencia entre las curvas cuando la pendiente es inclinada, ampliando la propagación de las curvas de fusión dentro de los mismos genotipos. El presente procedimiento proporciona una aproximación práctica por ordenador a la diferencia ortogonal entre las curvas de fusión, que, de este modo, representa mejor las variaciones en la distancia entre las curvas.

El presente ejemplo proporciona un nuevo procedimiento para el análisis de los datos de las curvas de fusión. Las curvas de fusión analizadas se pueden normalizar mediante diversos procedimientos conocidos, pero es preferente la sustracción del fondo exponencial novedosa descrita en el presente documento. En lugar de la distancia vertical a cada temperatura, es ventajosa una distancia ortogonal entre las curvas. El presente ejemplo obtiene la menor distancia entre las curvas, que es la menor distancia entre las curvas en cada punto. Cuando la fluorescencia disminuye rápidamente, en lugar de utilizar la distancia vertical entre las líneas, una medida ortogonal a la pendiente de las curvas es más útil y precisa para evaluar las diferencias entre las curvas. Cuando se miden las distancias ortogonales entre dos curvas, los resultados son a menudo diferentes dependiendo de la curva utilizada como referencia. La presente realización proporciona un procedimiento para estimar la distancia ortogonal entre dos curvas.

Para compensar el énfasis exagerado sobre las regiones de fusión de dos curvas de fusión que se comparan y el



que  $\|f_1 - f_2\|$  representa la medida de la distancia utilizada en la proporción que ordena los niveles de agrupamiento (número de grupos) por probabilidad. La medición de la diferencia ( $\|f_1 - f_2\|$ ) indica la medida de la distancia entre dos curvas de fusión normalizadas y se selecciona entre un grupo que comprende: la separación absoluta promedio, la separación cuadrática promedio y la separación máxima a una temperatura elegida. Se contempla que se pueden utilizar otros procedimientos conocidos para medir distancias entre dos curvas de fusión con respecto al conjunto de ecuaciones 8.

$$\text{Conjunto de ecuaciones (8)} \quad \min_{\{f_1 \in C_1, f_2 \in C_2\}} \|f_1 - f_2\|$$

En un punto en el que se han unido los subgrupos  $C_1$  y  $C_2$ , el valor ( $f_1$ ) representa una curva de fusión asociada con el subgrupo  $C_1$ , y el valor ( $f_2$ ) es una curva de fusión asociada con el subgrupo  $C_2$ . El valor ( $\min_{\{f_1 \in C_1, f_2 \in C_2\}}$ ) representa el valor más pequeño de la distancia entre todos los pares de curvas de fusión, en el que se toma un primer valor de un primer subgrupo, habiéndose unido el primer subgrupo, y se toma un segundo valor de un segundo subgrupo.

El presente ejemplo es un nuevo procedimiento para medir la distancia entre los niveles de agrupación para determinar la probabilidad de un nivel particular (número de grupos). La nueva medida de la distancia es la distancia mínima entre dos miembros cualesquiera, siempre que cada uno sea de un grupo diferente. A efectos de determinar la probabilidad de cualquier nivel de agrupación (es decir, 3 frente a 4 frente a 5 grupos), la proporción de la distancia hasta el siguiente grupo se divide por la distancia al grupo anterior. El procedimiento anterior proporciona la clasificación correcta de los genotipos cuando se varían parámetros, tales como la ubicación de los cursores. Cuando se utilizan promedios ponderados en lugar de la distancia entre las curvas más próximas entre sí en cada subgrupo para determinar la proporción, la elección del número más probable de grupos es mucho menos estable. Además, la presente realización se puede ejecutar de forma automática con una gran precisión.

Las figuras 3A-C muestran la capacidad de la nueva función de agrupamiento para asignar el número correcto de grupos de genotipo a una curva de fusión con múltiples muestras. Se amplificaron seis genotipos separados de ADN genómico humano del gen de la lipasa hepática (BioRad iCycler) utilizando volúmenes de reacción de 10  $\mu$ l en una placa de 96 pocillos. Se utilizó 1X de una mezcla madre de LightScanner® (Idaho Technology, Salt Lake City, UT). Las muestras se calentaron de 75°C a 94°C a 0,1°C/segundo en el instrumento de fusión LightScanner.

La figura 3A muestra una captura de pantalla de la curva de fusión original (panel superior) para la amplificación del gen de la lipasa hepática y la transición de fusión. El panel inferior muestra el resultado de la función de agrupación previa, junto con un menú desplegable que indica que la función de agrupación previa indica incorrectamente la presencia de sólo tres grupos de genotipos distintos en las muestras, que se representan como tres grupos de líneas de colores diferentes.

La figura 3B muestra una captura de pantalla de la curva de fusión original (panel superior) para la amplificación del gen de la lipasa hepática y la transición de fusión. El panel inferior muestra el resultado de la función de agrupación de la presente realización, junto con un menú desplegable que indica que la nueva función de agrupación indica correctamente la presencia de seis grupos de genotipos distintos en las muestras, que se representan como seis grupos separados de líneas de colores. La figura 3C muestra que el usuario puede identificar las muestras mediante el genotipo por su posicionamiento en la placa de reacción de 96 pocillos, mostrado a lo largo del gráfico de la fluorescencia frente a la temperatura de la figura 3B.

Se entiende que las representaciones de fusión se pueden analizar utilizando uno o más de los algoritmos de sustracción del fondo exponencial, la función de superposición de curvas, la función de representación de las diferencias y la función de agrupamiento, y que cada uno de estos procedimientos se pueden utilizar solos o en cualquier combinación y se pueden utilizar en combinación con otros procedimientos de análisis de representaciones de fusión.

### Ejemplo 1 – Genotipado de mutaciones y polimorfismos de la hemocromatosis (HFE)

Se conoce que las mutaciones y polimorfismos de los genes de la hemocromatosis interfieren con el metabolismo normal del hierro en los seres humanos. En este ejemplo ilustrativo, la detección y la identificación de genotipos de polimorfismos y mutaciones a través del análisis de la curva de fusión son más precisos con la sustracción del fondo exponencial que con la sustracción del fondo de la línea basal. El análisis se realizó con amplicones pequeños (78 pb y 40 pb) a efectos de aumentar la diferencia de  $T_f$  entre diferentes muestras de homocigotos. Se utilizan sondas no marcadas para el genotipado de SNP, que de otra manera no podrían genotiparse fácilmente mediante fusión de amplicones.

Se puede amplificar ADN genómico humano que representa distintos genotipos de hemocromatosis mediante varios instrumentos de PCR. Un procedimiento de ejemplo incluye volúmenes de reacción de 10  $\mu$ l con un LightCycler® 2.0 de Roche. Después de la amplificación por PCR, las muestras de HFE se calientan en un instrumento de fusión HR-1 (Idaho Technology) durante aproximadamente 115 segundos. El gráfico de la fluorescencia resultante frente a la temperatura se representa como la figura 4A. Las curvas de fusión derivadas se muestran en la figura 4B. En este

ejemplo ilustrativo, la sustracción del fondo exponencial se realiza sobre las curvas de fusión derivadas y las curvas de fusión derivadas normalizadas se muestran como la figura 4C. En comparación con la curva de fusión normalizada del fondo exponencial (figura 4C), la técnica de línea basal para la normalización de curvas de fusión es claramente deficiente. Tal como se muestra en las figuras 4D-4E, la técnica de la línea basal aplicada a una curva de fusión con sonda más amplicón combinados (figura 4D) da lugar a una ausencia de datos utilizables (figura 4E), aun cuando las técnicas de la línea basal y la sustracción del fondo exponencial utilizaban el mismo conjunto de datos de la curva de fusión (figuras 4B, 4D). La identificación del genotipo de la figura 4C para las muestras homocigota y heterocigota C282Y, heterocigota y homocigota H63D y de tipo natural se muestra en la figura 4F.

**10 Ejemplo 2 - Agrupación por derivación del factor V Leiden con sondas y amplicones combinados (placa de 384 pocillos)**

El factor V Leiden es el trastorno de la coagulación de la sangre hereditario más común en los Estados Unidos. Está presente en aproximadamente el 5% de la población caucásica y aproximadamente el 1,2% de la población afroamericana. El factor V Leiden como diana génica es importante para la detección de los SNP que están relacionados con la disposición al trastorno de la coagulación.

Se amplificaron muestras de ADN genómico del factor V Leiden humano que representaban diferentes genotipos del factor V Leiden en un ABI 9700 con un volumen de reacción de 10 µl y 12 µl de una capa de aceite. El ensayo incluyó un amplicón del factor V Leiden de aproximadamente 100 pb que tenía la secuencia 5'-CTGAAAGGTTACTTCAAGGACAAAATACCTGTATTCCTCGCCTGTCCAGGG ATCTGCTCTTAGAGATTAGAAGTAGTCCTATTAGCCCAGAGCCGATGTC-3' (SEQ ID NO: 1), que se amplificó mediante el cebador directo 5'-CTGAAAGGTTACTTCAAGGAC-3' (SEQ ID NO: 2) y el cebador inverso 5'-GACATCGCCTCTGGG-3' (SEQ ID NO: 3). El ensayo también incluía una sonda no marcada 3'-TGGACATAAGGAGCGGACAGGT-5' (SEQ ID NO: 4) configurada para hibridarse a la cadena directa del amplicón, tal como se indica mediante el subrayado. Las muestras de PCR resultantes se calentaron de 58°C a 88°C a 0,1°C/s en un instrumento de fusión LightScanner utilizando una placa de reacción de 384 pocillos. El procedimiento de fusión total requirió aproximadamente 5 minutos para su finalización.

El instrumento de fusión LightScanner midió y registró la fluorescencia en función de la temperatura. La curva de fusión original resultante del procedimiento se muestra en la figura 5A. Se calculó la derivada negativa de la curva de fusión de la figura 5A y se muestra en la figura 5B. En la figura 5B se muestra un diagrama representativo de la placa de 384 pocillos junto con la agrupación de los genotipos. La eliminación del fondo no se realizó y el agrupamiento realizado no genotipó correctamente varias de las muestras analizadas (compárese la placa de 384 pocillos con la de la figura 5D).

Se realizó una función de corrección lineal sobre el gráfico de la derivada de la figura 5B y se muestra en la figura 5C. En la figura 5C se muestra un diagrama representativo de la placa de 384 pocillos junto con la agrupación de los genotipos. Es evidente a partir de la placa de 384 pocillos que el agrupamiento realizado después de la corrección lineal no genotipó correctamente varias de las muestras analizadas. La corrección lineal realizada sobre los datos de la curva de fusión derivada incluye la supresión de la pendiente de extremo a extremo mediante la sustracción de la función lineal.

La figura 5D representa los datos de la curva de fusión derivada de la figura 5B después de realizar la sustracción del fondo exponencial sobre el conjunto de datos. Sólo después de la sustracción del fondo exponencial se obtiene la correcta agrupación de genotipos. La precisión de la agrupación de genotipos es visible dentro de la placa de 384 pocillos (figura 5D), en la que las letras "S-N-P" indicando los genotipos por posición de la placa, y la identificación adecuada es visible para la región del amplicón y la sonda del gen diana del factor V Leiden.

**50 Ejemplo 3 – Agrupación por derivación del factor V Leiden sólo con la sonda (placa de 96 pocillos)**

Se amplificaron muestras de ADN genómico del factor V Leiden humano que representaban diferentes genotipos del factor V Leiden en un ABI 9700 con un volumen de reacción de 10 µl y 12 µl de una capa de aceite. El ensayo incluyó un amplicón del factor V Leiden y una sonda no marcada, tal como se ha descrito anteriormente en el ejemplo 2, en presencia de 1X LCGreen Plus®. Las muestras de PCR resultantes se calentaron de 58°C a 88°C a 0,1°C/s en un instrumento de fusión LightScanner utilizando una placa de reacción de 96 pocillos. El procedimiento de fusión total requirió aproximadamente 5 minutos para su finalización.

El instrumento de fusión LightScanner midió y registró la fluorescencia en función de la temperatura. La curva de fusión original resultante del procedimiento se muestra en la figura 6A. Se realizó un gráfico de la derivada negativa en función de la temperatura a partir de los datos de la curva de fusión de la figura 6A y se muestra en la figura 6B. Dos conjuntos de líneas de cursores verticales están presentes a aproximadamente 59°C y 60°C, y 71°C y 72°C, respectivamente. Se generan pendientes con la técnica de la sustracción del fondo exponencial utilizando estas líneas de cursores. Después del cálculo automático y el ajuste, se sustrae el fondo exponencial del gráfico de la derivada negativa original. El resultado de la sustracción del fondo exponencial específico para la región de la sonda se muestra en la figura 6C. Aunque los cursores de la pendiente vertical están situados a cada lado de la región de

la sonda, el exponencial utilizado para la región de la sonda también se puede utilizar para la región del amplicón. De hecho, la localización absoluta de dónde se obtienen las pendientes no es determinante del exponencial, siempre que las pendientes se encuentren en posiciones en las que no se produce ninguna fusión de la muestra. Los datos de la región de la sonda (figura 6C) se agrupan automáticamente mediante la función de agrupación del genotipo. Las muestras de ácido nucleico de la placa de 96 pocillos se agrupan en tres genotipos y se muestran en la figura 6D.

Con respecto a los ejemplos 4 y 5, se utilizaron los siguientes procedimientos e instrumentos. Para el análisis se utilizaron ADN genómico humano de muestras de genotipos conocidos del factor V Leiden y muestras de ADN genómico heterocigoto con mutaciones seleccionadas de la fibrosis quística.

Las  $T_f$  previstas de las sondas fueron inferiores a las  $T_f$  observadas, quizás debido a la estabilización del colorante. Se ajustó la temperatura de fusión de diferentes dúplex de sonda/alelo mediante la longitud de la sonda, la posición del apareamiento y el contenido de dU frente a dT en la sonda. Se evitó la extensión de sondas no marcadas durante la PCR mediante la incorporación de un 3'-fosfato durante la síntesis. De manera alternativa, se pueden utilizar otros mecanismos de bloqueo en 3', de manera ilustrativa, mediante la aportación adicional de un apareamiento adicional de dos bases adicionales al extremo 3' de la sonda. Cuando se utiliza una polimerasa negativa 5'-exonucleasa, las sondas se deben diseñar para fundir a una temperatura más baja que la temperatura de extensión de la PCR.

Se puede utilizar la asimetría de cebadores, ilustrativamente en proporciones de 1:5 a 1:10, para producir suficiente producto de doble cadena para la fusión del amplicón y suficiente producto de una cadena para la hibridación de la sonda. La PCR para el factor V realizada en un formato de 384 pocillos utilizó volúmenes de 5  $\mu$ l e incluía 20 ng de ADN genómico en Tris 50 mM, pH 8,3 con  $MgCl_2$  3 mM, 0,2 mM de cada dNTP, 500  $\mu$ g/ml de BSA, 1X LCGreen® PLUS (Idaho Technology), 0,2 U de KlenTaqI® (AB Peptides) y 70 ng de anticuerpo TaqStart® (Clontech). La PCR se realizó en un ciclador térmico 9700 (ABI) con una desnaturalización inicial a 94°C durante 10 s, seguido de 50 ciclos de 94°C durante 5 s, 57°C durante 2 s y 72°C durante 2 s. Después de la PCR, las muestras se calentaron hasta 94°C durante 1 s y, a continuación, se enfriaron hasta 10°C antes de la fusión.

Se realizó la PCR para la amplificación de los exones de CFTR 10 y 11 en volúmenes de 10  $\mu$ l e incluía 50 ng de ADN genómico en Tris 50 mM, pH 8,3 con  $MgCl_2$  2 mM, 0,2 mM de cada dNTP, 500  $\mu$ g/ml de BSA, 1X LCGreen I (Idaho Technology) y 0,4 U de Taq polimerasa (Roche). La PCR se realizó en tubos capilares en un LightCycler (Roche) con una desnaturalización inicial de 95°C durante 10 segundos, seguido de 45 ciclos de 95°C durante 1 s, 54°C durante 0 s y 72°C durante 10 s. Después de la amplificación, las muestras se calentaron hasta 95°C durante 0 s y se enfriaron rápidamente hasta 40°C antes de la fusión.

Cuando se amplificaron las muestras de ácido nucleico en placas de 384 pocillos, se realizó la obtención de la fusión en una versión prototipo del LightScanner® (Idaho Technology). Se sustituyeron los diodos emisores de luz de 470 nm estándar por diodos emisores de luz de 450 nm (Bright-LED Optoelectronics). Además, los filtros ópticos se cambiaron a filtros de excitación de 425-475 nm y filtros de emisión de paso largo de 485 nm (Omega Optical). La placa se calentó de 55 a 88°C a 0,1°C/s con un intervalo de captura de 300 ms, una exposición de 15 ms y una potencia LED del 100%, dando lugar a aproximadamente 25 puntos/°C.

