



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 271 946**

51 Int. Cl.:

C12N 15/53 (2006.01)

C12N 9/06 (2006.01)

C12Q 1/68 (2006.01)

A61K 48/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **95919289 .9**

86 Fecha de presentación : **25.05.1995**

87 Número de publicación de la solicitud: **0755450**

87 Fecha de publicación de la solicitud: **29.01.1997**

54 Título: **cDNA para metilenotetrahidrofolato reductasa humana.**

30 Prioridad: **26.05.1994 GB 9410620**

45 Fecha de publicación de la mención BOPI:
16.04.2007

45 Fecha de la publicación del folleto de la patente:
16.04.2007

73 Titular/es: **McGill University**
845 Sherbrooke Street West
Montreal, Québec H3A 2T5, CA

72 Inventor/es: **Rozen, Rima y**
Goyette, Philippe

74 Agente: **Torner Lasalle, Elisabet**

ES 2 271 946 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

cDNA para metilenotetrahidrofolato reductasa humana.

5 Antecedentes de la invención

(a) Ámbito de la invención

La invención se refiere a una sonda de cDNA para metilenotetrahidrofolato reductasa humana (MTHFR) y a sus
10 usos.

(b) Descripción del estado de la técnica

Los derivados del ácido fólico son coenzimas para varias y decisivas reacciones de transferencia de un solo carbono
15 entre las que se incluyen reacciones que intervienen en la biosíntesis de purinas, timidilato y metionina. La metilene-
trahidrofolato reductasa (MTHFR; EC 1.5.1.20) cataliza la reducción por unión a NADPH (NADPH = Nicotinamida-
Adenina Dinucleótido Fosfato) de 5,10-metilenotetrahidrofolato a 5-metiltetrahidrofolato, que es un cosustrato para la
metilación de homocisteína a metionina. La enzima de hígado porcino, una flavoproteína, ha sido purificada hasta la
homogeneidad, y es un homodímero de subunidades de 77 kDa. La proteólisis parcial del péptido porcino ha revelado
20 dos dominios espacialmente distintos: un dominio N-terminal de 40 kDa y un dominio C-terminal de 37 kDa. Este
último dominio contiene el sitio de fijación para el regulador alostérico S-adenosilmetionina.

La deficiencia hereditaria de la MTHFR, que es un trastorno recesivo autosómico, es el error ingénito más común
del metabolismo del ácido fólico. Un bloqueo en la producción de metiltetrahidrofolato conduce a elevados niveles
25 de homocisteína con niveles de metionina que van desde los bajos hasta los normales. Los pacientes con severas
deficiencias de la MTHFR (con una actividad de un 0-20% en los fibroblastos) pueden tener fenotipos variables. Se ha
informado en este grupo sobre retraso del desarrollo, retardo mental, anomalías motoras y de la marcha, neuropatía
periférica, ataques y trastornos psiquiátricos, si bien era asintomático al menos un paciente con severa deficiencia de
la MTHFR. Los cambios patológicos en la forma severa incluyen los cambios vasculares que se han encontrado en
30 otros estados con homocisteína elevada, así como reducidos niveles de neurotransmisores y metionina en el sistema
nervioso central. Una más leve deficiencia de la MTHFR (con una actividad de un 35-50%) ha sido descrita en
pacientes con enfermedad arterial coronaria (véase lo expuesto más adelante). Es probable la heterogeneidad genética,
considerando las diversas características clínicas, los niveles variables de actividad enzimática y los diferentes perfiles
de terminación de la reductasa en las células de los pacientes.

La enfermedad arterial coronaria (CAD) es responsable de un 25% de las muertes de los canadienses. Los factores
de riesgo cardiovascular (sexo masculino, historial familiar, tabaquismo, hipertensión, dislipoproteinemia y diabetes)
constituyen aproximadamente de un 60 a un 70% de nuestra capacidad para discriminar a los pacientes de CAD de
los sujetos sanos. Se ha demostrado también que la elevada homocisteína en plasma constituye un factor de riesgo
40 independiente para la enfermedad cardiovascular.

La homocisteína es un aminoácido que contiene sulfhidrilo y es formado por la desmetilación de metionina. La
homocisteína es normalmente metabolizada para ser así convertida en cisteína (transsulfuración), o bien es remetilada
para ser así convertida en metionina. Los errores ingénitos del metabolismo (como en la severa deficiencia de la
MTHFR) que ocasionan extremas elevaciones de la homocisteína en plasma, con homocistinuria, van asociados a
45 enfermedad vascular prematura y a extensos fenómenos trombóticos arteriales y venosos. Las más leves elevaciones de
la homocisteína en plasma (como en la leve deficiencia de la MTHFR) han sido asociadas al desarrollo de enfermedad
vascular periférica, enfermedad cerebrovascular y CAD prematura.

La remetilación de la homocisteína para su conversión en metionina requiere el intermedio de ácido fólico 5-
metiltetrahidrofolato, que es producido a partir de folato de 5,10-metilenotetrahidrofolato por medio de la acción de
5,10-metilenotetrahidrofolato reductasa (MTHFR). La deficiencia de la MTHFR redundante en una incapacidad para
metabolizar la homocisteína para su conversión en metionina; siendo las consecuencias metabólicas de todo ello una
elevada homocisteína en plasma y una reducida metionina. Las severas deficiencias de la MTHFR (menos de un 20%
55 de la actividad de los controles) que han sido descritas anteriormente van asociadas a síntomas neurológicos de inicio
temprano (retardo mental, neuropatía periférica, ataques, etc.) y a cambios ateroscleróticos y tromboembolismo. Más
leves deficiencias de la MTHFR (un 35-50% de la actividad de los controles), con una forma termolábil de la enzima,
se ven en pacientes con enfermedad cardiovascular sin obvias anomalías neurológicas.

En un estudio efectuado con 212 pacientes con enfermedad arterial coronaria demostrada, la forma termolábil de
la MTHFR fue encontrada en un 17% del grupo con CAD y un 5% de los controles. En un subsiguiente informe sobre
339 sujetos que fueron sometidos a angiografía coronaria, se halló una correlación entre la MTHFR termolábil y el
grado de estenosis arterial coronaria. De nuevo, tradicionales factores de riesgo (edad, sexo, tabaquismo, hipertensión,
etc.) no estaban significativamente asociados a la MTHFR termolábil. Todos los estudios sobre la MTHFR fueron
65 llevados a cabo mediante análisis enzimáticos de la MTHFR en linfocitos, con mediciones de la actividad antes y
después del tratamiento térmico para determinar la termolabilidad de la enzima.

Puesto que el 5-metiltetrahidrofolato, que es el producto de la reacción de la MTHFR, es la forma primaria de folato circulatorio, una deficiencia de la MTHFR podría conducir a otros tipos de trastornos. Por ejemplo, la administración periconcepcional de folato a mujeres reduce la aparición y la recurrencia de defectos del tubo neural en su descendencia. Los defectos del tubo neural constituyen un grupo de malformaciones del desarrollo (meningomielocelo, anencefalia, encefalocele) que surgen debido a la incapacidad de cierre del tubo neural. Se han descrito elevados niveles de homocisteína en plasma en madres de niños con defectos del tubo neural. La elevada homocisteína en plasma podría ser debida a una deficiencia de la MTHFR, como se ha descrito anteriormente para la enfermedad cardiovascular.

Los neuroblastomas son tumores derivados de células de la cresta neural. Se ha descrito que muchos estos tumores presentan delecciones de la región lp36 del cromosoma humano, que es la región del genoma en la cual se ha mapeado la MTHFR. Es posible que las delecciones/mutaciones de la MTHFR sean responsables de la formación de este tipo de tumor o contribuyan a la misma. Las anomalías de la MTHFR pueden también contribuir a la formación de otros tipos de tumores tales como los tumores colorrectales, puesto que se ha demostrado que los altos niveles de folato dietario están inversamente asociados al riesgo de contraer adenomas colorrectales.

La actividad de MTHFR es necesaria para la metilación de la homocisteína para su conversión en metionina. La metionina es necesaria para la formación de S-adenosilmetionina, que es dador primario de metilo para la metilación del DNA, de las proteínas, de los lípidos, de los neurotransmisores, etc. Las anomalías en la MTHFR podrían conducir a unos más bajos niveles de metionina de S-adenosilmetionina, así como a elevados niveles de homocisteína. El desbaratamiento de los procesos de metilación podría redundar en una amplia variedad de estados tales como neoplasias, anomalías del desarrollo, trastornos neurológicos, etc.

La publicación *Biochemical Medicine and Metabolic Biology* 43/3, 1990, páginas 234 a 242, pretende aislar y purificar la proteína MTHFR. Sin embargo, esto ha sido posteriormente desacreditado, y se piensa que de hecho se trata de una ACADSB (acil-CoA deshidrogenasa) de cadena corta o ramificada.

Nat.Gen. 7/2, junio 1994, pp. 195-200, describe MTHFR, pero no MDHFR que tiene una mutación en el bp 167 (bp = par de bases), sino tan sólo en el bp 482 y en el bp 559.

A pesar de que el gen de MTHFR en *Escherichia coli* (*metF*) ha sido aislado y secuenciado, los estudios moleculares de la enzima en organismos superiores han venido estando limitados en ausencia de la disponibilidad de un cDNA eucariótico. En esta comunicación informamos sobre el aislamiento de un cDNA humano para MTHFR, que es su asignación cromosómica, y sobre la identificación de mutaciones en pacientes con deficiencia de la MTHFR. Este informe representa la primera descripción molecular de mutaciones en la deficiencia de la MTHFR.

Sería muy deseable contar con una sonda de cDNA para metilenotetrahidrofolato reductasa (MTHFR). Esta sonda podría ser usada para la identificación de anomalías secuenciales en individuos con severa o leve deficiencia de la MTHFR, incluyendo los pacientes cardiovasculares y los pacientes con síntomas o tumores neurológicos. La sonda sería también usada en terapia genética, aislamiento del gen y estudios de expresión para producir la proteína MTHFR. La sonda también aportaría la secuencia de aminoácidos de la proteína MTHFR humana, lo cual resultaría útil para la terapia de la deficiencia de la MTHFR mediante enfoques bioquímicos o farmacológicos.

Sería muy deseable contar con una descripción molecular de las mutaciones en la deficiencia de la metilenotetrahidrofolato reductasa.

Breve exposición de la invención

Un objetivo de la presente invención es el de aportar una sonda de cDNA para metilenotetrahidrofolato reductasa (MTHFR) humana.

Otro objetivo de la presente invención es el de aportar una descripción molecular de las mutaciones en la deficiencia de la metilenotetrahidrofolato reductasa.

Otro objetivo de la presente invención es el de aportar una secuencia de ácido nucleico y de aminoácidos para metilenotetrahidrofolato reductasa humana.

Otro objetivo de la presente invención es el de aportar una potencial terapia para los individuos con deficiencia de la metilenotetrahidrofolato reductasa.

Otro objetivo de la presente invención es el de aportar un sistema para la síntesis de la proteína MTHFR *in vitro*.

Un objetivo adicional de la presente invención es el de aportar una tecnología/un protocolo para la identificación de cambios secuenciales en el gen de MPHFR.

Así, en un primer aspecto, la presente invención aporta una sonda de cDNA para una metilenotetrahidrofolato reductasa (MTHFR) humana que tiene una secuencia de nucleótidos de la ID SEC N°: 1 o una secuencia de nucleótidos que codifica una frecuencia de aminoácidos de la ID SEC N°: 2, comprendiendo dicha sonda una secuencia que se hibrida en condiciones de rigor con dichas secuencias de MTHFR.

En un segundo aspecto, la invención también aporta un método de diagnóstico de una deficiencia de la MTHFR en un paciente, comprendiendo dicho método los pasos de:

5 a) amplificar una muestra de DNA de un paciente o someter a transcripción inversa a una muestra de DNA de un paciente para así convertirla en DNA y amplificar dicho DNA; y

10 b) analizar el DNA amplificado del paso a) para determinar si el mismo contiene al menos una anomalía de secuencia con respecto a la metileno-tetra-hidrofolato reductasa (MTHFR) humana que tiene una secuencia de nucleótidos de la ID SEC N°: 1 o una secuencia que codifique la secuencia de aminoácidos de la ID SEC N°: 2, siendo dichas anomalías indicativas de deficiencia genómica de la MTHFR.

La invención también aporta el uso de un vector recombinante que comprende la ID SEC N°: 1 para la fabricación de un medicamento para la expresión de la proteína MTHFR (ID SEC N°: 2) bajo el control de un promotor adecuado.

15 En un aspecto adicional, la invención aporta una proteína MTHFR humana que es codificada por la secuencia de nucleótidos de la ID SEC N°: 1 y tiene la secuencia de aminoácidos de la ID SEC N°: 2.

20 La invención aporta adicionalmente un ácido nucleico de metileno-tetra-hidrofolato reductasa (MTHFR) humana que tiene una secuencia de nucleótidos de la ID SEC N°: 1 y codifica la secuencia de aminoácidos de la ID SEC N°: 2.

La invención también aporta el uso de la proteína de la reivindicación 6 o del ácido nucleico de la reivindicación 7 para la fabricación de un medicamento para tratar pacientes con deficiencia de la MTHFR.

Breve descripción de los dibujos

25 Las Figs. 1A a 1B ilustran la primera secuencia codificante de cDNA para metileno-tetra-hidrofolato reductasa (MTHFR);

30 la Fig. 2 es la alineación de aminoácidos para metileno-tetra-hidrofolato reductasa (MTHFR) humana, de los genes *metF* de *E. coli* (ECOMETF) y *S. typhimurium* (STYMETF) y de un marco de lectura abierto no identificado en *Saccharomyces cerevisiae* que es transcrito divergentemente desde un gen de reparación por escisión (*ysRAD1*);

35 Las Figs. 3A y 3B ilustran los análisis con enzimas de secuenciación y de restricción para la sustitución de Arg por Ter;

las Figs. 4A y 4B ilustran los análisis con enzimas de secuenciación y de restricción para la sustitución de Arg por Gln;

40 las Figs. 5A y 5B ilustran los análisis con enzimas de cambio secuencial y de restricción para la sustitución de alanina por valina;

la Fig. 6 ilustra la secuencia disponible total del cDNA para MTHFR humana;

45 las Figs. 7A y 7B ilustran el análisis de la expresión del cDNA para MTHFR en *E. coli*, siendo respectivamente (7A) el análisis Western blot de tejidos y extractos bacterianos, y (7B) el análisis de la termolabilidad de extractos bacterianos;

las Figs. 8A a 8D ilustran la identificación de una mutación de sitio de corte y empalme 5' que conduce a una delección en marco de 57 bp del cDNA;

50 las Figs. 9A a 9D ilustran el análisis diagnóstico con endonucleasa de restricción de 4 mutaciones;

las Figs. 10A a 10D ilustran los análisis de hibridación con oligonucleótidos aleloespecíficos de 2 mutaciones; y

55 la Fig. 11 ilustra la región de homología entre la metileno-tetra-hidrofolato reductasa (MTHFR) humana y la dihidrofolato reductasa (DHFR) humana.

Descripción detallada de la invención

60 Secuenciación de péptidos de MTHFR porcina

MTHFR porcina nativa homogénea fue sometida a digestión con tripsina para generar un fragmento N-terminal de 40 kDa y un fragmento C-terminal de 31 kDa; siendo el fragmento de 31 kDa un producto proteolítico del fragmento de 37 kDa. Los fragmentos fueron separados mediante SDS-PAGE (SDS-PAGE = electroforesis en gel de poliacrilamida en presencia de dodecilsulfato sódico) y electroeluidos, y los fragmentos desnaturalizados fueron sometidos a digestión con lisil endopeptidasa (LysC). Los péptidos resultantes fueron separados mediante HPLC (HPLC = cromatografía de líquidos de alta resolución) de fase inversa y fueron sometidos a análisis secuencial mediante degradación de Edman (los detalles están contenidos en Goyette P *et al.*, *Nature Genetics*, 1994, 7:195-200).

Aislamiento y secuenciación de cDNAs

Fueron sintetizados dos oligonucleótidos degenerados sobre la base de la secuencia de un péptido de MTHFR porcina de 30 aminoácidos (el primer péptido subrayado en la Fig. 2). Estos fueron usados para generar un producto de PCR (PCR = reacción en cadena de la polimerasa) de 90 bp que codifica el péptido predicho, mediante reacciones de transcripción inversa-PCR de 500 ng de RNA poliA+ de hígado de cerdo. Entonces fue sintetizado a partir de esta corta secuencia de cDNA un cebador de la PCR porcino-específico (no degenerado, antisentido). Usando este cebador y un cebador para brazos de fago, fue sometida a selección por PCR una biblioteca de cDNA λ gt10 de hígado humano (Clontech); suponiendo esta técnica la generación de cantidades de lisado de fago (50.000 pfu) (pfu = unidades formadoras de placas) que fueron hervidas por espacio de 5 minutos y fueron luego usadas directamente en reacciones PCR con estos dos cebadores. Los fragmentos de PCR fueron entonces secuenciados directamente (conjunto de materiales y utensilios Cycle Sequencing^{MF}, GIBCO) (MF = Marca de Fábrica), y fue identificado un clon positivo por comparación de la secuencia de aminoácidos deducida con la secuencia del péptido porcino (salvando las variaciones interespecies). El material positivo fue entonces cultivado de nuevo en placas de cultivo a más baja densidad y sometido a selección con el producto de PCR positivo radiomarcado mediante hibridación en placas hasta que fue identificada una placa perfectamente aislada. El DNA fago fue purificado y el inserto fue entonces subclonado en pBS+ (Bluescript) y secuenciado en ambas hebras (conjunto de materiales y utensilios Cycle Sequencing^{MF}, GIBCO y Sequenase^{MF}, Pharmacia). La secuencia de aminoácidos deducida del cDNA humano fue alineada con las secuencias de péptidos porcinos, los genes *metF* de *E. coli* (*ecometf*, número de acceso VO1502) y *S. typhimurium* (*stymetf*, número de acceso XO7689) y con un marco de lectura abierto previamente no identificado en *Saccharomyces cerevisiae* que es transcrito divergentemente con respecto al gen de reparación por escisión, *ysRAD1* (número de acceso KO2070). Las alineaciones iniciales fueron llevadas a cabo usando BestFit^{MF} en el paquete informático GCG, y estas alineaciones fueron ajustadas manualmente para maximizar las homologías.

En resumen, fueron diseñados cebadores oligonucleotídicos degenerados para amplificar una secuencia que corresponde a un segmento de 30 aminoácidos de un péptido porcino de la región N-terminal de la enzima (el primer péptido porcino en la Fig. 2). Fue obtenido un fragmento de cDNA porcino de 90 bp mediante transcripción inversa/PCR de RNA de hígado de cerdo. La secuenciación del fragmento de PCR confirmó su identidad por comparación de la secuencia de aminoácidos deducida con la secuencia de péptidos porcinos. Un cebador oligonucleotídico no degenerado, basado en la secuencia interna del cDNA porcino, fue usado en conjunción con cebadores para los brazos del fago para someter a selección a una biblioteca de cDNA λ gt10 de hígado humano por PCR. El inserto del clon positivo fue aislado y secuenciado. La secuencia constaba de 1266 bp con un marco de lectura abierto continuo.

Homología con la MTHFR de otras especies

La secuencia de aminoácidos deducida del cDNA humano fue alineada con los genes *metF* de *E. coli* y *S. typhimurium*, así como con un ORF (ORF = marco de lectura abierto) previamente no identificado en *Saccharomyces cerevisiae* que es transcrito divergentemente con respecto al gen de reparación por escisión, *ysRAD1* (Fig. 2). Las secuencias que son homólogas a 5 péptidos porcinos están subrayadas en la Fig. 2. Tres segmentos (los residuos 61-94, 219-240 y 337-351) corresponden a la secuencia de péptidos internos del dominio N-terminal de 40 kDa de la enzima de hígado porcino. Los residuos 374-393 corresponden a la parte del lado de cabeza del péptido LysC del dominio C-terminal de la enzima de hígado porcino que es marcada cuando la enzima es irradiada con luz ultravioleta en presencia de (³H-metil)AdoMet; y como se predice mediante los estudios de marcación AdoMet, este péptido está situado en un extremo (N-terminal) del dominio de 37 kDa. Fue también identificada una quinta región de homología (residuos 359-372), pero la localización del péptido porcino dentro de la proteína nativa no había sido determinada anteriormente.

La metilenotetrahidrofolato reductasa (MTHFR) es una enzima que interviene en el metabolismo de los aminoácidos, que es decisivo para mantener unas adecuadas reservas de metionina, así como para asegurar que la concentración de homocisteína no alcance niveles tóxicos. El alto grado de conservación de la secuencia, desde la *E. coli* hasta el *Homo sapiens*, atestigua la importancia de la MTHFR en estas especies. La enzima en *E. coli* (codificada por el locus *metF*) es un péptido de 33 kDa que fija el FAD reducido (FAD = dinucleótido de flavina y adenina) y cataliza la reducción de metilenotetrahidrofolato para su conversión en metiltetrahidrofolato. La enzima *metF* se diferencia de la enzima de mamífero en que no puede ser reducida por NADPH o NADH, y su actividad no es regulada alostéricamente por la S-adenosilmetionina. La enzima porcina nativa es susceptible de experimentar corte triptico entre el dominio N-terminal de 40 kDa y el dominio C-terminal de 37 kDa, y este corte redundante en la pérdida de regulación alostérica por la adenosilmetionina, pero no redundante en pérdida de actividad catalítica. Puesto que la homología entre las enzimas bacteriana y de mamífero está dentro del dominio N-terminal, esta región debe contener el sitio de fijación de flavina y los residuos que son necesarios para fijar el sustrato de folato y catalizar su reducción. La estructura del dominio de la enzima humana no ha sido dilucidada, si bien se ha informado de que la enzima humana tiene una masa molecular de 150 kDa y es probable que sea un homodímero de 77 kDa.

Nosotros predecimos que el punto de corte entre los dos dominios está situado entre los residuos 351 y 374 de la secuencia humana, sobre la base de la localización de los péptidos obtenidos de los dominios aislados de la enzima porcina. Se prevé que esta región, que contiene la altamente cargada secuencia KRREED, tendrá la más alta hidrofiliidad y probabilidad superficial de entre todas las regiones de la secuencia humana deducida.

Ha sido secuenciado el extremo N-terminal de la proteína porcina, y la región que codifica esta parte de la proteína falta en el cDNA humano. Estimamos que a este cDNA le faltan solamente unos pocos residuos en el extremo N-terminal, puesto que la masa molecular predicha de la secuencia deducida en el lado de cabeza con respecto al putativo sitio de corte (KRREED) es de 40 kDa, y la masa molecular medida del dominio N-terminal porcino es también de 40 kDa. Cuando son alineadas las secuencias bacteriana, de levadura y humana, la secuencia humana deducida contiene una extensión N-terminal de 40 aminoácidos; y sospechamos que esta extensión contiene determinantes para la fijación de NADPH. Muchas oxidoreductasas dependientes de nucleótidos piridínicos contienen tales determinantes en el extremo N-terminal de la proteína.

