(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 273 934**

(51) Int. Cl.:
G01N 33/576 (2006.01)
A61K 39/29 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Número de solicitud europea: **02006640 .3**
(86) Fecha de presentación : **23.08.1991**
(87) Número de publicación de la solicitud: **1227323**
(87) Fecha de publicación de la solicitud: **31.07.2002**

(54) Título: **Antígeno del virus de la hepatitis no-A, no-B y métodos diagnósticos.**

(30) Prioridad: **27.08.1990 US 573643**
21.11.1990 US 616369
21.08.1991 US 748564

(45) Fecha de publicación de la mención BOPI:
16.05.2007

(45) Fecha de la publicación del folleto de la patente:
16.05.2007

(73) Titular/es: **Bioprocess Pty. Ltd.**
Suite 145, National Innovation Centre
The Australian Technology Park
Eveleigh, NSW 1430, AU
F. Hoffmann-La Roche Ltd.

(72) Inventor/es: **Zebedee, Suzanne;**
Inchauspe, Geneviève;
Nasoff, Marc S. y
Prince, Alfred M.

(74) Agente: **Isern Jara, Jorge**

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).

Venta de fascículos: Oficina Española de Patentes y Marcas. Pº de la Castellana, 75 – 28071 Madrid

DESCRIPCIÓN

Antígeno del virus de la hepatitis no-A, no-B y métodos diagnósticos.

Campo técnico

La presente invención esta relacionada con composiciones que contienen una proteína estructural de NANBV, útil en los métodos y equipos diagnósticos que comprenden dicha composición, así como con la utilización de dichos equipos y composiciones con propósitos diagnósticos.

Antecedentes de la invención

Se cree que la hepatitis no-A, no-B (NANBH) está causada por un virus trasmisible que se ha denominado tanto virus de la hepatitis C (HCV) como virus de la hepatitis no-A, no-B (NANBV). Aunque esta enfermedad transmisible se descubrió hace años, todavía se está realizando la caracterización completa del agente causal.

Se han obtenido aislamientos del NANBV, y se han clonado y secuenciado en el ámbito molecular porciones o la totalidad del genoma viral de los diferentes aislamientos. Choo *et al., Science*, 244:359-362 (1989); Choo *et al., Proc. Natl. Acad. Sci. USA*, 88:2451-2455 (1991); Takamizawa *et al., J. Virol*., 65:1105-1113 (1991); Kato *et al., Proc. Natl. Acad. Sci. USA*, 87:9524-9528 (1990); y Takeuchi *et al., Nucl. Acids Res*., 18:4626 (1990). Las similitudes en la secuencia de bases nucleotídicas entre los diferentes aislamientos del NANBV sugieren que estos son parte de una familia de virus relacionados. Okamoto *et al., Japan J. Exp. Med*., ,60:163-177 (1990)y Ogata *et al., Proc. Natl. Acad. Sci. USA*, 88:3392-3396 (1991). Las propiedades del genoma del NANBV sugieren que los NANBV pueden ser un pariente muy lejano de la familia de los flavivirus. Sin embargo, las similitudes tanto en el tamaño como la hidropaticidad de las proteínas estructurales sugieren que los virus NANB podrían estar también relacionados de forma lejana con la familia de los pestivirus. Miller *et al., Proc. Natl. Acad. Sci*., 87:2057-2061 (1990);y Okamoto *et al., Japan J. Exp. Med*., 60:163-177 (1990).

Las dificultades en la caracterización taxonómica de los aislamientos del NANBV y la falta de información respecto a las proteínas codificadas por el genoma del NANBV han dificultado la identificación de productos génicos relevantes útiles como marcadores diagnósticos.

El genoma del NANBV está comprendido por una molécula de RNA de cadena positiva que codifica una única poliproteína. Basándose en la homología con otros virus relacionados caracterizados, se cree que los productos génicos del NANBV incluyen tanto proteínas estructurales como no estructurales. A partir de estas homologías, se puede predecir que el NANBV expresa el producto génico de una única poliproteína a partir del genoma viral completo, que entonces se escinde en proteínas estructurales y no estructurales funcionalmente diferenciadas. Este tipo de morfogénesis viral impide la identificación positiva de las proteínas virales maduras individuales, hasta que éstas no se hayan aislado y caracterizado físicamente. Ya que no se ha desarrollado ningún sistema de cultivo *in vitro* para propagar los virus NANBV, no se han aislado productos génicos estructurales o no estructurales (proteínas) del NANBV a partir de especímenes biológicos o células infectadas por el NANBV. Por lo tanto, todavía está pendiente la identificación de las proteínas del NANBV, de su papel en el ciclo de vida del virus y de su papel en la enfermedad. En particular, los marcadores antigénicos de la enfermedad inducida por el NANBV todavía no se han caracterizado completamente.

Un producto génico del NANBV, concretamente el antígeno C100-3, derivado de porciones de los genes no estructurales designados NS3 y NS4, se ha expresado como una proteína de fusión y se ha utilizado para detectar anticuerpos anti-C100-3 en pacientes con varias formas de hepatitis NANB. Véase, por ejemplo, Kuo *et al., Science*, 244:362-364 (1989); y la solicitud de patente internacional Nº PCT/US88/04125. Un ensayo diagnóstico basado en el antígeno C100-3 está comercialmente disponible en Ortho Diagnostics, Inc. (Raritan, NJ). Este ensayo del C100-3 representa el estado actual de la detección de las infecciones por NANBV. Sin embargo, se ha descrito que el inmunoensayo basado en el antígeno C100-3 detecta preferentemente los anticuerpos en sueros de pacientes infectados de manera crónica. La seroconversión con C100-3 aparece generalmente a partir de entre cuatro y seis meses tras el inicio de la hepatitis, y en algunos casos el C100-3 no consigue detectar anticuerpo alguno en presencia de una infección por NANBV. Alter *et al., New Eng. J. Med*., 321:1538-39 (1989); Alter *et al., New Eng. J. Med*., 321:1.94-1500 (1989); y Weiner *et al., Lancet*, 335:1-3 (1990). McFarlane *et al., Lancet*, 335:754-757 (1990), describieron resultados de falsos positivos cuando el inmunoensayo basado en el C100-3 se utilizaba para analizar los anticuerpos en pacientes con hepatitis autoinmune crónica activa. Utilizando el inmunoensayo basado en el C100-3, Grey *et al., Lancet*, 335:609-610 (1990), describieron resultados de falso positivo en sueros de pacientes con enfermedad hepática causada por varias condiciones diferentes del NANBV.

En la actualidad no está disponible un inmunoensayo del NANBV que pueda detectar de forma precisa la seroconversión en una etapa temprana tras la infección, o que pueda identificar una infección aguda por el NANBV.

La cepa Hutch del HCV es un aislamiento con interés clínico comparado con la cepa Donn (HCV-1) ya que la HCV-H aumenta hasta títulos extremadamente altos en el paciente.

Resumen de la invención

Una cepa Hutch (HCV-H) del virus de la hepatitis no-A, no-B (NANBV) denominada el aislamiento Hutch c59 (o HCV-Hc59) se ha propagado a través de pases en animales, y el genoma viral completo se ha clonado y secuenciado. Cuando se utiliza el término "subgrupo" la presente especificación se refiere a un grupo del NANBV que están definidos serológicamente por cepas concretas, como la cepa Hutch c59. Los datos de la secuencia muestran que hay diferencias tanto a nivel nucleotídico como aminoacídico si se compara con las cepas del NANBV previamente descritas. Véanse las secuencias de los siguientes aislamientos de HCV, en los que la designación del aislamiento se muestra en paréntesis para su comparación, Okamoto *et al., Japan J. Exp. Med*., 60:163-177, 1990 (HC-J1, HC-J4); Takeuchi *et al., Nucleic Acids Res*.,18:4626, 1990 (HCV-JH); Choo *et al., Proc. Natl. Acad. Sci. USA*, 88:2451-2455, 1991 (HCV-1); Kato *et al., Proc. Natl. Acad. Sci. USA*, 87:9524-9528, 1990 (HCV-J); Takamizawa *et al., J. virol*., 65:1105-1113, 1991 (HCV-BK); patente estadounidense Nº 5.032.511 de Takahashi et al.; Ogata *et al., Proc. Natl. Acad. Sci. USA*, 88:3392-3396, 1991 (HCV-Hh); y la solicitud internacional Nº PCT/US88/04125.

Aquí se muestra que las secuencias identificadas codifican proteínas estructurales del NANBV. También se muestra aquí que las proteínas estructurales del NANBV incluyen epítopos antigénicos útiles para el diagnóstico de anticuerpos inmunoreactivos con proteínas estructurales del NANBV, y para su utilización en vacunas para inducir anticuerpos neutralizantes contra el NANBV. Concretamente, los antígenos del NANBV de esta invención son antígenos del aislamiento Hutch c59 del NANBV.

La secuencia nucleotídica que codifica la porción amino terminal de la poliproteína de los genes estructurales de la cepa Hutch del NANBV la contiene el Id. de Sec. Nº:1. Por comparación con otros aislamientos del NANBV, los flavivirus y los pestivirus, se cree que la secuencia de nucleótidos que contiene el Id. de Sec. Nº:1 codifica proteínas estructurales del NANBV, concretamente la cápside y porciones de la cubierta.

Los antígenos estructurales descritos aquí están presentes en la presunta proteína de la cápside contenida en el Id. de Sec. Nº:1 desde los residuos de las posiciones aminoacídicas 1-120, y están presentes en la porción amino terminal de la presunta proteína de la cubierta contenida en el Id. de Sec. Nº:1 desde los residuos de las posiciones 121 a 326.

Las secuencias de residuos nucleotídicos y aminoacídicos están definidas aquí a partir de un número de posición de residuo base o aminoácido inicial hasta un número de posición de residuo o base final. Se entenderá que todas estas secuencias incluyen tanto los números de posición inicial y final.

La secuencia completa del genoma del aislamiento Hutch c59 también se ha determinado y descrito. Así, la especificación describe un segmento de DNA que codifica el genoma viral del aislamiento Hutch c59 del NANBV contenido en el Id. de Sec. Nº:46 desde las posiciones nucleotídicas 1 a 9416.

La especificación también describe un segmento de DNA que codifica una proteína estructural del NANBV que comprende un antígeno estructural del NANBV, preferiblemente un antígeno de la cápside. Un antígeno de la cápside particularmente preferible incluye una secuencia de residuos aminoacídicos representada por el Id. de Sec. Nº:1 desde el residuo 1 al residuo 20, desde el residuo 21·al residuo 40, desde el residuo 2·al residuo 40, o desde el residuo 1 al residuo 74. El segmento de DNA descrito aquí preferiblemente incluye la secuencia de bases nucleotídicas representada por el Id. de Sec. Nº:1 desde la base en posición 1·a la base en posición 60, desde la base en posición 61·a la base en posición 120, desde la base en posición 4·a la base en posición 120, o desde la base en posición 1 a la base en posición 222, respectivamente.

También se describe aquí un polinucleótido que comprende una secuencia de nucleótidos que codifica porciones de la poliproteína del aislamiento Hutch c59, en particular porciones de las regiones específicas de secuencia de c59 en las regiones V, V_1, V_2 o V_3.

También se describe una molécula de DNA recombinante que comprende un vector, preferiblemente un vector de expresión, ligado operativamente a un segmento de DNA de la presente invención. Una molécula de DNA recombinante preferida es el pGEX-3X-690:691, pGEX-3X-690:694, pGEX-3X-693:691, PGEX-3X-15:17, pGEX-3X-15:18, pGEX-2T-15:17, pGEX-2T-CAP-A, pGEX-2T-CAP-B o pGEX-2T-CAP-A-B.

Se contempla que una proteína estructural del NANBV que comprende una secuencia de residuos aminoacídicos que define un antígeno estructural del NANBV, preferiblemente un antígeno de la cápside, y más preferiblemente una que incluye la secuencia de residuos aminoacídicos contenida en el ID. de SEC. Nº:1 desde el residuo 1 al residuo 20, desde el residuo 21 al residuo 40, desde el residuo 2 al residuo 40, o desde el residuo 1 al residuo 74. También se contemplan las proteínas de fusión que comprenden una proteína estructural del NANBV de esta invención.

La invención proporciona un método para la detección de la seroconversión asociada con la infección por NANBV en las etapas tempranas tras la infección, que comprende:

(a) obtener una mezcla de inmunoreacción acuosa mezclando una muestra de fluido corporal con un antígeno de la cápside del NANBV;

(b) mantener dicha mezcla de inmunoreacción acuosa durante un periodo de tiempo suficiente que permita que los anticuerpos contra el antígeno de la cápside del NANBV presentes en la muestra de fluido corporal inmunoreaccionen con dicho antígeno de la cápside del NANBV para formar un producto de inmunoreacción; y

(c) detectar la presencia de dicho producto de inmunoreacción formado y por lo tanto detectar seroconversión temprana.

La invención también está relacionada con la utilización del antígeno de la cápside del NANBV para la preparación de una composición diagnóstica para la detección de la seroconversión asociada con la infección por NANBV en las etapas tempranas tras la infección.