La fusión de los exones de CFTR se llevó a cabo en el instrumento de fusión de alta resolución HR-1 (Idaho Technology) con una adquisición de 24 bits de la temperatura y la fluorescencia. Después de la PCR en el LightCycler, cada tubo capilar se transfirió al HR-1 y se fundió de 50°C a 90°C con una pendiente de 0,3°C/s, dando lugar a 65 puntos/°C.

Las curvas de fusión se pueden analizar en cualquier software adecuado conocido en la técnica. Un paquete de software de ejemplo para implementar los procedimientos de análisis de fusión de las diversas realizaciones de la presente invención es LabVIEW (National Instruments). La normalización y la sustracción del fondo se realizaron, en primer lugar, mediante el ajuste de un exponencial al fondo que rodea las transiciones de fusión de interés. Se obtuvieron gráficos de la derivada de las transiciones de fusión de la sonda mediante la estimación polinómica de Salvitsky-Golay. Las curvas de fusión de los productos de la PCR se compararon sobre gráficos de diferencias de curvas de fusión normalizadas con la temperatura superpuesta. Las curvas de fusión normalizadas estaban superpuestas con la temperatura (para eliminar pequeños errores de temperatura entre pocillos o pruebas) mediante la selección de un intervalo de fluorescencia (ilustrativamente baja fluorescencia/alta temperatura, habitualmente 5-10% de fluorescencia) y el desplazamiento de cada curva a lo largo del eje X para superponer mejor una muestra patrón dentro de este intervalo. Los gráficos de diferencias de las curvas normalizadas superpuestas con la temperatura se obtuvieron tomando la diferencia de fluorescencia de cada curva de la curva de tipo natural promedio a todos los puntos de temperatura. Estos procedimientos analíticos se han aplicado previamente a la exploración de mutaciones y emparejamientos de HLA.

La agrupación jerárquica imparcial aglomerativa de los datos de las curvas de fusión se realizó mediante procedimientos anteriores programados a medida en LabVIEW. La distancia entre las curvas se tomó como el promedio del valor absoluto de la diferencia de fluorescencia entre las curvas sobre todos los puntos de temperatura.

El número de grupos se identificó de forma automática mediante la selección de la mayor proporción de distancias entre los niveles de agrupación consecutiva. Los procedimientos de agrupación representan medios menos precisos para agrupar genotipos que los nuevos procedimientos de agrupación descritos en el presente documento (figuras 9A-D).

5

#### **Ejemplo 4 – Exploración y genotipado de CFTR**

Se han elegido varios exones del gen regulador de la transconductancia de la fibrosis quística (CFTR) para demostrar la exploración y el genotipado simultáneos de múltiples variantes. Se analizaron tres SNP en dos regiones del exón 11 del gen de CFTR con dos sondas no marcadas, secuencias en parte TCTTGGAGAA (SEQ. ID NO: 5) y AGGTCAACGA (SEQ. ID NO: 6). Dos de las mutaciones estaban separadas sólo por seis bases, lo que permitía que una de las sondas pudiera cubrir ambas mutaciones (figura 7A). Se amplificaron cinco réplicas de cada genotipo y se analizaron. Las curvas de fusión normalizadas después de la SFE (figura 7B) muestran regiones de fusión de la sonda (56-74°C) y de fusión del producto de la PCR (80-83°C). En una observación casual, no está claro a partir de la curva de fusión normalizada qué información se puede extraer. Sin embargo, cuando la región de la sonda se muestra como un gráfico de la derivada (figura 7C), las transiciones de fusión de todos los alelos comunes con ambas sondas son evidentes. Ambas sondas no marcadas se emparejaban con la secuencia de tipo natural, pero una de las sondas era más corta y contenía dU en lugar de dT para disminuir su temperatura de fusión. La sonda más estable cubría un único SNP, dando lugar a dos alelos separados por la Tf, siendo ambos más estables que todos los alelos de la sonda menos estable. La sonda menos estable cubría dos SNP, dando lugar a tres picos para los genotipos comunes. El desapareamiento específico y su posición dentro de la sonda afectan a la estabilidad del dúplex, permitiendo un diseño de sonda que distinga alelos múltiples. En la figura 7D se muestra un gráfico de las diferencias del producto de la PCR. Las muestras mutantes heterocigotas, de tipo natural y homocigotas son claramente diferentes. Sin embargo, es difícil distinguir entre diferentes heterocigotos solo mediante la fusión del producto de la PCR. La agrupación jerárquica imparcial agrupaba todos los heterocigotos juntos (datos no mostrados). Los tres heterocigotos están todos en la misma clase de SNP (12), dando lugar a los mismos desapareamientos de heterodúplex (C:A y T:G) y emparejamientos de homodúplex (C:G y A:T). Aunque las estabildades previstas de los tres heterocigotos utilizando la termodinámica del vecino más próximo (13, 14) no son idénticas, el genotipado definitivo requería la utilización de sondas. El punto fuerte de la fusión del producto es identificar fácilmente la presencia de heterocigotos, mientras que las sondas no marcadas discriminan adicionalmente entre heterocigotos e identifican más fácilmente variantes de homocigotos. Tal como se ilustra con este ejemplo, la combinación del genotipado y la exploración da lugar a la observación de transiciones de fusión del amplicón (producto de la PCR) y la sonda no marcada.

#### **Ejemplo 5 – Genotipado de CFTR**

También se analizaron tres SNP y dos deleciones en el exón 10 del gen de CFTR con dos sondas no marcadas. La sonda con la mayor Tf, secuencia en parte TTCTCAGTTT (SEQ. ID NO: 7) cubría un único SNP, mientras que la sonda con la Tf más baja, secuencia en parte TATCATCTTTG (SEQ. ID NO: 8), cubría dos SNP y dos deleciones (figura 8A). Las curvas de fusión normalizadas después de la SFE (figura 8B) muestran regiones de fusión de la sonda a baja temperatura (56-67°C), fusión de la sonda a alta temperatura (67-75°C) y fusión del producto de la PCR (80-83°C). Cuando las regiones de las sondas se muestran como un gráfico de la derivada (figura 8C), los cinco genotipos heterocigotos siguen trayectorias únicas que los distinguen del tipo natural y entre sí. Cuatro de los heterocigotos muestran picos resueltos, mientras que uno se identifica por un pico ancho que resulta de un desapareamiento relativamente estable (un desapareamiento A:G cerca de un extremo de la sonda en una región rica en AT). La discriminación de alelos no requiere una Tf única para cada alelo, sólo que las curvas sean diferentes en alguna región de la transición de fusión. En la figura 8D se muestra un gráfico de las diferencias de la transición de fusión del producto de la PCR. El doble heterocigoto muestra la mayor desviación del tipo natural, ya que están presentes dos desapareamientos dentro del producto de la PCR. Los cuatro heterocigotos individuales son fácilmente distinguibles del tipo natural. A diferencia del exón 11, todos los heterocigotos se pudieron genotipar mediante la fusión del producto de la PCR o la sonda. La consideración de ambas regiones a menudo proporciona una confirmación independiente del genotipo.

#### **Ejemplo 6 - Genotipado del amplicón completo de $\beta$ globina**

La  $\beta$  globina presenta una diana génica con SNP conocidos que son importantes para el análisis de hemoglobinopatías, las más destacadas son las mutaciones HbC y HbS. Se amplificaron muestras de ADN genómico humano de diferentes genotipos de  $\beta$  globina en un LightCycler de Roche utilizando volúmenes de reacción de 10  $\mu$ l. Se utilizó 1X LCGreen® de Idaho Technology en la PCR para amplificar un amplicón de 45 pb: 5'-CCATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCC-3' (SEQ. ID NO: 9). Las muestras del producto de la PCR se calentaron de aproximadamente 72°C a aproximadamente 88°C a 0,3°C/s en el instrumento de fusión HR-1. El tiempo requerido para la fusión es de aproximadamente 60 segundos.

El instrumento de fusión HR-1 midió y registró la fluorescencia de las muestras en función de la temperatura. La curva de fusión original para el amplicón completo de  $\beta$  globina se muestra en la figura 9A. A continuación, se

65

normaliza la curva de fusión original para el amplicón completo de  $\beta$  globina mediante el procedimiento de sustracción del fondo exponencial descrito en el presente documento, lo que da lugar a una curva de fusión que es la fluorescencia relativa de las muestras en función de la temperatura en la figura 9B. La función de agrupación de genotipos se realiza sobre las curvas de fusión normalizadas de la figura 9B. La figura 9C muestra los genotipos agrupados de las curvas de fusión normalizadas de la figura 9B.

### Ejemplo 7 – Análisis de los gráficos de diferencias del tipo natural del factor V Leiden

Se amplificaron muestras de ADN genómico del factor V Leiden humano que representaban 60 muestras de tipo natural diferentes del factor V Leiden en un ABI 9700 con un volumen de reacción de 10  $\mu$ l y 12  $\mu$ l de una capa de aceite. El ensayo incluyó un amplicón del factor V Leiden y una sonda no marcada, tal como se ha descrito anteriormente en el ejemplo 2, en presencia de 1X LCGreen Plus®. Las muestras de PCR resultantes se calentaron de 58°C a 88°C a 0,1°C/s en un instrumento de fusión LightScanner utilizando una placa de reacción de 96 pocillos. El procedimiento de fusión total requirió aproximadamente 5 minutos para su finalización.

El instrumento de fusión LightScanner midió y registró la fluorescencia en función de la temperatura. La aplicación de varias funciones en los gráficos de diferencias proporciona un valor de diferencia de fluorescencia en función de la temperatura. Las curvas de fusión para 60 tipos naturales del factor V Leiden se muestran en la figura 10A después de realizar la función de sustracción del fondo exponencial. La curva de fusión normalizada aparece como una curva negra gruesa, que es característica de las similitudes sustanciales de las 60 muestras de tipo natural.

La figura 10B representa el resultado gráfico de la técnica previa de la representación de diferencias verticales, que representa la diferencia de fluorescencia vertical entre las 60 muestras de tipo natural en función de la temperatura. Esta técnica previa de representación de diferencias verticales presenta un artefacto interesante cuando las curvas de fusión tienen una pendiente suficientemente inclinada. El artefacto se manifiesta como una "burbuja" aproximadamente entre 80,25°C y 81,5°C (figura 10B), que es la parte más inclinada de la pendiente de la curva de fusión de datos originales.

El artefacto del gráfico de las diferencias verticales representa la amplificación artificial de pequeñas diferencias entre las 60 muestras de tipo natural. La figura 10C representa los resultados de la función del gráfico de las diferencias ortogonales. Dado que la función del gráfico de diferencias está diseñado para medir con precisión la diferencia entre las muestras, la estrecha proximidad de la diferencia de fluorescencia representada en función de la temperatura es indicativa de la similitud significativa entre las 60 muestras de tipo natural (figura 10C). El gráfico de la diferencia de fluorescencia frente a la temperatura de las 60 muestras de tipo natural aparece como una sola línea gruesa aproximadamente centrada en la diferencia de fluorescencia cero.

En referencia a la figura 11, se contempla que las realizaciones de la presente invención se pueden implementar en un sistema informático -100-. El sistema informático -100- comprende un instrumento de fusión de ácido nucleico -102-, un dispositivo de almacenamiento de memoria -104- y una unidad de procesamiento central (CPU) -106-. El dispositivo de almacenamiento de memoria -104- y la CPU -106- pueden estar integrados (no mostrado) dentro de la carcasa del instrumento de fusión -102- o pueden disponerse externamente (figura 11), ilustrativamente como un ordenador portátil o de sobremesa. El sistema -100- puede incluir de manera alternativa una interfaz gráfica de usuario (GUI) -108- para la visualización, manipulación y el análisis de la curva de fusión de alta resolución. El usuario -110- puede interactuar directamente con el instrumento de fusión, la GUI o cualquier otro aspecto del sistema -100-. El dispositivo de almacenamiento de memoria -104- incluye un código legible por ordenador que es la realización exacta o equivalente a las realizaciones de la invención dada a conocer en el presente documento. El código legible por ordenador incluye instrucciones para realizar los procedimientos descritos en el presente documento y se ejecuta mediante cualquier dispositivo de procesamiento central adecuado conocido en la técnica. El software está disponible con la mayoría de los instrumentos de fusión y, a menudo, permite la visualización de las transiciones de fusión de la sonda y del producto como picos derivados, habitualmente mediante la estimación polinomial Salvitsky-Golay de la pendiente en cada punto. Varios dispositivos centrales de procesamiento ejecutan diferentes lenguajes de codificación y los lenguajes de codificación a menudo son distintos en su estructura y ejecución. Las diversas realizaciones se proporcionan a modo de ejemplo y el alcance de la presente invención no pretende estar limitado por los ejemplos y ecuaciones proporcionadas en el presente documento.

### LISTADO DE SECUENCIAS

<110> university of Utah Research Foundation Palais, Robert A Wittwer, Carl T

# ES 2 615 377 T3

<120> ANÁLISIS DE CURVAS DE FUSIÓN CON SUSTRACCIÓN DEL FONDO EXPONENCIAL

<130> IDH-34160-PCT

<150> US 60/719,250

<151> 2005-09-20

<160> 9

<170> PatentIn version 3.3

<210> 1

<211> 100

<212> ADN

<213> Homo sapiens

<400> 1

```
ctgaaagggtt acttcaagga caaaatacct gtattcctcg cctgtccagg gatctgctct    60
tacagattag aagtagtcct attagcccag aggcgatgct                               100
```

<210> 2

<211> 21

<212> ADN

<213> Homo sapiens

<400> 2

```
ctgaaagggtt acttcaagga c          21
```

<210> 3

<211> 15

<212> ADN

<213> Homo sapiens

<400> 3

```
gacatgcct ctggg          15
```

<210> 4

<211> 22

<212> ADN

<213> Homo sapiens

<400> 4

```
tggacaggcg aggaatacag gt        22
```

<210> 5

<211> 10

<212> ADN

<213> Homo sapiens

<400> 5

```
tcttgagaa          10
```

<210> 6

<211> 10

<212> ADN

<213> Homo sapiens



# ES 2 615 377 T3

<400> 6  
aggtcaacga 10

<210> 7  
<211> 10  
<212> ADN  
<213> Homo sapiens

<400> 7  
ttctcagttt 10

<210> 8  
<211> 11  
<212> ADN  
<213> Homo sapiens

<400> 8  
tatcatctt g 11

<210> 9  
<211> 45  
<212> ADN  
<213> Homo sapiens

<400> 9  
ccatggtgca cctgactcct gaggagaagt ctgccgttac tgccc 45

**REIVINDICACIONES**

1. Procedimiento para generar una curva de fusión corregida de una muestra de ácido nucleico, que comprende medir la fluorescencia de la muestra de ácido nucleico en función de la temperatura para producir una curva de fusión original, comprendiendo la muestra un ácido nucleico y una molécula que se une al ácido nucleico para formar un complejo detectable por fluorescencia, comprendiendo la curva de fusión original una señal de fluorescencia de fondo y una señal de la muestra de ácido nucleico; medir un primer y un segundo valores de pendiente a partir de puntos de la curva de fusión original situados en una región de la curva de fusión original en la que no tiene lugar la fusión del ácido nucleico de muestra; utilizar el primer valor de pendiente y el segundo valor de pendiente para buscar un exponencial representativo del ruido de fondo; y separar la señal de la fluorescencia de fondo de la señal de la muestra de ácido nucleico mediante la utilización de un algoritmo de función exponencial para generar una curva de fusión corregida, incluyendo el algoritmo de función exponencial el exponencial representativo del ruido de fondo y comprendiendo la curva de fusión corregida la señal de la muestra de ácido nucleico.

2. Procedimiento, según la reivindicación 1, en el que la señal de fluorescencia de fondo se calcula mediante el ajuste de un exponencial descendente a una curva de fluorescencia de fondo frente a temperatura, derivándose el exponencial descendente del primer valor de pendiente tomado de un punto anterior a la transición de fusión del ácido nucleico y el segundo valor de pendiente tomado de un punto posterior a la transición de fusión del ácido nucleico.

3. Procedimiento, según la reivindicación 1, que comprende, además, realizar una función de diferencia para identificar diferencias entre la curva de fusión corregida y una curva de fusión corregida de una segunda muestra de ácido nucleico.

4. Procedimiento, según la reivindicación 1, que comprende, además, agrupar los genotipos de una pluralidad de muestras de ácido nucleico adicionales, en el que el agrupamiento se consigue dinámicamente mediante la asociación de la distancia mínima entre los perfiles de fusión de las muestras de diferentes grupos.