El extremo C-terminal de la secuencia humana contiene un péptido que es marcado cuando la proteína es irradiada con luz ultravioleta en presencia de AdoMet tritiado. La secuencia de cDNA sobre la que aquí informamos contiene tan sólo aproximadamente 7 kDa de la masa predicha de 37 kDa de este dominio, lo que indica que este cDNA está asimismo truncado en el extremo terminal 3'. Tampoco han sido detectados los de una serie de péptidos del dominio porcino C-terminal. Como podría ser de esperar, dado que las enzimas procarióticas no parecen ser reguladas alostéricamente por AdoMet, no hay homologías significativas entre la región C-terminal de este cDNA y las secuencias *metF* procarióticas. La alineación que se muestra en la Fig. 2 pone de manifiesto que las secuencias homólogas terminan justo antes del putativo sitio de corte de la enzima humana.

Asignación cromosómica

Para la localización del gen humano fue usada hibridación *in situ* con cromosomas humanos metafásicos. El análisis de la distribución de 200 granos de plata relevó una significativa agrupación de granos, y concretamente de 40 granos, en la región p36.3-36.2 del cromosoma 1 ($p < 0,0001$), siendo la mayoría de granos, y concretamente 25 granos, observados dentro de lp36.3.

El aislamiento del cDNA humano nos ha permitido situar el gen en el cromosoma lp36.3. La observación de una fuerte señal en ese cromosoma con poco fondo sugiere en gran medida un único locus sin pseudogenes. El análisis Southern blot de DNA humano reveló fragmentos de aproximadamente 10 kb, prediciendo un gen de tamaño medio, puesto que este cDNA codifica aproximadamente la mitad de la secuencia codificante.

Adicionales secuencias de cDNA y constructos para el análisis de la expresión

Una biblioteca de cDNA de carcinoma de colon humano (obsequio de la Dra. Nicole Beauchemin, de la McGill University) fue sometida a selección mediante hibridación en placas con el cDNA de 1,3 kb original para obtener adicionales secuencias codificantes. Fue aislado un cDNA de 2,2 kb que contenía 1,3 kb de secuencia solapada con el cDNA original y 900 bp adicionales en el extremo 3' (Fig. 6). La secuencia de aminoácidos es idéntica a la del cDNA original para el segmento solapado (codones 1-415) exceptuando el codón 177 (ASP), que era un codón GLY en el cDNA original. El análisis de 50 cromosomas de control reveló un codón ASP en esta posición. El cDNA tiene un marco de lectura abierto de 1980 bp, a 100 bp de la región UTR 3' y una cola poli A.

Se llevó a cabo secuenciación en ambas hebras para todo el cDNA. Adicionales secuencias del lado 5' (de 800 bp) fueron obtenidas de una biblioteca de cDNA de riñón humano (Clontech), pero estas secuencias no contenían adicionales secuencias codificantes y fueron por consiguiente usadas solamente para la mutagénesis por PCR (como se describe más adelante) y no para el análisis de la expresión. Los dos cDNAs (de 2,2 kb y 800 bp) fueron ligados usando el sitio *EcoRI* en el bp 199, y fueron insertados en el vector Bluescript^{MF} (Stratagene). El cDNA de 2,2 kb fue subclonado en el vector de expresión pTrc99A (Pharmacia) usando el sitio *NcoI* en el bp 11 y el sitio *XbaI* en la región poliligadora de los vectores tanto Bluescript^{MF} como pTrc99A. La secuenciación fue llevada a cabo de una a otra parte de los sitios de clonación para verificar el constructo de tipo salvaje.

Utilidad de la invención en la identificación de mutaciones

I. Identificación de dos primeras mutaciones en la deficiencia severa de la MTHFR

RNA total de fibroblastos de la piel de pacientes con deficiencia de la MTHFR fue sometido a transcripción inversa y amplificado mediante PCR para sus análisis por el método del polimorfismo conformacional de cadena sencilla (SSCP) (Orita, M. *et al.*, *Genomics*, 1989, 5:8874-8879). Fueron diseñados cebadores para generar fragmentos de 250-300 bp y cubrir las secuencias de cDNA disponibles con pequeñas regiones de solapamiento para cada fragmento en ambos extremos. La primera mutación identificada mediante SSCP era una sustitución de C por T en el bp 559 en el paciente 1554; convirtiendo esta sustitución un codón de arginina en un codón de terminación (Fig. 3A). Puesto que la mutación abolió un sitio *FokI*, fue usada digestión de restricción para la confirmación del cambio y para someter a selección a adicionales pacientes con respecto a esta mutación; habiendo sido identificado de esta manera un segundo paciente (1627) (Fig. 3B). La configuración por SSCP para el paciente 1554 y la configuración por digestión de restricción para ambos pacientes era coherente con un estado mutante homocigótico o con un combinado genético que constaba de la mutación sin sentido con una segunda mutación que no producía RNA detectable alguno (alelo nulo). Para la confirmación son necesarios estudios con los progenitores.

La segunda sustitución (Fig. 4A) era una transición de G a A en el bp 482 en el paciente 1834 que convirtió un residuo de arginina en un residuo de glutamina. La sustitución creó un sitio *PstI* que fue usado para verificar

la sustitución y para identificar a un segundo paciente (1863) con este cambio (Fig. 4B). El análisis del SSCP y la configuración por digestión de restricción eran coherentes con un estado heterocigótico para ambos pacientes. El codón de arginina que se ve afectado por este cambio es un residuo conservado evolutivamente, como se muestra en la Fig. 2. Esta observación, en conjunción con el hecho de que el cambio de codón no es conservativo, constituye un fuerte argumento en el sentido de indicar que la sustitución es un cambio patológico en lugar de un polimorfismo benigno. Además, fueron analizados 35 controles (de orígenes étnicos similares a los de los probandos) para detectar esta sustitución mediante análisis Southern blot de DNA digerido con *Pst*I; y resultaron todos ellos negativos.

Fue estudiada la familia del paciente 1834. Se comprobó que tanto el hermano sintomático como la madre del probando llevaban esta sustitución, mientras que el padre fue negativo con respecto al cambio (Fig. 4B). En la familia del 1863, se comprobó que la madre del probando era portadora, mientras que el padre y un hermano no afectado eran negativos.

Líneas celulares

La línea celular 1554 es de un varón del pueblo Hopi que ingresó a los tres meses de edad con homocistinuria, ataques, deshidratación, enturbiamiento corneal, hipotonía y sepsis por *Candida*. La distribución de folato en los fibroblastos cultivados puso de manifiesto una relación de *Pediococcus cerevisiae/Lactobacillus casei* (PC/LC) de 0,52 (Control 0,14). No había actividad de metilenoetetrahidrofolato reductasa (MTHFR) mensurable (valores de control = 9,7 y 15,1 nmoles/h/mg de proteína; actividad residual tras tratamiento de los extractos de control a 55°C por espacio de 20 min. = 28% y 31%).

La línea celular 1627 es de un varón de la tribu Choctaw que presentaba malnutrición, apnea, incapacidad para desarrollarse, deshidratación y homocistinuria a las cinco semanas de edad. Posteriormente se comprobó que este individuo tenía trombosis del seno sagital superior e hidrocefalia. La relación PC/LC era de 0,61 y la actividad específica de MTHFR era de 0,1 nmoles/h/mg de proteína. Hay consanguinidad por cuanto que se piensa que las abuelas materna y paterna son "parientes lejanos".

La línea celular 1779 es de un varón francocanadiense con homocistinuria que primeramente tenía debilidad de los miembros, incoordinación, parestesias y lapsos de memoria a la edad de 15 años, y quedó confinado a la silla de ruedas poco después de cumplir los veinte. Su hermano (línea celular 1834) también tiene homocistinuria, pero tiene 37 años de edad y es asintomático. La actividad específica de MTHFR era de 0,7 y 0,9 nmoles/h/mg de proteína para las líneas celulares 1779 y 1834, respectivamente; y la actividad residual tras tratamiento térmico a 55°C era de un 0,9% y de un 0% para las líneas celulares 1779 y 1834, respectivamente.

La línea celular 1863 es de un varón blanco que fue diagnosticado a los 21 años de edad debido a un progresivo deterioro de la marcha, espasticidad, degeneración de la materia blanca cerebral y homocistinuria. Este sujeto tenía un hermano que murió a los 21 años de edad de enfermedad neurodegenerativa. La actividad específica de MTHFR en extractos de fibroblastos era de 1,76 nmoles/h/mg de proteína, y la actividad enzimática residual tras tratamiento a 55°C era de un 3,6%.

Análisis de mutación

Fueron diseñados cebadores a partir de la secuencia de cDNA para generar fragmentos de 250-300 bp que estaban solapados en 50-75 bp en cada extremo. La parejas de cebadores fueron usadas en transcripción inversa-PCR de 5 µg de RNA fibroblástico total del paciente. Los productos de PCR fueron analizados mediante un protocolo de SSCP rápido no isotópico (PhartSystem^{MF}, Pharmacia) que usa tinción directa con plata para la detección de hebras únicas. Todos los productos de PCR de los pacientes que presentaban un desplazamiento sobre geles de SSCP fueron purificados mediante NuSieve (FMC Bioproducts) y secuenciados directamente (conjunto de materiales y utensilios Cycle Sequencing^{MF}, GIBCO) para identificar el cambio. Si el cambio afectaba a un sitio de restricción, un producto de PCR era entonces sometido a digestión con la apropiada endonucleasa de restricción y analizado sobre geles de poliacrilamida. Para efectuar la selección para detectar la mutación de Arg a Gln en los controles, 5 µg de DNA digerido con *Pst*I fueron pasados por geles de agarosa al 0,8% y analizados mediante análisis Southern blot usando el cDNA radiomarcado mediante técnicas estándar.

II. Siete mutaciones adicionales en el locus de la metilenoetetrahidrofolato reductasa (MTHFR) con correlaciones de genotipo:fenotipo en la deficiencia severa de la MTHFR

Se informa a continuación sobre la caracterización de 7 adicionales mutaciones en este locus: 6 mutaciones de sentido equivocado y un defecto de sitio de corte y empalme en el lado 5' que activa un sitio de corte y empalme críptico en la secuencia codificante. También presentamos un análisis preliminar de la relación entre genotipo y fenotipo para todas las 9 mutaciones hasta aquí identificadas en este locus. Una mutación sin sentido y 2 mutaciones de sentido equivocado (de prolina a leucina y de treonina a metionina) en el estado homocigótico van asociadas a una extremadamente baja actividad (0-3%) y al inicio de síntomas dentro del primer año. Otras mutaciones de sentido equivocado (de arginina a cisteína y de arginina a glutamina) van asociadas a una más alta actividad enzimática y a un más tardío inicio de los síntomas.

Se describen 7 mutaciones adicionales en el locus de MTHFR y se examina la asociación entre genotipo, actividad enzimática y fenotipo clínico en la severa deficiencia de la MTHFR.

Descripción de los pacientes

Se han descrito en la literatura publicada los hallazgos clínicos y de laboratorio relativos a los pacientes. La actividad residual de MTHFR fue medida previamente en fibroblastos cultivados en confluencia.

A la paciente 354, que era una muchacha afroamericana, le fue diagnosticado a los 13 años de edad leve retardo mental. A su hermana, la paciente 355, le fueron diagnosticados a los 15 años de edad anorexia, temblor, alucinaciones y abandono progresivo. En la paciente 354, la actividad residual de MTHFR era de un 19%, y en su hermana, la paciente 355, dicha actividad residual era de un 14% de los valores de control. La actividad residual tras calentamiento presentaba una termoestabilidad equivalente a la de la enzima de control.

La paciente 1807, que es una muchacha japonesa cuyos progenitores son primos hermanos, presentaba retraso en andar y en el habla hasta los 2 años de edad, ataques a los 6 años de edad y un trastorno de la marcha con neuropatía periférica a los 16 años de edad. La actividad residual de MTHFR era de un 3%, y la enzima era termolábil.

A la paciente 735, que era una muchacha afroindia, le fueron diagnosticados a los 7 meses de edad microcefalia, deterioro progresivo del desarrollo mental, apnea y coma. La actividad residual de MTHFR era de un 2% de los niveles de control. No fueron determinadas las propiedades térmicas.

Al paciente 1084, que era un varón caucásico, le fue diagnosticado a los 3 meses de edad un fibrosarcoma infantil. Este sujeto resultó ser hipotónico y devino apnéico. El sujeto murió a los 4 meses de edad. No era detectable una actividad residual de MTHFR. No fueron determinadas las propiedades térmicas.

El paciente 356, que es el primer paciente con deficiencia declarada de la MTHFR, es un varón italoamericano que se presentó a los 16 años de edad con debilidad muscular, marcha anormal y movimientos espasmódicos de las extremidades superiores. La actividad residual de MTHFR era de un 20% de los valores de control; quedando la actividad rápida y exponencialmente inactivada a 55°.

Al paciente 458, que era un varón caucásico, le fueron diagnosticados a los 12 años de edad ataxia y rendimiento escolar marginal. La actividad residual de MTHFR era de aproximadamente un 10%, y la actividad era termolábil.

De la paciente 1396, que era una hembra caucásica, se decía que era torpe y presentaba un trastorno global del aprendizaje en la infancia. A la edad de 14 años, la paciente desarrolló ataxia, pie caído e incapacidad de andar. La paciente desarrolló trombosis de las venas profundas y émbolos pulmonares bilaterales. La actividad residual de MTHFR era de un 14%, y la enzima era termolábil.

Estructura genómica y cebadores intrónicos

La nomenclatura de los exones está basada en la secuencia de cDNA disponible en Goyette *et al.* (*Nature Genetics*, 1994, 7:195-200). El exón 1 ha sido designado arbitrariamente como la región de cDNA que va desde el bp 1 hasta el primer intrón. La identificación de los intrones fue llevada a cabo mediante amplificación de DNA genómico usando secuencias cebadoras de cDNA. Los productos de PCR que eran de tamaño mayor que los tamaños de cDNA previstos fueron secuenciados directamente.

Detección de las mutaciones

Los exones específicos (véase la Tabla 1 con respecto a las secuencias de los cebadores) fueron amplificados mediante PCR a partir de DNA genómico y analizados mediante el protocolo del SSCP. El análisis del SSCP fue llevado a cabo con el sistema Phastgel^{MF} (Pharmacia), un protocolo de SSCP rápido no isotópico como el anteriormente descrito (Goyette P *et al.*, *Nature Genetics*, 1994, 7:195-200), o con productos de PCR marcados con ³⁵S pasados por acrilamida al 6%: geles de glicerol al 10% a temperatura ambiente (6 vatios, durante la noche). En algunos casos, el uso de endonucleasas de restricción para dividir el producto de PCR antes del análisis del SSCP facilitaba la detección de desplazamientos de banda. Los fragmentos de PCR con movilidad alterada fueron secuenciados directamente (conjunto de materiales y utensilios Cycle Sequencing^{MF}, GIBCO). Si el cambio de secuencia afectaba a un sitio para endonucleasa de restricción, el producto de PCR era entonces sometido a digestión con la enzima apropiada y analizado mediante PAGE (PAGE = electroforesis en gel de poliacrilamida). De otro modo, se llevaba a cabo hibridación con oligonucleótidos aleloespecíficos (ASO) sobre una transferencia de puntos del exón amplificado por PCR.

TABLA 1

Cebadores de la PCR para amplificación de DNA y análisis de mutación de MTHFR

Exón	Tipo de Cebador	Secuencia del Cebador (5'→3')	Situación	Tamaño del Fragmento (bp)
1	Con sentido	AGCCTCAACCCCTGCTTGGAGG	C	271
	Antisentido	TGACAGTTTGCTCCCCAGGCAC	I	
4	Con sentido	TGAAGGAGAAGGTGTCTGCCGGA	C	198
	Antisentido	AGGACGGTGCGGTGAGAGTGG	I	
5	Con sentido	CACCTGTGTTGGCATGGATGATG	I	392
	Antisentido	GGCTGCTCTTGGACCCTCCTC	I	
6	Con sentido	TGCTTCCGGCTCCCTCTAGCC	I	251
	Antisentido	CCGCCGCTCCCAAGAACAAG	I	

TABLA 2

Resumen de genotipos, actividad enzimática, edad al inicio y origen de los pacientes con deficiencia de la MTHFR

Paciente ^a	Cambios de los bp ^b	Cambios de los aminoácidos	% de actividad	Edad al inicio	Origen
1807	C764T/C764T	Pro→Leu/Pro→Leu	3	dentro del 1 ^{er} año	Japonés
735	C692T/C692T	Thr→Met/Thr→Met	2	7 meses	Afroindio
1084	C692T/C692T	Thr→Met/Thr→Met	0	3 meses	Caucásico
1554	C559T/C559T	Arg→Ter/Arg→Ter	0	1 mes	Nativo americano (Hopi)
1627	C559T/C559T	Arg→Ter/Arg→Ter	1	1 mes	Nativo americano (Choctaw)
356	C985T/C985T	Arg→Cys/Arg→Cys	20	3 años	Italoamericano
458	C1015T/G167A	Arg→Cys/Arg→Gln	10	11 años	Caucásico
1396	C1081T/G167A	Arg→Cys/Arg→Gln	14	14 años	Caucásico
1779 ^c	G482A/?	Arg→Gln/?	6	15 años	Francocanadiense
1834 ^c	G482A/?	Arg→Gln/?	7	Asintomático a los 37 años	Francocanadiense
1863	G482A/?	Arg→Gln/?	14	21 años	Caucásico
354 ^d	792 + 1G→A/?	5' splice site?	19	13 años	Afroamericano
355 ^d	792 + 1G→A/?	5' splice site?	14	11 años	Afroamericano

^a Los pacientes 1554, 1627, 1779, 1834 y 1863 fueron descritos previamente por Goyette et al. (1994)

^b ? = mutación no identificada.

^c Los pacientes 1779 y 1834 son hermanos.

^d Los pacientes 354 y 355 son hermanos.

(1) *Mutación del sitio de corte y empalme 5'*

La amplificación de cDNA, bp 653-939, a partir de RNA fibroblástico total transcrito inversamente, reveló 2 bandas en las hermanas 354 y 355: un fragmento de PCR más pequeño (de 230 bp) además del alelo de 287 bp normal (Fig. 8A). La Fig. 8A es el análisis por electroforesis en gel de poliacrilamida de productos de amplificación de cDNA, bp 653-939, a partir de RNA transcrito inversamente. Los controles tienen el fragmento de 287 bp esperado, mientras que las pacientes 354 y 355 tienen un fragmento de 230 bp adicional. La secuenciación del fragmento menor identificó una delección en marco de 57 bp que retiraría 19 aminoácidos (Fig. 8B). La Fig. 8B es la secuenciación directa de los productos de PCR de la paciente 354. La delección de 57 bp abarca las bp 736-792 del cDNA. Se ve en el punto de arranque de la delección en el lado 5' un sitio de corte y empalme 5' casi perfecto (en caja). El análisis de la secuencia junto al punto de arranque de la delección en el lado 5' en el fragmento no suprimido reveló una casi perfecta secuencia de consenso del sitio de corte y empalme 5' (AG/gcatgc). Esta observación sugería la presencia de una mutación de corte y empalme en el sitio de corte y empalme 5' natural que podría activar este sitio críptico, para generar el alelo suprimido. La secuencia a continuación del punto de arranque de la delección, en el alelo mutante, correspondía exactamente a la secuencia del siguiente exón. La amplificación de DNA genómico, usando los mismos cebadores de amplificación que fueron usados para RNA transcrito inversamente, generó un producto de PCR de 1,2 kb que indicaba la presencia de un intrón. La secuenciación directa de este fragmento de PCR en la paciente 354 identificó una sustitución G → A heterocigótica en el dinucleótido GT conservado del intrón en el sitio de corte y empalme 5' (Fig. 8C). La Fig. 8C es la secuenciación del sitio de corte y empalme 5' en un control y en la paciente 354. La paciente porta una sustitución G → A heterocigótica en el sitio de corte y empalme 5' (en caja). Las secuencias intrónicas están en minúsculas. Esta sustitución abolió un sitio para la endonucleasa de restricción *HphI*, lo cual fue usado para confirmar la mutación en las 2 hermanas (Fig. 8D). La Fig. 8D es el análisis efectuado con endonucleasa de restricción *HphI* sobre productos de PCR de DNA para el exón 4 de las pacientes 354 y 355 y de 3 controles (C). El producto de PCR de 198 bp tiene 2 sitios *HphI*. Los productos de digestión para el alelo de control son de 151, 24 y 23 bp. Los productos de digestión para el alelo mutante son de 175 y 23 bp debido a la pérdida de un sitio *HphI*. Los fragmentos de 24 y 23 bp se han salido del gel.

(2) *Pacientes con sustituciones codificantes homocigóticas*

El análisis del SSCP del exón 4 para la paciente 1807 reveló un fragmento que migraba anormalmente, el cual fue secuenciado directamente para revelar una sustitución C → T homocigótica (bp 764) que convertía un residuo de prolina en un residuo de leucina. Este cambio crea un sitio para la endonucleasa de restricción *MnII* que fue usado para confirmar el estado homocigótico de la mutación (Fig. 9A). La Fig. 9A es el análisis de restricción con *MnII* de productos de PCR del exón 4 para la paciente 1807 y 3 controles (C). Fragmentos esperados: alelo de control, 90, 46, 44, 18 bp; alelo mutante, 73, 46, 44, 18, 17 bp. Una banda adicional en el fondo del gel es el cebador. Cincuenta cromosomas caucásicos de control independientes y 12 cromosomas japoneses de control fueron verificados mediante análisis de restricción, y fueron todos ellos negativos para esta mutación. La homocigosidad en este paciente es probablemente debida a la consanguinidad de los progenitores.

Las pacientes 735 y 1084 tenían la misma mutación en el exón 4, en un estado homocigótico: una sustitución C → T (bp 692) que convertía un residuo de treonina conservado evolutivamente en un residuo de metionina, y abolía un sitio para la endonucleasa de restricción *NlaIII*. Fue usada hibridación con oligonucleótidos aleloespecíficos con el exón 4 amplificado (Figs. 10A y 10B) para confirmar la mutación en estas 2 pacientes y para someter a selección a 60 cromosomas independientes que resultaron ser todos ellos negativos. La Fig. 10A es la hibridación del oligonucleótido mutante (692T) con los productos de PCR del exón 4 de las pacientes 735, 1084 y 30 controles. Solamente se hibridó con esta sonda el DNA de las pacientes 735 y 1084. La Fig. 10B es la hibridación del oligonucleótido normal (692C) con la transferencia de puntos separada de A. Todos los DNAs de control se hibridaron con esta sonda.