Además, la presente invención esta relacionada con un equipo útil para el diagnóstico de la seroconversión asociada con la infección por NANBV en las etapas tempranas tras la infección en una muestra de fluido corporal que comprende un antígeno de la cápside del NANBV seleccionada del grupo que consiste en:

(a) un antígeno CAP-A de la cápside del NANBV con la secuencia de aminoácidos desde el residuo 1 al 20 del ID. de SEC. Nº:1;

(b) un antígeno CAP-B de la cápside del NANBV con la secuencia de aminoácidos desde el residuo 21 al 40 del ID. de SEC. Nº:1; y

(e) un antígeno CAP-N de la cápside del NANBV con la secuencia de aminoácidos desde el residuo 1 al 74 del ID. de SEC. Nº:1.

La especificación también describe moléculas de anticuerpo que inmunoreaccionan con el aislamiento Hutch c59 del NANBV, pero que no inmunoreaccionan con los aislamientos HCV-1, HCV-BK, HCV-J, HC-J1, HC-J4, HCV-JH o HCV-Hh del NANBV, es decir, moléculas de anticuerpo específicas de c59. Además se describe un cultivo de células transformadas con una molécula de DNA recombinante descrita aquí y los métodos para producir una proteína estructural del NANBV de esta invención utilizando el cultivo.

También se contempla una composición que comprende una proteína estructural del NANBV. La composición se caracteriza preferiblemente por estar esencialmente libre de (a) antígenos procarióticos, y (b) otras proteínas relacionadas con el NANBV.

Además también se contempla un sistema de diagnóstico en forma de equipo que comprende, en una cantidad suficiente para realizar al menos un ensayo, una composición con una proteína estructural del NANBV, un polipéptido o una proteína de fusión de esta invención, como un reactivo empaquetado de forma separada. Preferiblemente, el sistema de diagnóstico contiene la proteína de fusión fijada a una matriz sólida.

Además se contempla un método, preferiblemente un método *in vitro*, para el análisis en una muestra de fluido corporal de la presencia de anticuerpos contra al menos uno de los antígenos estructurales del NANBV descritos aquí. El método comprende la formación de una mezcla de inmunoreacción mediante la mezcla (puesta en contacto) de la muestra de fluido corporal con un reactivo inmunológico como una proteína, polipéptido o proteína de fusión estructural del NANBV de esta invención. La mezcla de inmunoreacción se mantiene durante un periodo de tiempo suficiente que permita que cualquiera de los anticuerpos presentes inmunoreaccionen con la mezcla de reactivo inmunológico para formar un producto de inmunoreacción, cuyo producto, al ser detectado, es indicativo de la presencia de anticuerpos contra la proteína estructural del NANBV.

Preferiblemente, el reactivo inmunológico está fijado a una matriz sólida cuando se utiliza este método.

La especificación también describe un inóculo (o una vacuna) que comprende una cantidad inmunológicamente efectiva de una proteína, polipéptido o proteína de fusión estructural del NANBV de esta invención, dispersada en un transportador farmacéuticamente aceptable y/o diluyente. El inóculo está esencialmente libre de (a) antígenos procarióticos y (b) otras proteínas relacionadas con el NANBV.

También se describe un método profiláctico para el tratamiento de una infección, comprendiendo este método la administración de un inóculo como los descritos aquí.

Breve descripción de las figuras

La Figura 1 es una representación esquemática del genoma del HCV-Hc59 y localización de los clones de cDNA del HCV-Hc59 numerados del cero al 39. Se muestra el alineamiento con la proteína codificada por los flavivirus así como los presuntos dominios en el genoma codificado por el HCV. Las regiones de homología de aminoácidos con el Ns3 del virus del Dengue Tipo 2 y el virus del jaspeado del clavel (CARMv) se indican mediante cajas rayadas y blancas, respectivamente.

Descripción detallada de la invención

A. *Definiciones*

Aminoácido: Todos los residuos aminoacídicos identificados aquí están en la configuración L natural. DE forma concordante con la nomenclatura estándar de los polipéptidos, *J. Biol. Chem*., 243:3557-59, (1969), las abreviaturas de los residuos aminoacídicos se muestran en la siguiente tabla de correspondencia:

TABLA DE CORRESPONDENCIA

Símbolo 3-Letras	Aminoácido
Tyr	L-tirosina
Gly	L-glicina
Phe	L-fenilalanina
Met	L-metionina
Ala	L-alanina
Ser	L-serina
Ile	L-isoleucina
Leu	L-leucina
Thr	L-treonina
Val	L-valina
Pro	L-prolina
Lys	L-lisina
His	L-histidina
Gln	L-glutamina
Glu	ácido L-glutámico
Glx	Gln o Glu
Trp	L-triptófano
Arg	L-arginina
Asp	ácido L-aspártico
Asn	L-asparagina
Asx	Asp o Asn
Cys	L-cisteína
Xaa	Desconocido u otro

Debe tenerse en cuenta que todas las secuencias de residuos aminoacídicos, normalmente denominadas aquí "secuencias de residuos", están representadas aquí por la fórmula cuya orientación de izquierda a derecha es la dirección convencional de amino terminal a carboxi terminal.

Antígeno: Un polipéptido o proteína que es capaz de unirse específicamente (inmunoreaccionar) a un anticuerpo y de formar un producto de inmunoreacción (inmunocomplejo). El punto del antígeno en el que se une el anticuerpo se denomina determinante antigénico o epítopo.

Nucleótido: una unidad monomérica de DNA o RNA que está formado por una porción azúcar (pentosa), un fosfato, y una base heterocíclica nitrogenada. La base está unida a la porción azúcar a través del carbono glicosídico (carbono 1' de la pentosa) y esta combinación de base y azúcar es un nucleósido. Cuando el nucleósido contiene un grupo fosfato unido en la posición 3' o 5' de la pentosa se denomina un nucleótido. Una secuencia de nucleótidos unidos de forma operativa se denomina generalmente aquí "secuencia de bases" y se representa aquí mediante una fórmula cuya orientación de izquierda a derecha es la dirección convencional del extremo 5' al extremo 3'.

Dúplex de DNA: Una molécula de ácido nucleico de doble cadena que comprende dos cadenas de polinucleótido sustancialmente complementario hibridadas entre ellas mediante la formación de un puente de hidrógeno entre cada uno de los nucleótidos complementarios presentes en un par de bases del dúplex. Como los nucleótidos que forman un par de bases pueden ser una base ribonucleotídica o una base desoxiribonucleotídica, la frase "dúplex de DNA" se refiere tanto a un dúplex de DNA-DNA que comprende dos cadenas de DNA (dsDNA), o un dúplex de RNA-DNA que comprende una cadena de DNA y una cadena de RNA.

Par de bases (pb): una pareja de adenina (A) con timina (T), o de citosina (C) con guanina (G) en una molécula de DNA de doble cadena. En el RNA, el uracilo (U)sustituye a la timina.

Secuencia de Nucleótidos Complementarios: una secuencia de nucleótidos en una molécula de cadena sencilla de DNA o RNA que es lo suficientemente complementaria a otra cadena sencilla para hibridar específicamente con ella (no al azar) con la consiguiente formación de puentes de hidrógeno.

Hibridación: el emparejamiento de secuencias nucleotídicas complementarias (cadenas de ácido nucleico) para formar un dúplex, heterodúplex o complejo que contiene más de dos ácidos nucleicos de cadena sencilla mediante el establecimiento de puentes de hidrógeno entre pares de bases complementarias. Es una interacción específica, es decir que no es al azar, entre polinucleótidos complementarios que puede inhibirse de forma competitiva.

Producto de Hibridación: El producto formado cuando un polinucleótido hibrida con un ácido nucleico de cadena sencilla o doble. Cuando un polinucleótido hibrida con un ácido nucleico de cadena doble, el producto de hibridación formado se denomina una triple hélice o molécula de ácido nucleico de cadena triple. Moser *et al., Science*, 238:645-50 (1987).

Análogo de Nucleótido: un nucleótido purínico o pirimidínico que es estructuralmente diferente de una A, T, G, C o U, pero que es lo suficientemente similar para poder sustituir a un nucleótido normal en una molécula de ácido nucleico. La inosina (I) es un nucleótido que puede formar puentes de hidrógeno con cualquiera de los demás nucleótidos, A, T, G, C o U. Además, se sabe que las bases metiladas pueden participar en una hibridación de ácidos nucleicos.

B. *Segmentos de DNA*

En los organismos vivos, la secuencia de residuos aminoacídicos de una proteína o polipéptido está relacionada de forma directa a través del código genético con la secuencia de ácido desoxiribonucleico (DNA) del gen estructural que codifica la proteína. Por lo tanto, un gen estructural puede definirse en términos de la secuencia de residuos aminoacídicos, es decir, la proteína o polipéptido para el cual ésta codifica.

Una característica importante y bien conocida del código genético es su redundancia. Es decir, para la mayoría de aminoácidos utilizados para formar las proteínas, más de un triplete nucleótidos codificante (codón) puede codificar o designar un residuo aminoacídico concreto. Por lo tanto, una serie de secuencias nucleotídicas diferentes pueden codificar una secuencia de residuos aminoacídicos particular. Tales secuencias nucleotídicas se consideran funcionalmente equivalentes ya que pueden resultar en la producción de la misma secuencia de residuos aminoacídicos en todos los organismos.

De forma ocasional, una variante metilada de una purina o pirimidina puede incorporarse en una secuencia nucleotídica dada. Sin embargo, estas metilaciones no afectan en modo alguno la codificación.

La especificación también describe un segmento de DNA aislado que comprende una secuencia de bases nucleotídicas que codifica una proteína estructural del NANBV que comprende un antígeno estructural del NANBV como un antígeno de la cápside, un antígeno de la cubierta, o ambas. Preferiblemente, el antígeno estructural está inmunológicamente relacionado con la cepa Hutch del NANBV.

Más preferiblemente, el antígeno estructural del NANBV codificado tiene una secuencia de residuos aminoacídicos que se corresponde, y preferiblemente es idéntica, a la secuencia de residuos aminoacídicos contenida en el ID. de SEC. Nº:1.

En una realización, el presunto antígeno de la cápside incluye una secuencia de residuos aminoacídicos contenida en el ID. de SEC. Nº:1 desde el residuo 1 al residuo 20, desde el residuo 21 al residuo 40, desde el residuo 2 al residuo 40, o desde el residuo 1 al residuo 74. En otra realización, el antígeno de la cápside incluye la secuencia contenida en el ID. de SEC. Nº:1 desde el residuo 69 al residuo 120.

Los segmentos de DNA descritos aquí incluyen una secuencia de bases representadas por la secuencia de bases contenida en el ID. de SEC. Nº:1 desde la base en posición 1 hasta la base en posición 222, desde la base en posición 205 360, desde la base en posición 361 hasta la base en posición 528, o desde la base en posición 361 hasta la base en posición 978.

La longitud de la secuencia de bases nucleotídicas descritas aquí no es superior a 3.000 bases, preferiblemente no es superior a 1.000 bases.

La secuencia de residuos aminoacídicos de una proteína estructural del NANBV particularmente preferibles está contenida en el ID. de SEC. Nº:2 desde el residuo 1 al residuo 315, en el ID. de SEC. Nº:3 desde el residuo 1 al residuo 252, en el ID. de SEC. Nº:4 desde el residuo 1 al residuo 252 y en el ID. de SEC. Nº:6 desde el residuo 1 al residuo 271.

Un segmento de DNA purificado como el descrito aquí está sustancialmente libre de otros ácidos nucleicos que no contienen las secuencias de bases nucleotídicas especificadas aquí para un segmento de DNA de esta invención, mientras que el segmento de DNA está presente en forma de una composición que contiene el segmento de DNA purificado, o como una solución, suspensión o formulación particulada. Por sustancialmente libre se entiende que

el segmento de DNA está presente en al menos un 10% del ácido nucleico total presente en peso, preferiblemente superior a un 50 por ciento, y más preferiblemente superior a un 90 por ciento del ácido nucleico total en peso.

Un segmento de DNA descrito aquí contiene una secuencia de bases nucleotídicas que define un gen estructural capaz de expresar una proteína de fusión. La frase "proteína de fusión" se refiere a una proteína con una porción polipeptídica ligada de forma operativa mediante un enlace peptídico a una segunda porción polipeptídica que define un antígeno estructural del NANBV como se describe aquí.

Una primera porción polipeptídica preferible tiene una secuencia de residuos aminoacídicos que corresponde a una secuencia como la contenida en el ID. de SEC. Nº:2 desde alrededor del residuo 1 a alrededor del residuo 221, y deriva de la proteína glutatión-S-transferasa (GST)

Una segunda porción polipeptídica preferible que define un antígeno estructural del NANBV en una proteína de fusión incluye una secuencia de residuos aminoacídicos representada por la secuencia contenida en el ID. de SEC. Nº:1 desde el residuo 1 al residuo 20, desde el residuo 21 al residuo 40, desde el residuo 2 al residuo 40, desde el residuo 1 al residuo 74 o desde el residuo 69 al residuo 120.