5. Procedimiento, según la reivindicación 1, en el que la señal de fluorescencia de fondo se calcula utilizando la ecuación

$$e^{a(T_R - T_L)} = B'(T_R)/B'(T_L),$$

en la que

$$a = \frac{\ln (B'(T_R)/B'(T_L))}{(T_R - T_L)}$$

B'(T<sub>L</sub>) es el primer valor de pendiente, y B'(T<sub>R</sub>) es el segundo valor de pendiente, y

la separación de las señales de fluorescencia de fondo comprende la utilización de la ecuación

$$M(T) = F(T) - Ce^{a(T - T_L)}$$

en la que

M(T) es la señal de la muestra de ácido nucleico, F(T) es la curva de fusión original, y

$$C = B'(T_L)/a.$$

6. Procedimiento, según la reivindicación 1, en el que la muestra de ácido nucleico es un producto de doble cadena de una reacción PCR.

7. Procedimiento, según la reivindicación 1, en el que la muestra de ácido nucleico está genotipada para variaciones de secuencia conocidas y explorada para variaciones de secuencias desconocidas.

8. Procedimiento, según la reivindicación 1, en el que la curva de fusión corregida comprende una transición de fusión para un producto de PCR y una sonda no marcada.

9. Procedimiento, según la reivindicación 1, en el que la muestra comprende un amplicón múltiplex.

10. Procedimiento, según la reivindicación 1, en el que una o más variantes de secuencia pueden estar presentes en

el ácido nucleico, identificándose las variantes a partir de la curva de fusión o la derivada de la misma.

5 11. Procedimiento, según la reivindicación 3, que comprende, además, realizar una función de superposición de curvas con la curva de fusión corregida de la muestra y una curva de fusión de una segunda muestra corregida, requiriendo la función de superposición de curvas un único punto de datos.

10 12. Procedimiento, según la reivindicación 3, en el que la función de diferencia aproxima la distancia ortogonal entre la curva de fusión corregida y la curva de fusión de la segunda muestra corregida mediante la utilización de la siguiente ecuación:  $f_1(T) - f_2(T) = \max \{\sqrt{(1 + f_1'(T)^2)}, \sqrt{(1 + f_2'(T)^2)}\}$ .

13. Procedimiento, según la reivindicación 3, en el que la función de diferencia compensa el énfasis exagerado en las regiones de fusión de la curva de fusión corregida y la curva de fusión de la segunda muestra corregida.

15 14. Procedimiento, según la reivindicación 3, en el que la función de diferencia compensa el menor énfasis en las regiones de fusión primarias de la curva de fusión corregida y la curva de fusión de la segunda muestra corregida.

20 15. Procedimiento, según la reivindicación 7, en el que la reacción PCR es asimétrica, produciendo más de una primera cadena de producto que de una segunda cadena de producto, realizándose la PCR en presencia de una sonda no marcada configurada para hibridarse a la primera cadena de producto.

16. Procedimiento, según la reivindicación 13, en el que la función de diferencia genera una medida de la diferencia que refleja las variaciones dependientes de la secuencia entre la primera curva de fusión de muestra y la curva de fusión de la segunda muestra.

25 17. Procedimiento, según la reivindicación 13, en el que la función de diferencia identifica un genotipo que corresponde a la primera muestra y un genotipo que corresponde a la segunda muestra.

30 18. Procedimiento, según la reivindicación 17, que comprende, además, agrupar los genotipos de una muestra de ácido nucleico, en el que el agrupamiento se consigue dinámicamente mediante la asociación de la distancia mínima entre las curvas de fusión de muestra de grupos distintos.

19. Sistema para generar una curva de fusión corregida, que comprende:

35 un instrumento de fusión de alta resolución para calentar un complejo detectable por fluorescencia mientras se controla su fluorescencia, comprendiendo el complejo un ácido nucleico y una especie fluorescente, estando el instrumento de fusión adaptado para medir y registrar la temperatura de la muestra y la fluorescencia de la muestra para determinar la fluorescencia de la muestra en función de la temperatura de la muestra para producir un perfil de fusión, comprendiendo el perfil de fusión una señal de fluorescencia de fondo y una señal de fluorescencia de muestra;

40 una unidad central de procesamiento (CPU) para llevar a cabo instrucciones ejecutables por ordenador; y un dispositivo de almacenaje de memoria para almacenar instrucciones ejecutables por ordenador que, cuando son ejecutadas por la CPU, provocan que la CPU lleve a cabo un proceso para analizar las variaciones de secuencia del ácido nucleico, en el que el procedimiento incluye generar una curva de fusión corregida mediante el procedimiento, según la reivindicación 1.

45 20. Sistema, según la reivindicación 19, que comprende, además: realizar una función de diferencia para identificar diferencias entre la curva de fusión de muestra corregida y una curva de fusión de la segunda muestra corregida, habiéndose separado previamente la señal de fluorescencia de fondo de la primera curva de fusión de muestra corregida y la curva de fusión de la segunda muestra corregida.

50 21. Sistema, según la reivindicación 17, que comprende, además, agrupar los genotipos de una pluralidad de muestras de ácido nucleico, alcanzándose el agrupamiento dinámicamente mediante la asociación de la distancia mínima entre las curvas de fusión de muestra de grupos distintos.

55 22. Sistema, según la reivindicación 19, en el que el sistema se configura para el genotipado para mutaciones conocidas y la exploración para mutaciones no conocidas.

23. Sistema, según la reivindicación 19, en el que la función exponencial se calcula utilizando la ecuación

60 
$$e^{a(T_R - T_L)} = B'(T_R)/B'(T_L),$$

$$a = \frac{\ln(B'(T_R)/B'(T_L))}{(T_R - T_L)}$$

65  $B'(T_L)$  es el primer valor de pendiente, y

$B'(T_R)$  es el segundo valor de pendiente, y

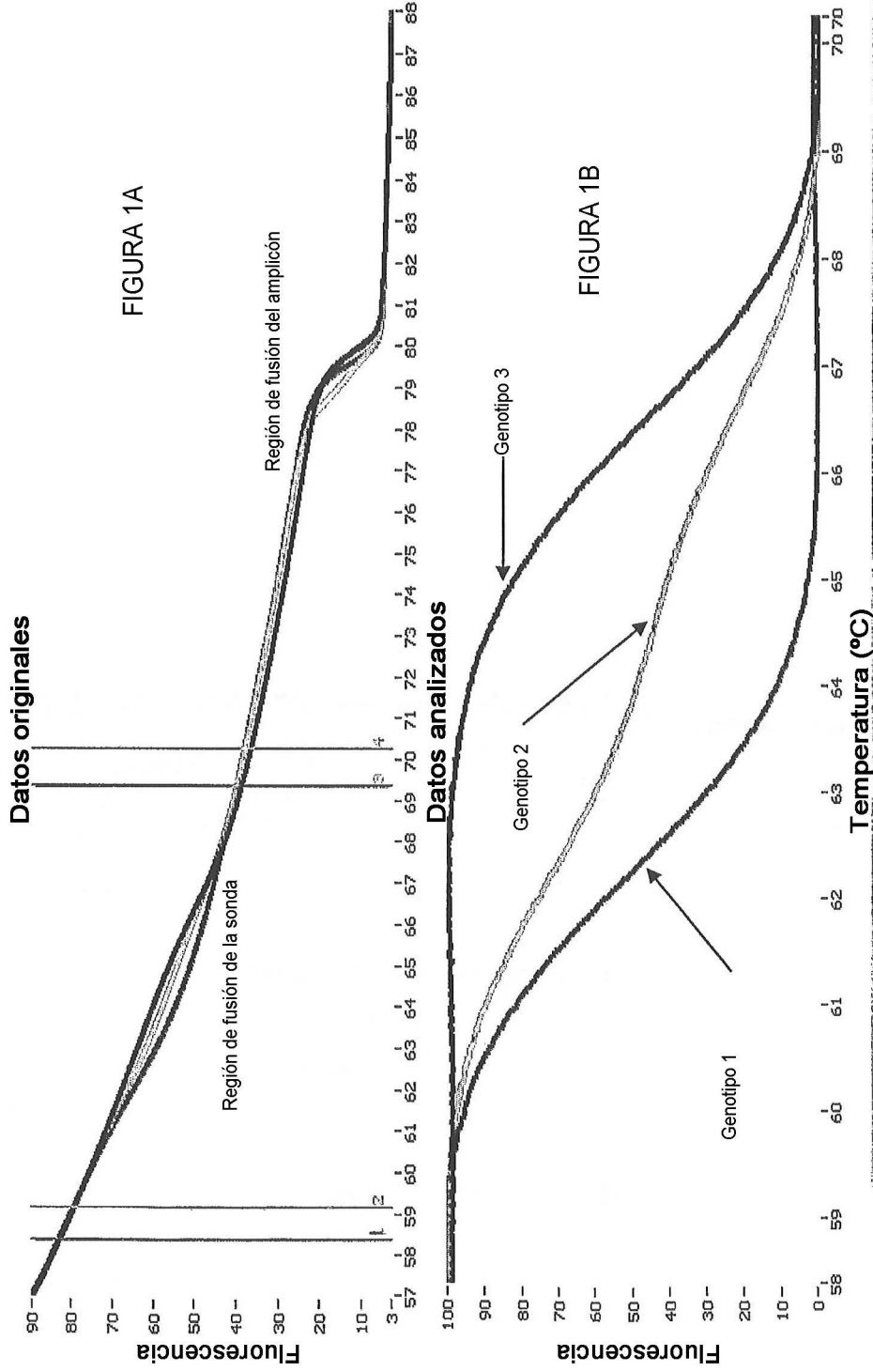
la sustracción comprende la utilización de la ecuación

5 
$$M(T) = F(T) - Ce^{a(T-T_L)}$$

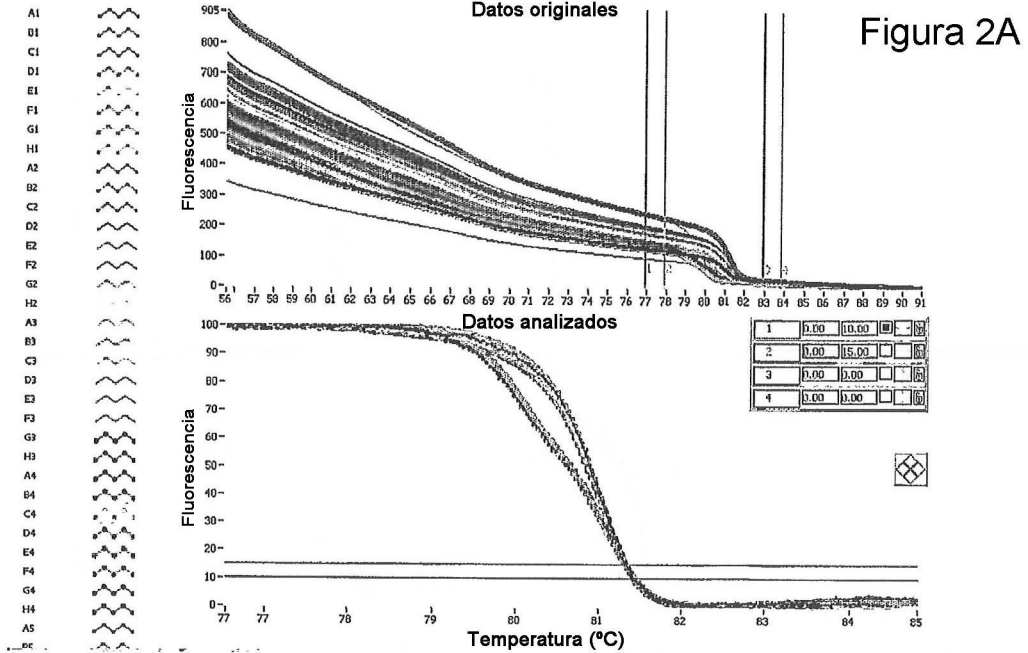
en la que

10  $M(T)$  es la señal de la muestra,  
 $F(T)$  es el perfil de fusión, y

$$C = B'(T_L)/a.$$



Por favor, agrupe el amplíción con los cursores inferiores (1 y 2) y los cursores superiores (3 y 4)



Muestras

- A1
- B1
- C1
- D1
- E1
- F1
- G1
- H1
- A2
- B2
- C2
- D2
- E2
- F2
- G2
- H2
- A3
- B3
- C3
- D3
- E3
- F3
- G3
- H3
- A4
- B4
- C4
- D4
- E4
- F4
- G4
- H4
- A5
- B5
- C5
- D5
- E5
- F5
- G5
- H5
- A6
- B6
- C6
- D6
- E6
- F6
- G6
- H6
- A7
- B7
- C7
- D7
- E7
- F7
- G7