El paciente 356 presentó un desplazamiento en el análisis del SSCP del exón 5. La secuenciación directa reveló una sustitución C → T homocigótica (bp 985) que convertía un residuo de arginina conservado evolutivamente en cisteína, y la sustitución abolió un sitio para la endonucleasa de restricción *AccII*. Esto fue usado para confirmar el estado homocigótico de la mutación en el paciente 356 (Fig. 9B) y su presencia en el estado heterocigótico en ambos progenitores. Cincuenta cromosomas de control independientes, analizados de la misma manera, dieron negativo para esta mutación. La Fig. 9B es el análisis de restricción con *AccII* de productos de PCR del exón 5 para el paciente 356, su padre (F), su madre (M) y 3 controles (C). Fragmentos esperados: alelo de control, 129, 105, 90, 68 bp; alelo mutante, 195, 129, 68 bp.

(3) *Pacientes que son combinados genéticos*

El paciente 458 es un heterocigoto combinado de una mutación en el exón 5 y una mutación en el exón 1. La sustitución en el exón 5 (C → T en el bp 1015) redundó en la sustitución de un residuo de cisteína por un residuo de arginina, y esto abolió un sitio para la endonucleasa de restricción *HhaI*, lo cual fue usado para confirmar la mutación en el paciente 458 (Fig. 9C) y para demostrar que los 50 cromosomas de control eran negativos. La Fig. 9C es el análisis de restricción con *HhaI* de productos de PCR del exón 5 para el paciente 458 y 4 controles (C). Fragmentos esperados: alelo de control, 317 y 75 bp; alelo mutante 392 bp. El fragmento de 75 bp no aparece en la Fig. 9C. La segunda mutación era una sustitución G → A heterocigótica (bp 167) que convertía un residuo de arginina en un residuo de glutamina. La hibridación con oligonucleótidos aleloespecíficos con el exón 1 amplificado confirmó el estado heterocigótico de esta mutación en el paciente 458 e identificó a un segundo paciente (1396) que portaba

esta mutación también en estado heterocigótico (Figs. 10C y 10D). La Fig. 10C es la hibridación del oligonucleótido mutante (167A) con productos de PCR del exón 1 de los pacientes 458, 1396 y 31 controles. La Fig. 10D es la hibridación del oligonucleótido normal (167G) con la transferencia de puntos separada de C. Ninguno de los 62 cromosomas de control llevaba esta mutación.

La segunda mutación en la paciente 1396 fue identificada en el exón 6: una sustitución C → T heterocigótica (bp 1081) que convertía un residuo de arginina en un residuo de cisteína y abolía un sitio para la endonucleasa de restricción *HhaI*. El análisis de restricción confirmó la sustitución heterocigótica en la paciente 1396 (Fig. 9D) y puso de manifiesto que los 50 cromosomas de control eran negativos. La Fig. 9D es el análisis de restricción con *HhaI* de productos de PCR del exón 6 para la paciente 1396 y 2 controles (C). Fragmentos esperados: alelo de control, 152, 86, 13 bp; alelo mutante 165, 86 bp. El fragmento de 13 bp se ha salido del gel.

(4) Adicionales cambios de secuencia

El análisis del exón 6 con *HhaI*, que ha sido mencionado anteriormente, reveló un polimorfismo del DNA. Este cambio es una sustitución T → C en el bp 1068 que no altera al aminoácido (serina), pero crea un sitio de reconocimiento para *HhaI*. En un 9% de los cromosomas analizados se encontró T en el bp 1068. El análisis de la secuencia identificó 2 discrepancias con la secuencia de cDNA publicada: una sustitución G → A en el bp 542 que convierte la glicina en un codón de aspartato, y un cambio C → T en el bp 1032 que no altera al aminoácido (treonina). Puesto que todos los DNAs analizados (> 50 cromosomas) llevaban la A en el bp 542 y la T en el bp 1032, es probable que la secuencia del cDNA original contuviese algunos artefactos de clonación.

Correlación genotipo:fenotipo

La Tabla 2 resume el estado actual de las mutaciones en la severa deficiencia de la MTHFR. En 8 pacientes han sido identificadas ambas mutaciones, y en 5 pacientes (3 familias) ha sido identificada solamente 1 mutación. En conjunto, la correlación entre el genotipo, la actividad enzimática y el fenotipo es del todo coherente. Cinco pacientes con inicio de síntomas dentro de su primer año de vida tenían ≤ 3% de la actividad de los controles. Tres de estos pacientes tenían mutaciones de sentido equivocado en el estado homocigótico: dos pacientes con la sustitución de treonina por metionina (C692T) y un paciente con la sustitución de prolina por leucina (C764T). La mutación sin sentido (C559T) en el estado homocigótico en los pacientes 1554 y 1627 (anteriormente descritos en Goyette P *et al.*, *Nature Genetics*, 1994, 7:195-200) va también asociada a una severa forma neonatal, como era de esperar.

Los otros pacientes en la Tabla 2 tenían ≥ 6% de la actividad de los controles e inicio de síntomas dentro o después de la 2ª década de vida; siendo la única excepción el paciente 1834, como se ha descrito anteriormente (Goyette P *et al.*, *Nature Genetics*, 1994, 7:195-200). Los tres pacientes (356, 458 y 1396) con mutaciones de sentido equivocado (G167A, C985T, C1015T y C1081T) son similares a los anteriormente descritos (pacientes 1779, 1834 y 1863) que tenían una sustitución de arginina por glutamina y una segunda mutación no identificada (Goyette P *et al.*, *Nature Genetics*, 1994, 7:195-200). Las hermanas con la mutación de corte y empalme en el lado 5' y una segunda mutación no identificada también tenían niveles de actividad situados dentro de la misma gama de valores e inicio de síntomas en la segunda década, pero la actividad es probablemente debida al segundo alelo no identificado.

Discusión

Los pacientes son de distintos orígenes étnicos. A pesar de que los pacientes 1554 y 1627 son ambos nativos americanos, las mutaciones se dan en distintos haplotipos, sugiriendo una mutación recurrente en lugar de identidad por descendencia. Puesto que la sustitución se da en un dinucleótido CpG, que es un "punto caliente" para la mutación, la mutación recurrente constituye una hipótesis razonable. Es difícil valorar si algunas mutaciones son específicas de la población puesto que las cifras son demasiado pequeñas en la actualidad.

La deficiencia de la MTHFR es el error ingénito más común del metabolismo del folato y constituye una causa importante de la homocisteinemia hereditaria. El reciente aislamiento de un cDNA para MTHFR ha permitido el análisis mutacional en este locus, con los objetivos de definir dominios importantes para la enzima y de correlacionar el genotipo con el fenotipo en los pacientes con deficiencia de la MTHFR.

Nuestra definición de una sustitución que causa enfermedad, a diferencia de un polimorfismo benigno, está basada en 3 factores: (1) la ausencia del cambio en al menos 50 cromosomas de control independientes; (2) la presencia del aminoácido en la enzima bacteriana, que atestigua su importancia evolutiva; y (3) la cuestión de si el cambio de aminoácido es conservativo. A pesar de que la expresión de las sustituciones es necesaria para demostrar formalmente que las mismas no son benignas, los anteriores criterios nos permiten postular que los cambios que se describen en este informe es probable que afecten a la actividad.

Las 7 mutaciones aquí descritas (6 sustituciones de un solo aminoácido y una mutación de sitio de corte y empalme 5') hacen que el total sea de 9 mutaciones identificadas hasta la fecha en la severa deficiencia de la MTHFR y completan el análisis de mutaciones para 8 pacientes. La identificación de cada mutación en solamente una o dos familias apunta al sorprendente grado de heterogeneidad genética en este locus. Siete de las 9 mutaciones están situadas en dinucleótidos CpG, que son propensos a experimentar eventos mutacionales.

Mutación de sitio de corte y empalme 5'

La sustitución G → A en el dinucleótido GT del sitio de corte y empalme 5' en las pacientes 354 y 355 redunda en una delección en marco de 57 bp de la secuencia codificante, la cual debería suprimir 19 aminoácidos de la proteína. Esta delección se da como resultado de la activación de un sitio de corte y empalme 5' críptico (AG/gc) incluso a pesar de que este sitio críptico no tiene una perfecta secuencia de consenso de sitio de corte y empalme 5' (AG/gt). Sin embargo, los GC (en lugar de los GT) como los 2 primeros nucleótidos de un intrón han sido descritos en varios sitios de corte y empalme que se dan de manera natural, tal como en los genes para la protrombina humana y la adenina fosforribosiltransferasa y dos veces dentro del gen para la mayor subunidad de RNA polimerasa II de ratón. Los restantes nucleótidos del sitio críptico se ajustan a una secuencia de consenso de sitio de corte y empalme normal con sus variaciones previstas (A₆₀G₇₉/g₁₀₀t₁₀₀a₅₉a₇₁g₈₂t₅₀). Es poco probable que la enzima sometida a delección y resultante de este mRNA empalmado de manera aberrante tenga actividad alguna; estando conservados en la enzima de *E. coli* 8 de los 19 aminoácidos suprimidos. A pesar de que las 2 pacientes presentan una actividad residual de la enzima que es del orden de un 20% de la de los controles, la actividad es probablemente debida al segundo alelo no identificado en estas pacientes.

6 mutaciones de sentido equivocado

La sustitución Arg → Cys (C1081T) en la paciente 1396 está dentro de una secuencia hidrofílica de la que anteriormente se ha postulado que es la región ligadora entre los dominios catalítico y regulador de MTHFR (Goyette P *et al.*, *Nature Genetics*, 1994, 7:195-200). Estos 2 dominios son fácilmente separables mediante leve tripsinización de la enzima porcina. El dominio ligador, que es una región altamente cargada, es probable que esté situado en la superficie exterior de la proteína y sea por consiguiente más accesible a la proteólisis. Debido al hecho de que la sustitución Arg → Cys convierte un residuo hidrofílico cargado en un residuo polar no cargado, la misma no puede ser considerada como un cambio conservativo, y podría afectar a la estabilidad de la enzima.

Las 2 sustituciones Arg → Cys identificadas en los pacientes 356 y 458 (C985T y C1015T, respectivamente) puede ser que intervengan en la fijación del cofactor FAD. Los anteriores trabajos sobre los que se informa en la literatura demostraron que el calentamiento de extractos de fibroblastos a 55° en ausencia del cofactor FAD inactivaba completamente la MTHFR. La adición de FAD a la mezcla de reacción antes de la termoinactivación les restituyó cierta actividad enzimática a los extractos de control y a los extractos de algunos pacientes, mientras que los extractos de los pacientes 356 y 458 no se vieron afectados. Sobre la base de estas observaciones, se sugirió que estos 2 pacientes tenían mutaciones que afectaban a una región de la proteína que intervenía en la fijación del FAD. Las 2 mutaciones se encuentran muy cerca una de otra, a 11 aminoácidos de distancia. En el paciente 356, el residuo Arg es conservado evolutivamente en *E. coli* y se encuentra en un tramo de 9 aminoácidos conservados, lo que sugiere un papel decisivo para este residuo; y el residuo Arg alterado en el paciente 458 no es conservado evolutivamente. El análisis de la estructura cristalina de la acil-CoA deshidrogenasa de cadena mediana (MCAD), que es una flavoproteína, ha definido residuos decisivos que intervienen en la fijación de FAD. Dos residuos consecutivos de la proteína MCAD, el Met165 y el Trp166, que intervienen en interacciones con FAD, pueden ser también identificados en la MTHFR, en el lado de cola respectivamente a 3 y 4 aminoácidos del residuo Arg alterado en el paciente 458.

La sustitución Thr → Met (C692T) se encuentra en una región de alta conservación con la enzima *E. coli* y en una región de buena homología con la dihidrofolato reductasa humana (DHFR) (Fig. 11). En la Fig. 11, = es identidad; • es homología; y ◇ es identidad con la enzima DHFR bovina. Un asterisco (*) indica la situación de la sustitución Thr → Met. Considerando el fenotipo de temprana iniciación de los pacientes, puede suponerse que el residuo de treonina es decisivo para la actividad o que el mismo contribuye a un importante dominio de la proteína. Esta región de homología en la DHFR contiene un residuo, el Thr136, del que se ha descrito que interviene en la fijación del folato. Este residuo Thr en la DHFR se alinea con un residuo Ser en la MTHFR que es un aminoácido con similares propiedades bioquímicas. La sustitución Thr → Met está situada hacia el lado de cola a 8 aminoácidos de este codón Ser, en el centro de la región de homología entre las 2 enzimas. Establecemos por consiguiente la hipótesis de que la sustitución Thr → Met puede alterar la fijación del sustrato de folato.

Las sustituciones G167A (Arg → Gln) y C764T (Pro → Leu) afectan ambas a aminoácidos no conservados. No puede determinarse en la actualidad su importancia en el desarrollo de la deficiencia de la MTHFR. Todas las mutaciones identificadas hasta la fecha están situadas en el extremo 5' de la secuencia codificante, que es la región de la que se piensa que codifica el dominio catalítico de la MTHFR. El análisis de las mutaciones ha sido útil para empezar a abordar los aspectos relativos a la relación existente entre la estructura y las propiedades funcionales de la enzima, así como para empezar a entender los diversos fenotipos en la severa deficiencia de la MTHFR.

III. Identificación de la mutación A → V

Se usaron el análisis del SSCP y la secuenciación directa de fragmentos PCR para identificar una sustitución de C a T en el bp 677, cuya sustitución convierte un residuo de alanina en un residuo de valina (Fig. 5A). Los cebadores para el análisis del cambio A → V son los siguientes: 5'-TGAAGGAGAA GGTGTCTGCG GGA-3' (exónico) y 5'-AGGACGGTGC GGTGAGAGTG-3' (intrónico); generando estos cebadores un fragmento de 198 bp. La Fig. 5A ilustra la secuencia de dos individuos que son un homocigoto para el residuo de alanina y un homocigoto para el residuo de valina. Están ilustradas las hebras antisentido. Esta alteración crea un sitio *HinfI* (Fig. 5B), lo cual fue usado para someter a selección a 114 cromosomas francocanadienses no seleccionados; siendo la frecuencia alélica de

la sustitución de 0,38. La sustitución crea una secuencia de reconocimiento para *HinfI*, lo cual da lugar a una digestión del fragmento de 198 bp que queda así convertido en un fragmento de 175 bp y un fragmento de 23 bp; habiéndose salido del gel el fragmento mencionado en último lugar. La Fig. 5B ilustra los tres posibles genotipos. Las frecuencias de los 3 genotipos eran las siguientes: -/-, 37%; +/-, 51%; y +/+, 12% (el (+) indica la presencia del sitio de restricción para *HinfI* y de un residuo de valina).

El residuo de alanina es conservado en la MTHFR porcina, así como en los correspondientes genes *metF* y *stymetF* de *E. coli* y *S. typhimurium*, respectivamente. El alto grado de conservación de este residuo y su situación en una región de alta homología con las enzimas bacterianas aludían a su importancia en la estructura o función de la enzima. Además, la frecuencia del genotipo (+/+) era coherente con la frecuencia de la variante de MTHFR termolábil que está implicada en la enfermedad vascular.

Material clínico

Para determinar la frecuencia de la mutación A → V, fue analizado mediante PCR y digestión de restricción el DNA de 57 individuos de Quebec. Los individuos, que eran todos ellos francocanadienses, no fueron examinados clínica ni bioquímicamente.

Los 40 individuos que se analizan en la Tabla 3 habían sido descritos anteriormente en Engbersen *et al.* (*Am. J. Hum. Genet.*, 1995, 56:142-150). De los 13 pacientes cardiovasculares, 8 tenían arterioesclerosis cerebrovascular y 5 tenían arterioesclerosis periférica. Cinco tenían MTHFR termolábil, mientras que 8 tenían MTHFR termoestable (con una actividad residual de más de un 33% tras calentamiento). Los controles y los pacientes eran todos ellos caucásicos holandeses de entre 20 y 60 años de edad. Ninguno de estos individuos usaba vitaminas que pudiesen alterar los niveles de homocisteína. Los análisis enzimáticos y las determinaciones de la homocisteína estaban también descritos por Engbersen *et al.* (*Am. J. Hum. Genet.*, 1995, 56:142-150).

TABLA 3

Correlación entre el genotipo de MTHFR y la actividad de la enzima, la termolabilidad y el nivel de homocisteína en plasma

	-/- n=19	+/- n=9	+/+ n=9
actividad específica^{a,b} (nmoles de CH₂O/mg de proteína/h)	22.9±1.7 (11.8-33.8)	15.0±0.8 (10.2-18.8)	6.9±0.6 (2.6-10.2)
actividad residual tras calentamiento^{a,b} (%)	66.8±1.5 (55-76)	56.2±2.8 (41-67)	21.8±2.8 (10-35)
homocisteína en plasma^{a,c} (μM) (tras ayuno)	12.6±1.1 (7-21)	13.8±1.0 (9.6-20)	22.4±2.9 (9.6-42)
homocisteína en plasma^{a,c} ((μM) (post-carga de metionina)	41.3±5.0 ^d (20.9-110)	41±2.8 (29.1-54)	72.6±11.7 ^e (24.4-159)

p < 0,01 según análisis de varianza de una vía

p < 0,01 según la prueba t para muestras apareadas para todas las combinaciones

p < 0,05 según prueba t para muestras apareadas para el grupo +/+ frente al grupo +/- o al grupo -/-; p > 0,05 para el grupo +/- frente al grupo -/-.

n = 18 para este parámetro

n = 11 para este parámetro

La actividad de la enzima y la homocisteína en plasma fueron determinadas como se ha descrito anteriormente. Cada valor representa la media μ el error característico. La gama de valores está indicada entre paréntesis debajo de la media.

Correlación de la mutación A → V con la función alterada de la MTHFR

En un total de 40 pellets de linfocitos de pacientes con enfermedad vascular prematura y de controles fue llevado a cabo un análisis genotípico y fueron medidas la actividad y la termolabilidad de la enzima. Fueron seleccionados 13 pacientes vasculares de un estudio anterior (Engbersen *et al.*, *Am. J. Hum. Genet.*, 1995, 56:142-150), de entre los cuales se consideraba que 5 tenían MTHFR termolábil. De entre los de un gran grupo de referencia de 89 controles

fueron estudiados los 7 individuos que tenían MTHFR termolábil, y fueron seleccionados de entre los del mismo grupo de referencia 20 controles adicionales con MTHFR normal. La Tabla 3 documenta la relación existente entre los genotipos y la actividad específica de la enzima, la termolabilidad y el nivel de homocisteína en plasma. La actividad media de MTHFR para los individuos homocigóticos para la sustitución (+/+) era de aproximadamente un 30% de la actividad media para los individuos (-/-), que son homocigóticos para el residuo de alanina. Los heterocigotos tenían una actividad media de MTHFR que era de un 65% de la actividad de los individuos (-/-); siendo este valor intermedio entre los valores para los individuos (-/-) y (+/+). Las gamas de valores de las actividades presentaban cierto solapamiento para los genotipos heterocigóticos y (-/-), pero los individuos homocigóticos (+/+) virtualmente no presentaban solapamiento con los grupos anteriores. Un análisis de varianza de una vía produjo un valor $p < 0,0001$; demostrando una prueba t de Bonferroni en parejas que los tres genotipos eran significativamente distintos con $p < 0,01$ para las tres posibles combinaciones.

Los tres genotipos eran todos ellos significativamente distintos ($p < 0,01$) con respecto a la termolabilidad de la enzima. La actividad residual media tras termoinactivación por espacio de 5 minutos a 46° era de un 67%, un 56% y un 22% para los genotipos (-/-), (+/-) y (+/+), respectivamente. Mientras que el grado de termolabilidad presenta cierto solapamiento para los individuos (-/-) y los heterocigotos, los individuos con dos alelos mutantes presentaban una gama de valores claramente inferior. Cada individuo con el genotipo (+/+) presentaba una actividad residual $< 35\%$ tras calentamiento y una actividad específica $< 50\%$ de la del genotipo (-/-).

Las concentraciones de homocisteína total tras ayuno y 6 horas después de la carga de metionina fueron medidas en plasma mediante cromatografía de líquidos de alta resolución usando detección por fluorescencia. Los niveles de homocisteína en ayunas en los individuos (+/+) eran casi el doble del valor para los individuos (+/-) y (-/-). Las diferencias entre genotipos para la homocisteína en plasma se mantenían cuando la homocisteína fue medida a las 6 horas de la carga de metionina. Un análisis de varianza de una vía produjo un valor $p < 0,01$ para los niveles de homocisteína en ayunas y post-metionina. Una prueba t de Bonferroni en parejas demostró que solamente los individuos mutantes homocigóticos tenían niveles de homocisteína significativamente elevados ($p < 0,05$).

Mutagénesis por PCR para la expresión de la mutación A $\rightarrow$ V in vitro

Fue usada para crear la mutación de A a V mutagénesis por PCR usando como plantilla el vector Bluescript^{MF} con contenido de cDNA. Fue usada para reducir los errores de la PCR Vent^{MF} polimerasa (NEB). Fueron usados los cebadores siguientes: cebador 1, bp -200 a -178, sentido; cebador 2, bp 667 a 687, antisentido, que contiene un emparejamiento incorrecto, A, en el bp 677; cebador 3, 667 a 687, sentido, que contiene un emparejamiento incorrecto, T, en el bp 677; cebador 4, bp 1092 a 1114, antisentido. La PCR fue llevada a cabo usando los cebadores 1 y 2 para generar un producto de 887 bp, y usando los cebadores 3 y 4 para generar un producto de 447 bp. Los dos fragmentos de PCR fueron aislados mediante un gel de agarosa al 1,2% mediante GeneClean^{MF} (BIO 101). Usando los cebadores 1 y 4 y los 2 primeros fragmentos de PCR como plantilla, fue llevada a cabo una reacción PCR final para generar una banda de 1,3 kb que contenía la mutación. El fragmento de 1,3 kb fue sometido a digestión con *NcoI* y *MscI*, y fue insertado en el vector de expresión con contenido de cDNA de tipo salvaje substituyendo las secuencias entre el sitio *NcoI* en el bp 11 y el sitio *MscI* en el bp 943. Fueron secuenciados todo el fragmento de sustitución y los sitios de clonación para verificar que no habían sido introducidos por la PCR cambios adicionales.