Una proteína de fusión como la descrita aquí puede contener más de una porción polipeptídica que define un antígeno estructural del NANBV, como por ejemplo la combinación de dos porciones polipeptídicas que representan diferentes antígenos estructurales, como se muestra en la secuencia de residuos aminoacídicos contenidos en el ID. de SEC. Nº:1 desde el residuo 1 al residuo 120, o en el ID. de SEC. Nº:1 desde el residuo 1 al residuo 326.

Como se describe aquí, esta porción de un segmento de DNA que codifica una proteína de fusión, que codifica una porción polipeptídica que define un antígeno de la cápside del NANBV, incluye una secuencia de bases nucleotídicas que corresponde a una secuencia que codifica una secuencia de residuos aminoacídicos contenida en una ID. de SEC. Nº:1 desde el residuo 1 al residuo 20, desde el residuo 21 al residuo 40, desde el residuo 2 al residuo 40, o desde el residuo 1 al residuo 74, y más preferiblemente incluye una secuencia de bases nucleotídicas que corresponde a la secuencia de bases contenida en el ID. de SEC. Nº:1 desde la base 1 a la base 60, desde la base 61 a la base 120, desde la base 4 a la base 120, o desde la base 1 a la base 222, respectivamente.

También se describe aquí que un segmento de DNA está unido a un segmento de DNA complementario, formando así un segmento de DNA de doble cadena. Además, debe tenerse en cuenta que un segmento de DNA de doble cadena de esta invención preferiblemente tiene una cola de cadena sencilla cohesiva en uno o ambos extremos.

También se describe que un segmento de DNA comprende una secuencia de bases nucleotídicas que codifica el genoma del aislamiento Hutch del NANBV. Preferiblemente, el segmento de DNA tiene una secuencia de bases nucleotídicas que codifica la secuencia de residuos aminoacídicos de la poliproteína producida por el genoma del aislamiento Hutch c59, cuya secuencia de residuos aminoacídicos se muestra en el ID. de SEC. Nº:46 desde el residuo 1 al residuo 3011. Más preferiblemente, el segmento de DNA en esta realización tiene la secuencia nucleotídica que se muestra en el ID. de SEC. Nº:46 desde la base 1 a la base 9416.

Un segmento de DNA que codifica el genoma del aislamiento c59 es útil para la obtención de un estándar o control de hibridación en los métodos diagnósticos basados en la hibridación de ácidos nucleicos utilizando los polinucleótidos, para la obtención de antígenos o proteínas de fusión estructurales del NANBV mediante métodos de DNA recombinante, para la obtención de partículas de aislamiento c59 del NANBV infeccioso en cultivo, y similares, todo ello descrito aquí.

También se describe aquí un fragmento de un segmento de DNA de esta invención que corresponde a una porción de un genoma del NANBV o que codifica una porción del antígeno estructural del NANBV. Estos fragmentos, cuando están presentes en forma de cadena sencilla o se mencionan en el contexto de una cadena de un segmento de DNA de doble cadena, se denominan aquí polinucleótidos.

Cuando el polinucleótido se utiliza para codificar un antígeno estructural del NANBV o región de la poliproteína del aislamiento Hutch c59, el polinucleótido corresponde a la cadena codificante del genoma del NANBV como se describe aquí.

Así, la especificación describe un polinucleótido que comprende una secuencia de bases nucleotídicas que incluye una secuencia de bases nucleotídicas que codifica una secuencia de residuos aminoacídicos que corresponde a una porción de la poliproteína expresada por el aislamiento Hutch del NANBV. Preferiblemente, el polinucleótido codifica una secuencia que corresponde a una porción de la secuencia de residuos aminoacídicos del aislamiento c59 que se muestra en el ID. de SEC. Nº:46 desde el residuo 1 al residuo 3011.

Son particularmente preferibles las regiones del aislamiento Hutch c59 que son únicas, y por lo tanto proporcionan una forma de distinguir el aislamiento Hutch, y más preferiblemente el aislamiento c59, de otros aislamientos del NANBV basándose en las diferencias en la secuencia de residuos aminoacídicos o bases nucleotídicas. Las regiones del genoma del aislamiento c59 útiles para distinguir los aislamientos contienen diferencias en la secuencia de bases nucleotídicas, y preferiblemente definen diferencias en la secuencia de residuos aminoacídicos codificada, al comparar con la secuencia de residuos nucleotídicos o aminoacídicos del aislamiento a distinguir.

Las comparaciones representativas para identificar las diferencias en la secuencia del aislamiento Hutch se muestran aquí en los Ejemplos, y en particular en la Tabla 11.

Un segmento de DNA o polinucleótido descrito aquí puede obtenerse fácilmente a partir de virus obtenidos de la sangre de individuos infectados por el NANBV como los descritos aquí o pueden sintetizarse de novo mediante técnicas químicas.

La síntesis química *de novo* de un segmento de DNA o un polinucleótido puede realizarse utilizando cualquier método disponible, como por ejemplo, los métodos de fosfotriéster o fosfodiéster. Véase Narang *et al., Meth. Enzymol*., 68:90, (1979); la patente estadounidense Nº 4.356.270; Itakura *et al., Ann. Rev. Biochem*., 53:323-56 (1989), Brown *et al., Meth. Enzymo*l., 68:109, (1979); y Matteucci *et al., J. Am. Chem. Soc*., 103:3185 (1981).

Por supuesto, mediante la síntesis química de la porción de gen estructural, puede realizarse cualquier modificación deseada de forma sencilla mediante la sustitución de las bases apropiadas por aquellas que codifiquen un residuo aminoacídico nativo. Sin embargo, son preferibles los segmentos de DNA que incluyen las secuencias idénticas al segmento contenido en el ID. de SEC. Nº:1, 2, 3, 4 o 6.

La derivación de un polinucleótido a partir de un ácido nucleico involucra el clonaje de un ácido nucleico en e huésped adecuado mediante un vector de clonación, la replicación del vector y por lo tanto, la multiplicación de la cantidad deácido nucleico clonado; y luego el aislamiento de subfragmentos de los ácidos nucleicos clonados.

Para una descripción de la subclonación de fragmentos de ácido nucleico, véase Maniatis *et al., Molecular Cloninq: A Laboratory Manual*, Cold Spring Harbor Laboratory, págs. 390-401 (1982); y véanse las patentes estadounidenses Nº 4.416.988 y Nº 4.403.036.

En realizaciones preferibles una sonda de polinucleótidos tiene un tamaño inferior de alrededor de 200 nucleótidos en longitud, preferiblemente inferior a 100 nucleótidos, y más preferiblemente inferior a 30 nucleótidos.

Por "sustancialmente complementario" y sus equivalentes gramaticales en relación con una sonda, se entiende que existe una similitud de la secuencia de bases nucleotídicas suficiente entre una sonda polinucleotídica sujeto y una secuencia de ácido nucleico específica presente en un gen de interés, de tal forma que la sonda es capaz de hibridar con la secuencia específica bajo condiciones de hibridación y formar un dúplex formado por la sonda y la secuencia específica.

Por lo tanto, la secuencia de la sonda polinucleotídica puede no ser la secuencia exacta de la secuencia diana siempre que la sonda contenga una complementariedad sustancial con la secuencia diana. Por ejemplo, un polinucleótido no complementario puede estar unido a un extremo de la sonda, siendo el resto de la secuencia de la sonda sustancialmente complementaria a la secuencia diana. Tales polinucleótidos no complementarios pueden codificar una diana de restricción por endonucleasas o una diana de unión de proteínas. Alternativamente, pueden intercalarse en la sonda bases no complementarias o secuencias más largas, siempre que la secuencia de la sonda tenga suficiente complementariedad con la secuencia de la cadena diana como para no hibridar al azar con ella, y por lo tanto formar un producto de hibridación bajo condiciones de hibridación.

La sonda polinucleotídica se proporciona en forma de cadena sencilla para su máxima eficiencia, pero alternativamente puede ser de doble cadena. Si es de doble cadena, la sonda polinucleotídica se trata en primer lugar para separar sus cadenas antes de utilizarla en la hibridación.

Además, un segmento de DNA puede prepararse mediante la síntesis inicial de oligonucleótidos que corresponden a las porciones del segmento de DNA, ensamblándose entonces los oligonucleótidos mediante la hibridación y ligación en un segmento de DNA completo. Tales métodos también son bien conocidos en la materia. Véase por ejemplo, Paterson *et al., Cell*, 48:441-452 (1987); y Lindley *et al., Proc. Natl. Acad. Sci*., 85:9199-9203 (1988), en los que un péptido recombinante, del factor activado por neutrófilos, se produjo de la expresión de un gen sintetizado químicamente en *E. coli*.

Un segmento de DNA como el descrito aquí puede utilizarse para la preparación de moléculas de rDNA, en la construcción de vectores para la expresión de una proteína o proteína de fusión estructural del NANBV de esta invención, o como una sonda de hibridación para la detección de la presencia de secuencias específicas de ácido nucleico del NANBV en las muestras.

Cuando la utilización de un segmento de DNA es para la preparación de proteínas, el residuo aminoacídico especificado se considera importante, y la secuencia de bases nucleotídicas del segmento de DNA puede variar basándose en la redundancia del código genético, como es bien conocido, para proporcionar la secuencia de residuos aminoacídicos deseada.

C. *Moléculas de DNA Recombinante*

La presente especificación también describe un DNA recombinante, (rDNA) que incluye un segmento de DNA de la presente invención ligado de forma operativa a un vector. Un rDNA preferible de la presente invención está

caracterizado por ser capaz de expresar directamente, en un huésped compatible, una proteína o proteína de fusión estructural del NANBV de esta invención. Los segmento de DNA preferidos para su utilización en un rDNA son aquellos descritos anteriormente aquí.

Por "expresar directamente " se entiende que la cadena de polipéptido maduro de la proteína está formada a partir de únicamente la traducción, en oposición a la escisión proteolítica de dos o más residuos aminoacídicos terminales de una proteína precursora traducida mayor. Los rDNA preferibles de la presente invención son los plásmidos pGEX-3X-690:694, pGEX-3X-693:691, pGEX-3X-690:691, pGEX-3X-15:17, pGEX-3X-15:18, pGEX-2T-15:17, pGEX-2T-CAP-A; pGEX-2T-CAP-B y pGEX-2T-CAP-A-B, descritos en el Ejemplo 1.

Una molécula de DNA recombinante (rDNA) como la descrita aquí puede obtenerse mediante la ligación de forma operativa de un vector a un segmento de DNA de la presente invención. Ejemplos de moléculas de rDNA y los métodos para su preparación se describen en el Ejemplo 1.

En otra realización, una molécula de rDNA como la descrita aquí comprende un vector ligado de forma operativa a un segmento de DNA que comprende una secuencia de bases nucleotídicas que codifica el genoma del aislamiento Hutch del NANBV. Preferiblemente, la molécula de rDNA incluye una secuencia de bases nucleotídicas que codifica la secuencia de residuos aminoacídicos de la poliproteína producida por el genoma del aislamiento Hutch c59, cuya secuencia de residuos aminoacídicos se muestra en el ID. de SEC. Nº:46 desde el residuo 1 al residuo 3011. Más preferiblemente, la molécula de rDNA en esta realización incluye una secuencia de bases nucleotídicas que se muestra en el ID. de SEC. Nº:46 desde la base 1 a la base 9416.

Como se utiliza aquí, el término "vector" significa una molécula de DNA capaz de realizar una replicación autónoma en una célula, a la cual se puede ligar de forma operativa otro segmento de DNA para conseguir la replicación del segmento insertado. Los vectores típicos son los plásmidos, bacteriófagos y similares. Los vectores capaces de dirigir la expresión de una proteína o proteína de fusión estructural del NANBV se denominan aquí "vectores de expresión". Por lo tanto, una molécula de DNA recombinante (rDNA) es una molécula de DNA híbrida que comprende al menos dos secuencias nucleotídicas que normalmente no se encuentran juntas en la naturaleza.

La elección del vector al cual se liga de forma operativa un segmento de DNA de la presente invención depende directamente, como es bien conocido en la materia, de las propiedades funcionales deseadas, por ejemplo la expresión de proteínas, y de la célula huésped a transformar, siendo éstas limitaciones inherentes en la construcción de moléculas de DNA recombinantes. Sin embargo, un vector como el descrito aquí es al menos capaz de dirigir la replicación, y preferiblemente también la expresión, del gen estructural de la proteína recombinante o de fusión incluido en los segmentos de DNA a los que está ligado de forma operativa.

En las realizaciones preferidas, el vector descrito aquí incluye un replicón procariótico (ori); es decir, una secuencia de DNA con la capacidad de dirigir una replicación y mantenimiento autónomos de la molécula de DNA recombinante de forma extracromosómica en una célula huésped procariótica, como una célula huésped bacteriana, transformada con ésta. Tales replicones son bien conocidos en la materia. Además, estas realizaciones que incluyen un replicón procariótico también incluyen normalmente un gen cuya expresión confiere la resistencia a un fármaco en un huésped bacteriano transformado con éste. Los genes bacterianos típicos de resistencia a un fármaco para su utilización en estos vectores, son aquellos que confieren resistencia a la ampicilina o tetraciclina. Algunos vectores plasmídicos típicos son el pUC8, pUC9, pBR322 y pBR329, disponibles en Biorad Laboratories, (Richmond, CA).