**Figura 2B**

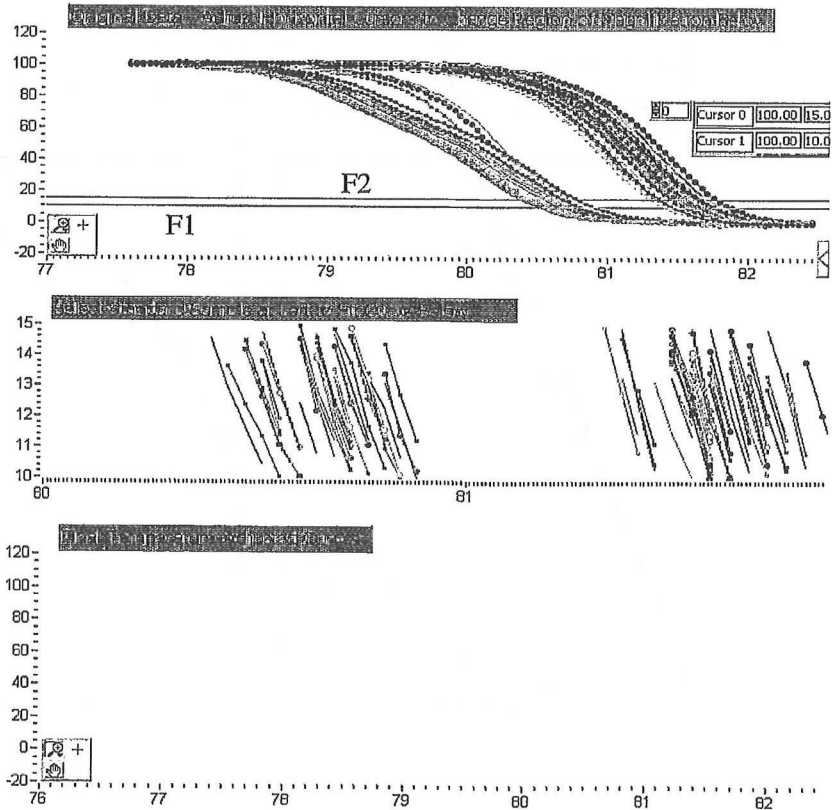


Figura 3A

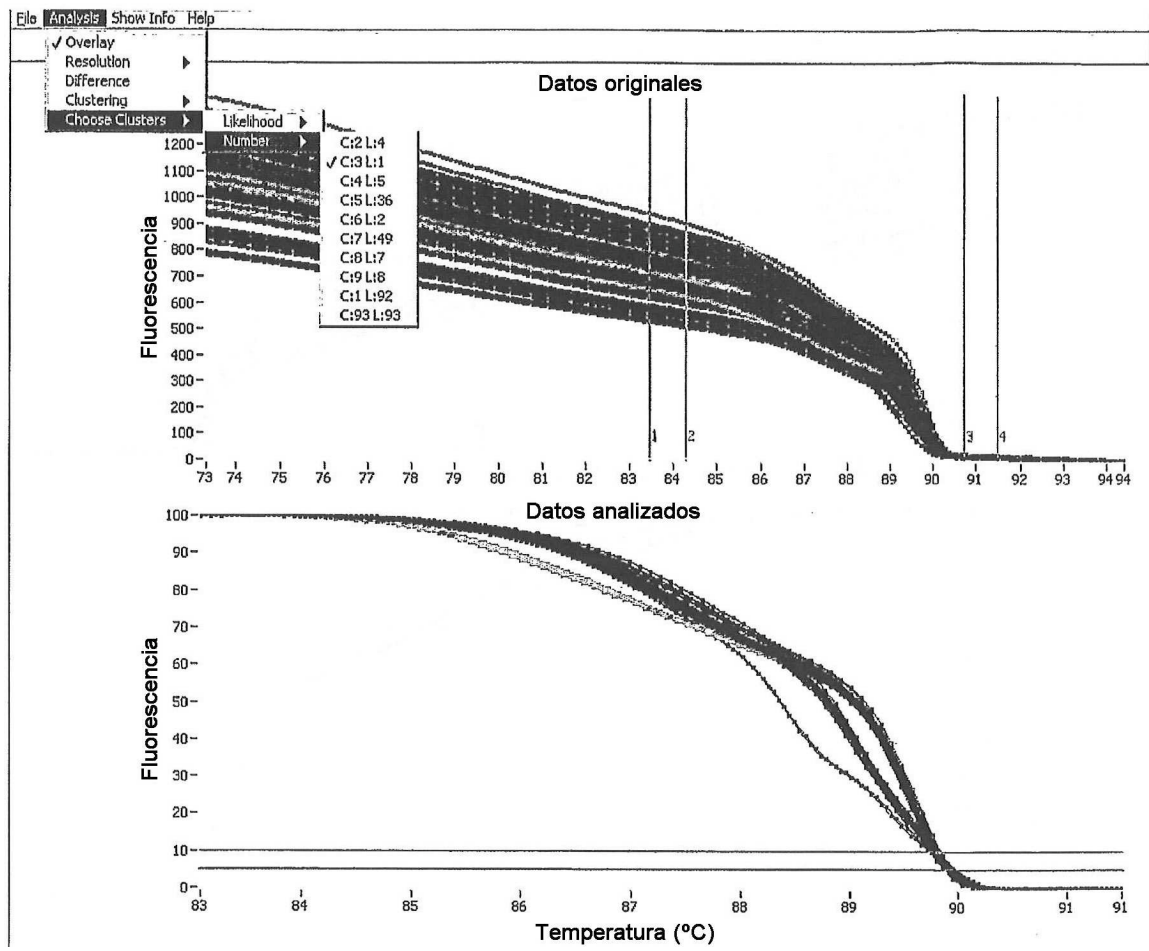


Figura 3B

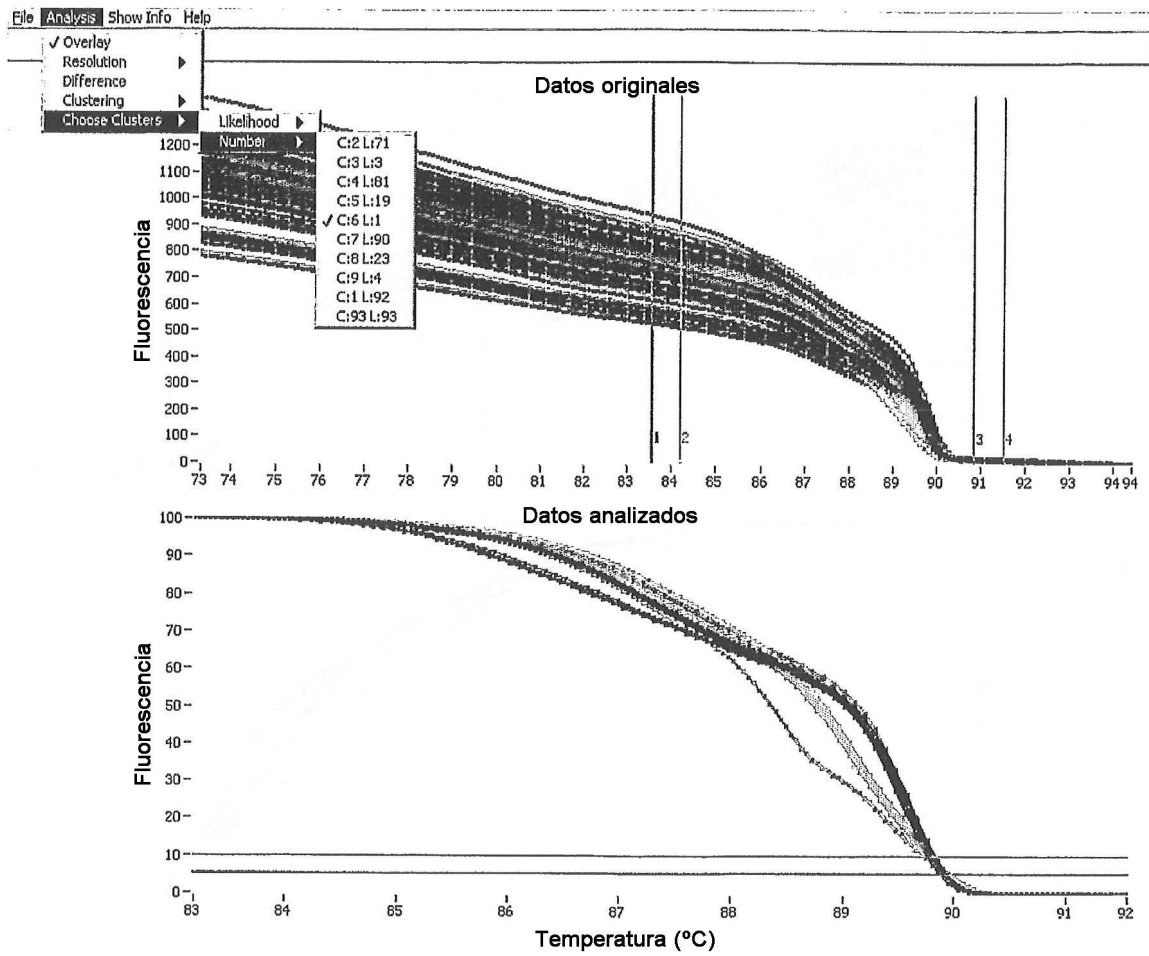




Figura 3C

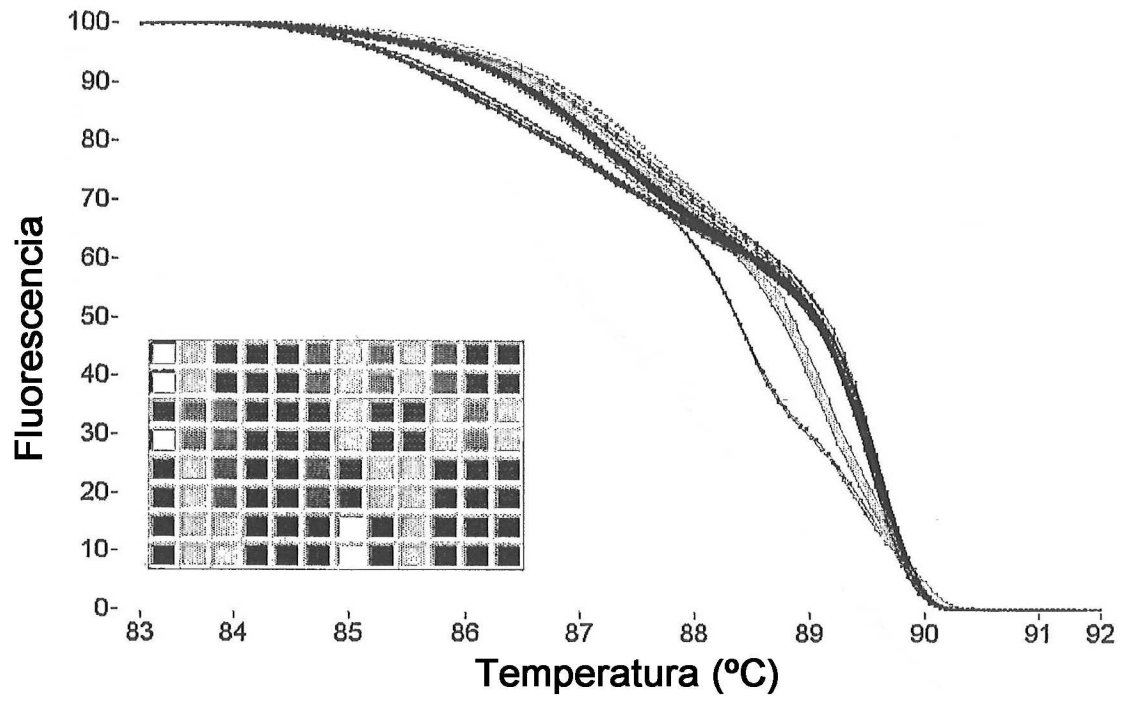
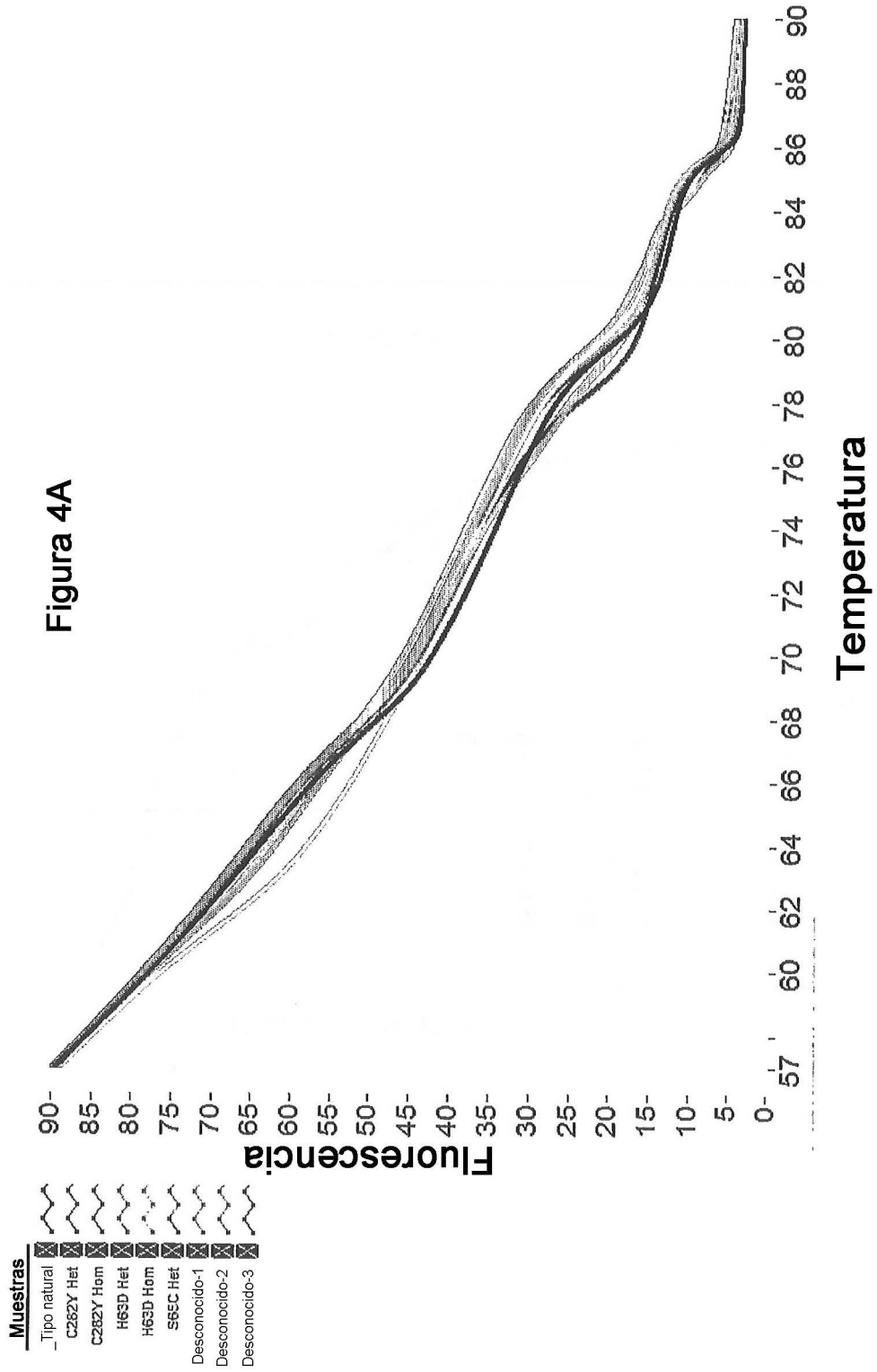
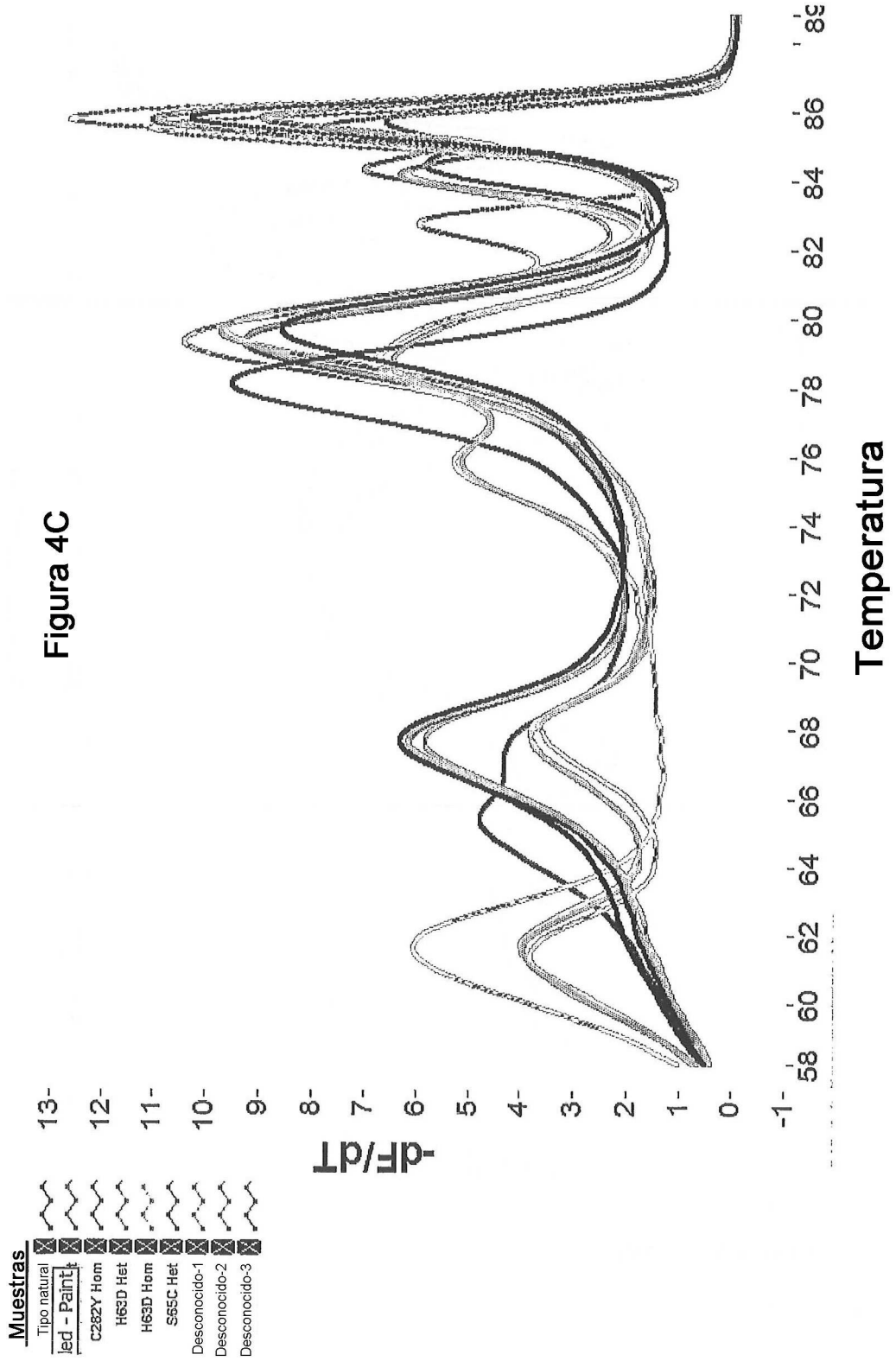
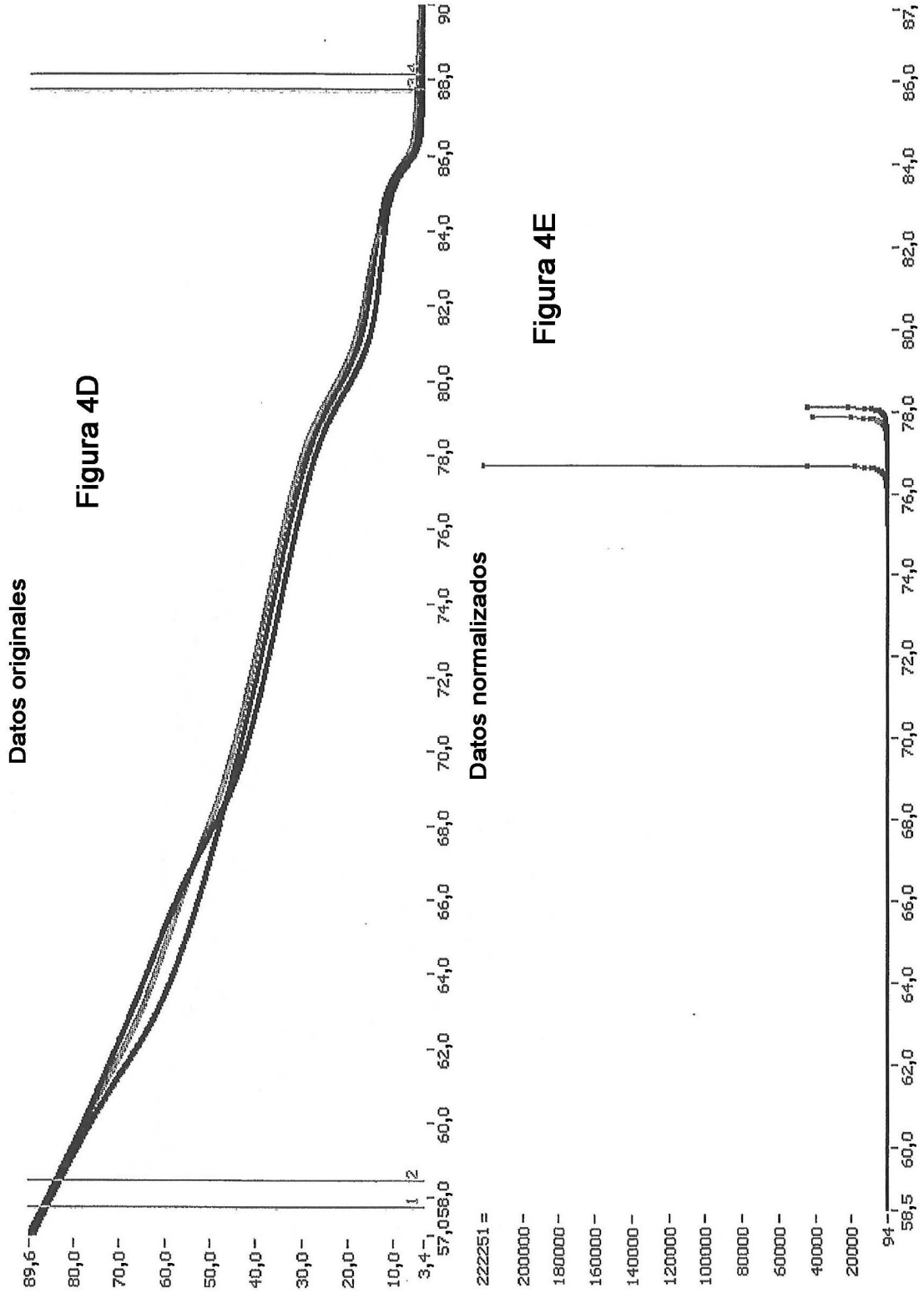