Análisis de la expresión de cDNA de tipo salvaje y mutagenizado

Se hicieron durante la noche a 37°C en 2 x medio YT con 0,05 mg/ml de ampicilina cultivos de vector con contenido de JM105^{MF} en solitario, vector + cDNA de MTHFR de tipo salvaje, o vector + cDNA mutagenizado. Cultivos recientes de 10 ml de cada uno fueron inoculados con aproximadamente 50 μl de cultivos llevados a cabo durante la noche para una densidad óptica de 0,05, y fueron cultivados a 37°C hasta una densidad óptica de 1 a 420 nm. Los cultivos fueron entonces inducidos por espacio de 2 horas con IPTG (IPTG = isopropil- β -D-tiogalactopiranosido) 1 mM y pelletizados. Las células fueron puestas nuevamente en suspensión en tampón TE con 2 $\mu\text{g/ml}$ de aprotinina y 2 $\mu\text{g/ml}$ de leupeptina (3,5 x el peso de las células en húmedo). Las suspensiones celulares fueron sonicadas en hielo 3 veces por espacio de 15 seg. cada vez para así romper y abrir las membranas celulares, y fueron luego centrifugadas por espacio de 30 minutos a 4°C para pelletizar los residuos celulares y las células no lisadas. El supernatante fue retirado y analizado para determinar la concentración de proteína con el análisis de proteínas Bio-Rad^{MF}. Fue llevado a cabo un análisis Western usando el conjunto de materiales y utensilios Amersham ECL^{MF} según las instrucciones del proveedor, usando antisuero generado contra MTHFR de hígado porcino purificada. Los análisis enzimáticos fueron llevados a cabo mediante procedimientos establecidos, y la termolabilidad fue valorada pretratando los extractos a 46°C por espacio de 5 minutos antes de determinar la actividad. Las actividades específicas (nmoles de formaldehído/h/mg de proteína) fueron calculadas para los 2 constructos que contenían cDNA tras sustracción de los valores obtenidos con vector solamente (para restar la actividad de fondo de MTHFR de *E. coli*).

El cDNA de MTHFR (2,2 kb) (Fig. 6) tiene un marco de lectura abierto de 1980 bp, que predice una proteína de 74,6 kDa. Se ha demostrado que la enzima de hígado porcino purificada tiene subunidades 77 kDa. El análisis Western (Fig. 7A) de varios tejidos humanos y de hígado porcino ha revelado un polipéptido de 77 kDa en todos los tejidos estudiados, así como un polipéptido adicional de aproximadamente 70 kDa en hígado fetal humano y en hígado porcino, sugiriendo la presencia de isozimas. Dos μg de proteína de extracto bacteriano fueron usados para los carriles 1-3. Los tejidos (carriles 4-8) fueron preparados por homogeneización en sucrosa 0,25M con inhibidores de proteasa (2 $\mu\text{g/ml}$ de aprotinina y otros tantos de leupeptina), siendo efectuada a continuación sonicación (3 x 15 seg.) en hielo.

Los extractos fueron centrifugados por espacio de 15 min. en una microcentrífuga a 14.000 g, y fueron usados 100 μ g de proteína supernatante para el análisis Western. h = humano; p = porcino.

El cDNA de tipo salvaje y un cDNA mutagenizado que contenía la sustitución A $\rightarrow$ V fueron expresados en *E. coli* para producir una proteína de aproximadamente 70 kDa (Fig. 7A) que comigra con el polipéptido más pequeño anteriormente mencionado. El tratamiento de los extractos a 46°C por espacio de 5 minutos reveló que la enzima que contenía la sustitución era significativamente más termolábil que la enzima de tipo salvaje ($p < 0,001$; Fig. 7B). Fueron llevados a cabo dos experimentos independientes (con 3-4 réplicas para cada constructo para cada experimento) para medir la actividad termoestable de los cDNAs de MTHFR de tipo salvaje y de MTHFR A $\rightarrow$ V mutagenizada. Los valores que se indican representan la media μ el error característico para cada experimento, como porcentaje de actividad residual tras el calentamiento. Las medias de las actividades específicas antes del calentamiento (expresadas como nmoles de formaldehído/h/mg de proteína) eran las siguientes: Exp. 1, 3,8 y 5,3 para MTHFR y MTHFR A $\rightarrow$ V, respectivamente; Exp. 2, 6,2 y 7,5 para MTHFR y MTHFR A $\rightarrow$ V, respectivamente. Los experimentos de expresión no estaban destinados a medir las diferencias de la actividad específica antes del calentamiento, puesto que la variación en los rendimientos de expresión podría contribuir a crear dificultades de interpretación. Curiosamente, sin embargo, la actividad específica para el constructo mutante era más alta en ambos experimentos. Es posible que la proteína mutante tenga una estabilidad incrementada en *E. coli*, o que los cuerpos de inclusión en nuestros extractos contribuyesen a crear diferencias en la recuperación de la enzima correctamente ensamblada.

Estos estudios han identificado una sustitución común en el gen de MTHFR que redundaba en termolabilidad *in vitro* e *in vivo*. La mutación, en el estado heterocigótico u homocigótico, está correlacionada con la reducida actividad de la enzima y la incrementada termolabilidad de la MTHFR en los extractos de linfocitos. Se observó una significativa elevación de la homocisteína en plasma en los individuos que eran homocigóticos para la mutación. No fueron observadas entre los heterocigotos y los individuos (-/-) diferencias estadísticamente significantes para los niveles de homocisteína, y esta observación no es sorprendente, puesto que la homocisteína en plasma puede verse influenciada por varios factores ambientales entre los que se incluyen la toma de folato, vitamina B₁₂, vitamina B₆ y metionina, así como por la variación genética en otros loci, tal como en el caso del gen de la cistationina- β -sintasa.

La sustitución de alanina por valina conserva la hidrofobicidad del residuo y va asociada a pequeños cambios de la actividad, en contraste con los cambios no conservativos, tales como el anteriormente descrito cambio de arginina por glutamina en la MTHFR, que va asociado a una mayor disminución de la actividad de la enzima y a hiperhomocisteinemia severa. El residuo de alanina está situado en una región de homología con los genes metF bacterianos. Hemos observado también la misma región de homología en el gen de la dihidrofolato reductasa (DHFR) humana (Fig. 11), a pesar de que el propio residuo de alanina no es conservado; y esta región de aminoácidos 130-149 de DHFR contiene T136, que se ha visto implicado en la fijación del folato en un análisis de la estructura cristalina de la DHFR humana recombinante. Es tentador especular con que esta región en la MTHFR también intervenga en la fijación del folato, y con que la enzima pueda ser estabilizada en presencia de folato. Esta hipótesis es compatible con la perfectamente documentada influencia del folato en los niveles de homocisteína y con la descrita corrección de la hiperhomocisteinemia leve mediante ácido fólico en individuos con enfermedad vascular prematura y en individuos con MTHFR termolábil.

A pesar de que nuestro cDNA no es lo suficientemente largo como para codificar el polipéptido MTHFR de mayor tamaño, el mismo es capaz de dirigir la síntesis de la isozima de menor tamaño. El codón de inicio ATG para este polipéptido está dentro de una buena secuencia de consenso para la iniciación de la traducción. Quedan por determinar las cuestiones de si la isozima está limitada al hígado y de cuál es su papel en este tejido.

Estos datos han identificado un cambio genético común en la MTHFR que redundaba en termolabilidad; no abordando nuestros experimentos directamente la relación existente entre este cambio y la enfermedad vascular. Sin embargo, este polimorfismo representa un análisis diagnóstico para la evaluación de la termolabilidad de la MTHFR en hiperhomocisteinemia. Son necesarios estudios comparativos a gran escala para evaluar la frecuencia de este cambio genético en varias formas de enfermedad arterial oclusiva y para examinar la interacción entre este marcador genético y los factores dietarios. Tienen que ser examinadas poblaciones perfectamente definidas, puesto que el limitado conjunto de datos sugiere hasta la fecha que pueden existir frecuencias alélicas específicas de la población. Más importante es sin embargo el hecho de que la identificación de un candidato a factor de riesgo genético para la enfermedad vascular, que puede verse influenciada por la toma de nutrientes, representa un paso decisivo en el diseño de terapias apropiadas para la forma homocisteinémica de la arterioesclerosis.

cDNA para MTHFR y su utilidad potencial

Un cDNA humano para MTHFR (2,2 kb) ha sido aislado como hemos descrito en Goyette *et al.* (*Nature Genetics*, 1994, 7:195-200) y en Frosst *et al.* (*Nature Genetics*, 1995, 10:111-113). El cDNA ha sido expresado *in vitro* para producir una proteína MTHFR de aproximadamente 70 kDa (Fross P *et al.*, *Nature Genetics*, 1995, 10:111-113).

Usando la secuencia del cDNA fueron identificadas mutaciones en pacientes con severa y leve deficiencia de la MTHFR (Goyette P *et al.*, *Nature Genetics*, 1994, 7:195-200; Goyette P *et al.*, *Am. J. Hum. Genet.*, 1995, 56:1052-1059; Frosst P *et al.*, *Nature Genetics*, 1995, 10:111-113).

La secuencia del cDNA es un necesario punto de partida para la detección de cambios de la secuencia de la MTHFR que identificasen a los individuos que están en situación de riesgo de contraer enfermedades cardiovasculares y neurológicas, así como de padecer otros trastornos afectados por el metabolismo del ácido fólico. Los análisis diagnósticos mediante análisis del DNA son más eficaces y precisos que las pruebas efectuadas mediante análisis enzimáticos/bioquímicos. Se requiere menos sangre, y se dispone de los resultados en un periodo de tiempo más corto. Las pruebas podrían ser llevadas a cabo como una operación rutinaria en cualquier laboratorio que lleve a cabo diagnósticos genético-moleculares, sin los reactivos especializados y sin la pericia especializada que son necesarios para un análisis basado en enzimas.

La segunda utilidad importante del cDNA radicaría en el diseño de protocolos terapéuticos para la corrección de la deficiencia de la MTHFR. Estos protocolos podrían implicar directamente al gen, como en las pruebas de terapia genética o en el uso de reactivos que pudiesen modificar la expresión génica. Como alternativa, la terapia podría requerir el conocimiento de la secuencia de aminoácidos (sacada de la secuencia del cDNA), como en el uso de reactivos que modificasen la actividad de la enzima. La identificación de secuencias y/o cambios de secuencia en regiones específicas del cDNA o de la proteína, tales como los sitios de fijación del FAD o los sitios de fijación del folato, es útil para diseñar protocolos terapéuticos que atañan a los susodichos nutrientes.

Utilidad de la invención en los estudios clínicos y diagnósticos

Fueron estudiados pacientes con enfermedad arterial coronaria en Montreal (n = 153) para examinar la frecuencia de la sustitución de alanina por valina. El catorce por ciento de los pacientes eran homocigóticos para esta mutación. Un análisis de 70 individuos de control (exentos de enfermedad cardiovascular) demostró que solamente un siete por ciento de estos individuos eran homocigóticos para la mutación de alanina a valina.

El análisis de los niveles de homocisteína en 123 hombres del susodicho grupo de pacientes indicó que el alelo mutante incrementaba significativamente los niveles de homocisteína haciendo que pasasen de ser de 10,2 micromoles/l en los hombres normales homocigóticos a ser de 11,5 y 12,7 en los heterocigotos y en los mutantes homocigóticos, respectivamente.

Han sido examinadas para determinar la presencia de la mutación de alanina a valina familias con un niño con espina bífida, que es un defecto del tubo neural. Aproximadamente las de un 16% de las madres que tenían un hijo con espina bífida eran homocigóticas para esta mutación, mientras que tan sólo los de un 5% de los individuos de control eran homocigóticos. Los padres de los niños que tenían espina bífida también presentaban una incrementada prevalencia del genotipo mutante homocigótico (10%), al igual como lo hacían los propios niños afectados (13%).

La Tabla 4 indica el efecto interactivo del ácido fólico con el cambio de alanina a valina de los mutantes homocigóticos. En un estudio de familias de Framingham, Massachusetts y Utah, los individuos que eran mutantes homocigóticos pero tenían niveles de folato de más de 5 ng/ml no presentaban incrementados niveles de homocisteína en comparación con los individuos que tenían el genotipo normal o heterocigótico. Sin embargo, los individuos que eran mutantes homocigóticos pero tenían niveles de folato de menos de 5 ng/ml tenían niveles de homocisteína que eran significativamente más altos que los de los otros genotipos.

TABLA 4

Niveles medios de homocisteína en ayunas y post-carga de metionina para distintos genotipos de MTHFR

Homocisteína en plasma	Normales (-/-)	Heterocigotos (+/-)	Homocigotos (++)	P _{tendencia}
N	58	61	30	
En ayunas*	9.4	9.2	12.1	0.02
Folato < 5 ng/ml	10.2	10.4	15.2	0.002
Folato ≥ 5 mg/ml	8.2	7.5	7.5	0.52
Post-carga de metionina	30.0	30.9	31.3	0.62

*Significante interacción entre los niveles de folato y el genotipo (p = 0,03)

El Ejemplo III proporciona datos preliminares para la intervención terapéutica mediante suplementación de ácido fólico a los individuos que son homocigóticos para el cambio de alanina a valina. Los datos sugieren que unos más altos niveles de folato en plasma conducirían a una normalización de los niveles de homocisteína en los individuos mutantes y podrían impedir que se produjesen trastornos que van asociados a los altos niveles de homocisteína, tales

como la enfermedad cardiovascular, los defectos del tubo neural y posiblemente otros trastornos. La suplementación de ácido fólico para los individuos mutantes podría también restablecer los niveles normales de metionina y S-adenosilmetionina. Esto sería relevante para los trastornos que se ven influenciados por la metilación, tales como las neoplasias, las anomalías del desarrollo, la enfermedad neurológica, etc.

5

Si bien la invención ha sido descrita en conexión con realizaciones específicas de la misma, se entenderá que la invención puede ser objeto de adicionales modificaciones, y que esta solicitud pretende englobar cualesquiera variaciones, usos o adaptaciones de la invención que sigan en general los principios de la invención, incluyendo todas aquellas variaciones que con respecto a la presente descripción estén enmarcadas dentro de la práctica conocida o habitual dentro de la técnica a la que pertenece la invención y que puedan ser aplicadas a las características esenciales que han sido expuestas anteriormente y a continuación dentro del alcance de las reivindicaciones adjuntas.

15

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

1. Sonda de cDNA para una metilenotetrahidrofolato reductasa (MTHFR) humana que tiene una secuencia de nucleótidos de la ID SEC N°: 1 o una secuencia de nucleótidos que codifica una secuencia de aminoácidos de la ID SEC N°: 2, comprendiendo dicha sonda una secuencia que se hibrida en condiciones de rigor con dichas secuencias de MTHFR.

2. Método de diagnosis de una deficiencia de la MTHFR en un paciente, comprendiendo dicho método los pasos de:

a) amplificar una muestra de DNA de un paciente o someter a transcripción inversa a una muestra de RNA de un paciente para así convertirla en DNA y amplificar dicho DNA; y

b) analizar el DNA amplificado del paso a) para determinar si el mismo contiene al menos una anomalía de secuencia con respecto a la metilenotetrahidrofolato reductasa (MTHFR) humana que tiene una secuencia de nucleótidos de la ID SEC N°: 1 o una secuencia que codifique la secuencia de aminoácidos de la ID SEC N°: 2, siendo dichas anomalías indicativas de deficiencia genómica de la MTHFR.

3. El método de la reivindicación 2, en el que dicha deficiencia de la MTHFR va asociada a un trastorno seleccionado de entre los miembros del grupo que consta de trastornos cardiovasculares y neurológicos, tumores y trastornos influenciados por el metabolismo del ácido fólico.

4. El método de la reivindicación 2, en el que dicha anomalía secuencial comprende una mutación seleccionada de entre los miembros del grupo que consta de

Situación de los bp	Sustitución en el DNA	Cambio de a.a.
167	CGA → CAA	R → Q
482	CGC → CAG	R → Q
559	CGA → TGA	R → X
677	GCC → GTC	A → V
692	ACG → ATG	T → M
764	CCC → CTC	P → L
792 + 1 sitio de corte y empalme 5'	GT → AT	Δ19a.a.
985	CGC → TGC	R → C
1015	CGC → TGC	R → C
1081	CGC → TGC	R → C

5. Uso de un vector recombinante que comprende la ID SEC N°: 1 para la fabricación de un medicamento para la expresión de la proteína MTHFR (ID SEC N°: 2) bajo el control de un promotor adecuado.

6. Proteína MTHFR humana que es codificada por la secuencia de nucleótidos de la ID SEC N°: 1 y tiene la secuencia de aminoácidos de la ID SEC N°: 2.

7. Ácido nucleico de metilenotetrahidrofolato reductasa (MTHFR) humana recombinante que tiene una secuencia de nucleótidos de la ID SEC N°: 1 y codifica la secuencia de aminoácidos de la ID SEC N°: 2.

8. Uso de la proteína de la reivindicación 6 o del ácido nucleico de la reivindicación 7 para la fabricación de un medicamento para tratar pacientes con deficiencia de la MTHFR.

9. El método de la reivindicación 2, en el que dicho paciente tiene menos de un año de edad.

10. El método de la reivindicación 2, en el que dicho paciente tiene al menos un año de edad.

11. El método de la reivindicación 2, en el que dicha deficiencia de la MTHFR va asociada a un incrementado riesgo de que aparezca un defecto del tubo neural en un descendiente de dicho paciente.

ES 2 271 946 T3

12. El método de la reivindicación 3, en el que dichos tumores son seleccionados de entre los miembros del grupo que consta de neuroblastomas y carcinomas colorrectales.

5 13. El método de la reivindicación 4, en el que dicha anormalidad secuencial consta de la anormalidad secuencial 677C-T.

10

15

20

25

30

35

40

45

50

55

60

65

ES 2 271 946 T3

10 20 30 40 50
 AAT TCC GGA GGC ATG GTG AAC GAA GGC AGA GGA AAC AGC AGC CTC AAC CCC
 TTA AGG CCT CGG TAC CAC TIG CTT CGG TCT CCT TIG TCG TCG GAG TIG GGG
 Asn Ser Gly Ala Met Val Asn Glu Ala Arg Gly Asn Ser Ser Leu Asn Pro

60 70 80 90 100
 TGC TTG GAG GGC AGT GGC AGC AGT GGC AGT GAG AGC TCC AAA GAT AGT TCG
 ACG AAC CTC CCG TCA CCG TCG TCA CCG TCA CTC TCG AGG TTT CTA TCA AGC
 Cys Leu Glu Gly Ser Ala Ser Ser Gly Ser Glu Ser Ser Lys Asp Ser Ser

110 120 130 140 150
 AGA TGT TCC ACC CCG GGC CTG GAC CCT GAG CGG CAT GAG AGA CTC CGG GAG
 TCT ACA AGG TGG GGC CCG GAC CTG GGA CTC GCT GTA CTC TCT GAG GGC CTC
 Arg Cys Ser Thr Pro Gly Leu Asp Pro Glu Arg His Glu Arg Leu Arg Glu

160 170 180 190 200
 AAG ATG AGG CGG CGA TTG GAA TCT GGT GAC AAG TGG TTC TCC CTG GAA TTC
 TTC TAC TCC GGC GCT AAC CTT AGA CCA CTG TTC ACC AAG AGG GAC CTT AAG
 Lys Met Arg Arg Arg Leu Glu Ser Gly Asp Lys Trp Phe Ser Leu Glu Phe

210 220 230 240 250
 TTC CCT CCT CGA ACT GCT GAG GGA GCT GTC AAT CTC ATC TCA AGG TTT GAC
 AAG GGA GGA GCT TGA CGA CTC CTT CGA CAG TTA GAG TAG AGT TCC AAA CTG
 Phe Pro Pro Arg Thr Ala Glu Gly Ala Val Asn Leu Ile Ser Arg Phe Asp

260 270 280 290 300
 CCG ATG GCA GCA GGT GGC CCC CTC TAC ATA GAC GTG ACC TGG CAC CCA GCA
 GCC TAC CGT CGT CCA CCG GGG GAG ATG TAT CTG CAC TGG ACC GTG GGT CGT
 Arg Met Ala Ala Gly Gly Pro Leu Tyr Ile Asp Val Thr Trp His Pro Ala

310 320 330 340 350
 GGT GAC CCT GGC TCA GAC AAG GAG ACC TCC TCC ATG ATG ATC GCC AGC ACC
 CCA CTG GGA CCG AGT CTG TTE CTC TGG AGG AGG TAC TAC TAG CCG TCG TGG
 Gly Asp Pro Gly Ser Asp Lys Glu Thr Ser Ser Met Met Ile Ala Ser Thr

FIG. 1A

```

360          370          380          390          400
  *          *          *          *          *
GEE GTG AAC TAC TGT GGC CTG GAG ACC ATC CTG CAC ATG ACC TGC TGC CGT
CGG CAC TTG ATG ACA CCG GAC CTC TGG TAG GAC GTG TAC TGG ACG ACG GCA
Ala Val Asn Tyr Cys Gly Leu Glu Thr Ile Leu His Met Thr Cys Cys Arg-

410          420          430          440          450
  *          *          *          *          *
CAG CGC CTG GAG GAG ATC ACG GGC CAT CTG CAC AAA GGT AAG CAG CTG GGC
GTC GCG GAC CTC CTC TAG TGC CCG GTA GAC GTG TTT CGA TTC GTC GAC CCG
Gln Arg Leu Glu Glu Ile Thr Gly His Leu His Lys Ala Lys Gln Leu Gly-

460          470          480          490          500          510
  *          *          *          *          *          *
CTG AAG AAC ATC ATG GCG CTG CCG GGA GAC CCA ATA GGT GAC CAG TGG GAA
GAC TTC TTG TAG TAC CCG GAC GGC CCT CTG GGT TAT CCA CTG GTC ACC CTT
Leu Lys Asn Ile Met Ala Leu Arg Gly Asp Pro Ile Gly Asp Gln Trp Glu-

          520          530          540          550          560
  *          *          *          *          *          *
GAG GAG GAG GGA GGC TTC AAC TAC GCA GTG GAC CTG GTG AAG CAC ATC CGA
CTC CTC CTC CCT CCG AAG TTG ATG CGT CAC CTG GAC CAC TTC GTG TAG GGT
Glu Glu Glu Gly Gly Phe Asn Tyr Ala Val Asp Leu Val Lys His Ile Arg-

          570          580          590          600          610
  *          *          *          *          *          *
AGT GAG TTT GGT GAC TAC TTT GAC ATC TGT GTG GCA GGT TAC CCC AAA GGC
TCA CTC AAA CCA CTG ATG AAA CTG TAG ACA CAC CGT CCA ATG GGG TTT CCG
Ser Glu Phe Gly Asp Tyr Phe Asp Ile Lys Val Ala Gly Tyr Pro Lys Gly-

          620          630          640          650          660
  *          *          *          *          *          *
CAC CCC GAA GCA GGG AGC TTT GAG GGT GAC CTG AAG CAC TTG AAG GAG AAG
GTG GGG CTT CGT CCC TCG AAA CTC CGA CTG GAC TTC GTG AAC TTC CTC TTC
His Pro Glu Ala Gly Ser Phe Glu Ala Asp Leu Lys His Leu Lys Glu Lys-

          670          680          690          700          710
  *          *          *          *          *          *
GTG TCT GCG GGA GCC GAT TTC ATC ATC ACG CAG CTT TTC TTT GAG GGT GAC
CAC AGA CGC CCT CCG CTA AAG TAG TAG TGC GTC GAA AAG AAA CTC CGA CTG
Val Ser Ala Gly Ala Asp Phe Ile Ile Thr Gln Leu Phe Phe Glu Ala Asp-