Aquellos vectores que incluyen un replicón procariótico también pueden incluir un promotor procariótico capaz de dirigir la expresión (transcripción y traducción) del gen que codifica una proteína o proteína de fusión estructural del NANBV en una célula huésped bacteriana, como *E. coli*, transformada con este. Un promotor es un elemento de control de la expresión formado por una secuencia de DNA que permite la unión de la polimerasa de RNA y que aparezca la subsiguiente iniciación de la transcripción. Las secuencias promotoras compatibles con huéspedes bacterianos se proporcionan normalmente en los vectores plasmídicos, que contienen dianas de restricción adecuadas para la inserción de un segmento de DNA de la presente invención. Un vector típico es el pPL-lambda disponible en Pharmacia, (Piscataway, NJ).

Los plásmidos pTTQ disponibles en Amersham (Arlington Heights, IL), y el plásmido pKK223-3 disponible en Pharmacia son vectores plasmídicos con un promotor bacteriano que es inducible por IPTG. Otros vectores de expresión para la producción en procariotas de un producto génico clonado en forma de una proteína de fusión son bien conocidos y están disponibles comercialmente.

Aunque los vectores de expresión pGEX-3X y pGEX-2T se han utilizado como ejemplo en la producción de las proteínas de fusión descritas aquí, puede utilizarse otros vectores de expresión funcionalmente equivalentes. Los vectores funcionalmente equivalentes contienen un promotor de la expresión que es inducible por IPTG para la expresión de proteínas de fusión en *E. coli*, y una configuración tal que tras la inserción del segmento de DNA en el vector se obtiene una proteína de fusión. Los vectores disponibles comercialmente que son funcionalmente equivalentes a los vectores pGEX-3X y pGEX-2T utilizados aquí incluyen el vector plasmídico pGEMEX-1 de Promega (Madison, WI) que produce una fusión entre la porción amino terminal de la proteína del gen 10 del T7 y el gen clonado insertado, el vector plasmídico pMAL de New England Biolabs (Beverly, MA) que produce una fusión con la proteína de unión

a la maltosa (MBF) codificada por el gen *mal* E, y los plásmidos pGEX-3X y pGEX-2T de Pharmacia que producen una fusión con la enzima glutatión-S-transferasa (GST) y el gen clonado insertado, respectivamente.

La construcción y la utilización de los vectores pGEX-3X y pGEX-2T han sido descritas por Smith *et al., Gene*, 67:31-40 (1988).

En las realizaciones particularmente preferidas, una proteína de fusión contiene una porción polipeptídica derivada de la GST como un dominio funcional añadido ligado de forma operativa a un antígeno estructural del NANBV de esta invención. Cualquier vector dirigido por un promotor inducible, como los vectores pTTQ, pKK223-3, pGEX-3X o pGEX-2T descritos anteriormente y similares, pueden utilizarse para la expresión de una proteína estructural GST-NANBV, denominada aquí una proteína de fusión GST:NANBV. Así, aunque los vectores pGEX-3X y pGEX-2T se describen en los ejemplos, las moléculas de DNA descritas aquí no se limitan a la construcción en estos vectores, ya que la descripción en una realización se dirige a un rDNA para la expresión de una proteína con antígenos estructurales del NANBV fusionados con la GST y no se centra en el vector *per se*.

Se han desarrollado una serie de métodos para unir de forma operativa los segmentos de DNA a vectores a través de extremos cohesivos complementarios. Por ejemplo, pueden añadirse fragmentos homopoliméricos complementarios al segmento de DNA a insertar y al DNA vector. El vector y el segmento de DNA se unen entonces mediante puentes de hidrógeno entre las colas homopoliméricas complementarias para formar moléculas de DNA recombinante.

Los enlazantes sintéticos que contienen una o más dianas de restricción proporcionan un método alternativo para la unión del segmento de DNA y los vectores. Un segmento de DNA generado mediante una digestión con endonucleasas de restricción se trata con la polimerasa de DNA del bacteriófago T4 o la polimerasa de DNA I de *E. coli*, unas enzimas que eliminan los extremos protuberantes 3' de cadena sencilla mediante su actividad exonucleolítica 3'-5' y rellenan los extremos 3' con su actividad polimerasa.

La combinación de estas actividades por lo tanto genera segmentos de DNA con extremos romos. Los segmentos con extremos romos se incuban entonces con un gran exceso molar de moléculas de enlazante en presencia de un enzima que es capaz de catalizar la ligación de moléculas de DNA con extremos romos, como la ligasa de DNA del bacteriófago T4. Así, los productos de la reacción son segmentos de DNA que contienen secuencias de enlazante poliméricas en sus extremos. Estos segmentos de DNA se escinden entonces con la enzima de restricción apropiada y se ligan a un vector de expresión que se ha escindido con una enzima que produce extremos compatibles con los del segmento de DNA.

Enlazantes sintéticos que contienen una serie de dianas de restricción por endonucleasas están disponibles comercialmente en varios distribuidores, lo que incluye International Biotechnologies, Inc., NewHaven, CN.

También se contempla en la presente invención los equivalentes de RNA de las moléculas de DNA recombinante anteriormente descritas.

D. *Células Transformadas y Cultivos*

La especificación también describe una célula huésped transformada con una molécula de DNA recombinante descritos aquí.

El término "célula huésped" incluye tanto huéspedes eucarióticos y procarióticos. Las moléculas de rDNA preferibles para su utilización en una célula transformada son las descritas aquí anteriormente, y preferiblemente son rDNA capaces de expresar una proteína recombinante o de fusión. Las realizaciones específicas preferibles de células transformadas son aquellas que contienen una molécula de rDNA con uno de los segmentos de DNA preferibles descritos aquí anteriormente, y en particular las células transformadas con el plásmido de rDNA pGEX-3X-690:694, pGEX-3X-693:691, pGEX-3X-690:691, pGEX-3X-15:17, pGEX-3X-15:18, pGEX-2T-15:17, pGEX-2T-CAP-A, pGEX-2T-CAP-B o pGEX-2T-CAP-A-B.

Las células bacterianas son células huésped procarióticas preferibles y normalmente son una cepa de *E. coli*, como por ejemplo, la cepa de *E. coli* DH5α disponible en Bethesda Research Laboratories, Inc., Bethesda, MD. La transformación de células huésped apropiadas con una molécula de DNA recombinante de la presente invención se consigue mediante métodos bien conocidos que normalmente dependen del tipo de vector utilizado. Respecto de la transformación de células huésped procarióticas, véase por ejemplo, Cohen *et al., Proc. Natl. Acad. Sci. USA*, 69:2110 (1972); y Sambrook *et al., Molecular Cloning, A Laboratory Manual*, 2ª Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989).

Las células transformadas correctamente, es decir, las células que contienen una molécula de DNA recombinante como las descritas aquí, pueden identificarse mediante técnicas bien conocidas. Por ejemplo, las células que resultan de la introducción de un rDNA como los descritos aquí, pueden aislarse como colonias separadas. Las células de estas colonias pueden recogerse, lisarse y examinar la presencia del rDNA en su contenido de DNA utilizando un método como el descrito por Southern, *J. Mol. Biol*., 98:503 (1975) o Berent *et al., Biotech*., 3:208 (1985).

Además del análisis directo de la presencia del rDNA, las células transformadas con el rDNA apropiado pueden identificarse mediante métodos inmunológicos bien conocidos cuando el rDNA es capaz de dirigir la expresión de una proteína estructural del NANBV. Por ejemplo, las células transformadas correctamente con un vector de expresión como los descritos aquí producen proteínas que muestran antigenicidad de la proteína estructural del NANBV. Se recogen muestras de las células que se sospecha se han transformado y se analiza la presencia de un antígeno estructural del NANBV utilizando anticuerpos específicos de ese antígeno. Tales anticuerpos se describen en más detalle aquí.

Así, además de las propias células huésped transformadas, la especificación también describe un cultivo de estas células, preferiblemente un cultivo monoclonal (clonalmente homogéneo), o un cultivo derivado de un cultivo monoclonal, en un medio de nutrientes. Preferiblemente, el cultivo también contiene una proteína que muestra antigenicidad de la proteína estructural del NANBV.

Los medios de nutrientes útiles para el cultivo de células huésped transformadas son bien conocidos en la materia y pueden obtenerse de varios distribuidores comerciales.

E. *Métodos para la Producción de Proteína, Polipéptidos y Proteínas de Fusión Estructurales del NANBV*

También se describe aquí un método para la producción de proteínas y proteínas de fusión recombinantes de esta invención.

El presente método implica la preparación de un cultivo que comprende un medio de nutrientes que contiene células huésped, preferiblemente células *E. coli*, transformada con una molécula de DNA recombinante que es capaz de expresar una proteína o una proteína de fusión estructural del NANBV. El cultivo se mantiene durante un periodo de tiempo suficiente para que las células transformadas expresen la proteína o proteína de fusión estructural del NANBV. La proteína expresada se recupera entonces a partir del cultivo.

Los vectores de expresión y las condiciones de cultivo de los vectores de expresión para la producción de proteínas estructurales del NANBV son generalmente bien conocidos en la materia. Tales vectores y condiciones de cultivo pueden alterarse sin afectar el espíritu de la presente invención. Sin embargo, los vectores preferibles son los diseñados específicamente para la producción de proteínas que no se encuentran normalmente en la célula huésped utilizada para expresar una proteína estructural del NANBV. Son ejemplos los vectores que contienen promotores inducibles para dirigir la expresión de segmentos de DNA que codifica la proteína estructural del NANBV. Los vectores con promotores inducibles por IPTG también son bien conocidos. Véanse por ejemplo los plásmidos pTTQ y pKK223-3 disponibles en Amersham y Pharmacia, respectivamente. Son particularmente preferibles los promotores inducibles por IPTG presentes en los vectores pGEX, pGEX-3X y pGEX-2T descritos aquí.

Utilizando vectores con promotores inducibles, la expresión de proteínas estructurales del NANBV requiere una fase de inducción al inicio del anteriormente descrito paso de mantenimiento para expresar la proteína, como es conocido y se describe en detalle en el Ejemplo 2.

Los métodos para la recuperación de la proteína expresada a partir de un cultivo son bien conocidos en la materia e incluyen el fraccionamiento de la porción del cultivo que contiene proteína utilizando técnicas bioquímicas bien conocidas. Por ejemplo, los métodos de gel filtración, cromatografía en gel, ultrafiltración, electroforesis, intercambio iónico, cromatografía de afinidad y similares, que son conocidos en el fraccionamiento de proteínas, pueden utilizarse para aislar las proteínas expresadas que se encuentran en el cultivo. Además, los métodos inmunoquímicos, como la inmunoafinidad, inmunoadsorción y similares pueden realizarse utilizando métodos bien conocidos.

Son particularmente preferibles los métodos de aislamiento que utilizan la presencia de la glutatión-S-transferasa (GST) para definir la porción polipeptídica como método para separar la proteína de fusión a partir de mezclas complejas de proteína. La adsorción por afinidad de una proteína de fusión que contiene la GST a una fase sólida que contiene glutatión adherido a ella, se puede conseguir como describen Smith *et al., Gene*, 67:31 (1988). Alternativamente, la porción polipeptídica de la proteína de fusión que contiene la GST puede separarse del antígeno estructural del NANBV mediante la escisión selectiva de la proteína de fusión en dianas de escisión proteolítica específica, de acuerdo con los métodos de Smith *et al*. Ejemplos de métodos de aislamiento se describen en los Ejemplos 5 y 6.

Además de prepararse mediante la utilización de un vector de expresión de rDNA, puede prepararse una proteína estructural del NANBV que comprenda un antígeno estructural del NANBV como un polipéptido sintético.

Los polipéptidos pueden sintetizarse mediante cualquiera de las técnicas que son conocidas para los expertos en la materia de los polipéptidos. Las técnicas de química sintética, como una síntesis tipo Merrifield en fase sólida, son preferibles por razones de pureza, especificidad antigénica, eliminación de productos colaterales no deseados, facilidad de producción y similares, y pueden realizarse de acuerdo con los métodos descritos en Merrifield *et al., J. Am. Chem. Soc*., 85:2149-2154 (1963) y Houghten *et al., Int. J. Pept. Prot. Res*., 16:311-320 (1980). Un resumen excelente de una serie de técnicas disponibles puede encontrarse en J. M. Steward y J. D. Young, "Solid Phase Peptide Synthesis", W. H. Freeman Co., San Francisco, 1969; M. Bodanszky et al., "Peptide Synthesis", John Wiley & Sons, 2ª Edición, 1976, y J. Meienhofer, "Hormonal Proteins and Peptides", Vol. 2, pág. 46, Academic Press (NY), 1983, para la síntesis de péptidos en fase sólida, y E. Schroder y K. Kubke, "The peptides", Vol. 1, Academic Press (New York), 1965, para la síntesis en solución clásica.

Los grupos protectores apropiados útiles en tal síntesis se describen en los textos anteriores y en J. F. W. McOmie, "Protective Groups in Organic Chemistry", Plenum Press, New York, 1973.

Un polipéptido sujeto incluye cualquier derivado químico de un polipéptido cuya secuencia de residuos aminoacídicos se muestra aquí. Por lo tanto, el presente polipéptido puede someterse a varios cambios que proporcionen ciertas ventajas para su utilización.