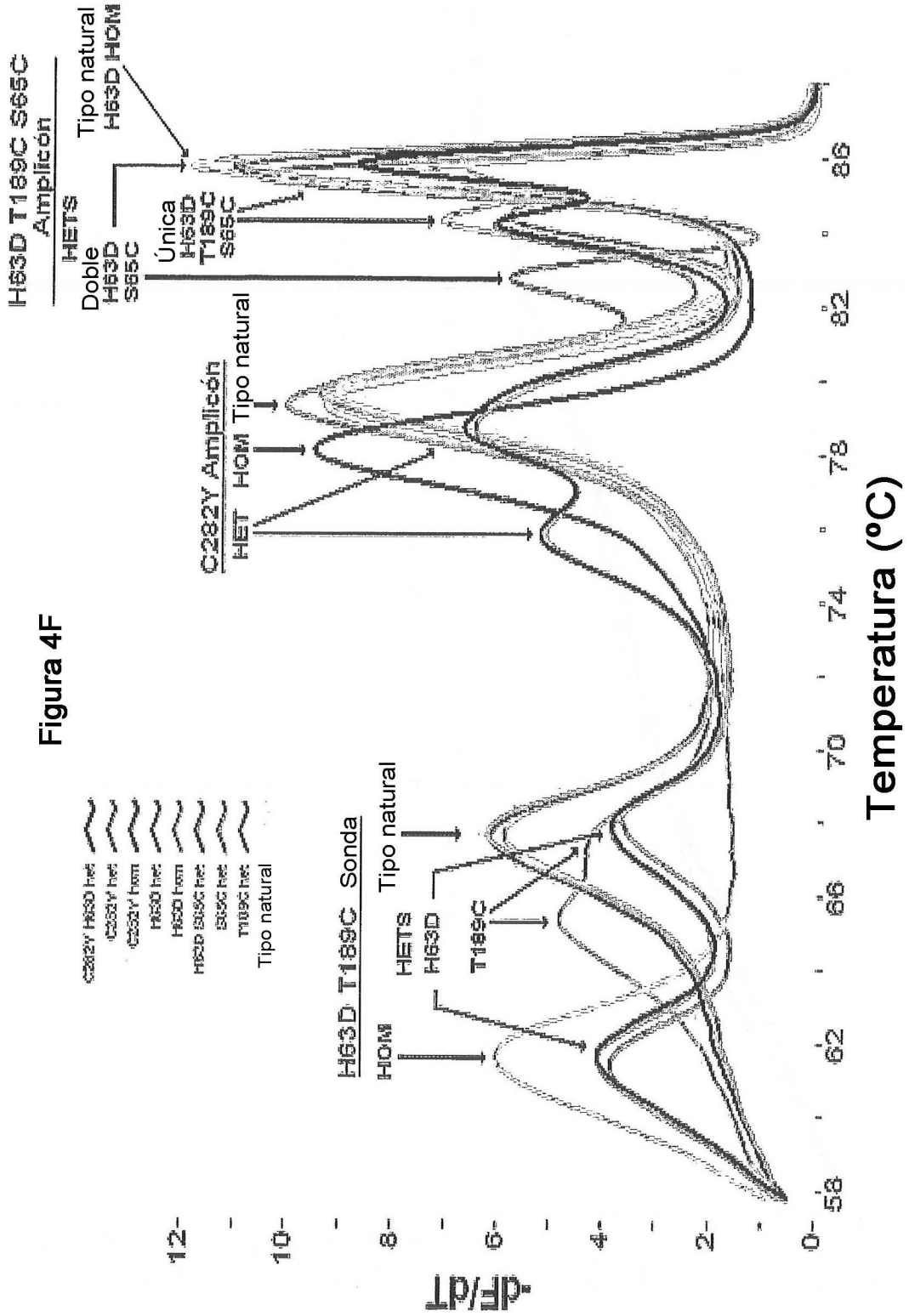
Figura 4A











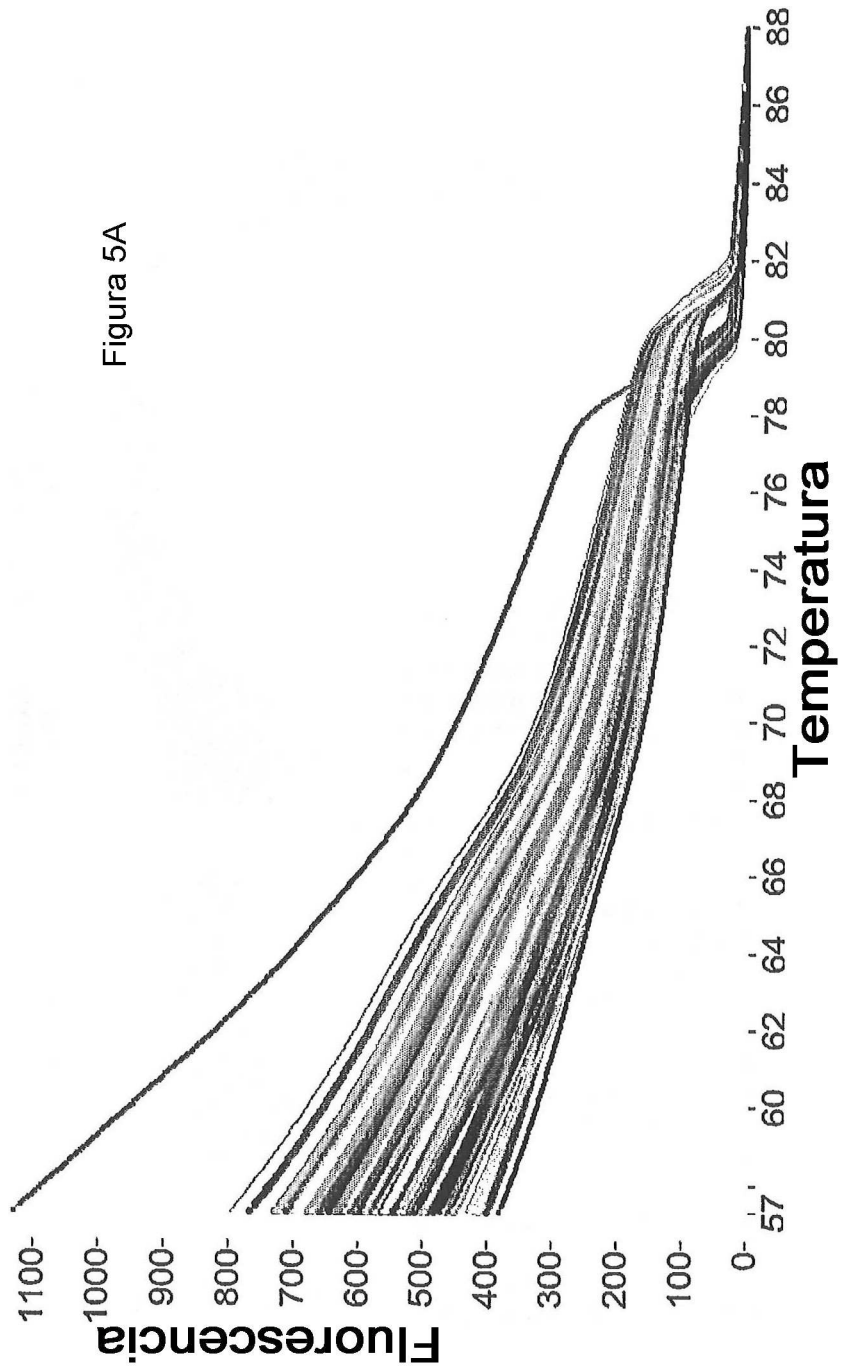




Figura 5B

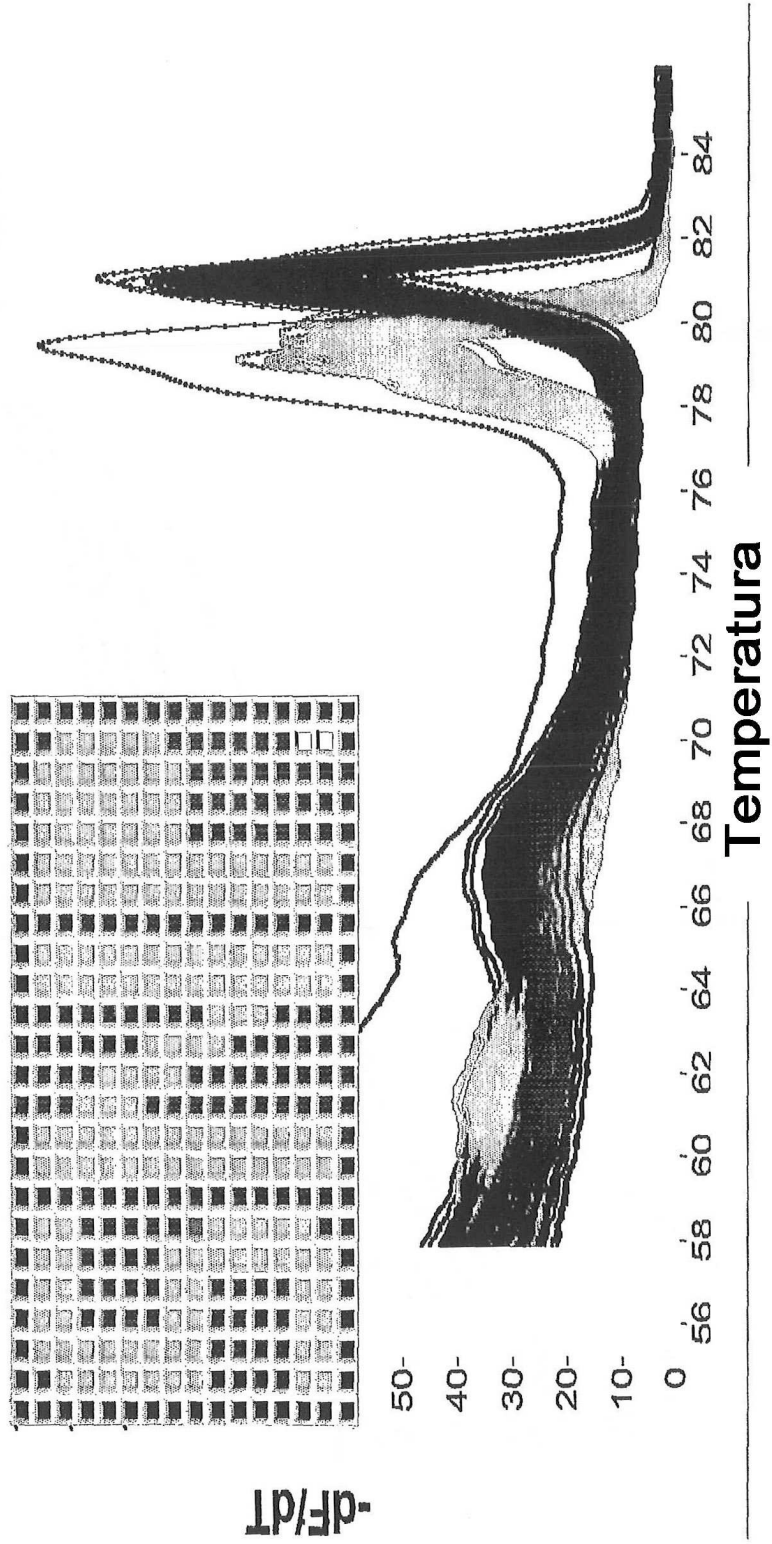




Figura 5C

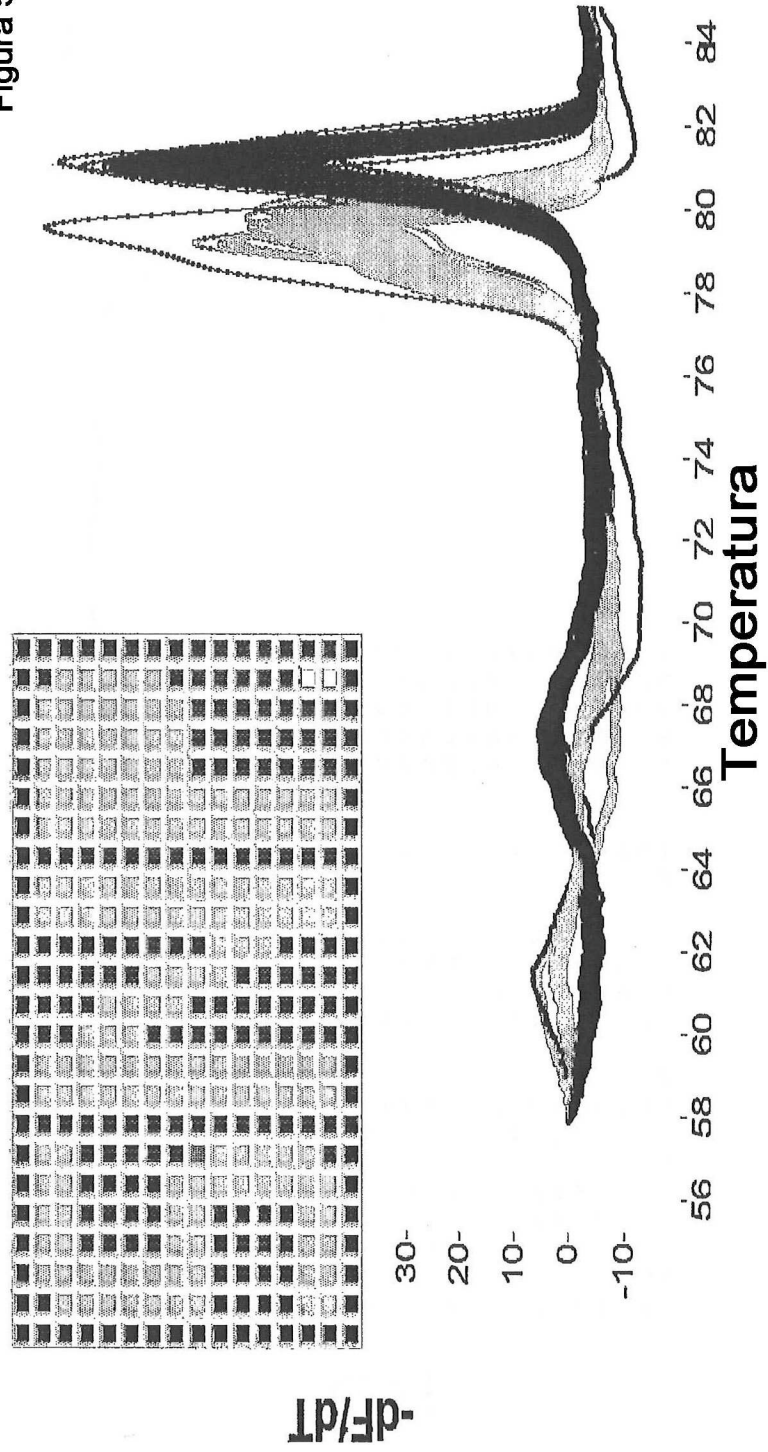


Figura 5D

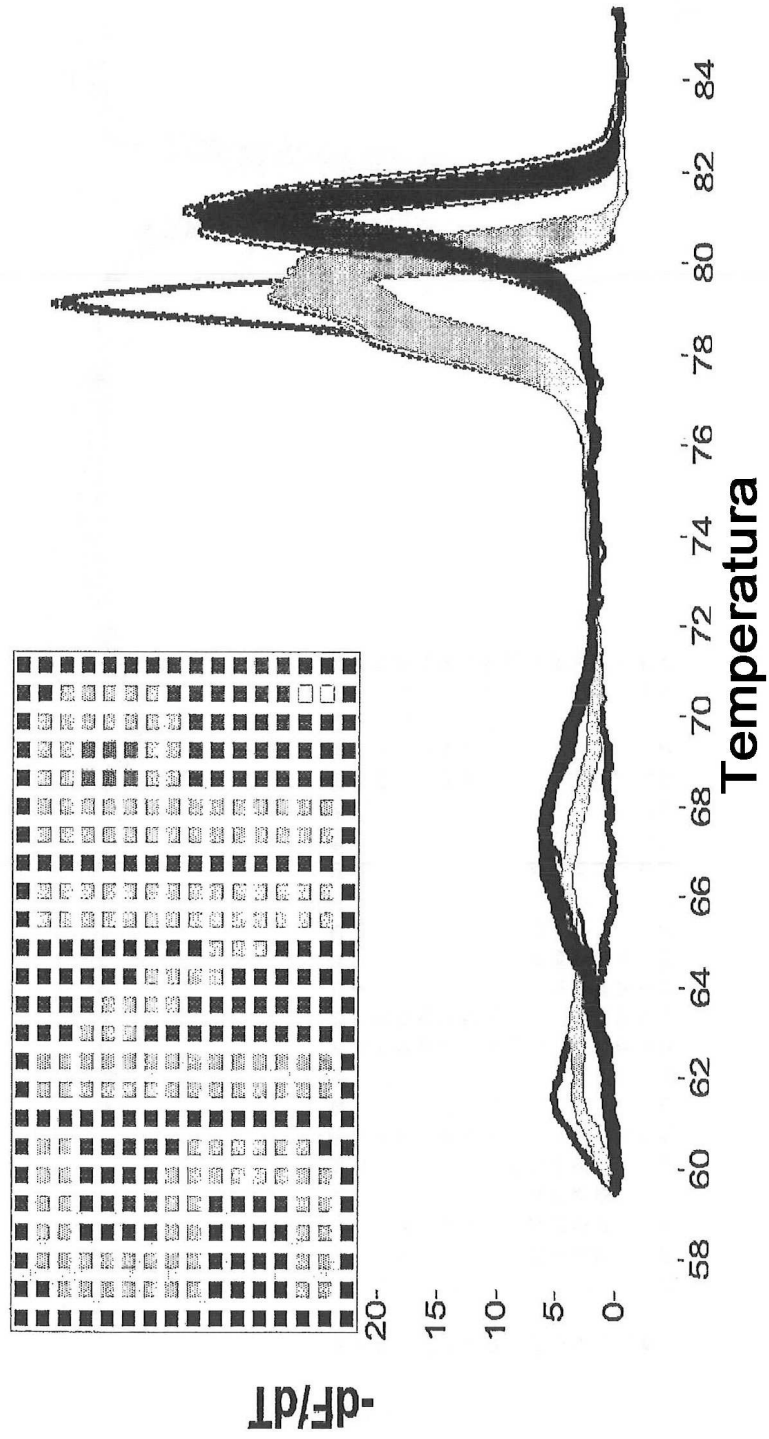
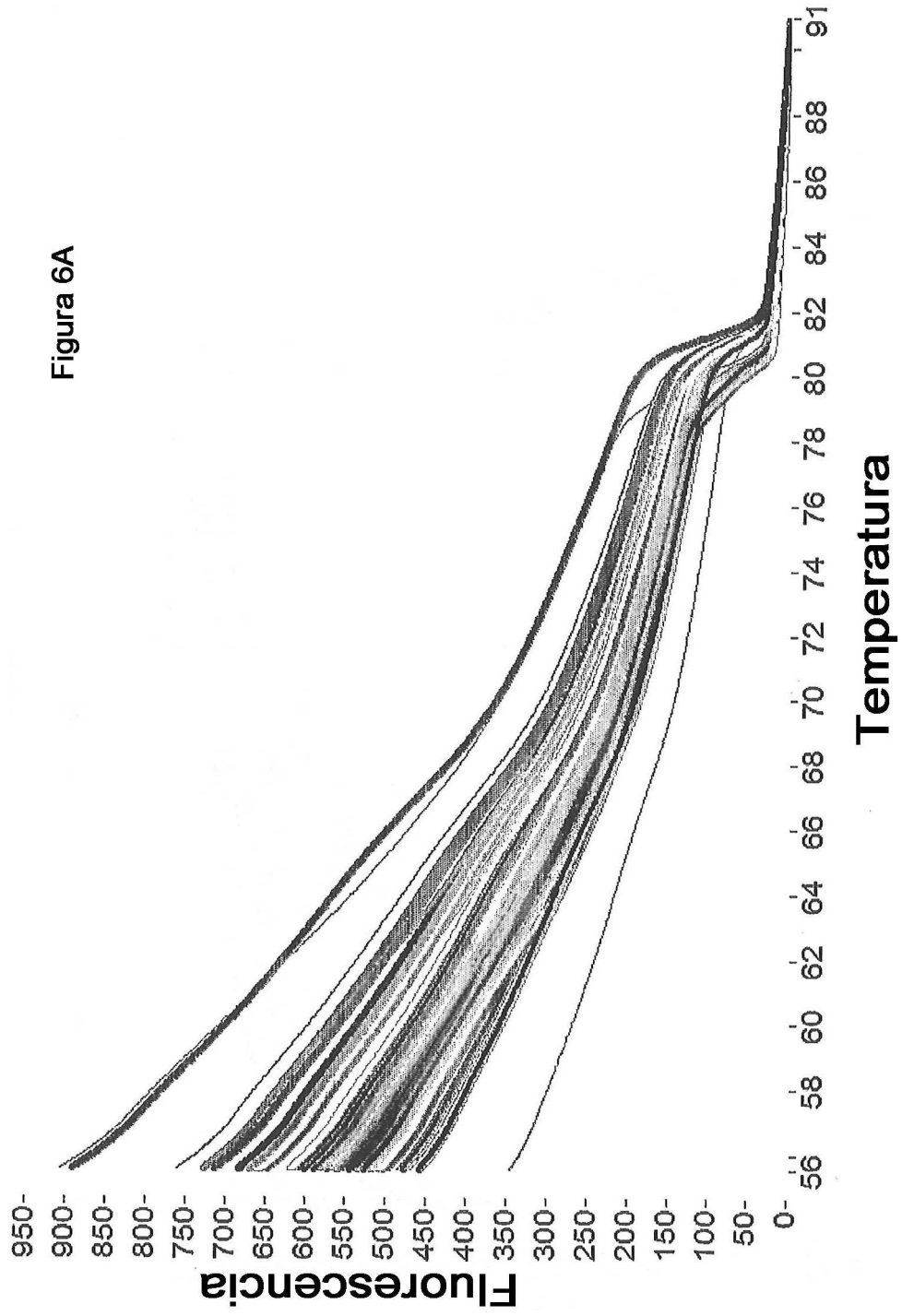