```

ES 1B

ES 2 271 946 T3

720	730	740	750	760
* * *	* * *	* * *	* * *	* * *
ACA TTC TTC CGC TTT GTG AAG GCA TGC ACC GAC ATG GGC ATC ACT TGC CCC				
TGT AAG AAG GCG AAA CAC TTC CGT ACG TGG CTG TAC CCG TAG TGA ACG GGG				
Thr Phe Phe Arg Phe Val Lys Ala Cys Thr Asp Met Gly Ile Thr Cys Pro-				

770	780	790	800	810
* * *	* * *	* * *	* * *	* * *
ATC GTC CCC GGG ATC TTT CCC ATC CAG GGC TAC CAC TCC CTT CCG CAG CTT				
TAG CAG GGG CCC TAG AAA GGG TAG GTC CCG ATG GTG AGG GAA GCC GTC GAA				
Ile Val Pro Gly Ile Phe Pro Ile Gln Gly Tyr His Ser Leu Arg Gln Leu-				

820	830	840	850	860
* * *	* * *	* * *	* * *	* * *
GTG AAG CTG TCC AAG CTG GAG GTG CCA CAG GAG ATC AAG GAC GTG ATT GAG				
CAC TTC GAC AGG TTC GAC CTC CAC GGT GTC CTC TAG TTC CTG CAC TAA CTC				
Val Lys Leu Ser Lys Leu Glu Val Pro Gln Glu Ile Lys Asp Val Ile Glu-				

870	880	890	900	910
* * *	* * *	* * *	* * *	* * *
CCA ATC AAA GAC AAC GAT GCT GGC ATC CGC AAC TAT GGC ATC GAG CTG GCC				
GGT TAG TTT CTG TTG CTA CGA CCG TAG GCG TTG ATA CCG TAG CTC GAC CCG				
Pro Ile Lys Asp Asn Asp Ala Ala Ile Arg Asn Tyr Gly Ile Glu Leu Ala-				

920	930	940	950	960
* * *	* * *	* * *	* * *	* * *
GTG AGC CTG TGC CAG GAG CTT CTG GTC AGT GGC TTG GTG CCA GGC CTC CAC				
CAC TCG GAC ACG GTC CTC GAA GAC CCG TCA CCG AAC CAC GGT CCG GAG GTG				
Val Ser Leu Cys Gln Glu Leu Leu Ala Ser Gly Leu Val Pro Gly Leu His-				

970	980	990	1000	1010	1020
* * *	* * *	* * *	* * *	* * *	* * *
TTC TAC ACC CTC AAC CGC GAG ATG GET ACC ACA GAG GTG CTG AAG CGC CTG					
AAG ATG TGG GAG TTG GCG CTC TAC CGA TGG TGT CTC CAC GAC TTC GCG GAC					
Phe Tyr Thr Leu Asn Arg Glu Met Ala Thr Thr Glu Val Leu Lys Arg Leu-					

1030	1040	1050	1060	1070
* * *	* * *	* * *	* * *	* * *
GGG ATG TGG ACT GAG GAC CCC AGG CGT CCC CTA CCC TGG GET CTC AGT GCC				
CCC TAC ACC TGA CTC CTG GGG TCC GCA GGG GAT GGG ACC CGA GAG TCA CCG				
Gly Met Trp Thr Glu Asp Pro Arg Arg Pro Leu Pro Trp Ala Leu Ser Ala-				

ES-1C

1080	1090	1100	1110	1120
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *
EAC CCC AAG CGC CGA GAG GAA GAT GTA CGT CCC ATC TTC TGG GCC TCC AGA				
GTG GGG TTC GCG GCT CTC CTT CTA CAT GCA GGG TAG AAG ACC CGG AGG TCT				
His Pro Lys Arg Arg Glu Glu Asp Val Arg Pro Ile Phe Trp Ala Ser Arg>				

1130	1140	1150	1160	1170
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *
CEA AAG AGT TAC ATC TAC CGT ACC CAG GAG TGG GAC GAG TTC CCT AAC GGC				
GGT TTE TCA ATG TAG ATG GCA TGG GTC CTC ACC CTG CTC AAG GGA TTG CCG				
Pro Lys Ser Tyr Ile Tyr Arg Thr Glu Glu Trp Asp Glu Phe Pro Asn Gly>				

1180	1190	1200	1210	1220
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *
CGC TGG GGC AAT TCC TCT TCC CCT GCC TTT GGG GAG CTG AAG GAC TAC TAC				
GCG ACC CCG TTA AGG AGA AGG GGA CCG AAA CCC CTC GAC TTC CTG ATG ATG				
Arg Trp Gly Asn Ser Ser Ser Pro Ala Phe Gly Glu Leu Lys Asp Tyr Tyr>				

1230	1240	1250	1260	1270
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *
CTC TTC TAC CTG AAG AGE AAG TCC CCC AAG GAG GAG CTG CTG AAG ATG TGG				
GAG AAG ATG GAC TTC TCG TTC AGG GGG TTC CTC CTC GAC GAC TTC TAC ACC				
Leu Phe Tyr Leu Lys Ser Lys Ser Pro Lys Glu Glu Leu Leu Lys Met Trp>				

1280	1290	1300	1310	1320
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *
GGG GAG GAG CTG ACC AGT GAA GCA AGT GTC TTT GAA GTC TTT GTT CTT TAC				
CCC CTC CTC GAC TGG TCA CTT CGT TCA CAG AAA CTT CAG AAA CAA GAA ATG				
Gly Glu Glu Leu Thr Ser Glu Ala Ser Val Phe Glu Val Phe Val Leu Tyr>				

1330	1340	1350	1360	1370
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *
CTC TCG GGA GAA CCA AAC CCG AAT GGT CAC AAA GTG ACT TGC CTG CCC TGG				
GAG AGE CCT CTT GGT TTG GCC TTA CCA GTG TTT CAC TGA ACG GAC GGG ACC				
Leu Ser Gly Glu Pro Asn Arg Asn Gly His Lys Val Thr Cys Leu Pro Trp>				

1380	1390	1400	1410	1420
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *
AAC GAT GAG CCC CTG GCG GGT GAG ACC AGC CTG CTG AAG GAG GAG CTG CTG				
TTG CTA CTC GGG GAC GCG CGA CTC TGG TCG GAC GAC TTC CTC CTC GAC GAC				
Asn Asp Glu Pro Leu Ala Ala Glu Thr Ser Leu Leu Lys Glu Glu Leu Leu>				

FILE 11

ES 2 271 946 T3

1430 1440 1450 1460 1470
 * * * * *
 CCG GTG AAC CGC CAG GGC ATC CTC ACC ATC AAC TCA CAG CCC AAC ATC AAC
 GGC CAC TTG GCG GTC CCG TAG GAG TGG TAG TIG AGT GTC GGG TTG TAG TTG
 Arg Val Asn Arg Gln Gly Ile Leu Thr Ile Asn Ser Gln Pro Asn Ile Asn.

1480 1490 1500 1510 1520 1530
 * * * * * *
 GGG AAG CCG TCC TCC GAC CCC ATC GTG GGC TGG GGC CCC AGC GGG GGC TAT
 CCC TTC GGC AGG AGG CTG GGG TAG CAC CCG ACC CCG GGG TCG CCC CCG ATA
 Gly Lys Pro Ser Ser Asp Pro Ile Val Gly Trp Gly Pro Ser Gly Gly Tyr.

 1540 1550 1560 1570 1580
 * * * * *
 GTC TTC CAG AAG GGC TAC TTA GAG TTT TTC ACT TCC CGC GAG ACA GCG GAA
 CAG AAG GTC TTC CCG ATG AAT CTC AAA AAG TGA AGG GCG CTC TGT CCG CTT
 Val Phe Gln Lys Ala Tyr Leu Glu Phe Phe Thr Ser Arg Glu Thr Ala Glu.

 1590 1600 1610 1620 1630
 * * * * *
 GCA CTT CTG CAA GTG CTG AAG AAG TAC GAG CTC CCG GTT AAT TAC CAC CTT
 CGT GAA GAC GTT CAC GAC TTC TTC ATG CTC GAG GCC CAA TTA ATG GTG GAA
 Ala Leu Leu Gln Val Leu Lys Lys Tyr Glu Leu Arg Val Asn Tyr His Leu.

 1640 1650 1660 1670 1680
 * * * * *
 GTC AAT GTG AAG GGT GAA AAC ATC ACC AAT GCC CCT GAA CTG CAG CCG AAT
 CAG TTA CAC TTC CCA CTT TTG TAG TGG TTA CCG GGA CTT GAC GTC GGC TTA
 Val Asn Val Lys Gly Glu Asn Ile Thr Asn Ala Pro Glu Leu Gln Pro Asn.

 1690 1700 1710 1720 1730
 * * * * *
 GCT GTC ACT TGG GGC ATC TTC CCT GGG CGA GAG ATC ATC CAG CCC ACC GTA
 CGA CAG TGA ACC CCG TAG AAG GGA CCC GCT CTC TAG TAG GTC GGG TGG CAT
 Ala Val Thr Trp Gly Ile Phe Pro Gly Arg Glu Ile Ile Gln Pro Thr Val.

 1740 1750 1760 1770 1780
 * * * * *
 GTG GAT CCC GTC AGC TTC ATG TTC TGG AAG GAC GAG GCC TTT GCC CTG TGG
 CAC CTA GGG CAG TCG AAG TAC AAG ACC TTC CTG CTC CCG AAA CCG GAC ACC
 Val Asp Pro Val Ser Phe Met Phe Trp Lys Asp Glu Ala Phe Ala Leu Trp.

FIG. 1E

1790	1800	1810	1820	1830
ATG GAG CGG TGG GGA AAG CTG TAT GAG GAG GAG TCC CCG TCC CGC ACC ATC				
TAA CTC GCC ACC CCT TTC GAC ATA CTC CTC CTC AGG GGC AGG GCG TGG TAG				
Ile Glu Arg Trp Gly Lys Leu Tyr Glu Glu Glu Ser Pro Ser Arg Thr Ile				

1840	1850	1860	1870	1880
ATC CAG TAC ATC CAC GAC AAC TAC TTC CTG GTC AAC CTG GTG GAC AAT GAC				
TAG GTC ATG TAG GTG CTG TTG ATG AAG GAC CAG TTG GAC CAC CTG TTA CTG				
Ile Gln Tyr Ile His Asp Asn Tyr Phe Leu Val Asn Leu Val Asp Asn Asp				

1890	1900	1910	1920	1930
TTC CCA CTG GAC AAC TGC CTC TGG CAG GTG GTG GAA GAC ACA TTG GAG CTT				
AAG GGT GAC CTG TTG ACG GAG ACC GTC CAC CAC CTT CTG TGT AAC CTC GAA				
Phe Pro Leu Asp Asn Cys Leu Trp Gln Val Val Glu Asp Thr Leu Glu Leu				

1940	1950	1960	1970	1980	1990
CTC AAC AGG CCC ACC CAG AAT GCG AGA GAA ACG GAG GGT CCA TGACCTGCG					
GAG TTG TCC GGG TGG GTC TTA CGC TCT CTT TGC CTC CGA GGT ACTGGGAAGC					
Leu Asn Arg Pro Thr Gln Asn Ala Arg Glu Thr Glu Ala Pro					

2000	2010	2020	2030	2040	2050
TCTTGAGGCT CTGCTTGGG GCACTCTCTG TCCCGCTTC CTCTCCACA GTGCTGCTTC					
AGGACTGCGG GAGCAACCT CGGTGAGGAC AGGGCGGAAG GAGGAGGTGT CACGACGAAG					

2060	2070	2080	2090	2100	2110
TCTTGGGAAC TCACTCTCC TTGCTGTCTC TCCCACTCCG GCTTCACTC CCACTCTGA					
AGAACCCTTG AGGTGAGAGG AAGCAAGAG AGGGTGGGG CCGAGGTGAG GGGGTGGACT					

2120	2130	2140	2150	2160	2170
CAATGGCAGC TAGACTGGAG TGAGGCTTCC AGGCTCTTCC TGGACCTGAG TCGSCCCAC					
GTACCGCTCG ATCTGACTC ACTCCGAAGG TCTGAGAAGG AACTGACTC AGCCGGGGTG					

2180	2190	2200	2210	2220
ATGGGAACCT AGTACTCTCT GCTCTAAAA AAAAAAAAAA AAAGGAATTC				
TACCTTTGGA TCATGAGAGA CGAGATTTT TTTTITTTTT TTTCTTAAG				

FIG 1F

AMVNE ARGNS SLNPC LEGSA SSGSE SSKDS SRCST PGLDP ERHER LREKM RRRLE S--GD~~KV~~ FSLEF
 ms FFHas qrdal nqsla evqGqin vSFEF
 ms FFHan qREal nqsla evqGqin vSFEF
 ms iRdLy haraspf iSLEF
 mthfr
 ecometf
 stymetf
 ysRADI

100
 FPPRT AEGAV NLISR FDRMA AGCPL YIDVT WHPAG DPGSD KEISS MMJAS TAVNY CGLET ILHMT
 FPPRT sEmeq tLwns iDRIs sIKPk fvsVT y--ga nsGer drThs i-ikg ik-dr tGLEa oPHIT
 FPPRT sEmeq tLwns iDRIs sIKPk fvsVT y--ga nsGer drThs v-ikg ik-er tGLEa oPHIT
 FPPkT eLGtr NLmer mHRMt ALDPL fItVT y--ga -gGtt aEktI t-IAS lAqqt Inipv cmHIT
 mthfr
 ecometf
 stymetf
 ysRADI

*
 CCRQR LEEIT GHLHK AKQLG LKNIM ALRGD -PIGDQ WEEEE GGFNY AVGLV KHIRS EFGDY FDICV
 CIdat pdElr tiard ywnng irhIv ALRGD lPpGsg kPE-- ---mY AsdLV tIk- EvaD- FDIsV
 CIdat rdElr tiard ywnng irhIv ALRGD lPpGsg kPE-- ---mY AadLV gIk- EvaD- FDIsV
 CtnTe kaild daldr cynaG irNI l ALRGn lPIGvv Wlvsg snrll nmrLf>
 200-
 AGYPK GHPEA GSFEA DLKHL KEKVS AGADE IITQL FFEAD IEFRE VKACT DMGIT CPIVP GIPPI
 AaYPE vHPEA kSaQA DLInL KrKvD AGAnr aITQF FFdve syl RF rdrCv saGId velIP GILPv
 AaYPE vHPEA kSaQA DLInL KrKvD AGAnr aITQF FFdve syl RF rdrCv saGId velIP GILPv
 mthfr
 ecometf
 stymetf

300-
 QGYHS LRQLV KLSKL EVPQE IKDVI EPIKD NDAAI RN-YGI ELAVS LCQEL LASGL VPGLH FYTLN
 snfkq akkfa dmtnv rIPaw maqmf dgl-D dDAet RklvGa niAmd mvkiL sreg- VkdFH FYTLN
 snfkq akkfa dmtnv rIPsw mslmf Egl-D nDAet RklvGa niAmd mvkiL sreg- VkdFH FYTLN
 mthfr
 ecometf
 stymetf

R-EMAT TEVLK RLGMW TEDPR RPLPW ALSAH PKRRE EDVRP IFWAS RPKSY IYRID EWDEE PNGRW
 RoEMsy a-ich tLGvr pgl>
 RoEMsy a-ich tLGvr pgl>
 400-
 GNSSS PAFGE LKDYI LFYLK SKSPK E mthfr

738-2

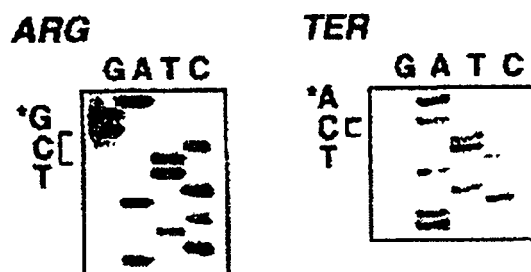


Fig. 3A

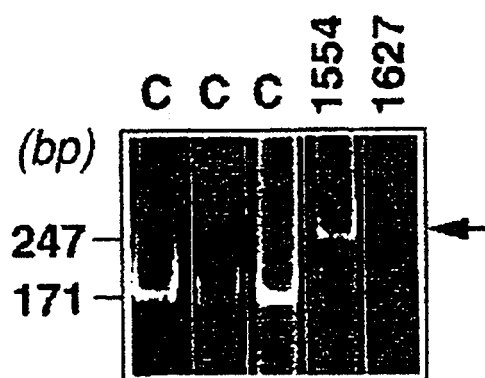


Fig. 3B

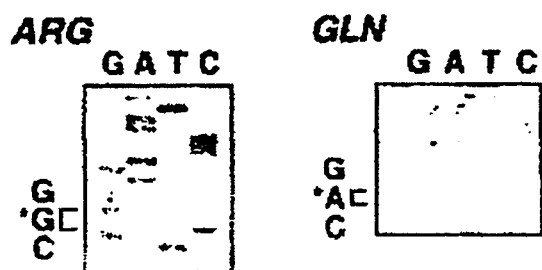


Fig. 4A

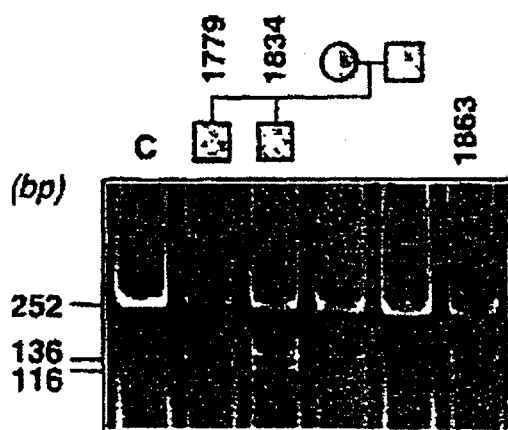
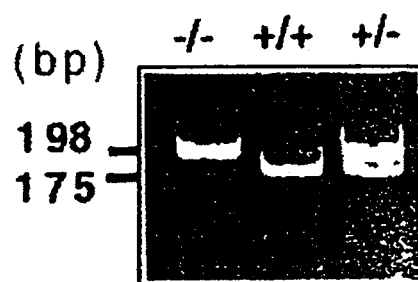
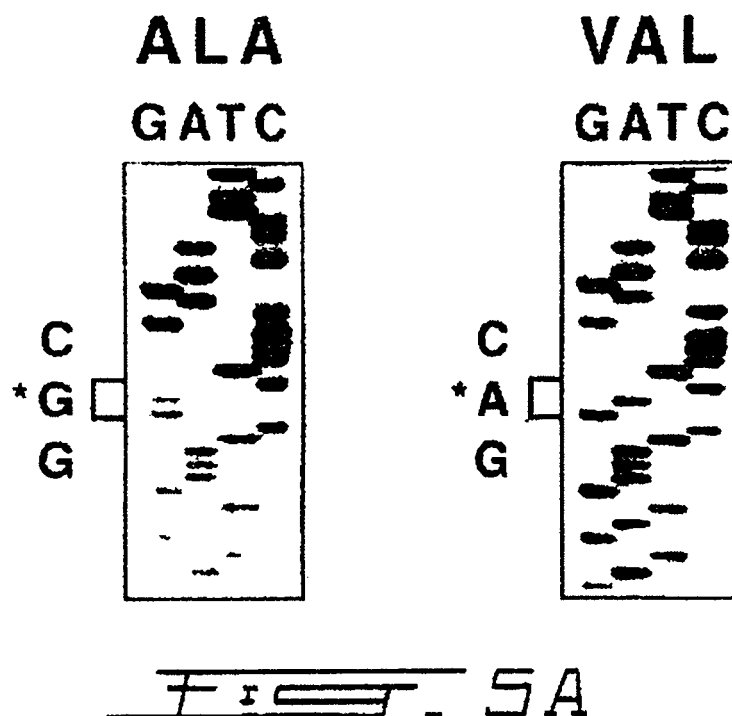