"Derivado químico" se refiere al polipéptido sujeto con uno o más residuos derivados químicamente mediante la reacción de un grupo lateral funcional. Tales moléculas derivadas incluyen por ejemplo, aquellas moléculas en las que los grupos amino libres se han derivado para formar clorhidratos amina, grupos p-toluenosulfonilo, grupos carbobenzoxi, grupos t-butiloxicarbonilo, grupos cloroacetilo o grupos formilo. Los grupos carboxilo libres pueden derivarse para formar sales, ésteres de metilo y etilo u otros tipos de ésteres o hidrazidas. Los grupos hidroxilo libres pueden derivarse para formar derivados O-acilo u O-alquilo. El nitrógeno imidazol de la histidina puede derivarse para formar N-im-bencilhistidina. También se incluyen como derivados químicos aquellos péptidos que contienen uno o más derivados de los que aparecen en la naturaleza de los veinte aminoácidos estándar. Por ejemplo: la 4-hidroxiprolina puede sustituir a la prolina; la 5-hidroxilisina puede sustituir a la lisina; la 3-metilhistidina puede sustituir a la histidina; la homoserina puede sustituir a la serina; y la ornitina puede sustituir a la lisina. Los polipéptidos de la presente invención también incluyen cualquier polipéptido con una o más adiciones en relación con la secuencia de un polipéptido, cuya secuencia se muestra aquí, siempre que la actividad necesaria se mantenga.

Los residuos adicionales también pueden añadirse en alguno de los extremos con el propósito de proporcionar un "enlazante" mediante el que los polipéptidos de esta invención pueden fijarse convenientemente a un marcaje, matriz sólida o transportador. Preferiblemente los residuos enlazante no formar antígenos estructurales del NANBV.

Los marcajes, matrices sólidas y transportadores que pueden utilizarse con los polipéptidos de esta invención se describen aquí a continuación.

Los enlazantes de residuos aminoacídicos normalmente son de al menos un residuo y pueden ser de 40 o más residuos, más frecuentemente de 1 a 10 residuos, pero no forman epítopos del NANBV. Los residuos aminoacídicos típicos utilizados para la unión son la tirosina, cisteína, lisina, ácido glutámico y aspártico, o similares. Además, un polipéptido sujeto puede diferir, a no ser que se especifique de otro modo, de la secuencia natural de la poliproteína del NANBV mediante la modificación de la secuencia por acilación en el extremo NH_2, por ejemplo, acetilación o amidación con ácido tioglicólico, mediante amidación carboxílica terminal, por ejemplo, con amonio, metilamina y similares.

Si se acopla a un transportador para formar lo que se conoce en la materia como un conjugado transportador-hapteno, un polipéptido de la presente invención es capaz de inducir anticuerpos que inmunoreaccionan con el NANBV. En vista del principio bien establecido de la reactividad cruzada inmunológica, la presente invención contempla las variantes antigénicamente relacionadas de los polipéptidos descritos aquí. Una "variante antigénicamente relacionada" es un polipéptido sujeto que es capaz de inducir la inmunoreacción de las moléculas de anticuerpo con un polipéptido descrito en esta especificación y con el NANBV.

Cualquier péptido descrito aquí puede utilizarse en forma de una sal farmacéuticamente aceptable. Los ácidos adecuados que son capaces de formar sales con los péptidos de la presente invención incluyen ácidos inorgánicos como el ácido clorhídrico, ácido bromhídrico, ácido perclórico, ácido nítrico, ácido tiociánico, ácido sulfúrico, ácido fosfórico acético, ácido propiónico, ácido glicólico, ácido láctico, ácido pirúvico, ácido oxálico, ácido malónico, ácido succínico, ácido maleico, ácido fumárico, ácido antranílico, ácido cinámico, ácido naftalenosulfónico, ácido sulfanílico o similares.

Las bases adecuadas capaces de formar sales con los péptidos de la presente invención incluyen las bases inorgánicas como el hidróxido sódico, hidróxido amónico, hidróxido potásico y similares; y bases orgánicas como las mono-, di- y tri-alquilaminas y arilaminas (por ejemplo trietilamina, diisopropilamina, metilamina, dimetilamina y similares),y etanolaminas opcionalmente sustituidas (por ejemplo etanolamina, dietanolamina y similares).

En general, los métodos de síntesis en fase sólida contemplados comprenden la adición secuencial de uno o más residuos aminoacídicos o de residuos aminoacídicos adecuadamente protegidos a una cadena peptídica en crecimiento. Normalmente, el grupo amino o el carboxilo del primer residuo aminoacídico está protegido mediante un grupo protector adecuado que puede eliminarse de forma selectiva. Un grupo protector que puede eliminarse de forma selectiva diferente se utiliza para los aminoácidos que contienen un grupo lateral reactivo como la lisina.

Utilizando la síntesis en fase sólida como ejemplo, el aminoácido protegido o derivado está unido a un soporte sólido inerte a través de su grupo carboxilo o amino no protegido. El grupo protector del grupo amino o carboxilo se elimina entonces de forma selectiva y el siguiente aminoácido en la secuencia con el grupo de cortesía (amino o carboxilo) adecuadamente protegido, se mezcla y se hace reaccionar bajo la condiciones adecuadas para formar el enlace amida con el residuo ya unido al soporte sólido. El grupo protector se elimina entonces de este residuo aminoacídico recién añadido, y entonces se añade el siguiente aminoácido (adecuadamente protegido), y así repetidamente. Después de que todos los aminoácidos deseados se han unido en la secuencia adecuada, cualquier grupo protector de los grupos laterales y terminales restantes (y del soporte sólido) se eliminan de forma secuencial o concurrente, para conseguir el polipéptido final.

F. *Composiciones de Proteína y Proteína de Fusión Estructural del NANBV*

En otra realización, la presente especificación describe una composición que contiene una proteína estructural del NANBV, preferiblemente aislada que comprende una secuencia de residuos aminoacídicos que define un antígeno estructural del NANBV como el descrito aquí.

Por aislado se entiende que una proteína estructural del NANBV de esta invención está presente en una composición como un constituyente proteico mayoritario, normalmente en cantidades superiores al 10% de la proteína total en la composición, pero preferiblemente en cantidades superiores al 90% de la proteína total en la composición.

Un antígeno estructural del NANBV, como se utiliza aquí, es una proteína estructural codificada por el genoma del NANBV y tiene las propiedades de un antígeno como el definido aquí, es decir, es capaz de inmunoreaccionar específicamente con un anticuerpo. Se ha intentado diseñar las proteínas estructurales del NANBV de la cápside y la cubierta, y se ha caracterizado parcialmente como se describe aquí que contienen los antígenos estructurales de la cápside y la cubierta del NANBV, respectivamente.

Un antígeno de la cápside del NANBV como el descrito aquí comprende una secuencia de residuos aminoacídicos que está inmunológicamente relacionada a nivel de secuencia con el presunto antígeno de la cápside de la cepa Hutch del NANBV, cuya secuencia está contenida en el ID. de SEC. Nº:1 desde el residuo 1 al residuo 120.

Por "inmunológicamente relacionado" se entiende que presenta suficiente homología en la secuencia aminoacídica entre las dos secuencias proteicas a comparar de forma que los anticuerpos específicos de una proteína inmunoreaccionan (reaccionan de forma cruzada) con la otra proteína. La reactividad cruzada inmunológica puede medirse mediante métodos bien conocidos, lo que incluye los métodos de inmunoensayo descritos aquí.

Como se utiliza aquí, la frase "proteína recombinante" se refiere a una proteína de al menos 20 residuos aminoacídicos de longitud, y preferiblemente al menos 50 residuos, que incluye una secuencia de residuos aminoacídicos, que corresponde, y preferiblemente es idéntica, a una porción de la proteína estructural del NANBV contenida en el ID. de SEC. Nº:1.

En realizaciones preferibles, una proteína estructural del NANBV incluye una secuencia de residuos aminoacídicos que está inmunológicamente relacionada y preferiblemente es idéntica a la secuencia contenida en el ID. de SEC. Nº:1 desde el residuo 1 al residuo 20, desde el residuo 21 al residuo 40, desde el residuo 2 al residuo 40, o desde el residuo 1 al residuo 74. La proteína estructural del NANBV con la secuencia indicada es particularmente preferible para su utilización en métodos y sistemas diagnósticos, ya que se ha demostrado aquí que los antígenos de la cápside contenidos en éstos son particularmente útiles para la detección de una infección aguda por NANBV. Las proteínas estructurales relacionadas con el NANBV incluyen una secuencia contenida en el ID. de SEC. Nº:1 desde el residuo 1 al residuo 120, desde el residuo 1 al residuo 176, y desde el residuo 1 al residuo 326. Son un ejemplo las proteínas descritas aquí con una secuencia contenida en ID. de SEC. Nº:2 desde el residuo 1 al residuo 315, en el ID. de SEC. Nº:3 desde el residuo 1 al residuo 252, en el ID. de SEC. Nº:4 desde el residuo 1 al residuo 252, o en el ID. de SEC. Nº:6 desde el residuo 1 al residuo 271.

En otra realización, una proteína estructural del NANBV incluye una secuencia de residuos aminoacídicos que está inmunológicamente relacionada, y preferiblemente es idéntica, a la secuencia contenida en el ID. de SEC. Nº:1 desde el residuo 69 al residuo 120. Un ejemplo de proteína estructural del NANBV tiene la secuencia de la proteína expresada codificada por el plásmido de rDNA pGEX-3X-693:691.

En otra realización, se contempla una proteína estructural del NANBV que comprende una secuencia de residuos aminoacídicos de acuerdo con un polipéptido descrito aquí.

En realizaciones preferibles, una proteína estructural del NANBV está esencialmente libre tanto de antígenos procarióticos (es decir, antígenos específicos de la célula huésped) como de otras proteínas relacionadas con el NANBV. Por "esencialmente libre" se entiende que la proporción entre el antígeno estructural del NANBV y el antígeno extraño, como el antígeno procariótico, u otras proteínas relacionadas con el NANBV, es de al menos 10:1, preferiblemente es de 100:1, y más preferiblemente es de 200:1.

La presencia y cantidad de proteína contaminante en un preparado de proteína estructural del NANBV puede determinarse mediante métodos bien conocidos. Preferiblemente, se somete una muestra de la composición a electroforesis en gel de poliacrilamida con dodecilsulfato sódico (SDS-PAGE) para separar la proteína estructural del NANBV de cualquier contaminante proteico presente. La proporción entre las cantidades de proteínas presentes en la muestra se determina entonces mediante barrido láser densitométrico suave, como es bien conocido en la materia. Véase Guilian *et al., Anal. Biochem.*, 129:277-287 (1983).

Una proteína estructural del NANBV puede obtenerse como una proteína aislada, y más preferiblemente esencialmente libre de antígenos procarióticos o antígenos no estructurales del NANBV mediante los métodos descritos aquí para la producción de proteínas estructurales del NANBV. Son particularmente preferibles los métodos que se basan en las propiedades de una región polipeptídica de una proteína de fusión, cuya región estápresente en la proteína de fusión para facilitar la separación de la proteína de fusión de las proteínas de la célula huésped en base a su afinidad.

Son un ejemplo las proteínas de fusión que contienen la GST, cuyas secuencias de residuos aminoacídicos están contenidas en los ID. de SEC. Nº:2, 3, 4 o 6, en los que cada región polipeptídica de la GST le proporciona a la proteína de fusión un dominio funcional con afinidad de unión al sustrato normal de la GST, es decir el glutatión. La purificación de una proteína de fusión con una región polipeptídica de la GST se describe en más detalle aquí.

En una realización relacionada, la invención describe un polipéptido que define un antígeno del NANBV. Así, la invención contempla un polipéptido que corresponde a una región de la poliproteína del NANBV que define un determinante antigénico del virus que es útil como antígeno del NANBV en ensayos serológicos o en un inóculo para la inducción de antisueros anti-NANBV, como se describe aquí.

Un polipéptido descrito aquí comprende una secuencia de aminoácidos de entre alrededor de 7 y alrededor de 200 residuos de longitud, preferiblemente entre alrededor de 20 y 150 residuos de longitud, que comprende la secuencia de residuos aminoacídicos definida por la secuencia nucleotídica de un polinucleótido descrito aquí.

Un polipéptido preferible comprende una secuencia de residuos aminoacídicos que incluye una secuencia de residuos aminoacídicos seleccionada de entre el grupo de secuencias que consiste en: desde el residuo 391 al residuo 404 del ID. de SEC. Nº:46, desde el residuo 246 al residuo 256 del ID. de SEC. Nº:46, desde el residuo 461 al residuo 466 del ID. de SEC. Nº:46, desde el residuo 473 al residuo 482 del ID. de SEC. Nº:46, y desde el residuo 2356 al residuo 2379 del ID. de SEC. Nº:46. En realizaciones particularmente preferibles, el polipéptido tiene una secuencia de residuos aminoacídicos que corresponde a la secuencia que se muestra en el ID. de SEC. Nº:46.