Figura 6A



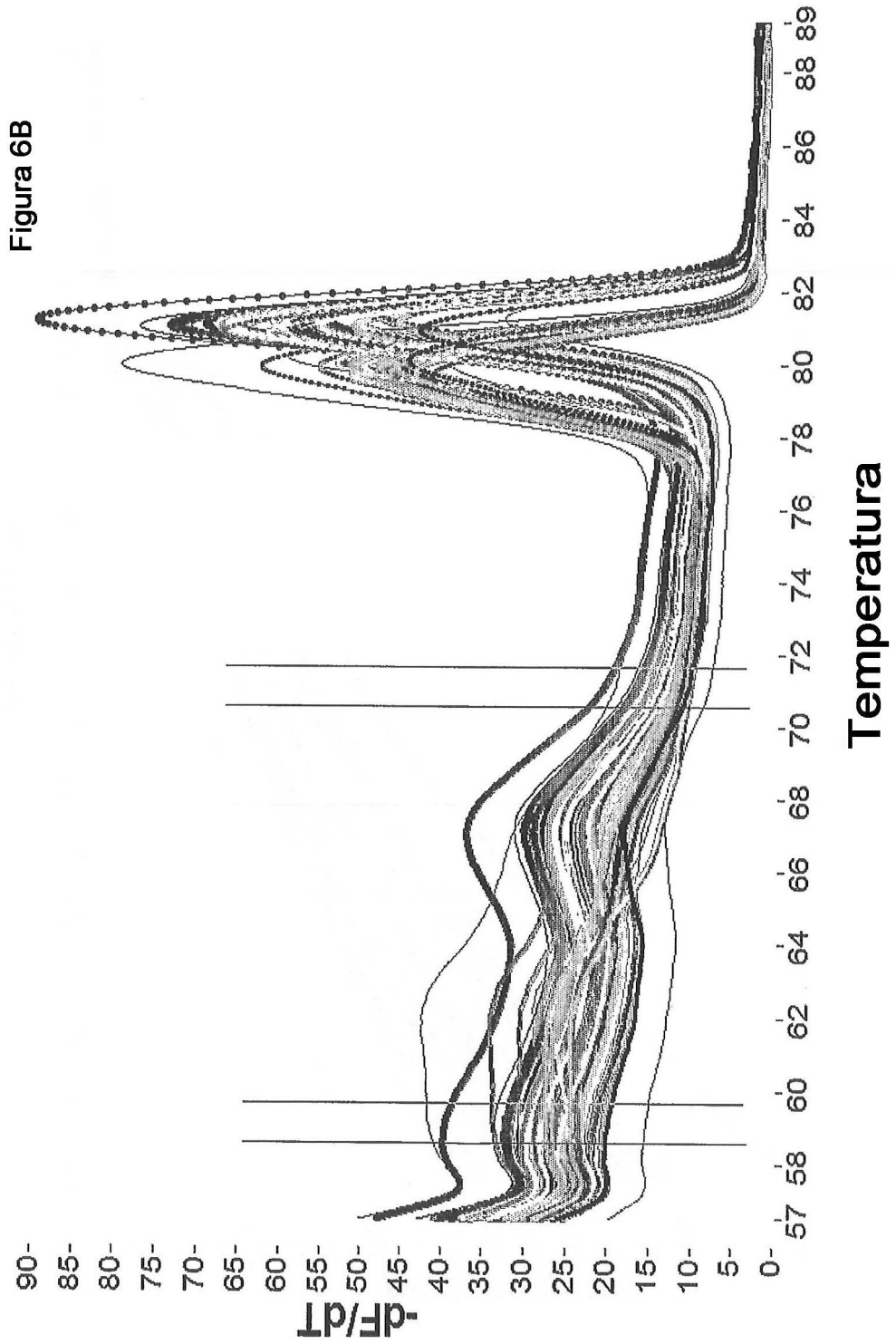


Figura 6C

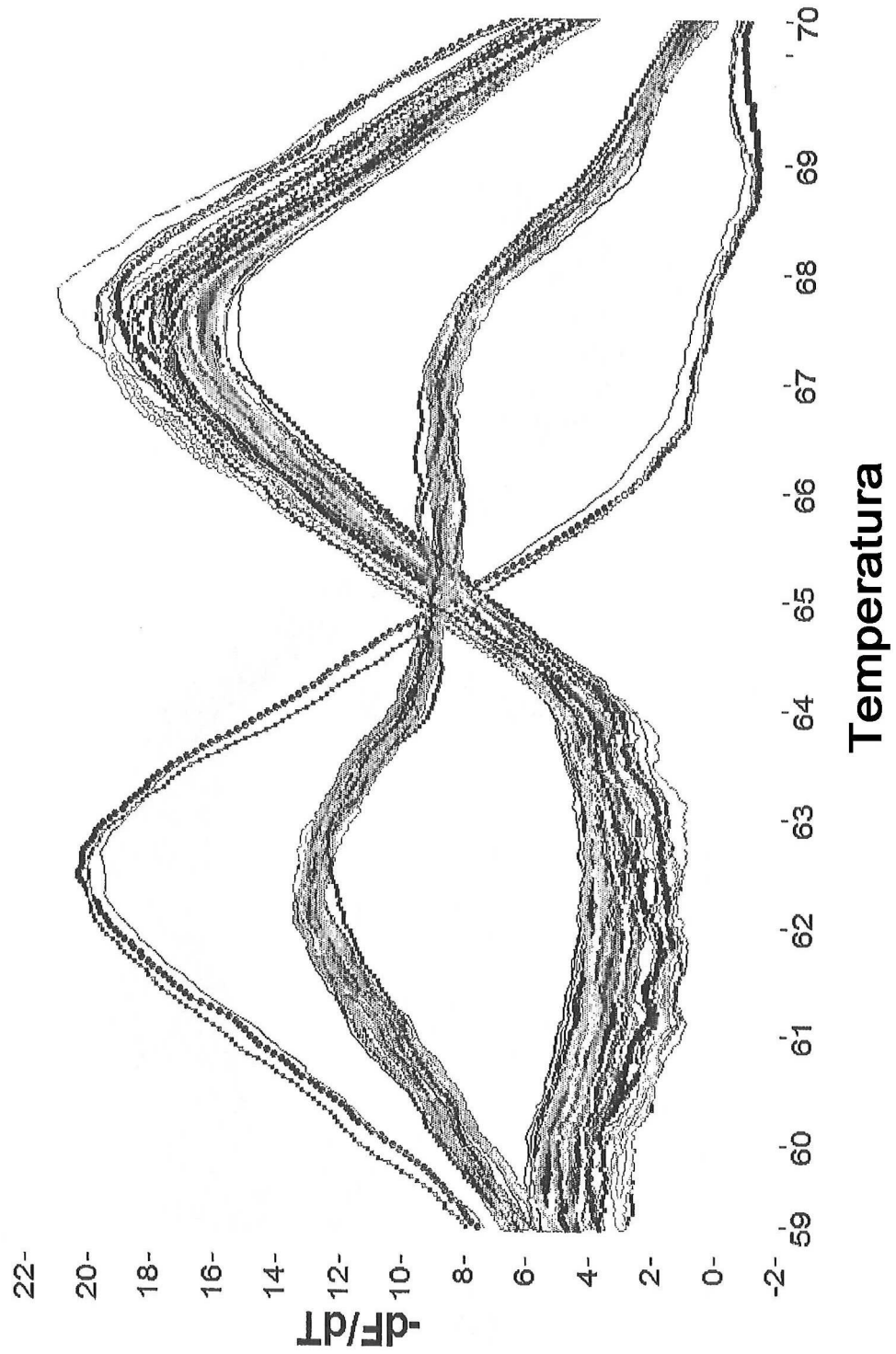




Figura 6D

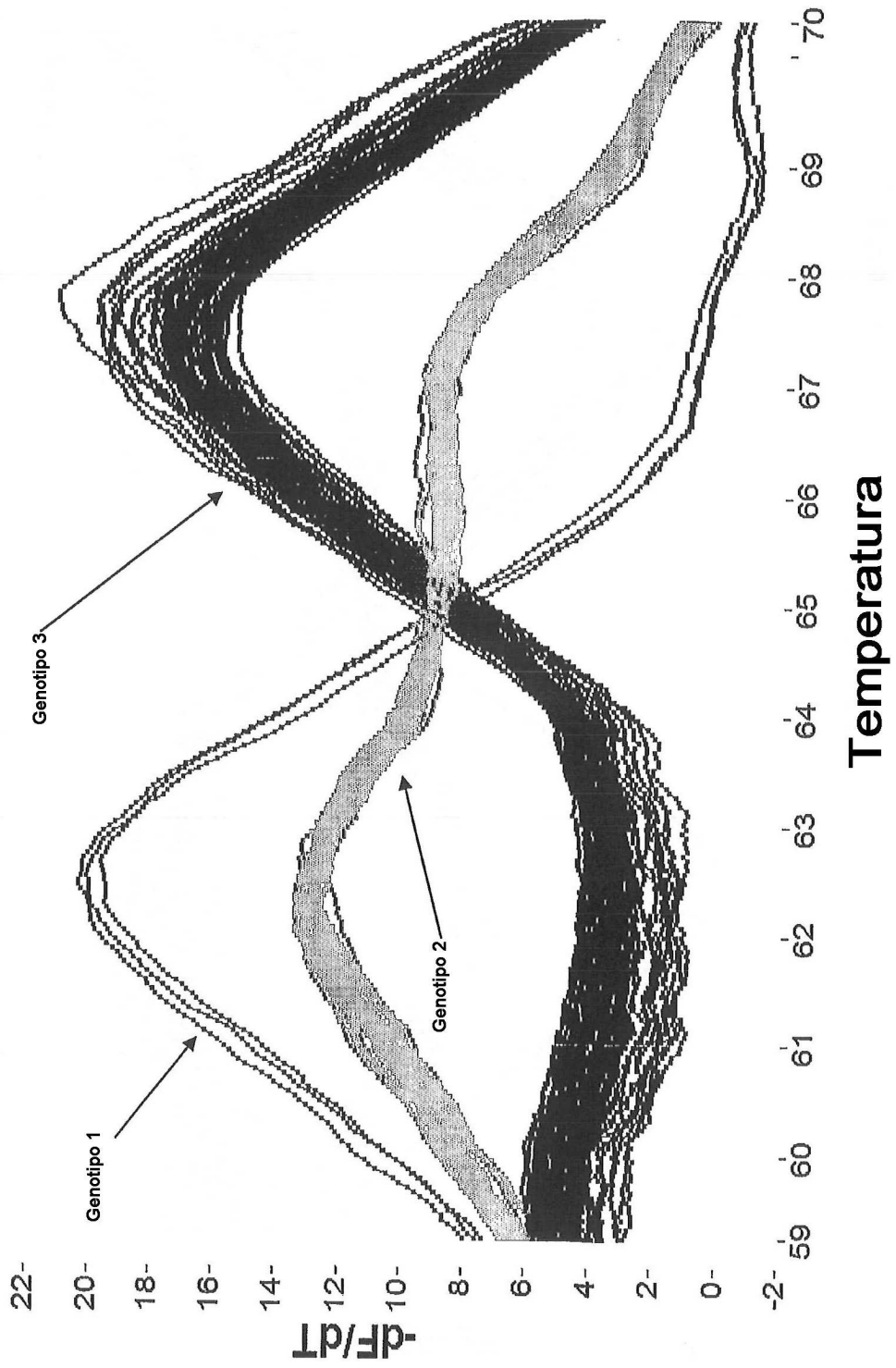


Figura 7A

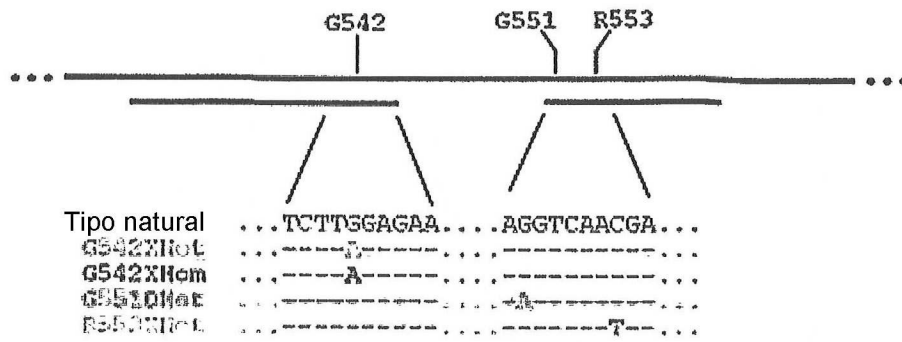


Figura 7B

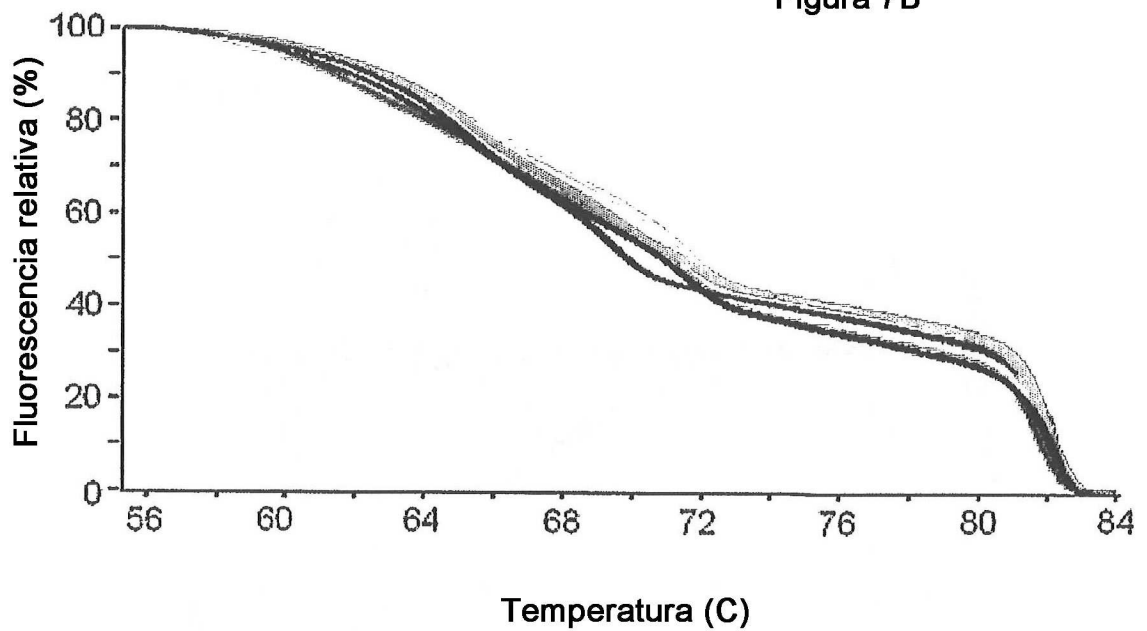


Figura 7C