Fig. 4B



AAT	TCC	GGA	GCC	ATG	GTG	AAC	GAA	GCC	AGA	GGA	AAC	AGC	AGC	CTC	AAC	CCC	TGC	TTG	GAG	60
				Met	Val	Asn	Glu	Ala	Arg	Gly	Asn	Ser	Ser	Leu	Asn	Pro	Cys	Leu	Glu	16
GGC	AGT	GCC	AGE	AGT	GGC	AGT	GAG	AGC	TCC	AAA	GAT	AGT	TGC	AGA	TGT	TCC	ACC	CCG	GGC	120
Gly	Ser	Ala	Ser	Ser	Gly	Ser	Glu	Ser	Ser	Lys	Asp	Ser	Ser	Arg	Cys	Ser	Thr	Pro	Gly	36
CTG	GAC	CTT	GAG	CGG	CAT	GAG	AGA	CTC	CGG	GAG	AAG	ATG	AGG	CGG	CGA	TTG	GAA	TCT	GGT	180
Leu	Asp	Pro	Glu	Arg	His	Glu	Arg	Leu	Arg	Glu	Lys	Met	Arg	Arg	Arg	Leu	Glu	Ser	Gly	56
GAC	AAG	TGG	TTC	TCC	CTG	GAA	TTC	TTC	CTT	CTT	CGA	AET	GCT	GAG	GGA	GCT	GTC	AAT	CTC	240
Asp	Lys	Trp	Phe	Ser	Leu	Glu	Phe	Phe	Pro	Pro	Arg	Thr	Ala	Glu	Gly	Ala	Val	Asn	Leu	76
ATC	TCA	AGG	TTT	GAC	CGG	ATG	GCA	GCA	GGT	GGC	CCC	CTC	TAC	ATA	GAC	GTG	ACC	TGG	CAC	300
Ile	Ser	Arg	Phe	Asp	Arg	Met	Ala	Ala	Gly	Gly	Pro	Leu	Tyr	Ile	Asp	Val	Thr	Trp	His	96
CCA	GCA	GGT	GAC	CTT	GGC	TCA	GAC	AAG	GAG	ACC	TCC	TCC	ATG	ATG	ATC	GCC	AGC	ACC	GCC	360
Pro	Ala	Gly	Asp	Pro	Gly	Ser	Asp	Lys	Glu	Thr	Ser	Ser	Met	Met	Ile	Ala	Ser	Thr	Ala	116
GTG	AAC	TAC	TGT	GGC	CTG	GAG	ACC	ATC	CTG	CAC	ATG	ACC	TGC	TGC	CGT	CAG	CGC	CTG	GAG	420
Val	Asn	Tyr	Cys	Gly	Leu	Glu	Thr	Ile	Leu	His	Met	Thr	Cys	Cys	Arg	Glu	Arg	Leu	Glu	136
GAG	ATC	ATG	GGC	CAT	CTG	CAC	AAA	GCT	AAG	CAG	CTG	GGC	CTG	AAG	AAC	ATC	ATG	GTC	CTG	480
Glu	Ile	Thr	Gly	His	Leu	His	Lys	Ala	Lys	Glu	Leu	Gly	Leu	Lys	Asn	Ile	Met	Ala	Leu	156
CGG	GGA	GAC	CCA	ATA	GGT	GAC	CAG	TGG	GAA	GAG	GAG	GAG	GGA	GGC	TTC	AAC	TAC	GCA	GTG	540
Arg	Gly	Asp	Pro	Ile	Gly	Asp	Glu	Trp	Glu	Glu	Glu	Glu	Gly	Gly	Phe	Asn	Tyr	Ala	Val	176
GAC	CTG	GTG	AAG	CAC	ATC	CGA	AGT	GAG	TTT	GGT	GAC	TAC	TTT	GAC	ATC	TGT	GTG	GCA	GGT	600
Asp	Leu	Val	Lys	His	Ile	Arg	Ser	Glu	Phe	Gly	Asp	Tyr	Phe	Asp	Ile	Cys	Val	Ala	Gly	196
TAC	CCC	AAA	GGC	CAC	CCC	GAA	GCA	GGG	AGC	TTT	GAG	GCT	GAC	CTG	AAG	CAC	TTG	AAG	GAG	660
Tyr	Pro	Lys	Gly	His	Pro	Glu	Ala	Gly	Ser	Phe	Glu	Ala	Asp	Leu	Lys	His	Leu	Lys	Glu	216
AAG	GTG	TCT	GCG	GGA	GCC	GAT	TTC	ATC	ATC	ACG	CAG	CTT	TTC	TTT	GAG	GCT	GAC	ACA	TTC	720
Lys	Val	Ser	Ala	Gly	Ala	Asp	Phe	Ile	Ile	Thr	Glu	Leu	Phe	Phe	Glu	Ala	Asp	Thr	Phe	236

ES GA

```

      *      *      *      *      *
TTC CGC TTT GTG AAG GCA TGC ACC GAC ATG GGC ACC ACT TGC CCC ATC GTC CCC GGG ATC 780
Phe Arg Phe Val Lys Ala Cys Thr Asp Met Gly Ile Thr Cys Pro Ile Val Pro Gly Ile 256

      *      *      *      *      *
TTT CCC ATC CAG GGC TAC CAC TCC CTT CGG CAG CTT GTG AAG CTG TCC AAG CTG GAG GTG 840
Phe Pro Ile Glu Gly Tyr His Ser Leu Arg Gln Leu Val Lys Leu Ser Lys Leu Glu Val 276

      *      *      *      *      *
CCA CAG GAG ATC AAG GAC GTG ATT GAG CCA ATC AAA GAC AAC GAT GGT GGC ATC CGC AAC 900
Pro Gln Glu Ile Lys Asp Val Ile Glu Pro Ile Lys Asp Asn Asp Ala Ala Ile Arg Asn 296

      *      *      *      *      *
TAT GGC ATC GAG CTG GCC GTG AGC CTG TGC CAG GAG CTT CTG GCC AGT GGC TTG GTG CCA 960
Tyr Gly Ile Glu Leu Ala Val Ser Leu Cys Gln Glu Leu Leu Ala Ser Gly Leu Val Pro 316

      *      *      *      *      *
GGC CTC CAC TTC TAC ACC CTC AAC CGC GAG ATG GCT ACC ACA GAG GTG CTG AAG CGC CTG 1020
Gly Leu His Phe Tyr Thr Leu Asn Arg Gln Met Ala Thr Thr Glu Val Leu Lys Arg Leu 336

      *      *      *      *      *
GGG ATG TGG ACT GAG GAC CCC AGG CGT CCC CTA CCC TGG GCT CTC AGT GCC CAC CCC AAG 1080
Gly Met Trp Thr Glu Asp Pro Arg Arg Pro Leu Pro Trp Ala Leu Ser Ala His Pro Lys 356

      *      *      *      *      *
CGC CGA GAG GAA GAT GTA CGT CCC ATC TTC TGG GCC TCC AGA CCA AAG AGT TAC ATC TAC 1140
Arg Arg Glu Gln Asp Val Arg Pro Ile Phe Trp Ala Ser Arg Pro Lys Ser Tyr Ile Tyr 376

      *      *      *      *      *
CGT ACC CAG GAG TGG GAC GAG TTC CTT AAC GGC CGC TGG GGC AAT TCC TCT TCC CTT GCC 1200
Arg Thr Gln Glu Trp Asp Glu Phe Pro Asn Gly Arg Trp Gly Asn Ser Ser Ser Pro Ala 396

      *      *      *      *      *
TTT GGG GAG CTG AAG GAC TAC TAC CTC TTC TAC CTG AAG AGC AAG TCC CCC AAG GAG GAG 1260
Phe Gly Glu Leu Lys Asp Tyr Tyr Leu Phe Tyr Leu Lys Ser Lys Ser Pro Lys Glu Glu 416

      *      *      *      *      *
CTG CTG AAG ATG TGG GGG GAG GAG CTG ACC AGT GAA GCA AGT GTC TTT GAA GTC TTT GTT 1320
Leu Leu Lys Met Trp Gly Glu Glu Leu Thr Ser Gln Ala Ser Val Phe Gln Val Phe Val 436

      *      *      *      *      *
CTT TAC CTC TCG GGA GAA CCA AAC CGG AAT GGT CAC AAA GTG ACT TGC CTG CCC TGG AAC 1380
Leu Tyr Leu Ser Gly Glu Pro Asn Arg Asn Gly His Lys Val Thr Cys Leu Pro Trp Asn 456

      *      *      *      *      *
GAT GAG CCC CTG GCG GGT GAG ACC AGE CTG CTG AAG GAG GAG CTG CTG CGG GTG AAC CGC 1440
Asp Glu Pro Leu Ala Ala Glu Thr Ser Leu Leu Lys Gln Glu Leu Leu Arg Val Asn Arg 476

```

FIG. 6B

```

      *           *           *           *           *
CAG GGC ATC CTC ACC ATC AAC TCA CAG CCC AAC ATC AAC GGG AAG CCG TCC TCC GAC CCC 1500
Gln Gly Ile Leu Thr Ile Asn Ser Gln Pro Asn Ile Asn Gly Lys Pro Ser Ser Asp Pro 496

      *           *           *           *           *
ATC GTG GGC TGG GGC CCC AGC GGG GGC TAT GTC TTC CAG AAG GGC TAC TTA GAG TTT TTC 1560
Ile Val Gly Trp Gly Pro Ser Gly Gly Tyr Val Phe Gln Lys Ala Tyr Leu Glu Phe Phe 516

      *           *           *           *           *
ACT TCC CGC GAG ACA GCG GAA GCA CTT CTG CAA GTG CTG AAG AAG TAC GAG CTC CGG GTT 1620
Thr Ser Arg Glu Thr Ala Glu Ala Leu Leu Gln Val Leu Lys Lys Tyr Glu Leu Arg Val 536

      *           *           *           *           *
AAT TAC CAC CTT GTC AAT GTG AAG GGT GAA AAC ATC ACC AAT GCC CTT GAA CTG CAG CCG 1680
Asn Tyr His Leu Val Asn Val Lys Gly Glu Asn Ile Thr Asn Ala Pro Glu Leu Gln Pro 556

      *           *           *           *           *
AAT GCT GTC ACT TGG GGC ATC TTC CTT GGG CGA GAG ATC ATC CAG CCC ACC GTA GTG GAT 1740
Asn Ala Val Thr Trp Gly Ile Phe Pro Gly Arg Glu Ile Ile Gln Pro Thr Val Val Asp 576

      *           *           *           *           *
CCC GTC AGC TTC ATG TTC TGG AAG GAC GAG GCC TTT GGC CTG TGG ATT GAG CCG TGG GGA 1800
Pro Val Ser Phe Met Phe Trp Lys Asp Glu Ala Phe Ala Leu Trp Ile Glu Arg Trp Gly 596

      *           *           *           *           *
AAG CTG TAT GAG GAG GAG TCC CTG TCC CGC ACC ATC ATC CAG TAC ATC CAC GAC AAC TAC 1860
Lys Leu Tyr Gln Glu Glu Ser Pro Ser Arg Thr Ile Ile Gln Tyr Ile His Asp Asn Tyr 616

      *           *           *           *           *
TTC CTG GTC AAC CTG GTG GAC AAT GAC TTC CCA CTG GAC AAC TGC CTC TGG CAG GTG GTG 1920
Phe Leu Val Asn Leu Val Asp Asn Asp Phe Pro Leu Asp Asn Cys Leu Trp Gln Val Val 636

      *           *           *           *           *
GAA GAC ACA TTG GAG CTT CTC AAC AGG CCC ACC CAG AAT GCG AGA GAA ACG GAG GCT CCA 1980
Glu Asp Thr Leu Glu Leu Leu Asn Arg Pro Thr Gln Asn Ala Arg Glu Thr Glu Ala Pro 656

      *           *           *           *           *
TGA CCC TGC GTC CTG ACG CCC TGC GTT GGA GCC ACT CCT GTC CCG CCT TCC TCC TCC ACA 2040
End

      *           *           *           *           *
GTG CTG CTT CTC TTG GGA ACT CCA CTC TCC TTC GTG TCT CTC CCA CCC CCG CCT CCA CTC 2100

      *           *           *           *           *
CCC CAC CTG ACA ATG GCA GCT AGA CTG GAG TGA GGC TTC CAG GGT CTT CCT GGA CCT GAG 2160

      *           *           *           *           *
TCG GCC CCA CAT GGG AAC CTA GTA CTC TCT GGT CTA AAA AAA AAA AAA AAA AAG GAA TT 2220

```

FIG. 6C

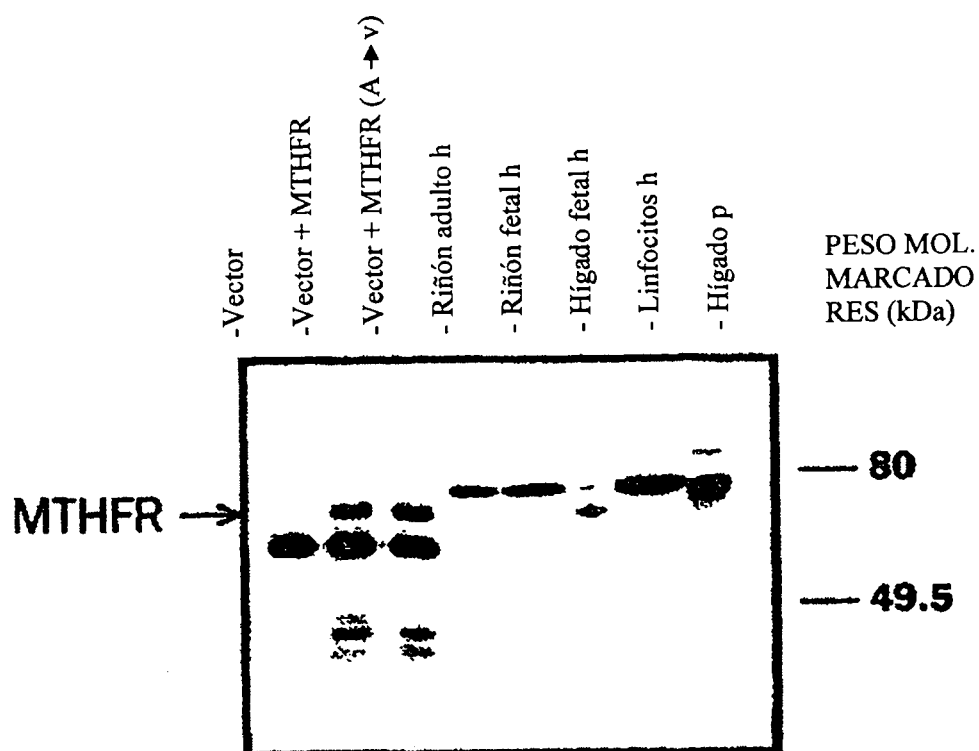


Fig. 7A

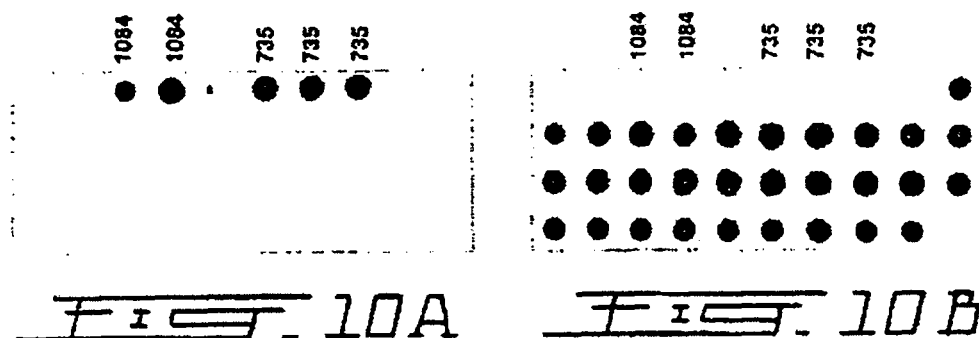


Fig. 10A

Fig. 10B

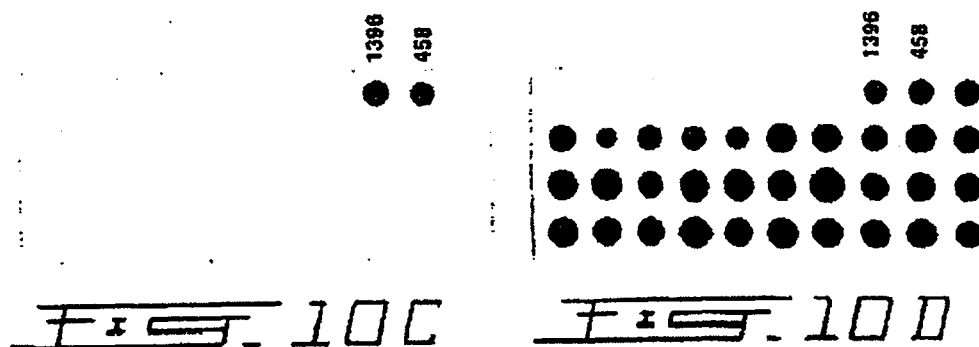


Fig. 10C

Fig. 10D

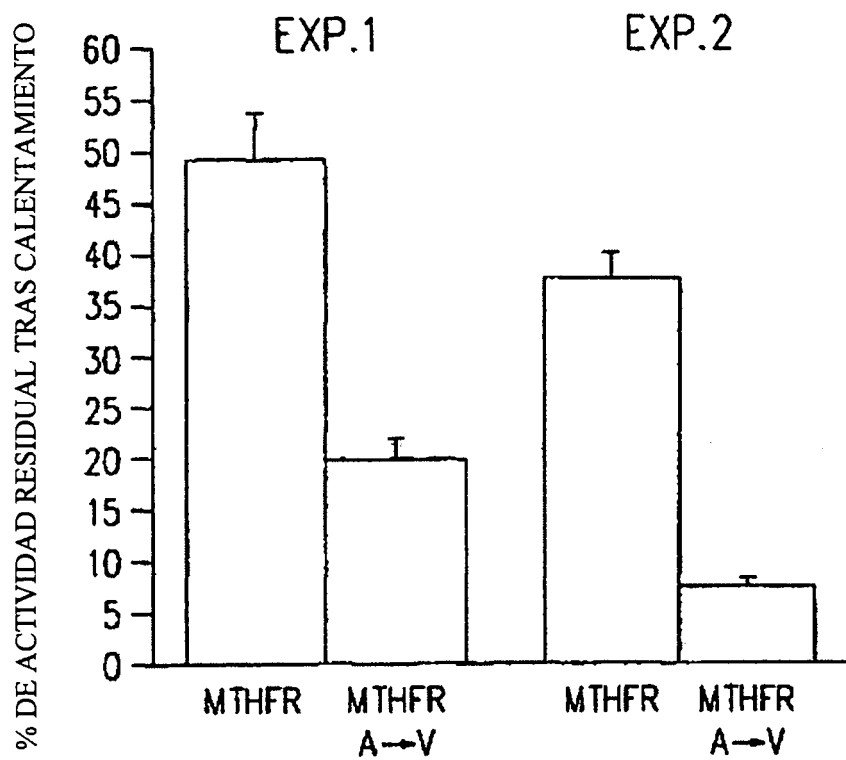
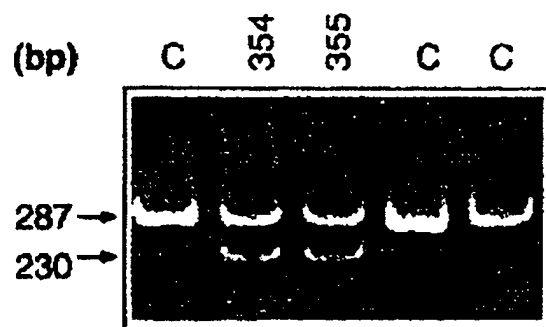


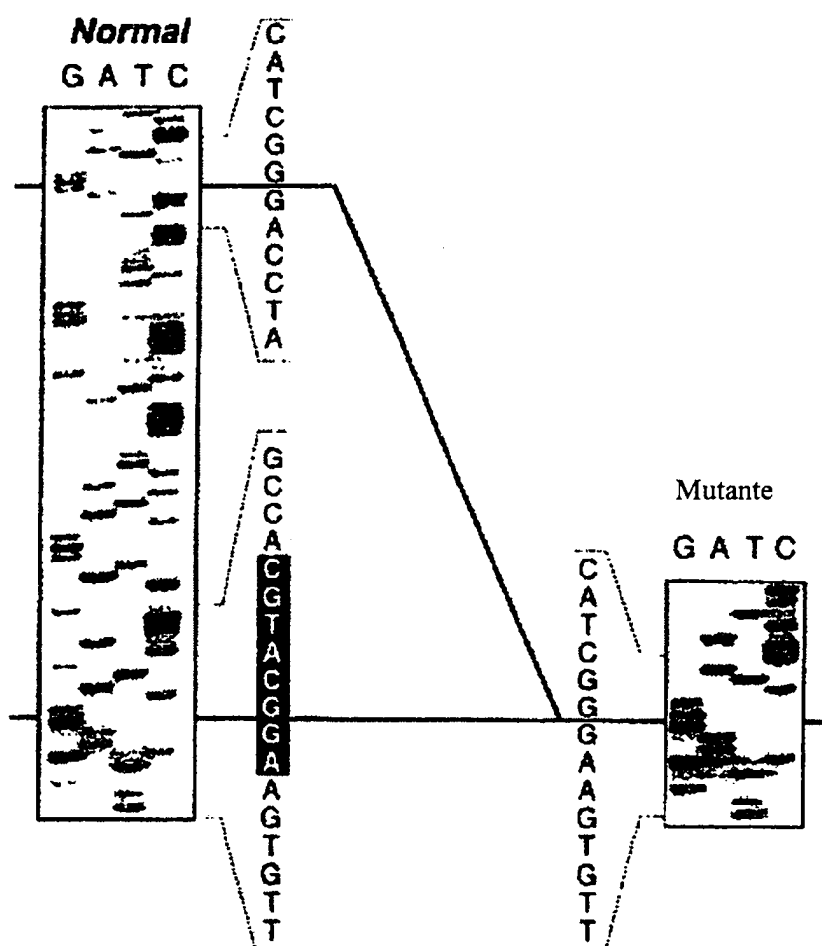
Fig. 7 B

MTHFR:	KHLKEKVSAGADFIIITQLFFEADTFFR
DHFR:	GHLKLFVT----R-IMQD-FESDTFFP

Fig. 11



FILE BA



Fraser BB

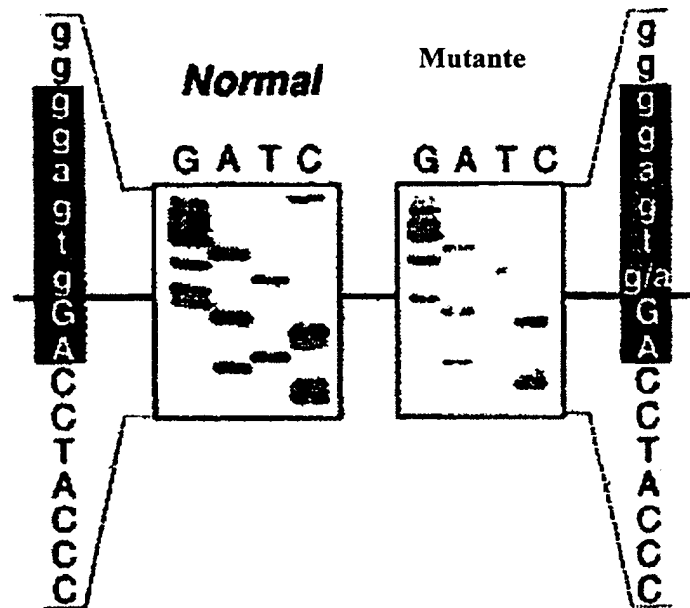


Fig. 8C

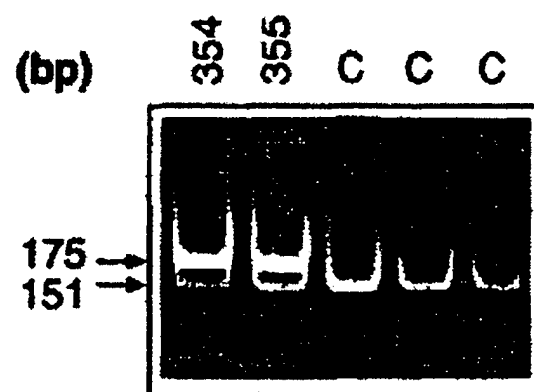
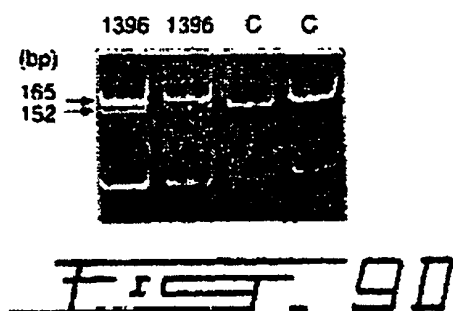
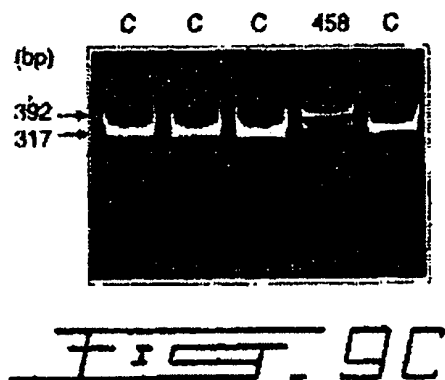
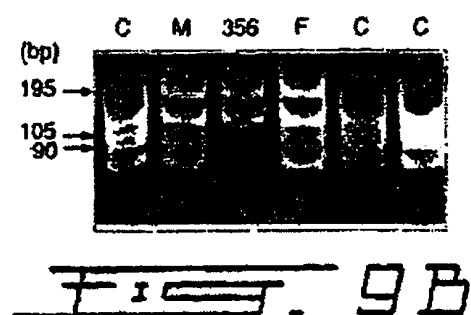
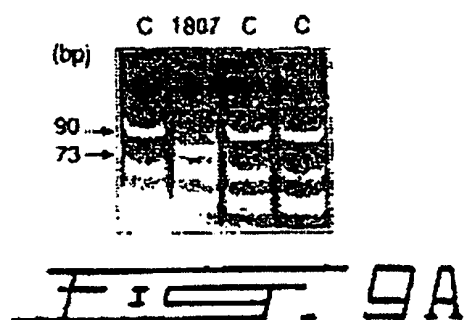