En tanto que un polipéptido es útil para la distinción de los aislamientos Hutch, la especificación describe un polipéptido con una longitud de entre alrededor de 7 y alrededor de 200 residuos aminoacídicos y que comprende una secuencia de residuos aminoacídicos que corresponde a una porción de la secuencia del aislamiento Hutch c59 del NANBV que se muestra en el ID. de SEC. Nº:46. En esta realización, el polipéptido tiene al menos una diferencia entre los residuos aminoacídicos de la secuencia si se compara con la secuencia de residuos aminoacídicos de un aislamiento del NANBV seleccionado de entre el grupo que consiste en el HCV-1, HCV-BK, HCV-J, HC-J1, HC-J4, HCV-JH y HCV-Hh.

Preferiblemente, un polipéptido es inmunoreactivo con los antisueros anti-cepa Hutch del NANBV cuando se mide en inmunoensayos serológicos estándar como los descritos aquí.

En otra realización, también se contempla una composición que comprende una proteína de fusión aislada de la presente invención, que comprende una proteína estructural del NANBV de esta invención ligada de forma operativa en uno o ambos extremos a otro polipéptido mediante un enlace peptídico. El polipéptido añadido puede ser cualquier polipéptido diseñado para aumentar los dominios funcionales presentes en la proteína de fusión. Los dominios funcionales añadidos se incluyen para proporcionar epítopos inmunogénicos adicionales, para añadir masa a la proteína de fusión, para alterar la solubilidad de la proteína de fusión, para proporcionar un método para el aislamiento basado en la afinidad de la proteína de fusión, y similares. Son un ejemplo de dominios funcionales añadidos las dianas de escisión específicas de la trombina o el factor Xa que se proporcionan cuando una proteína de fusión sujeto se produce en los vectores pGEX-3X o pGEX-2T, respectivamente, como se describe aquí. Un ejemplo adicional de dominio es el dominio proteico derivado de la GST, que permite un aislamiento rápido utilizando una cromatografía de afinidad a una fase sólida que contiene glutatión fijado en ella.

Un dominio que contiene una diana de escisión de la trombina o el factor Xa se utiliza aquí, en una realización, para permitir la producción de una proteína estructural del NANBV libre del dominio funcional de la GST. Es un ejemplo la proteína producida en el ejemplo 6, con la secuencia de residuos aminoacídicos contenida en el ID. de SEC. Nº:2 desde el residuo 226 al residuo 315. El dominio que contiene la diana de escisión del factor Xa también se utiliza en el vector de expresión de proteínas de fusión pMAL, descrito aquí y comercialmente disponible en New England Biolabs (Beverly, MA).

En una realización relacionada, una proteína estructural del NANBV se obtiene mediante la escisión con trombina de una proteína producida utilizando el vector pGEX-2T, como una proteína con una secuencia de residuos aminoacídicos contenida en el ID. de SEC. Nº:3 desde el residuo 225 al residuo 252, en el ID. de SEC. Nº:4 desde el residuo 225 al residuo 252, o en el ID. de SEC. Nº:6 desde el residuo 225 al residuo 271.

Una proteína de fusión descrita aquí incluye una secuencia de residuos aminoacídicos que corresponde desde su extremo amino a su extremo carboxi a la secuencia de residuos aminoacídicos contenida en el ID. de SEC. Nº:1desde el residuo 1 al residuo 20, desde el residuo 21 al residuo 40, desde el residuo 2 al residuo 40, desde el residuo 1 al residuo 74, desde el residuo 69 al residuo 120, desde el residuo 121 al residuo 176, o desde el residuo 121 al residuo 326. Una proteína de fusión preferible tiene una secuencia que corresponde, y más preferiblemente es idéntica, a la secuencia de residuos aminoacídicos en el ID. de SEC. Nº:2 desde el residuo 1 al residuo 315, en el ID. de SEC. Nº:3 desde el residuo 1 al residuo 252, en el ID. de SEC. Nº:4 desde el residuo 1 al residuo 252, o en el ID. de SEC. Nº:6 desde el residuo 1 al residuo 271. Otras proteínas de fusión preferibles están definidas por la secuencia de residuos aminoacídicos de la secuencia codificante de la proteína expresada presente en los plásmidos de rDNA pGEX-3X-690:694, pGEX-3X-690:691, pGEX-3X-693:691, pGEX-3X-15:17, pGEX-3X-15:18, pGEX-2T-15:17, pGEX-2T-CAP-A, pGEX-2T-CAP-B, y pGEX-2T-CAP-A-B.

La frase "proteína de fusión", cuando se utiliza aquí se refiere a una proteína aislada como se ha definido para una proteína estructural del NANBV como aquí se describe. Así, una proteína de fusión aislada es una composición con una proteína de fusión como la descrita aquí en cantidades superiores al 10 por ciento de la proteína total de la composición, y preferiblemente superiores al 90 por ciento de la proteína total de la composición.

Una proteína de fusión preferible es una proteína de fusión, es decir, una proteína de fusión que contiene una porción polipeptídica derivada de una proteína originada en una especie heteróloga de virus, organismo, patógeno o animal, es decir, una proteína que no pertenece al NANBV. Preferiblemente, una proteína de fusión heteróloga contiene una porción polipeptídica que no pertenece al NANBV que no está inmunológicamente relacionada con un antígeno estructural del NANBV descrito aquí.

Una proteína de fusión puede contener un dominio funcional que proporcione un epítopo inmunogénico o antigénico diferente del antígeno estructural del NANBV definido aquí, y que es preferiblemente derivado de un patógeno distinto o de varios patógenos. El dominio funcional es inmunogénico donde este dominio está presente formando una vacuna o inmunógeno polivalente, con el propósito de inducir anticuerpos inmunoreactivos tanto con la proteína estructural del NANBV como con un segundo patógeno. El dominio funcional es antigénico donde este dominio está presente formando un antígeno polivalente para su utilización en sistemas y métodos diagnósticos para la detección de al menos dos especies de anticuerpos.

Son de particular interés en esta realización las proteínas de fusión diseñadas incluyendo un dominio funcional que se derive de otros virus causantes de hepatitis, como el virus de la hepatitis B yel virus de la hepatitis A. Estos virus están bien caracterizados y contienen·los determinantes antigénicos y determinantes inmunogénicos adecuados para su utilización en proteínas de fusión como las descritas aquí, y proporcionan la ventaja de los reactivos bioquímicos con varios usos, tanto en aplicaciones diagnósticas como en vacunas. De forma adicional, el dominio funcional incluido puede contener secuencias aminoacídicas de otros patógenos, preferiblemente aquellos que también pueden infectar a individuos con hepatitis por NANBV, como el HIV.

Las proteínas o proteínas de fusión estructurales del NANBV preferibles que comprenden un antígeno estructural del NANBV de la presente invención están en forma no reducida, es decir, están sustancialmente libres de grupos sulfhidrilo a causa de los puentes intramoleculares Cys-Cys.

En composiciones preferibles, la proteína o proteína de fusión estructural del NANBV como se describe aquí, está presente, por ejemplo, en composiciones líquidas como las suspensiones o soluciones estériles, o como preparaciones isotónicas que contienen conservantes adecuados.

Una de estas composiciones útiles para la inducción de anticuerpos anti-proteína estructural del NANBV en un mamífero se denomina una vacuna y contiene una proteína o proteína de fusión estructural del NANBV como los descritos aquí.

H. *Composiciones de Anticuerpo*

Un anticuerpo para su utilización en el contexto de esta invención es una composición que contiene moléculas de anticuerpo que inmunoreaccionan con un antígeno estructural del NANBV, con el aislamiento Hutch del NANBV, preferiblemente el aislamiento c59, y con una proteína, polipéptido o proteína de fusión estructural del NANBV de la presente invención (moléculas de anticuerpo anti-proteína estructural del NANBV). Un anticuerpo preferible contiene moléculas de anticuerpo que inmunoreaccionan con un epítopo presente en un polipéptido con una secuencia de residuos aminoacídicos contenida en el ID. de SEC. Nº:1 desde el residuo 1 al residuo 326, preferiblemente que inmunoreacciona con un polipéptido con la secuencia contenida en el ID. de SEC. Nº:1 desde el residuo 1 al residuo 20, desde el residuo 21 al residuo 40,desde el residuo 2 al residuo 40, desde el residuo 1 al residuo 74, desde el residuo 49 al residuo 120, o desde el residuo 121 al residuo 326.

Además, es preferible que las moléculas de anticuerpo anti-proteína estructural del NANBV no inmunoreaccionen con los aislamientos HCV-1, HCV-BK, HCV-J, HC-J1, HC-J4, HCV-JH o HCV-Hh del NANBV o con el antígeno C-100-3 descrito aquí, y que están disponibles en el ensayo comercial disponible en Ortho Diagnostics, Inc.

Típicamente, un anticuerpo puede producirse mediante la inmunización de un mamífero con un inóculo que contiene aislamiento Hutch c59 o una proteína o polipéptido estructural del NANBV de esta invención, induciendo así en el mamífero moléculas de anticuerpo con inmunoespecificidad por los antígenos estructurales del NANBV descritos aquí. Las moléculas de anticuerpo se recogen entonces del mamífero y se aíslan hasta el punto deseado mediante técnicas bien conocidas como, por ejemplo, utilizando DEAE Sephadex para obtener la fracción de IgG.

Para aumentar la especificidad del anticuerpo, los anticuerpos pueden purificarse mediante cromatografía de inmunoafinidad utilizando la proteína estructural del NANBV inmunizante fijada en la fase sólida. El anticuerpo se pone en contacto con la proteína estructural del NANBV fijada en fase sólida durante un periodo de tiempo suficiente para que la proteína estructural del NANBV inmunoreaccione con las moléculas de anticuerpo para formar un inmunocomplejo fijado en fase sólida. Los anticuerpos unidos se separan del complejo mediante técnicas estándar.

Para obtener una composición de anticuerpo que no inmunoreacciona con el antígeno C-100-3 o el aislamiento del NANBV identificado anteriormente, los métodos de inmunoadsorción se utilizan para eliminar las inmunoespecificidades no deseadas. En general, los métodos de inmunoadsorción para eliminar las inmunoespecificidades son bien conocidos e involucran inicialmente la puesta en contacto de la composición de anticuerpo con una fase sólida con uno o más de los antígenos o aislamientos del NANBV fijados en ella, para formar una mezcla de inmunoadsorción. Preferiblemente, existe un exceso de antígeno o NANBV en la fase sólida en proporción a los anticuerpos en la composición que tiene las inmunoespecificidades no deseadas en la mezcla de inmunoadsorción.

La mezcla de inmunoadsorción se mantiene bajo condiciones de inmunoreacción y durante un periodo de tiempo suficiente para que se forme un inmunocomplejo en la fase sólida. A continuación, las fases líquidas y sólidas se separan, y la fase líquida seretiene teniendo las moléculas de anticuerpo no deseado separadas por inmunoadsorción en la fase sólida.

Es particularmente preferible una composición de anticuerpo que contiene antisueros específicos del c59, obtenidos mediante inmunización con el aislamiento Hutch c59 o preferiblemente con un polipéptido como los descritos aquí, seleccionado con una secuencia de residuos aminoacídicos específica del c59 como se ha definido aquí, y preferiblemente derivado de las regiones variables V, V1, V2 o V3 del NANBV. A continuación, la composición de anticuerpo producida se inmunoadsorbe para eliminar los anticuerpos inmunoreactivos con los aislamientos del NANBV diferentes del c59 como los descritos aquí.

Los anticuerpos producidos de este modo pueden utilizarse, *inter alia*, en métodos y sistemas diagnósticos de la presente invención para detectar antígenos estructurales del NANBV como los descritos aquí que estén presentes en una muestra corporal.

La palabra "inóculo" en sus diferentes formas gramaticales se utiliza aquí para describir una composición que contiene un antígeno estructural del NANBV de esta invención como un ingrediente activo, utilizado para la preparación de anticuerpos inmunoreactivos con antígenos estructurales del NANBV.

La obtención y utilización de un inóculo para la producción de un anticuerpo de esta invención tiene un gran paralelismo con las descripciones que se dan aquí de una vacuna, siempre que la vacuna también se haya diseñado para inducir la producción de anticuerpos, y es un ejemplo de la obtención y utilización de un inóculo. Una diferencia clave es que el inóculo está formulado para su utilización en un animal en lugar de en un humano, como es bien conocido.

Un anticuerpo preferible es un anticuerpo monoclonal, y puede utilizarse de la misma forma a la descrita aquí para los anticuerpos descritos aquí.

Un anticuerpo monoclonal está normalmente compuesto de anticuerpos producidos por clones de una única célula llamada hibridoma que secreta (produce) sólo un tipo de molécula de anticuerpo. La célula de hibridoma se forma mediante la fusión de una célula productora de anticuerpo y un mieloma u otra línea celular que se autoperpetúe. La obtención de tales anticuerpos se describieron inicialmente en Kohler y Milstein, *Nature* 256:495-497 (1975).

Puede analizarse la inmunoreactividad de los sobrenadantes del hibridoma preparados así con un antígeno estructural del NANBV, como la proteína estructural del NANBV utilizada en el inóculo para inducir la célula productora de anticuerpos. Otros métodos para producir anticuerpos monoclonales, la célula del hibridoma, y los cultivos celulares de hibridoma son también bien conocidos.