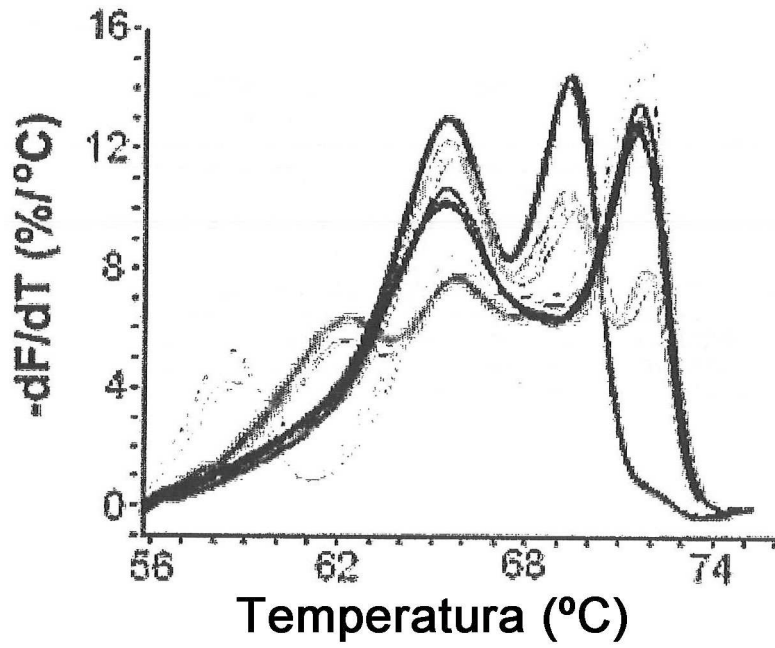


FIGURA 7D

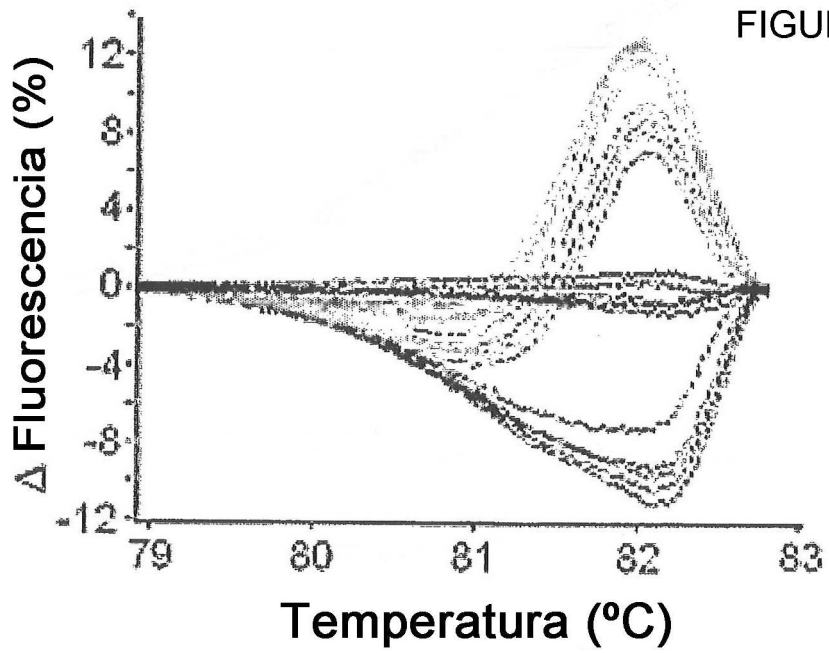




Figura 8A

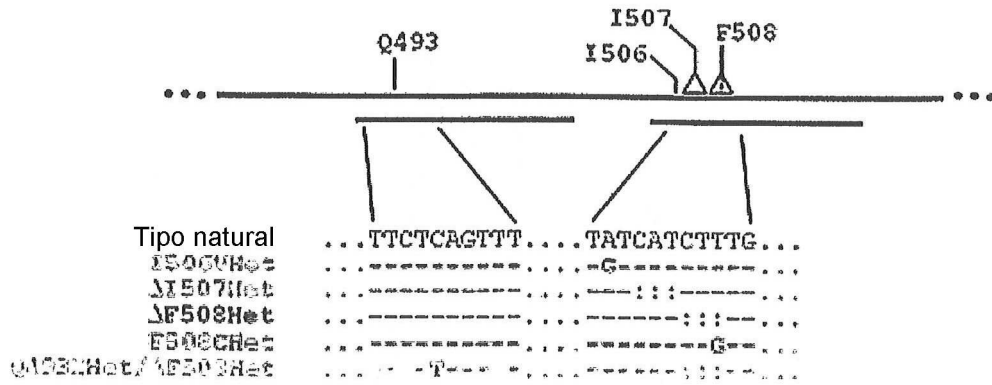


Figura 8B

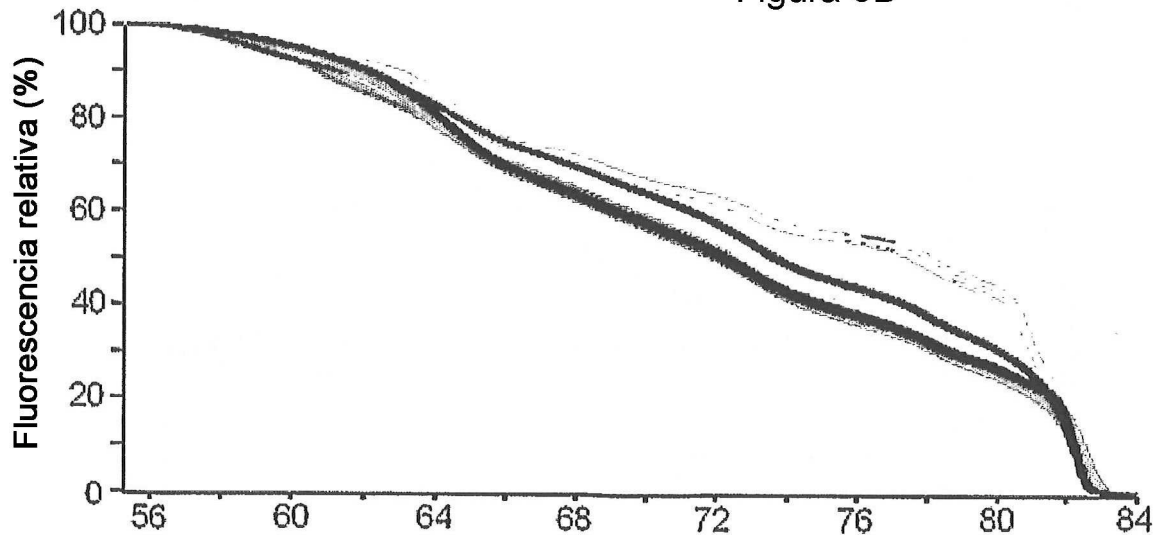


Figura 8C

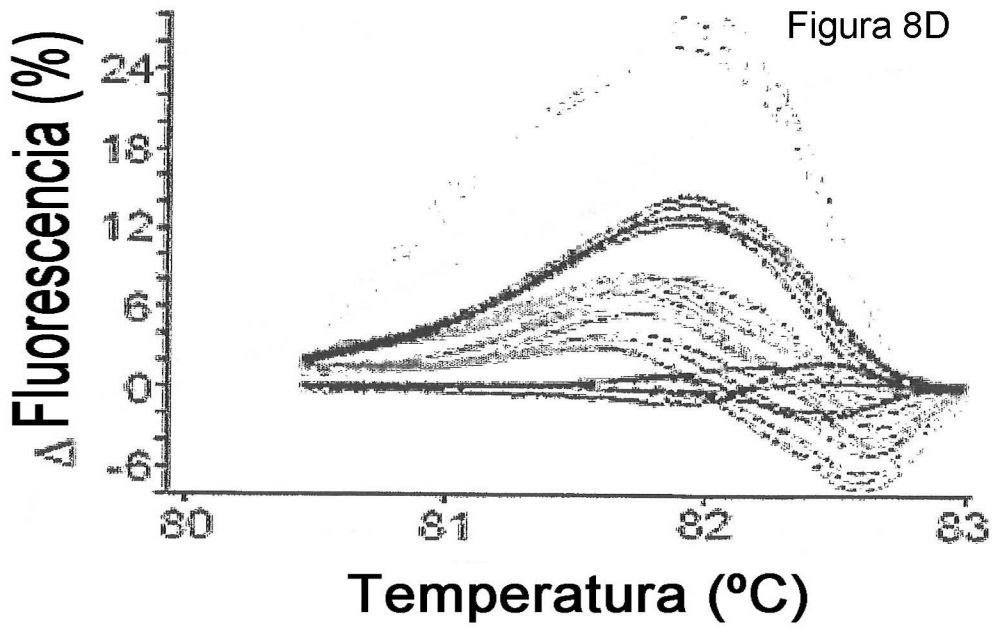
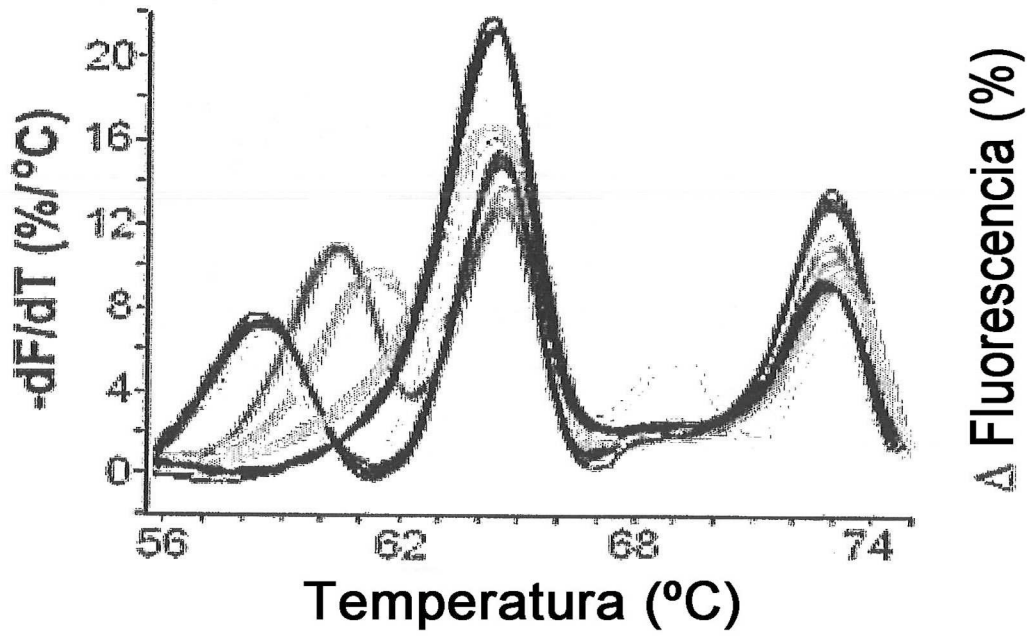