Fig. 8D



ES 2 271 946 T3

LISTA DE SECUENCIAS

(1) INFORMACIÓN GENERAL:

(i) SOLICITANTE:

- (A) NOMBRE: MCGILL UNIVERSITY
- (B) CALLE: 845 Sherbrooke Street West
- (C) CIUDAD: Montreal
- (D) ESTADO: Quebec
- (E) PAÍS: Canadá
- (F) CÓDIGO POSTAL (ZIP): H3A 1B1
- (G) TELÉFONO: 514-398-4200
- (H) TELEFAX: 514-398-8479

- (A) NOMBRE: ROZEN, Rima
- (B) CALLE: 9 Fairfield Cr.
- (C) CIUDAD: Montreal West
- (D) ESTADO: Quebec
- (E) PAÍS: Canadá
- (F) CÓDIGO POSTAL (ZIP): H4X 1R5

- (A) NOMBRE: GOYETTE, Philippe
- (B) CALLE: 4368 Harvard
- (C) CIUDAD: Montreal
- (D) ESTADO: Quebec
- (E) PAÍS: Canadá
- (F) CÓDIGO POSTAL (ZIP): H4A 2X1

(ii) TÍTULO DE LA INVENCION: cDNA PARA TILNOTETRAHIDROFOLATO REDUCTASA HUMANA

(iii) NÚMERO DE SECUENCIAS: 14

(iv) FORMA LEGIBLE EN ORDENADOR:

- (A) TIPO DE SOPORTE: Disco flexible
- (B) ORDENADOR: Compatible con PC IBM
- (C) SISTEMA OPERATIVO: PC-DOS/MS-DOS
- (D) SOPORTE LÓGICO INFORMÁTICO: PatentIn Lanzamiento N° 1.0, Versión N° 1.30 (EPO)

(v) DATOS DE LA SOLICITUD ACTUAL:

NÚMERO DE SOLICITUD: WO PCT/CA95/00314

(vi) DATOS DE LA SOLICITUD ANTERIOR:

- (A) NÚMERO DE SOLICITUD: GB 9410620.0
- (B) FECHA DE PRESENTACIÓN: 26 MAYO 1994

(2) INFORMACIÓN PARA LA ID SEC N°: 1:

(i) CARACTERÍSTICAS DE LA SECUENCIA

- (A) LONGITUD: 2220 pares de bases
- (B) TIPO: ácido nucleico
- (C) ORIENTACIÓN: doble

ES 2 271 946 T3

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICA: NO

(iv) ANTISENTIDO: NO

(xi) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) SITUACIÓN: 1...1980

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC N°: 1:

15	AAT TCC GGA GCC ATG GTG AAC GAA GCC AGA GGA AAC AGC AGC CTC AAC	48
	Asn Ser Gly Ala Met Val Asn Glu Ala Arg Gly Asn Ser Ser Leu Asn	
	1 5 10 15	
20	CCC TGC TTG GAG GGC AGT GCC AGC AGT GGC AGT GAG AGC TCC AAA GAT	96
	Pro Cys Leu Glu Gly Ser Ala Ser Ser Gly Ser Glu Ser Ser Lys Asp	
	20 25 30	
25	AGT TCG AGA TGT TCC ACC CCG GGC CTG GAC CCT GAG CGG CAT GAG AGA	144
	Ser Ser Arg Cys Ser Thr Pro Gly Leu Asp Pro Glu Arg His Glu Arg	
	35 40 45	
30	CTC CGG GAG AAG ATG AGG CGG CGA TTG GAA TCT GGT GAC AAG TGG TTC	192
	Leu Arg Glu Lys Met Arg Arg Arg Leu Glu Ser Gly Asp Lys Trp Phe	
	50 55 60	
35	TCC CTG GAA TTC TTC CCT CCT CGA ACT GCT GAG GGA GCT GTC AAT CTC	240
	Ser Leu Glu Phe Phe Pro Pro Arg Thr Ala Glu Gly Ala Val Asn Leu	
	65 70 75 80	
40	ATC TCA AGG TTT GAC CGG ATG GCA GCA GGT GGC CCC CTC TAC ATA GAC	288
	Ile Ser Arg Phe Asp Arg Met Ala Ala Gly Gly Pro Leu Tyr Ile Asp	
	85 90 95	
45	GTG ACC TGG CAC CCA GCA GGT GAC CCT GGC TCA GAC AAG GAG ACC TCC	336
	Val Thr Trp His Pro Ala Gly Asp Pro Gly Ser Asp Lys Glu Thr Ser	
	100 105 110	
50	TCC ATG ATG ATC GCC AGC ACC GCC GTG AAC TAC TGT GGC CTG GAG ACC	384
	Ser Met Met Ile Ala Ser Thr Ala Val Asn Tyr Cys Gly Leu Glu Thr	
	115 120 125	
55	ATC CTG CAC ATG ACC TGC TGC CGT CAG CGC CTG GAG GAG ATC ACG GGC	432
	Ile Leu His Met Thr Cys Cys Arg Gln Arg Leu Glu Glu Ile Thr Gly	
	130 135 140	
60	CAT CTG CAC AAA GCT AAG CAG CTG GGC CTG AAG AAC ATC ATG GCG CTG	480
	His Leu His Lys Ala Lys Gln Leu Gly Leu Lys Asn Ile Met Ala Leu	
	145 150 155 160	
65	CGG GGA GAC CCA ATA GGT GAC CAG TGG GAA GAG GAG GAG GGA GGC TTC	528
	Arg Gly Asp Pro Ile Gly Asp Gln Trp Glu Glu Glu Glu Gly Gly Phe	
	165 170 175	
70	AAC TAC GCA GTG GAC CTG GTG AAG CAC ATC CGA AGT GAG TTT GGT GAC	576
	Asn Tyr Ala Val Asp Leu Val Lys His Ile Arg Ser Glu Phe Gly Asp	
	180 185 190	

ES 2 271 946 T3

5	TAC TTT GAC ATC TGT GTG GCA GGT TAC CCC AAA GGC CAC CCC GAA GCA Tyr Phe Asp Ile Cys Val Ala Gly Tyr Pro Lys Gly His Pro Glu Ala 195 200 205	624
10	GGG AGC TTT GAG GCT GAC CTG AAG CAC TTG AAG GAG AAG GTG TCT GCG Gly Ser Phe Glu Ala Asp Leu Lys His Leu Lys Glu Lys Val Ser Ala 210 215 220	672
15	GGA GCC GAT TTC ATC ATC ACG CAG CTT TTC TTT GAG GCT GAC ACA TTC Gly Ala Asp Phe Ile Ile Thr Gln Leu Phe Phe Glu Ala Asp Thr Phe 225 230 235 240	720
20	TTC CGC TTT GTG AAG GCA TGC ACC GAC ATG GGC ATC ACT TGC CCC ATC Phe Arg Phe Val Lys Ala Cys Thr Asp Met Gly Ile Thr Cys Pro Ile 245 250 255	768
25	GTC CCC GGG ATC TTT CCC ATC CAG GGC TAC CAC TCC CTT CGG CAG CTT Val Pro Gly Ile Phe Pro Ile Gln Gly Tyr His Ser Leu Arg Gln Leu 260 265 270	816
30	GTG AAG CTG TCC AAG CTG GAG GTG CCA CAG GAG ATC AAG GAC GTG ATT Val Lys Leu Ser Lys Leu Glu Val Pro Gln Glu Ile Lys Asp Val Ile 275 280 285	864
35	GAG CCA ATC AAA GAC AAC GAT GCT GCC ATC CGC AAC TAT GGC ATC GAG Glu Pro Ile Lys Asp Asn Asp Ala Ala Ile Arg Asn Tyr Gly Ile Glu 290 295 300	912
40	CTG GCC GTG AGC CTG TGC CAG GAG CTT CTG GCC AGT GGC TTG GTG CCA Leu Ala Val Ser Leu Cys Gln Glu Leu Leu Ala Ser Gly Leu Val Pro 305 310 315 320	960
45	GGC CTC CAC TTC TAC ACC CTC AAC CGC GAG ATG GCT ACC ACA GAG GTG Gly Leu His Phe Tyr Thr Leu Asn Arg Glu Met Ala Thr Thr Glu Val 325 330 335	1008
50	CTG AAG CGC CTG GGG ATG TGG ACT GAG GAC CCC AGG CGT CCC CTA CCC Leu Lys Arg Leu Gly Met Trp Thr Glu Asp Pro Arg Arg Pro Leu Pro 340 345 350	1056
55	TGG GCT CTC AGT GCC CAC CCC AAG CGC CGA GAG GAA GAT GTA CGT CCC Trp Ala Leu Ser Ala His Pro Lys Arg Arg Glu Glu Asp Val Arg Pro 355 360 365	1104
60	ATC TTC TGG GCC TCC AGA CCA AAG AGT TAC ATC TAC CGT ACC CAG GAG Ile Phe Trp Ala Ser Arg Pro Lys Ser Tyr Ile Tyr Arg Thr Gln Glu 370 375 380	1152
65	TGG GAC GAG TTC CCT AAC GGC CGC TGG GGC AAT TCC TCT TCC CCT GCC Trp Asp Glu Phe Pro Asn Gly Arg Trp Gly Asn Ser Ser Ser Pro Ala 385 390 395 400	1200
	TTT GGG GAG CTG AAG GAC TAC TAC CTC TTC TAC CTG AAG AGC AAG TCC Phe Gly Glu Leu Lys Asp Tyr Tyr Leu Phe Tyr Leu Lys Ser Lys Ser 405 410 415	1248
	CCC AAG GAG GAG CTG CTG AAG ATG TGG GGG GAG GAG CTG ACC AGT GAA Pro Lys Glu Glu Leu Leu Lys Met Trp Gly Glu Glu Leu Thr Ser Glu 420 425 430	1296

ES 2 271 946 T3

5	GCA AGT GTC TTT GAA GTC TTT GTT CTT TAC CTC TCG GGA GAA CCA AAC Ala Ser Val Phe Glu Val Phe Val Leu Tyr Leu Ser Gly Glu Pro Asn 435 440 445	1344
10	CGG AAT GGT CAC AAA GTG ACT TGC CTG CCC TGG AAC GAT GAG CCC CTG Arg Asn Gly His Lys Val Thr Cys Leu Pro Trp Asn Asp Glu Pro Leu 450 455 460	1392
15	GCG OCT GAG ACC AGC CTG CTG AAG GAG GAG CTG CTG CGG GTG AAC CGC Ala Ala Glu Thr Ser Leu Leu Lys Glu Glu Leu Leu Arg Val Asn Arg 465 470 475 480	1440
20	CAG GGC ATC CTC ACC ATC AAC TCA CAG CCC AAC ATC AAC GGG AAG CCG Gln Gly Ile Leu Thr Ile Asn Ser Gln Pro Asn Ile Asn Gly Lys Pro 485 490 495	1488
25	TCC TCC GAC CCC ATC GTG GGC TGG GGC CCC AGC GGG GGC TAT GTC TTC Ser Ser Asp Pro Ile Val Gly Trp Gly Pro Ser Gly Gly Tyr Val Phe 500 505 510	1536
30	CAG AAG GCC TAC TTA GAG TTT TTC ACT TCC CGC GAG ACA GCG GAA GCA Gln Lys Ala Tyr Leu Glu Phe Phe Thr Ser Arg Glu Thr Ala Glu Ala 515 520 525	1584
35	CTT CTG CAA GTG CTG AAG AAG TAC GAG CTC CGG GTT AAT TAC CAC CTT Leu Leu Gln Val Leu Lys Lys Tyr Glu Leu Arg Val Asn Tyr His Leu 530 535 540	1632
40	GTC AAT GTG AAG GGT GAA AAC ATC ACC AAT GCC CCT GAA CTG CAG CCG Val Asn Val Lys Gly Glu Asn Ile Thr Asn Ala Pro Glu Leu Gln Pro 545 550 555 560	1680
45	AAT GCT GTC ACT TGG GGC ATC TTC CCT GGG CGA GAG ATC ATC CAG CCC Asn Ala Val Thr Trp Gly Ile Phe Pro Gly Arg Glu Ile Ile Gln Pro 565 570 575	1728
50	ACC GTA GTG GAT CCC GTC AGC TTC ATG TTC TGG AAG GAC GAG GCC TTT Thr Val Val Asp Pro Val Ser Phe Met Phe Trp Lys Asp Glu Ala Phe 580 585 590	1776
55	GCC CTG TGG ATT GAG CGG TGG GGA AAG CTG TAT GAG GAG GAG TCC CCG Ala Leu Trp Ile Glu Arg Trp Gly Lys Leu Tyr Glu Glu Glu Ser Pro 595 600 605	1824
60	TCC CGC ACC ATC ATC CAG TAC ATC CAC GAC AAC TAC TTC CTG GTC AAC Ser Arg Thr Ile Ile Gln Tyr Ile His Asp Asn Tyr Phe Leu Val Asn 610 615 620	1872
65	CTG GTG GAC AAT GAC TTC CCA CTG GAC AAC TGC CTC TGG CAG GTG GTG Leu Val Asp Asn Asp Phe Pro Leu Asp Asn Cys Leu Trp Gln Val Val 625 630 635 640	1920
	GAA GAC ACA TTG GAG CTT CTC AAC AGG CCC ACC CAG AAT GCG ACA GAA Glu Asp Thr Leu Glu Leu Leu Asn Arg Pro Thr Gln Asn Ala Arg Glu 645 650 655	1968

ES 2 271 946 T3

ACG GAG GCT CCA TGACCTGCG TCCTGACGCC CTGCGTTGGA GCCACTCCTG 2020
 Thr Glu Ala Pro
 660

5 TCCCGCCTTC CTCTCCACA GTGCTGCTTC TCTTGGGAC TCCACTCTCC TTCGTGTCTC 2080

TCCACCCCG GCCTCCACTC CCCCACCTGA CAATGGCAGC TAGACTGGAG TGAGGCTTCC 2140

10 AGGCTCTTCC TGGACCTGAG TCGGCCCCAC ATGGGAACCT AGTACTCTCT GCTCTAAAAA 2200

AAAAAAAAAA AAAGGAATTC 2220

(2) INFORMACIÓN PARA LA ID SEC N°: 2:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 660 aminoácidos
- (B) TIPO: aminoácido
- (D) TOPOLOGIA: lineas

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC N°: 2:

Asn	Ser	Gly	Ala	Met	Val	Asn	Glu	Ala	Arg	Gly	Asn	Ser	Ser	Leu	Asn	1	5	10	15
Pro	Cys	Leu	Glu	Gly	Ser	Ala	Ser	Ser	Gly	Ser	Glu	Ser	Ser	Lys	Asp	20	25	30	
Ser	Ser	Arg	Cys	Ser	Thr	Pro	Gly	Leu	Asp	Pro	Glu	Arg	His	Glu	Arg	35	40	45	
Leu	Arg	Glu	Lys	Met	Arg	Arg	Arg	Leu	Glu	Ser	Gly	Asp	Lys	Trp	Phe	50	55	60	
Ser	Leu	Glu	Phe	Phe	Pro	Pro	Arg	Thr	Ala	Glu	Gly	Ala	Val	Asn	Leu	65	70	75	80
Ile	Ser	Arg	Phe	Asp	Arg	Met	Ala	Ala	Gly	Gly	Pro	Leu	Tyr	Ile	Asp	85	90	95	
Val	Thr	Trp	His	Pro	Ala	Gly	Asp	Pro	Gly	Ser	Asp	Lys	Glu	Thr	Ser	100	105	110	
Ser	Met	Met	Ile	Ala	Ser	Thr	Ala	Val	Asn	Tyr	Cys	Gly	Leu	Glu	Thr	115	120	125	
Ile	Leu	His	Met	Thr	Cys	Cys	Arg	Gln	Arg	Leu	Glu	Glu	Ile	Thr	Gly	130	135	140	
His	Leu	His	Lys	Ala	Lys	Gln	Leu	Gly	Leu	Lys	Asn	Ile	Met	Ala	Leu	145	150	155	160
Arg	Gly	Asp	Pro	Ile	Gly	Asp	Gln	Trp	Glu	Glu	Glu	Glu	Gly	Gly	Phe	165	170	175	
Asn	Tyr	Ala	Val	Asp	Leu	Val	Lys	His	Ile	Arg	Ser	Glu	Phe	Gly	Asp	180	185	190	

ES 2 271 946 T3

Tyr Phe Asp Ile Cys Val Ala Gly Tyr Pro Lys Gly His Pro Glu Ala
 195 200 205
 Gly Ser Phe Glu Ala Asp Leu Lys His Leu Lys Glu Lys Val Ser Ala
 210 215 220
 Gly Ala Asp Phe Ile Ile Thr Gln Leu Phe Phe Glu Ala Asp Thr Phe
 225 230 235 240
 Phe Arg Phe Val Lys Ala Cys Thr Asp Met Gly Ile Thr Cys Pro Ile
 245 250 255
 Val Pro Gly Ile Phe Pro Ile Gln Gly Tyr His Ser Leu Arg Gln Leu
 260 265 270
 Val Lys Leu Ser Lys Leu Glu Val Pro Gln Glu Ile Lys Asp Val Ile
 275 280 285
 Glu Pro Ile Lys Asp Asn Asp Ala Ala Ile Arg Asn Tyr Gly Ile Glu
 290 295 300
 Leu Ala Val Ser Leu Cys Gln Glu Leu Leu Ala Ser Gly Leu Val Pro
 305 310 315 320
 Gly Leu His Phe Tyr Thr Leu Asn Arg Glu Met Ala Thr Thr Glu Val
 325 330 335
 Leu Lys Arg Leu Gly Met Trp Thr Glu Asp Pro Arg Arg Pro Leu Pro
 340 345 350
 Trp Ala Leu Ser Ala His Pro Lys Arg Arg Glu Glu Asp Val Arg Pro
 355 360 365
 Ile Phe Trp Ala Ser Arg Pro Lys Ser Tyr Ile Tyr Arg Thr Gln Glu
 370 375 380
 Trp Asp Glu Phe Pro Asn Gly Arg Trp Gly Asn Ser Ser Ser Pro Ala
 385 390 395 400
 Phe Gly Glu Leu Lys Asp Tyr Tyr Leu Phe Tyr Leu Lys Ser Lys Ser
 405 410 415
 Pro Lys Glu Glu Leu Leu Lys Met Trp Gly Glu Glu Leu Thr Ser Glu
 420 425 430
 Ala Ser Val Phe Glu Val Phe Val Leu Tyr Leu Ser Gly Glu Pro Asn
 435 440 445
 Arg Asn Gly His Lys Val Thr Cys Leu Pro Trp Asn Asp Glu Pro Leu
 450 455 460
 Ala Ala Glu Thr Ser Leu Leu Lys Glu Glu Leu Leu Arg Val Asn Arg
 465 470 475 480
 Gln Gly Ile Leu Thr Ile Asn Ser Gln Pro Asn Ile Asn Gly Lys Pro
 485 490 495

Ser Ser Asp Pro Ile Val Gly Trp Gly Pro Ser Gly Gly Tyr Val Phe
 500 505 510
 5 Gln Lys Ala Tyr Leu Glu Phe Phe Thr Ser Arg Glu Thr Ala Glu Ala
 515 520 525
 10 Leu Leu Gln Val Leu Lys Lys Tyr Glu Leu Arg Val Asn Tyr His Leu
 530 535 540
 Val Asn Val Lys Gly Glu Asn Ile Thr Asn Ala Pro Glu Leu Gln Pro
 545 550 555 560
 15 Asn Ala Val Thr Trp Gly Ile Phe Pro Gly Arg Glu Ile Ile Gln Pro
 565 570 575
 20 Thr Val Val Asp Pro Val Ser Phe Met Phe Trp Lys Asp Glu Ala Phe
 580 585 590
 Ala Leu Trp Ile Glu Arg Trp Gly Lys Leu Tyr Glu Glu Glu Ser Pro
 595 600 605
 25 Ser Arg Thr Ile Ile Gln Tyr Ile His Asp Asn Tyr Phe Leu Val Asn
 610 615 620
 30 Leu Val Asp Asn Asp Phe Pro Leu Asp Asn Cys Leu Trp Gln Val Val
 625 630 635 640
 35 Glu Asp Thr Leu Glu Leu Leu Asn Arg Pro Thr Gln Asn Ala Arg Glu
 645 650 655
 Thr Glu Ala Pro
 660

- 40 (2) INFORMACIÓN PARA LA ID SEC N°: 3:
- (i) CARACTERÍSTICAS DE LA SECUENCIA:
 45 (A) LONGITUD: 2219 pares de bases
 (B) TIPO: ácido nucleico
 (C) ORIENTACIÓN: doble
 (D) TOPOLOGÍA: lineal
 50 (ii) TIPO MOLECULA: DNA (genomic)
 (iii) HIPOTÉTICA: NO
 55 (iv) ANTISENTIDO: NO
 (ix) CARACTERÍSTICA:
 (A) NOMBRE/CODIGO: CDS
 (B) SITUACIÓN: 13...1983
 60 (xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC N°: 3:
 65 AATTCGGAG CC ATG GTG AAC GAA GCC AGA GGA AAC AGC AGC CTC AAC 48
 Met Val Asn Glu Ala Arg Gly Asn Ser Ser Leu Asn
 665 670