También se contempla en esta invención la célula del hibridoma, y los cultivos que contienen una célula del hibridoma que produce un anticuerpo monoclonal de esta invención.

Debe entenderse que además de los ingredientes transportadores anteriormente mencionados, la formulación farmacéutica descrita aquí puede incluir, según sea apropiado, uno o más ingredientes transportadores adicionales como diluyentes, tampones, agregantes, agentes activos de superficie, espesantes, lubricantes, conservantes (incluyendo los antioxidantes) y similares, y sustancias incluidas con el propósito de dar lugar a una formulación isotónica con la sangre del recipiente. Normalmente, se añade un conservante como el mertiolato (a una dilución 1:5000 de una solución al 1%) para eliminar el riesgo de contaminación microbiana, incluso si se utilizan técnicas estériles en la elaboración del inóculo.

I. *Sistemas de Diagnóstico y Métodos*

1. *Sistemas de Diagnóstico*

La presente invención contempla un sistema de diagnóstico para buscar la presencia de anticuerpos anti-NANBV o antígenos estructurales del NANBV en una muestra corporal de acuerdo con los métodos diagnósticos descritos aquí.

Un sistema de diagnóstico en forma de equipo incluye, en una cantidad suficiente para que al menos un ensayo de acuerdo con los métodos descritos aquí, una proteína estructural de NANBV, un polipéptido o proteína de fusión o una combinación de los mismos descritos aquí, o una composición de anticuerpos anti-NANBV descrita aquí, como

un reactivo empaquetado de forma separada. Normalmente, también se incluyen instrucciones de uso del reactivo empaquetado.

"Instrucciones de uso" incluye normalmente una expresión tangible de la descripción de la concentración del reactivo o al menos un parámetro del método de ensayo como las cantidades relativas del reactivo y la muestra a ser mezclada, periodos de tiempo de mantenimiento de las mezclas de muestras/reactivo, temperatura, condiciones de los tampones y similares.

En realizaciones preferidas, un sistema de diagnóstico de la presente invención incluye además un marcaje o medio indicativo capaz de señalizar la formación de un complejo que contienen un antígeno estructural del NANBV, una proteína recombinante o un anticuerpo anti-NANBV.

Tal como se utiliza aquí, los términos "marcaje" y "medio indicativo" en sus diferentes formas gramaticales se refiere a átomos sencillos y moléculas que están directa o indirectamente involucradas en la producción de una señal detectable para indicar la presencia de un complejo. Cualquier marcaje o medio indicativo puede unirse a o incorporarse en una especie de reactivo como un anticuerpo o anticuerpo monoclonal, o puede utilizarse de forma separada, y aquellos átomos o moléculas pueden utilizarse solos o junto a otros reactivos adicionales. Tales marcajes son bien conocidos en la química del diagnóstico clínico.

El marcaje puede ser un agente marcador fluorescente que se une químicamente a anticuerpos o antígenos sin desnaturalizarlos para formar un fluorocromo (colorante) que es un trazador inmunofluorescente útil. Los agentes marcadores fluorescentes adecuados son los fluorocromos como el isocianato de fluoresceína (FIC}, isotiocianato de fluoresceína (FITC), cloruro de 5-dimetilamina-1-naftalensulfonilo (DANSC), isotiocianato de tetrametilrodamina (TRITC), lisamina, rodamina 8200 cloruro de sulfonilo (RB 200 SC), una unión de quelato lantánido (por ejemplo Eu, Tb, Sm) y similares. Una descripción de las técnicas de análisis de inmunofluorescencia se encuentra en DeLuca, "Immunofluorescence Analysis", en *Antibody As a Tool*, Marchalonis, *et al*., eds., John Wiley & Sons, Ltd., págs. 189-231 (1982).

En realizaciones preferidas, el marcaje es una enzima, como la peroxidasa de rábano picante (HRP), oxidasa de glucosa, fosfatasa alcalina o similares. En los casos en que el marcaje principal es una enzima como HRP u oxidasa de glucosa, son necesarios reactivos adicionales para visualizar el hecho de que se ha formado un complejo anticuerpo-antígeno (inmunoreactante). Tales reactivos adicionales para la HRP incluyen peróxido de hidrógeno y un precursor de colorante de oxidación como la dianinobencidina. Otro reactivo útil con HRP es el ácido 2,2'-azinobis-(3-etilbenzotiazolina-6-sulfónico) (ABTS)

Los elementos radioactivos son también agentes marcadores útiles y se utilizan aquí de forma ilustrativa. Un ejemplo de agente de radiomarcaje es un elemento radioactivo que produce emisiones de rayos gamma. Los elementos que emiten rayos gamma por sí mismos, como el ^{124}I, ^{125}I, ^{128}I, ^{131}I y ^{51}Cr representan una clase de grupos indicadores de elementos radioactivos que producen emisiones de rayos gamma. Particularmente preferido es el ^{125}I. Otro grupo de medio de marcaje útil son aquellos elementos como el ^{11}C, ^{18}H, ^{15}O y ^{13}N que por sí mismos emiten positrones. Los positrones emitidos de esta forma, producen rayos gamma cuando se encuentran con electrones presentes en el cuerpo del animal. También es útil un emisor beta, como el Indio 111, ^{3}H, ^{35}S, ^{14}C, o ^{32}P.

Se han descrito marcajes adicionales en el campo y son adecuados para utilizar en los sistemas de diagnóstico de esta invención. Por ejemplo, se puede utilizar la afinidad específica encontrada entre los pares de moléculas, una como una marcaje fijado al agente específico de unión y el otro como medio de detección de la presencia del marcaje. Pares ejemplares son biotina:avidina, en la que la biotina es el marcaje, y peroxidasa:anti-peroxidasa (PAP), en la que la peroxidasa es el marcaje.

La unión de marcajes, es decir el marcaje de polipéptidos y proteínas, es bien conocido en la materia. Por ejemplo, las moléculas de anticuerpo producidas mediante un hibridoma pueden marcarse mediante la incorporación metabólica de aminoácidos que contienen un radioisótopo proporcionado como un componente del medio de cultivo. Véase, por ejemplo, Galfre *et al., Meth. Enzymol*, 73:3-46 (1981). Se pueden aplicar concretamente las técnicas de conjugación de proteínas o acoplamiento a través de grupos funcionales activados. Véase, por ejemplo, Aurameas, *et al., Scand. J. Immunol*., Vol. 8 Supl. 7:7-23 (1978) Rodwell *et al., Biotech* 3:889-894 (1984), y Patente EEUU Nº. 4.493.795.

Los sistemas de diagnóstico pueden también incluir, preferiblemente como un paquete separado, un agente de unión específico. Un "agente de unión específico" es una entidad molecular capaz de unir de forma selectiva una especie de reactivo, que a su vez es capaz de reaccionar con un producto descrito aquí, pero que no es en sí un producto de expresión de proteína descrito aquí.

Ejemplos de agentes de unión específicos son moléculas de anticuerpo como los anti-IgG humana o anti-IgM humana, proteínas del complemento o fragmentos de las mismas, proteína A,·y similares. Preferiblemente el agente de unión específico puede unirse al anticuerpo anti-NANBV para ser detectado cuando el anticuerpo está presente como parte de un inmunocomplejo.

En realizaciones preferidas el agente de unión específico se marca. Sin embargo, cuando el sistema de diagnóstico incluye un agente de unión específico, el agente se utiliza normalmente como un medio de amplificación o reactivo.

En estas realizaciones, el agente de unión específico marcado es capaz de unirse de forma específica al medio de amplificación cuando el medio de amplificación está unido al complejo que contiene una especie de reactivo.

Los equipos de diagnóstico de la presente invención pueden utilizarse en un formato "ELISA" para detectar la presencia o cantidad de anticuerpos en una muestra de fluido corporal como suero, plasma o saliva. "ELISA" se refiere a un ensayo inmunosorbente unido a enzima que utiliza un anticuerpo o antígeno anido a una fase sólida y un conjugado enzima-antígeno o enzima-anticuerpo para detectar y cuantificar la cantidad de un antígeno o anticuerpo presente en una muestra. Una descripción de la técnica de ELISA se encuentra en el capítulo 22 de la 4ª Edición de *Basic and Clinical Immunology* por D. P. Sites *et al*., publicado por Lange Medical Publications de Los Altos, CA en 1982 y en Las Patentes Estadounidenses Nº. 3.654.090; Nº. 3.850.752; y Nº. 4.016.043.

Así, en realizaciones preferidas, la proteína estructural de NANBV, polipéptido, proteína de fusión o los anticuerpos anti-NANBV descritos aquí pueden fijarse a una matriz sólida para formar un soporte sólido que se empaqueta de forma separada en los sistemas de diagnóstico en cuestión.

El reactivo está normalmente fijado a la matriz sólida por adsorción a partir de un medio acuoso, aunque pueden utilizarse otros modos de fijación bien conocidos en la materia.

Las matrices sólidas útiles son bien conocidas en la materia. Tales materiales incluyen el dextrano reticulado disponible bajo el nombre comercial de SEPHADEX de Pharmacia Fine Chemicals (Piscataway, NJ); agarosa; cuentas de poliestireno de 1 micra a alrededor de 5 milímetros de diámetro disponible de Abbott Laboratories de North Chicago, IL; cloruro de polivinilo, poliestireno, poliacrilamida reticulada, redes basadas en nitrocelulosa o nilón como láminas, tiras o palas; o tubos, placas o los pocillos de una placa microtitulada como las de poliestireno o cloruro de polivinilo.

La proteína estructural del NANBV, polipéptido, proteína de fusión, anticuerpo anti-NANBV, polinucleótidos, agente de unión específico o reactivo de amplificación de cualquier sistema de diagnóstico descrito aquí, puede proporcionarse en forma de solución, como una dispersión líquida o como un polvo sustancialmente seco, por ejemplo, en forma liofilizada. Cuando el medio indicador es un enzima, el sustrato del enzima puede también proporcionarse en un paquete separado del sistema. También pueden incluirse como elementos empaquetados de forma separada en este sistema de ensayo diagnóstico un soporte sólido, como la placa microtitulada descrita antes, y uno o más tampones.

Los paquetes que aquí descritos en relación a los sistemas diagnósticos son aquellos utilizados habitualmente en los sistemas diagnósticos. Tales paquetes incluyen botellas y viales de vidrio y plástico (por ejemplo, polietileno; polipropileno y policarbonato), cubiertas de plástico y cubiertas laminadas de plástico y similares.

2. *Métodos de Diagnóstico*

La presente invención contempla cualquier método diagnóstico que propicie la detección de anticuerpos anti-proteína estructural del NANBV o antígeno estructural del NANBV en una muestra corporal utilizando una proteína estructural de NANBV, polipéptido, proteína de fusión o anticuerpo anti-antígeno estructural del NANBV descrito aquí como un reactivo inmunoreactivo para formar un producto de inmunoreacción cuya cantidad está relacionada, directa o indirectamente, con la cantidad de material a ser detectado en la muestra. Los entendidos en la materia entenderán que existen muchos procedimientos en la química del diagnóstico clínico bien conocidos, en que un reactivo inmunoquímico de esta invención puede utilizarse para formar un producto de inmunoreacción cuya cantidad está relacionada con la cantidad de anticuerpo o antígeno presente en una muestra corporal.

Pueden utilizarse diferentes protocolos heterogéneos y homogéneos, ya sean competitivos o no competitivos, en la realización de un método de ensayo de la presente invención.

Para detectar la presencia de anticuerpos anti-proteína estructural del NANBV en un paciente, se pone en contacto una muestra corporal, preferiblemente una muestra de fluido corporal como la sangre; plasma, suero, orina o saliva del paciente, en una mezcla bajo condiciones de ensayos biológicos con una molécula antigénica del NANBV descrita aquí, como una proteína estructural del NANBV, y preferiblemente con un polipéptido o proteína de fusión de la presente invención, para formar una mezcla de inmunoreacción. La mezcla se mantiene entonces durante un periodo de tiempo suficiente para permitir la formación de un producto de inmunoreacción de molécula antigénica del NANBV-molécula de anticuerpo (inmunocomplejo). La presencia, y preferiblemente la cantidad, de complejo puede detectarse entonces como se ha descrito aquí. La presencia del complejo es indicativo de anticuerpos anti-NANBV en la muestra.

En realizaciones preferidas, la presencia del producto de inmunoreacción formado entre moléculas antigénicas del NANBV y anticuerpos del paciente, se detecta utilizando un reactivo de unión específico como se ha discutido aquí. Por ejemplo, el producto de inmunoreacción se mezcla primero con un agente de unión específico marcado para formar una mezcla de marcaje. Un agente de unión específico marcado comprende un agente de unión específico y un marcaje como los que se han descrito aquí. La mezcla de marcaje se mantiene entonces bajo condiciones compatibles con la unión específica y durante un periodo de tiempo suficiente para permitir que un producto de inmunoreacción presente, se una con el agente de unión específico marcado y forme un producto marcado. La presencia, y preferiblemente la cantidad, de un producto marcado formado se detecta entonces para indicar la presencia o cantidad de un producto de inmunoreacción.