Figura 9A

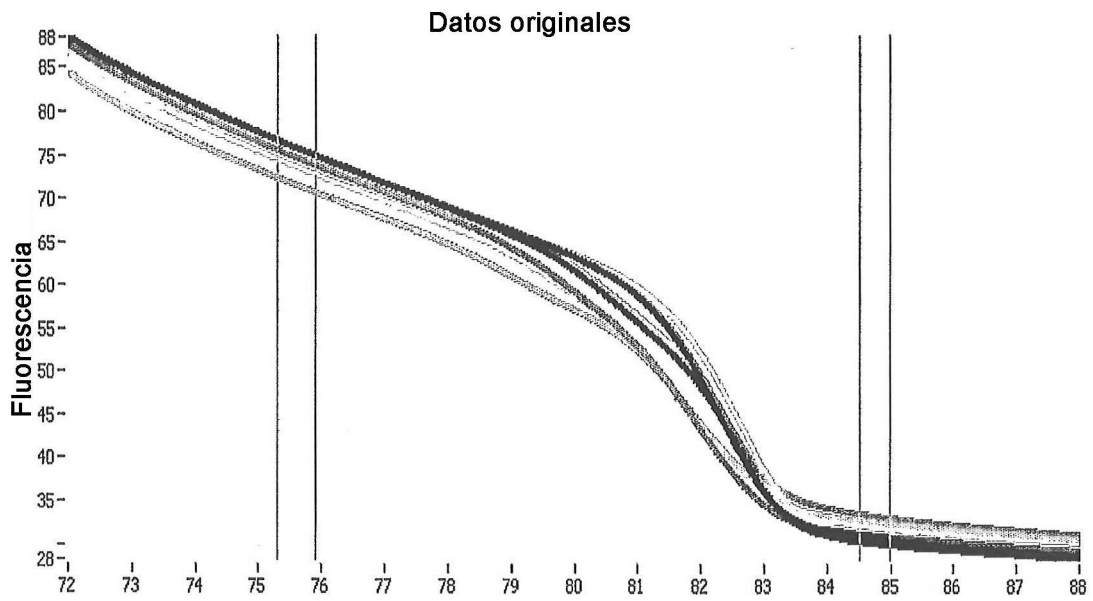
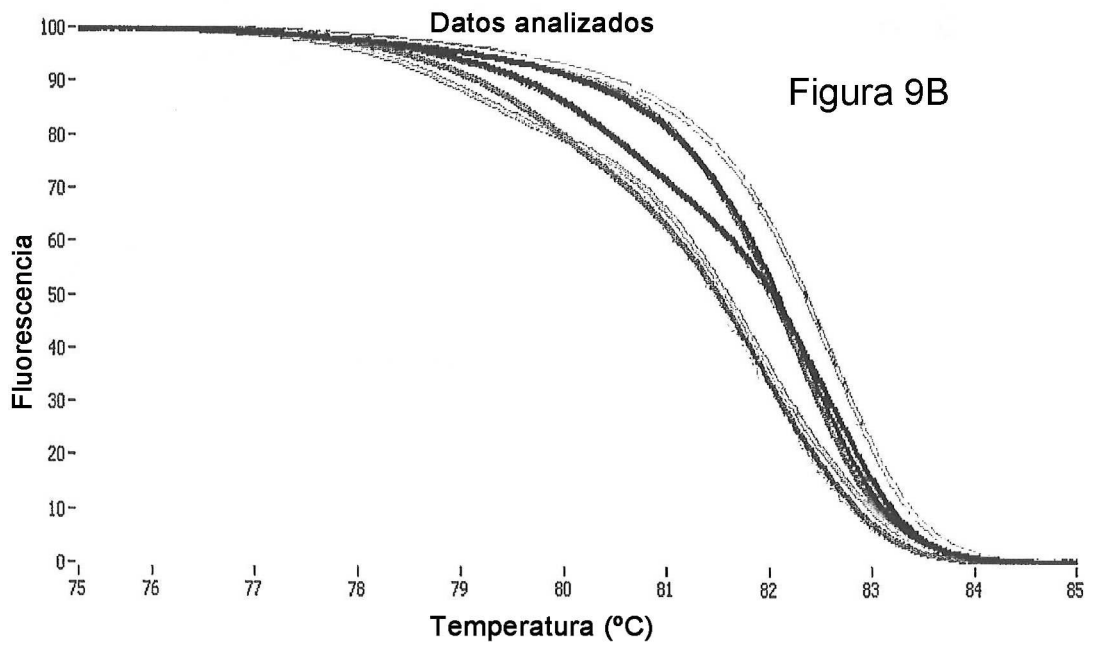
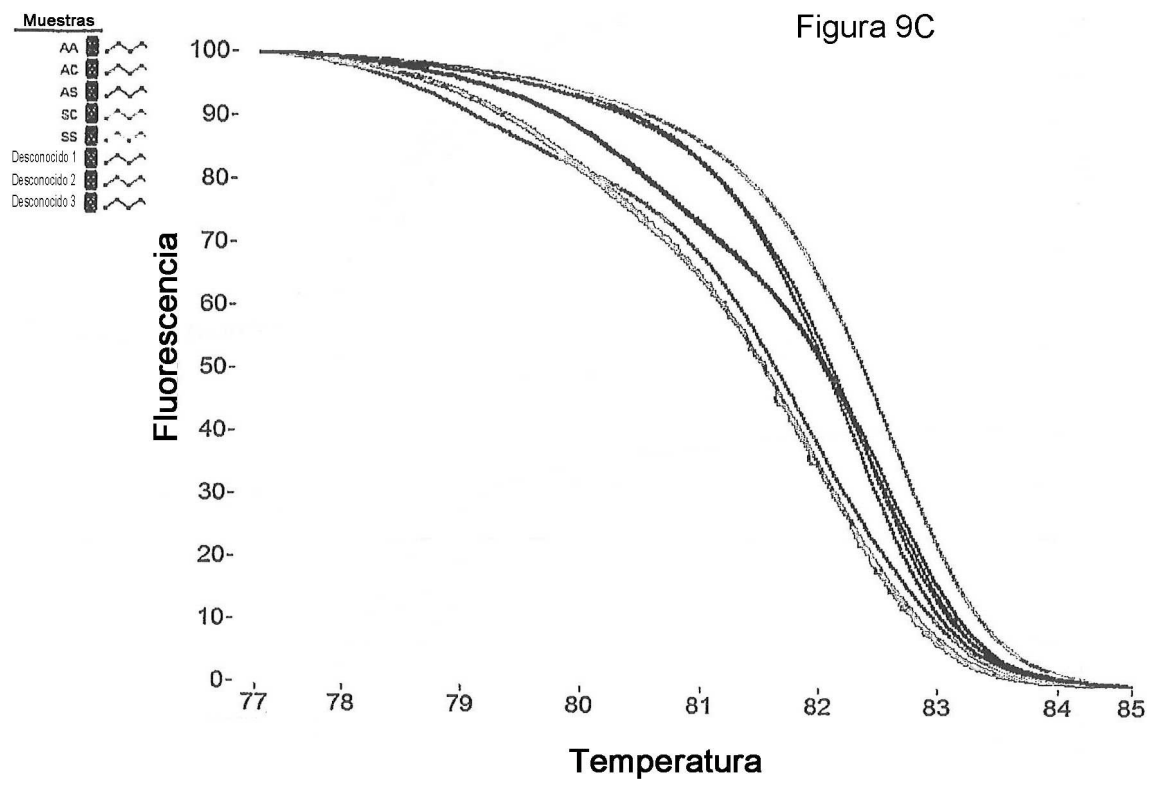
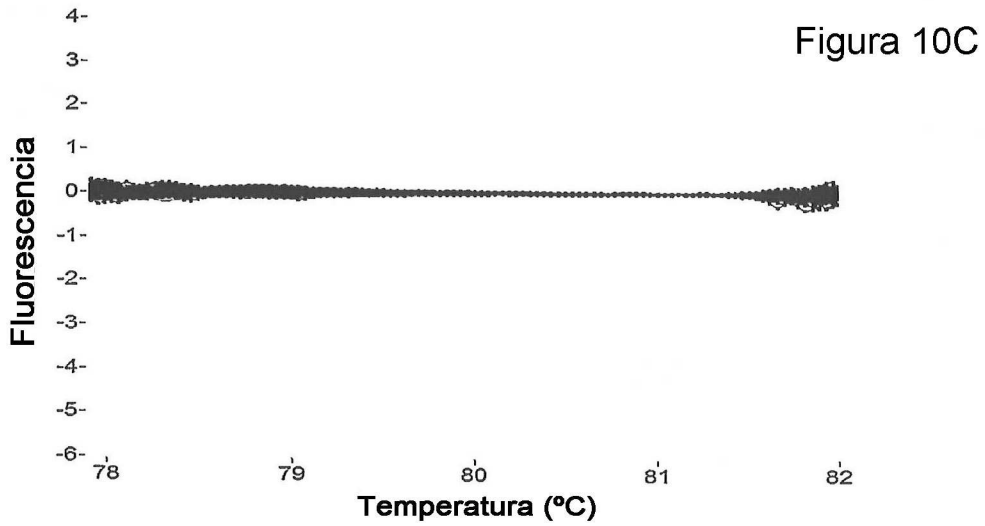
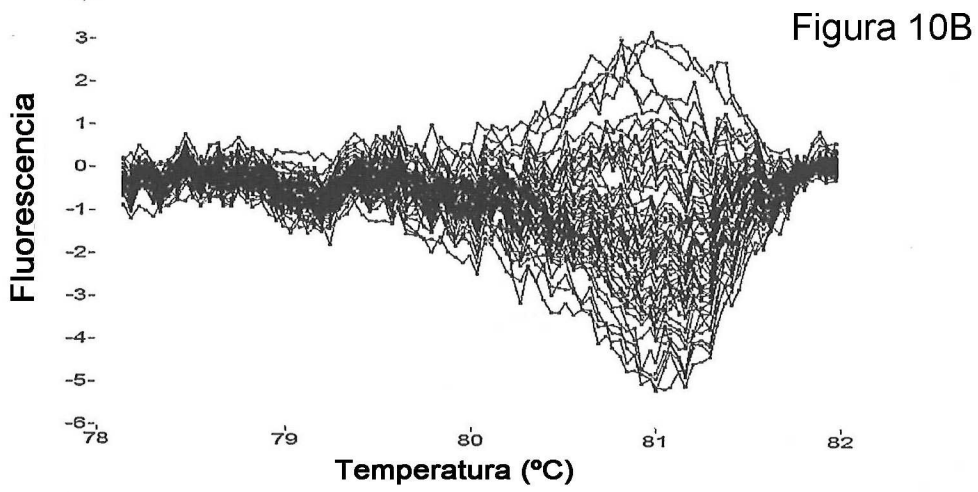
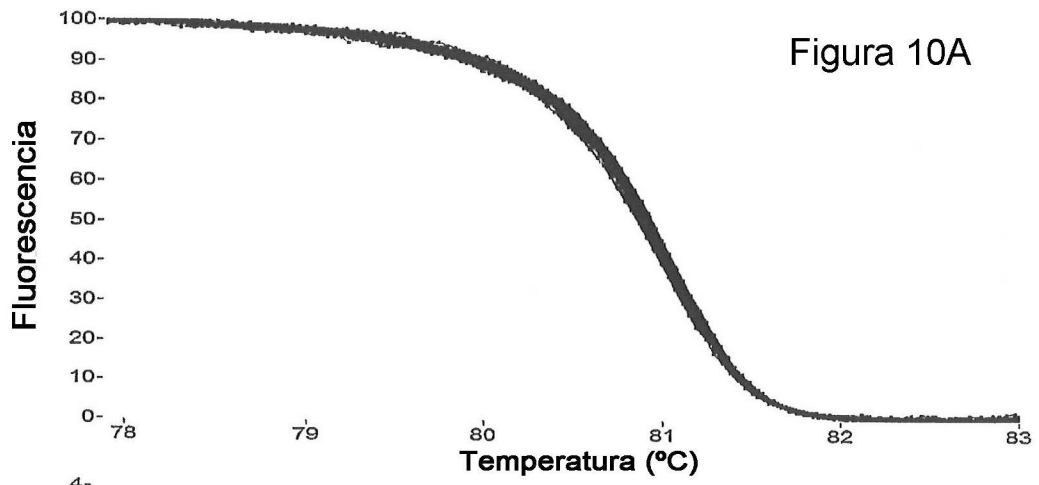


Figura 9B







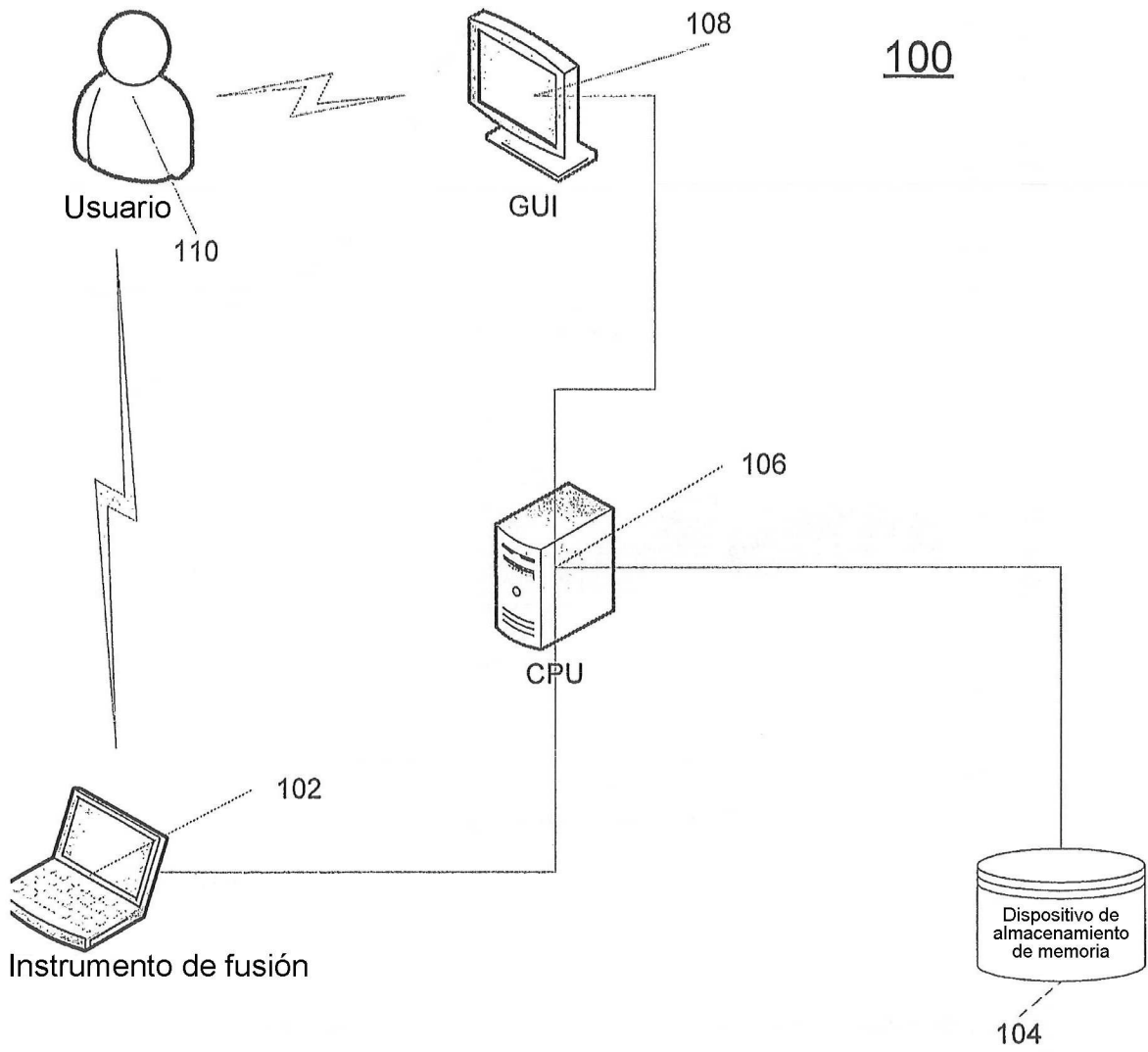


FIG. 11