ES 2 271 946 T3

5	CCC Pro	TGC Cys	TTG Leu	GAG Glu	GGC Gly	AGT Ser	CCC Ala	AGC Ser	AGT Ser	GGC Gly	AGT Ser	GAG Glu	AGC Ser	TCC Ser	AAA Lys	GAT Asp	96
			675					680					685				
	AGT Ser	TCG Ser	AGA Arg	TGT Cys	TCC Ser	ACC Thr	CCG Pro	GGC Gly	CTG Leu	GAC Asp	CCT Pro	GAG Glu	CGG Arg	CAT His	GAG Glu	AGA Arg	144
			690				695					700					
10	CTC Leu	CGG Arg	GAG Glu	AAG Lys	ATG Met	AGG Arg	CGG Arg	CGA Arg	TTG Leu	GAA Glu	TCT Ser	GGT Gly	GAC Asp	AAG Lys	TGG Tyr	TTC Phe	192
			705			710					715					720	
15	TCC Ser	CTG Leu	GAA Glu	TTC Phe	TTC Phe	CCT Pro	CCT Pro	CGA Arg	ACT Thr	GCT Ala	GAG Glu	GGA Gly	GCT Ala	GTC Val	AAT Asn	CTC Leu	240
					725					730					735		
20	ATC Ile	TCA Ser	AGG Arg	TTT Phe	GAC Asp	CGG Arg	ATG Met	GCA Ala	GCA Ala	GGT Gly	GGC Gly	CCC Pro	CTC Leu	TAC Tyr	ATA Ile	GAC Asp	288
				740				745						750			
25	GTG Val	ACC Thr	TGG Trp	CAC His	CCA Pro	GCA Ala	GGT Gly	GAC Asp	CCT Pro	GGC Gly	TCA Ser	GAC Asp	AAG Lys	GAG Glu	ACC Thr	TCC Ser	336
			755					760					765				
30	TCC Ser	ATG Met	ATG Met	ATC Ile	GCC Ala	AGC Ser	ACC Thr	GCC Ala	GTG Val	AAC Asn	TAC Tyr	TGT Cys	GGC Gly	CTG Leu	GAG Glu	ACC Thr	384
			770				775					780					
35	ATC Ile	CTG Leu	CAC His	ATG Met	ACC Thr	TGC Cys	TGC Cys	CGT Arg	CAG Gln	CGC Arg	CTG Leu	GAG Glu	GAG Glu	ATC Ile	ACG Thr	GGC Gly	432
			785			790				795						800	
40	CAT His	CTG Leu	CAC His	AAA Lys	GCT Ala	AAG Lys	CAG Gln	CTG Leu	GGC Gly	CTG Leu	AAG Lys	AAC Asn	ATC Ile	ATG Met	GCG Ala	CTG Leu	480
				805					810						815		
45	CGG Arg	GGA Gly	GAC Asp	CCA Pro	ATA Ile	GGT Gly	GAC Asp	CAG Gln	TGG Trp	GAA Glu	GAG Glu	GAG Glu	GAG Glu	GGA Gly	GGC Gly	TTC Phe	528
				820					825					830			
50	AAC Asn	TAC Tyr	GCA Ala	GTG Val	GAC Asp	CTG Leu	GTG Val	AAG Lys	CAC His	ATC Ile	CGA Arg	AGT Ser	GAG Glu	TTT Phe	GGT Gly	GAC Asp	576
			835					840					845				
55	TAC Tyr	TTT Phe	GAC Asp	ATC Ile	TGT Cys	GTG Val	GCA Ala	GGT Gly	TAC Tyr	CCC Pro	AAA Lys	GGC Gly	CAC His	CCC Pro	GAA Glu	GCA Ala	624
			850				855					860					
60	GGG Gly	AGC Ser	TTT Phe	GAG Glu	GCT Ala	GAC Asp	CTG Leu	AAG Lys	CAC His	TTG Leu	AAG Lys	GAG Glu	AAG Lys	GTG Val	TCT Ser	GCG Ala	672
			865			870				875						880	
65	GGA Gly	GCC Ala	GAT Asp	TTC Phe	ATC Ile	ATC Ile	ACG Thr	CAG Gln	CTT Leu	TTC Phe	TTT Phe	GAG Glu	GCT Ala	GAC Asp	ACA Thr	TTC Phe	720
					885				890						895		
70	TTC Phe	CGC Arg	TTT Phe	GTG Val	AAG Lys	GCA Ala	TGC Cys	ACC Thr	GAC Asp	ATG Met	GGC Gly	ATC Ile	ACT Thr	TGC Cys	CCC Pro	ATC Ile	768
				900					905					910			

ES 2 271 946 T3

5	GTC CCC GGG ATC TTT CCC ATC CAG GGC TAC CAC TCC CTT CGG CAG CTT Val Pro Gly Ile Phe Pro Ile Gln Gly Tyr His Ser Leu Arg Gln Leu 915 920 925	816
10	GTG AAG CTG TCC AAG CTG GAG GTG CCA CAG GAG ATC AAG GAC GTG ATT Val Lys Leu Ser Lys Leu Glu Val Pro Gln Glu Ile Lys Asp Val Ile 930 935 940	864
15	GAG CCA ATC AAA GAC AAC GAT GCT GCC ATC CGC AAC TAT GGC ATC GAG Glu Pro Ile Lys Asp Asn Asp Ala Ala Ile Arg Asn Tyr Gly Ile Glu 945 950 955 960	912
20	CTG GCC GTG AGC CTG TGC CAG GAG CTT CTG GCC AGT GGC TTG GTG CCA Leu Ala Val Ser Leu Cys Gln Glu Leu Leu Ala Ser Gly Leu Val Pro 965 970 975	960
25	GGC CTC CAC TTC TAC ACC CTC AAC CGC GAG ATG GCT ACC ACA GAG GTG Gly Leu His Phe Tyr Thr Leu Asn Arg Glu Met Ala Thr Thr Glu Val 980 985 990	1009
30	CTG AAG CGC CTG GGG ATG TGG ACT GAG GAC CCC AGG CGT CCC CTA CCC Leu Lys Arg Leu Gly Met Trp Thr Glu Asp Pro Arg Arg Pro Leu Pro 995 1000 1005	1056
35	TGG GCT CTC AGT GCC CAC CCC AAG CGC CGA GAG GAA GAT GTA CGT CCC Trp Ala Leu Ser Ala His Pro Lys Arg Arg Glu Glu Asp Val Arg Pro 1010 1015 1020	1104
40	ATC TTC TGG GCC TCC AGA CCA AAG AGT TAC ATC TAC CGT ACC CAG GAG Ile Phe Trp Ala Ser Arg Pro Lys Ser Tyr Ile Tyr Arg Thr Gln Glu 1025 1030 1035 1040	1152
45	TGG GAC GAG TTC CCT AAC GGC CGC TGG GGC AAT TCC TCT TCC CCT GCC Trp Asp Glu Phe Pro Asn Gly Arg Trp Gly Asn Ser Ser Ser Pro Ala 1045 1050 1055	1200
50	TTT GGG GAG CTG AAG GAC TAC TAC CTC TTC TAC CTG AAG AGC AAG TCC Phe Gly Glu Leu Lys Asp Tyr Tyr Leu Phe Tyr Leu Lys Ser Lys Ser 1060 1065 1070	1248
55	CCC AAG GAG GAG CTG CTG AAG ATG TGG GGG GAG GAG CTG ACC AGT GAA Pro Lys Glu Glu Leu Leu Lys Met Trp Gly Glu Glu Leu Thr Ser Glu 1075 1080 1085	1296
60	GCA AGT GTC TTT GAA GTC TTT GTT CTT TAC CTC TCG GGA GAA CCA AAC Ala Ser Val Phe Glu Val Phe Val Leu Tyr Leu Ser Gly Glu Pro Asn 1090 1095 1100	1344
65	CGG AAT GGT CAC AAA GTG ACT TGC CTG CCC TGG AAC GAT GAG CCC CTG Arg Asn Gly His Lys Val Thr Cys Leu Pro Trp Asn Asp Glu Pro Leu 1105 1110 1115 1120	1392
	GCG GCT GAG ACC AGC CTG CTG AAG GAG GAG CTG CTG CGG GTG AAC CGC Ala Ala Glu Thr Ser Leu Leu Lys Glu Glu Leu Leu Arg Val Asn Arg 1125 1130 1135	1440

ES 2 271 946 T3

	CAG GGC ATC CTC ACC ATC AAC TCA CAG CCC AAC ATC AAC GGG AAG CCG	1488
	Gln Gly Ile Leu Thr Ile Asn Ser Gln Pro Asn Ile Asn Gly Lys Pro	
	1140 1145 1150	
5	TCC TCC GAC CCC ATC GTG GGC TGG GGC CCC AGC GGG GGC TAT GTC TTC	1536
	Ser Ser Asp Pro Ile Val Gly Trp Gly Pro Ser Gly Gly Tyr Val Phe	
	1155 1160 1165	
10	CAG AAG GCC TAC TTA GAG TTT TTC ACT TCC CGC GAG ACA GCG GAA GCA	1584
	Gln Lys Ala Tyr Leu Glu Phe Phe Thr Ser Arg Glu Thr Ala Glu Ala	
	1170 1175 1180	
15	CTT CTG CAA GTG CTG AAG AAG TAC GAG CTC CGG GTT AAT TAC CAC CTT	1632
	Leu Leu Gln Val Leu Lys Lys Tyr Glu Leu Arg Val Asn Tyr His Leu	
	1185 1190 1195 1200	
20	GTC AAT GTG AAG GGT GAA AAC ATC ACC AAT GCC CCT GAA CTG CAG CCG	1680
	Val Asn Val Lys Gly Glu Asn Ile Thr Asn Ala Pro Glu Leu Gln Pro	
	1205 1210 1215	
25	AAT GCT GTC ACT TGG GGC ATC TTC CCT GGG CGA GAG ATC ATC CAG CCC	1728
	Asn Ala Val Thr Trp Gly Ile Phe Pro Gly Arg Glu Ile Ile Gln Pro	
	1220 1225 1230	
30	ACC GTA GTG GAT CCC GTC AGC TTC ATG TTC TGG AAG GAC GAG GCC TTT	1776
	Thr Val Val Asp Pro Val Ser Phe Met Phe Trp Lys Asp Glu Ala Phe	
	1235 1240 1245	
35	GCC CTG TGG ATT GAG CGG TGG GGA AAG CTG TAT GAG GAG GAG TCC CCG	1824
	Ala Leu Trp Ile Glu Arg Trp Gly Lys Leu Tyr Glu Glu Glu Ser Pro	
	1250 1255 1260	
40	TCC CGC ACC ATC ATC CAG TAC ATC CAC GAC AAC TAC TTC CTG GTC AAC	1872
	Ser Arg Thr Ile Ile Gln Tyr Ile His Asp Asn Tyr Phe Leu Val Asn	
	1265 1270 1275 1280	
45	CTG GTG GAC AAT GAC TTC CCA CTG GAC AAC TGC CTC TGG CAG GTG GTG	1920
	Leu Val Asp Asn Asp Phe Pro Leu Asp Asn Cys Leu Trp Gln Val Val	
	1285 1290 1295	
50	GAA GAC ACA TTG GAG CTT CTC AAC AGG CCC ACC CAG AAT GCG AGA GAA	1968
	Glu Asp Thr Leu Glu Leu Leu Asn Arg Pro Thr Gln Asn Ala Arg Glu	
	1300 1305 1310	
55	ACG GAG GCT CCA TGA CCCTGCGTCC TGACGCCCTG CGTTGGAGCC ACTCCTGTCC	2023
	Thr Glu Ala Pro *	
	1315	
60	CGCCTTCCTC CTCACAGTG CTGCTTCTCT TGGGAACCTC ACTCTCCTTC GTGTCTCTCC	2083
65	CACCCCGGCC TCCACTCCCC CACCTGACAA TGGCAGCTAG ACTGGAGTGA GGCTTCAGG	2143
	CTCTTCCTGG ACCTGAGTCG GCCCCACATG GGAACCTAGT ACTCTCTGCT CTAACAAAAA	2203
	AAAAAATAA GGAATT	2219

(2) INFORMACIÓN PARA LA ID SEC N°: 4:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 656 aminoácidos

(B) TIPO: aminoácido

(D) TIPO DE MOLÉCULA: proteína

ES 2 271 946 T3

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC Nº: 4:

5	Met Val Asn Glu Ala Arg Gly Asn Ser Ser Leu Asn Pro Cys Leu Glu	1 5 10 15
	Gly Ser Ala Ser Ser Gly Ser Glu Ser Ser Lys Asp Ser Ser Arg Cys	20 25 30
10	Ser Thr Pro Gly Leu Asp Pro Glu Arg His Glu Arg Leu Arg Glu Lys	35 40 45
15	Met Arg Arg Arg Leu Glu Ser Gly Asp Lys Trp Phe Ser Leu Glu Phe	50 55 60
	Phe Pro Pro Arg Thr Ala Glu Gly Ala Val Asn Leu Ile Ser Arg Phe	65 70 75 80
20	Asp Arg Met Ala Ala Gly Gly Pro Leu Tyr Ile Asp Val Thr Trp His	85 90 95
25	Pro Ala Gly Asp Pro Gly Ser Asp Lys Glu Thr Ser Ser Met Met Ile	100 105 110
	Ala Ser Thr Ala Val Asn Tyr Cys Gly Leu Glu Thr Ile Leu His Met	115 120 125
30	Thr Cys Cys Arg Gln Arg Leu Glu Glu Ile Thr Gly His Leu His Lys	130 135 140
35	Ala Lys Gln Leu Gly Leu Lys Asn Ile Met Ala Leu Arg Gly Asp Pro	145 150 155 160
	Ile Gly Asp Gln Trp Glu Glu Glu Glu Gly Gly Phe Asn Tyr Ala Val	165 170 175
40	Asp Leu Val Lys His Ile Arg Ser Glu Phe Gly Asp Tyr Phe Asp Ile	180 185 190
45	Cys Val Ala Gly Tyr Pro Lys Gly His Pro Glu Ala Gly Ser Phe Glu	195 200 205
	Ala Asp Leu Lys His Leu Lys Glu Lys Val Ser Ala Gly Ala Asp Phe	210 215 220
50	Ile Ile Thr Gln Leu Phe Phe Glu Ala Asp Thr Phe Phe Arg Phe Val	225 230 235 240
55	Lys Ala Cys Thr Asp Met Gly Ile Thr Cys Pro Ile Val Pro Gly Ile	245 250 255

ES 2 271 946 T3

Phe Pro Ile Gln Gly Tyr His Ser Leu Arg Gln Leu Val Lys Leu Se
 260 265 270
 5 Lys Leu Glu Val Pro Gln Glu Ile Lys Asp Val Ile Glu Pro Ile Ly
 275 280 285
 10 Asp Asn Asp Ala Ala Ile Arg Asn Tyr Gly Ile Glu Leu Ala Val Se
 290 295 300
 Leu Cys Gln Glu Leu Leu Ala Ser Gly Leu Val Pro Gly Leu His Phe
 305 310 315 320
 15 Tyr Thr Leu Asn Arg Glu Met Ala Thr Thr Glu Val Leu Lys Arg Leu
 325 330 335
 Gly Met Trp Thr Glu Asp Pro Arg Arg Pro Leu Pro Trp Ala Leu Ser
 340 345 350
 20 Ala His Pro Lys Arg Arg Glu Glu Asp Val Arg Pro Ile Phe Trp Ala
 355 360 365
 25 Ser Arg Pro Lys Ser Tyr Ile Tyr Arg Thr Gln Glu Trp Asp Glu Phe
 370 375 380
 Pro Asn Gly Arg Trp Gly Asn Ser Ser Ser Pro Ala Phe Gly Glu Leu
 385 390 395 400
 30 Lys Asp Tyr Tyr Leu Phe Tyr Leu Lys Ser Lys Ser Pro Lys Glu Glu
 405 410 415
 35 Leu Leu Lys Met Trp Gly Glu Glu Leu Thr Ser Glu Ala Ser Val Phe
 420 425 430
 Glu Val Phe Val Leu Tyr Leu Ser Gly Glu Pro Asn Arg Asn Gly His
 435 440 445
 40 Lys Val Thr Cys Leu Pro Trp Asn Asp Glu Pro Leu Ala Ala Glu Thr
 450 455 460
 45 Ser Leu Leu Lys Glu Glu Leu Leu Arg Val Asn Arg Gln Gly Ile Leu
 465 470 475 480
 Thr Ile Asn Ser Gln Pro Asn Ile Asn Gly Lys Pro Ser Ser Asp Pro
 485 490 495
 50 Ile Val Gly Trp Gly Pro Ser Gly Gly Tyr Val Phe Gln Lys Ala Tyr
 500 505 510
 55 Leu Glu Phe Phe Thr Ser Arg Glu Thr Ala Glu Ala Leu Leu Gln Val
 515 520 525
 Leu Lys Lys Tyr Glu Leu Arg Val Asn Tyr His Leu Val Asn Val Lys
 530 535 540
 60 Gly Glu Asn Ile Thr Asn Ala Pro Glu Leu Gln Pro Asn Ala Val Thr
 545 550 555 560
 65 Trp Gly Ile Phe Pro Gly Arg Glu Ile Ile Gln Pro Thr Val Val Asp
 565 570 575

ES 2 271 946 T3

	Pro	Val	Ser	Phe	Met	Phe	Trp	Lys	Asp	Glu	Ala	Phe	Ala	Leu	Trp	Ile
					580				585					590		
5	Glu	Arg	Trp	Gly	Lys	Leu	Tyr	Glu	Glu	Glu	Ser	Pro	Ser	Arg	Thr	Ile
					595			600					605			
10	Ile	Gln	Tyr	Ile	His	Asp	Asn	Tyr	Phe	Leu	Val	Asn	Leu	Val	Asp	Asn
					610			615				620				
15	Asp	Phe	Pro	Leu	Asp	Asn	Cys	Leu	Trp	Gln	Val	Val	Glu	Asp	Thr	Leu
					625			630			635				640	
20	Glu	Leu	Leu	Asn	Arg	Pro	Thr	Gln	Asn	Ala	Arg	Glu	Thr	Glu	Ala	Pro
					645					650					655	

(2) INFORMACIÓN PARA LA ID SEC N°: 5:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 22 pares de bases
- (B) TIPO: ácido nucleico
- (C) ORIENTACIÓN: única
- (D) TIPOLOGIA: lineal

(ii) TIPO DE MOLÉCULA: cDNA

(iii) HIPOTÉTICA: SÍ

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC N°: 5:

AGCCTCAACC CCTGCTTGGA GG

22

(2) INFORMACIÓN PARA LA ID SEC N°: 6:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 22 pares de bases
- (B) TIPO: ácido nucleico
- (C) ORIENTACIÓN: única
- (D) TIPOLOGIA: lineal

(ii) TIPO DE MOLÉCULA: cDNA

(iii) HIPOTÉTICA: SÍ

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC N°: 6:

TGACAGTTTG CTCCCCAGGC AC

22

(2) INFORMACIÓN PARA LA ID SEC N°: 7:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 23 pares de bases
- (B) TIPO: ácido nucleico
- (C) ORIENTACIÓN: única
- (D) TIPOLOGIA: lineal

ES 2 271 946 T3

(ii) TIPO DE MOLÉCULA: cDNA

(iii) HIPOTÉTICA: SÍ

5 (iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC N°: 7:

TGAAGGAGAA GGTGTCTGCG GGA

23

10

(2) INFORMACIÓN PARA LA ID SEC N°: 8:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

15

(A) LONGITUD: 21 pares de bases

(B) TIPO: ácido nucleico

(C) ORIENTACIÓN: única

20

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: cDNA

(iii) HIPOTÉTICA: SÍ

25

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC N°: 8:

AGGACGGTCG GGTGAGAGTG G

21

30

(2) INFORMACIÓN PARA LA ID SEC N°: 9:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

35

(A) LONGITUD: 23 pares de bases

(B) TIPO: ácido nucleico

(C) ORIENTACIÓN: única

40

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: cDNA

(iii) HIPOTÉTICA: SÍ

45

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC N°: 9:

CACTGTGGTT GGCATGGATG ATG

23

50

(2) INFORMACIÓN PARA LA ID SEC N°: 10:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

55

(A) LONGITUD: 21 pares de bases

(B) TIPO: ácido nucleico

(C) ORIENTACIÓN: única

60

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: cDNA

(iii) HIPOTÉTICA: SÍ

65

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC N°: 10:

ES 2 271 946 T3

GGCTGCTCTT GGACCCTCCT C

21

(2) INFORMACIÓN PARA LA ID SEC Nº: 11:

- 5 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 21 pares de bases
 - (B) TIPO: ácido nucleico
 - 10 (C) ORIENTACIÓN: única
 - (D) TIPOLOGIA: lineal
- (ii) TIPO DE MOLÉCULA: cDNA
- 15 (iii) HIPOTÉTICA: SÍ
- (iv) ANTISENTIDO: NO
- (xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC Nº: 11:
- 20 TGCTTCCGGC TCCCTCTAGC C

21

(2) INFORMACIÓN PARA LA ID SEC Nº: 12:

- 25 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 22 pares de bases
 - (B) TIPO: ácido nucleico
 - 30 (C) ORIENTACIÓN: única
 - (D) TIPOLOGIA: lineal
- (ii) TIPO DE MOLÉCULA: cDNA
- 35 (iii) HIPOTÉTICA: SÍ
- (iv) ANTISENTIDO: SI
- (xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC Nº: 12:
- 40 CCTCCCGCTC CCAAGAACAA AG

22

(2) INFORMACIÓN PARA LA ID SEC Nº: 13:

- 45 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 23 pares de bases
 - (B) TIPO: ácido nucleico
 - 50 (C) ORIENTACIÓN: única
 - (D) TIPOLOGIA: lineal
- (ii) TIPO DE MOLÉCULA: cDNA
- 55 (iii) HIPOTÉTICA: SÍ
- (iv) ANTISENTIDO: NO
- (xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC Nº: 13:
- 60 TGAAGGAGAA GGTGTCTGCG GGA

23

(2) INFORMACIÓN PARA LA ID SEC Nº: 14:

- 65 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 23 pares de bases

ES 2 271 946 T3

(B) TIPO: ácido nucleico

(C) ORIENTACIÓN: única

(D) TIPOLOGIA: lineal

5

(ii) TIPO DE MOLÉCULA: cDNA

(iii) HIPOTÉTICA: SÍ

10

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID SEC N°: 14:

AGGACGGTGC GGTGAGAGTG

20

15

20

25

30

35

40

45

50

55

60

65