En realizaciones preferidas, los métodos diagnósticos de la presente invención se practican de una forma en que el inmunocomplejo se forma y se detecta en una fase sólida, como se ha descrito para los sistemas diagnósticos aquí.

Así, en un método diagnóstico preferido, la proteína estructural del NANBV o polipéptido se fija a una matriz sólida para formar la fase sólida. Es más preferido que el agente de unión específico sea la proteína A, o un anti-Ig humana, como IgG o IgM, que pueda formar un complejo con anticuerpos anti-proteína estructural del NANBV que forma inmunocomplejo en la fase sólida con la proteína estructural del NANBV. Más preferido es el uso de agentes de unión específica marcados en las que el marcaje es un isótopo radioactivo, un enzima, biotina o un marcador fluorescente como un lantánido como se ha descrito para los sistemas de diagnóstico, o detallados en las referencias mostradas más adelante.

En esta realización de fase sólida, se prefiere de forma particular el uso de una proteína recombinante que contiene el antígeno definido por la secuencia de residuos aminoacídicos contenida en el ID. de SEC. Nº:1 desde el residuo 1 al residuo 20, desde el residuo 21 al residuo 40, desde el residuo 2 al residuo 40, o desde el residuo 1 al residuo 74, como en el caso de las proteínas de fusión descritas en el Ejemplo 7.

En otro método diagnóstico preferido, la molécula antigénica del NANBV de la invención está fijada a una matriz sólida como se ha descrito antes, y se someten diluciones de las muestras biológicas al paso de inmunocomplejo mediante el contacto de las diluciones de la muestra con la superficie sólida y eliminando los materiales no unidos. Debido a la multivalencia de los anticuerpos presentes en las muestras biológicas de individuos infectados (bivalente para IgG, pentavalente para IgM), la posterior adición a esta mezcla de proteína estructural del NANBV marcada, polipéptido o proteína de fusión descritos aquí, se unirá a la fase sólida por el anticuerpo de la muestra sirviendo como puente entre las moléculas antigénicas del NANBV de la fase sólida tal como se ha descrito y, las moléculas soluble s marcadas. La presencia del marcaje en la fase sólida indica la presencia y preferiblemente la cantidad de anticuerpo específico en la muestra. Un experto en la materia puede determinar un rango de diluciones y determinar a partir de esta la concentración de antígeno marcado en la fase sólida. La muestra biológica y las moléculas antigénicas del NANBV marcadas descritas aquí, pueden mezclarse antes, o de forma simultánea al contacto de la muestra biológica con la fase sólida permitiendo la formación del complejo trimolecular en la fase sólida utilizando la propiedad de enlace de los anticuerpos específicos bivalentes o multivalentes. Como un marcaje particularmente útil, las moléculas antigénicas del NANBV biotiniladas descritas aquí pueden ser el antígeno marcado, permitiendo la posterior detección mediante la adición de un enzima estreptavidina, o un complejo de enzima-avidina, seguido del sustrato adecuado. Los enzimas como la peroxidasa de rábano picante, la fosfatasa alcalina, la β-galactosidasa o la ureasa se usan de forma frecuente y éstas, y otras, entre varios sustratos adecuados están disponibles comercialmente. Los marcajes preferidos con un marcador que permite la detección directa del complejo formado incluyen el uso de un isótopo radioactivo, como por ejemplo, yodo, o un quelato de lantánido como el Europio.

Las técnicas de inmunoensayo enzimático, tanto los ensayos directos o de competición utilizando formatos de ensayo homogéneos o heterogéneos, han sido extensivamente descritas en la materia. Se pueden encontrar ejemplos de las técnicas en Maggio; *Enzyme Immunoassay*, CRC Press; Cleveland, OH (1981); y Tijssen, "Practice and Theory of Enzyme Immunoassays", Elsevier, Amsterdam (1988).

Las condiciones de ensayos biológicos son aquellos que mantienen la actividad biológica de las moléculas antigénicas del NANBV y los anticuerpos anti-proteína estructural del NANBV en la mezcla de inmunoreacción. Estas condiciones incluye un rango de temperatura de alrededor de 4ºC a alrededor de 45ºC, preferiblemente de alrededor de 37ºC, a un rango de pH entre alrededor de 5 y alrededor de 9, preferiblemente alrededor de 7, y una fuerza iónica que varía de la del agua destilada en alrededor de un cloruro sódico molar, preferiblemente alrededor del salino fisiológico. Los métodos para optimizar tales condiciones son bien conocidos en la materia.

También se contemplan los ensayos inmunológicos capaces de detectar la presencia de la formación de producto de inmunoreacción sin utilizar un marcaje. Tales métodos emplean un "medio de detección", tales medios son conocidos en sí mismos en la química del diagnóstico clínico.

Ejemplos de medios de detección incluyen los métodos conocidos como biosensores e incluyen métodos con biosensores basados en la detección de cambios en la reflectividad de la superficie (resonancia de plasmón en superficie), cambios en la absorción de una onda evanescente mediante fibra óptica o cambios en la propagación de las ondas acústicas en la superficie.

Otra realización contempla la detección del producto de inmunoreacción utilizando una fluorometría resuelta por tiempo (TR-FIA), en el que el marcaje utilizado es capaz de producir una señal detectable por TR-FIA. Los marcajes adecuados típicos para la TR-FIA son los agentes de acomplejamiento de metales como es un quelato de lantánido formado por un lantánido y una beta-dicetona aromática, el lantánido se une al antígeno o anticuerpo mediante un análogo de EDTA por lo que se forma un complejo de lantánido fluorescente.

El principio de la fluorescencia resuelta por tiempo se describe en Soini *et al., Clin. Chem*., 25:353-361 (1979), y se ha aplicado ampliamente en los inmunoensayos. Véase por ejemplo, Halonen *et al., Current Topics in Microbiology and Immunology*, 104: 133-146 (1985): Suonpaa *et al., Clinica Chimica Acta*, 145:341-348 (1985): Lovgren *et al., Talanta*, 31:909-916 (1984); Patentes Estadounidenses Nº. 4.374.120 y 4.569.790; y Solicitud de Patente Internacional publicada Nº. EPO 139 675 y W087/02708. Un lantánido preferido para utilizar en la TR-FIA es el Europio.

Los reactivos y sistemas para practicar la tecnología TR-FIA está disponible a través de proveedores comerciales (Pharmacia Diagnostics, Uppsala, Suecia).

Son particularmente preferidos los inmunoensayos de fase sólida descritos aquí en el Ejemplo 7, realizados como un "Western Blot" típico.

Los presentes métodos de diagnóstico pueden practicarse en combinación con otros métodos separados para la detección de la aparición de anticuerpos anti-NANBV en especies infectadas con NANBV. Por ejemplo, una composición de esta invención puede utilizarse junto con el antígeno comercialmente disponible Cl00-3 (Ortho Diagnostics, Inc., Raritan, N.J.) en ensayos para determinar la presencia de cualquiera de las dos o ambas especies de anticuerpo inmunoreactivas con los dos antígenos.

Ejemplos

Los siguientes ejemplos se proporcionan con el mero propósito de ilustrar y no limitar de ninguna manera el alcance de la invención.

Ejemplo 1

Producción de Moléculas de DNA Recombinante

A. *Aislamiento de Clones NANBV y Análisis de Secuencia*

(1) *Aislamiento de RNA del NANBV y Preparación de cDNA*

Como fuente de viriones NANB, se recogió sangre de un chimpancé infectado con la cepa Hutchinson (Hutch) que presentaba una fase aguda de NANBH. El plasma se clarificó mediante centrifugación y filtración. Los viriones NANB se aislaron entonces a partir del plasma clarificado mediante cromatografía de inmunoafinidad en una columna de IgG NANBV (cepa Hutch) acoplada a la proteína G- sefarosa. El RNA del NANBV se eluyó a partir de las cuentas de sefarosa mediante remojo en tiocianato de guanidina y el RNA eluído se concentró entonces a través de un amortiguador de cloruro de cesio (CsCl). Sambrook *et al., Molecular Cloninq: A Laboratory Manual*, Sambrook *et al*., eds. Segunda Edición, Cold Spring Harbor Laboratory Press, NY (1989).

El RNA purificado del NANBV en cantidades del orden de picogramos, se utilizó como molde en una mezcla de reacción de extensión de cebadores, que contienen cebadores aleatorios y oligo dT, dNTP, y transcriptasa inversa para formar la primera cadena de los cDNA. Las primeras cadenas de los cDNA resultantes se utilizaron como moldes para la síntesis de una segunda cadena de cDNA en una mezcla de reacción que contienen DNA polimerasa I y RNAsa H para formar un cDNA de doble cadena (ds) (Sambrook *et al., supra*). Los cDNA ds sintetizados, se amplificaron utilizando un sistema adaptador de cebadores sintéticos asimétrico en el que los cebadores directo y reverso se hibridaron entre ellos y se ligaron en los extremos de los cDNA del NANBV de doble cadena con la ligasa T4 bajo condiciones de extremos romos para formar moléculas adaptadoras de cDNA. La amplificación mediante la reacción en cadena de la polimerasa (PCR) se realizó como se describe a continuación mezclando las moléculas adaptadoras de cDNA con el mismo cebador adaptador de sentido positivo, dNTP y polimerasa Taq (promega Biotec, Madison, WI) para preparar los cDNA del NANBV amplificados. Las secuencias de cDNA del NANBV amplificados resultantes se utilizaron como moldes para posteriores amplificaciones en una reacción de PCR con cebadores oligonucleótidos específicos de NANBV.

(2*) Síntesis de Oligonucleótidos para utilizar en el clonaje del NANBV*

Los oligonucleótidos se seleccionaron para corresponder con la secuencia 3' de Hepatitis C que codifica de forma putativa las proteínas estructurales de la cápside y de la cubierta del NANBV (secuencia HCJ1: Okamoto *et al., Jap. J. Exp. Med*., 60:167-177, 1990). Los oligonucleótidos seleccionados se sintetizaron en un Ensamblador de genes (Gene Assembler) de Pharmacia de acuerdo con las instrucciones del fabricante, se purificaron mediante electroforesis en gel de poliacrilamida y poseen las secuencias de bases de nucleótido y Nº de ID de SEC consecutivos empezando con 15 y acabando con 23 como se muestra en la Tabla 1.

TABLA 1

Oligonucleótidos sintéticos

Designación de oligonucleótidos[a]	Región Putativa del NANBV	Secuencia de Oligonucleótidos	Id de Sec.
690 (+)	Cápside 1-21	ATGAGCACGATTCCCAAACCT	15
693 (+)	Cápside 146-162	GAGGAAGACTTCCGAGC	16
694 (-)	Cápside 208-224	GTCCTGCCCTCGGGCCG	17
691 (-)	Cápside 340-359	ACCCAAATTGCGCGACCTACG	18
14 (+)	Cubierta 356-374	TGGGTAAGGTCATCGATAC	19
15 (+)	Cubierta 361-377	AAGGTCATCGATACCCT	20
18 (-)	Cubierta 512-529	AGATAGAGAAAGAGCAAC	21
16 (-)	Cubierta 960-981	GGACCAGTTCATCATCATATAT	22
17 (-)	Cubierta 957-976	CAGTTCATCATCATATCCCA	23

[a] los oligonucleótidos están definidos de forma numérica y su polaridad está indicada como (+) y (-) indicando la secuencia que corresponde a la cadena codificante sentido y anti-sentido, respectivamente. Todas las secuencias se listan en la orientación 5' a 3'.

(3) *Amplificación por PCR del cDNA del NANBV*

La amplificación por PCR se realizó mezclando las secuencias de cDNA amplificadas adaptadas al cebador preparadas en el Ejemplo 1A(1) con los oligonucleótidos sintéticos 690 y 694 como cebador (pares de cebadores 690:694). La mezcla de reacción de PCR resultante contenía el molde de cDNA amplificado adaptado al cebador, oligonucleótidos 690 y 694, dNTP, sales (Kcl y $MgCl_2$) y polimerasa Taq. La amplificación por PCR del cDNA se llevó a cabo manteniendo la mezcla a una temperatura de hibridación de 37ºC durante 30 ciclos. Las alícuotas de las muestras de la primera ronda de amplificación se reamplificaron a una temperatura de hibridación de 55ºC durante 30 ciclos bajo condiciones similares.

(4) *Preparación de Vectores que contienen DNA ds amplificado por PCR*

Las alícuotas de la segunda ronda de amplificación por PCR se sometieron a electroforesis en un gel de acrilamida al 5%. Tras la separación de los productos de reacción de PCR, la región del gel que contiene los fragmentos de DNA que corresponden al producto amplificado esperado 690:694 de aproximadamente 224 pb, se escindió y purificó siguiendo las técnicas estándar de electroelución (Sambrook *et al., supra*). Los fragmentos purificados se expusieron a una quinasas y se clonaron en el vector de clonaje plasmídico en el sitio SmaI del polienlazante para formar un plásmido que contiene el segmento de DNA 690:694 unido de forma operativa al pUC18.

La mezcla resultante que contiene pUC18 y un segmento de DNA que corresponde a la región de la secuencia 690:694 se transformó entonces en la cepa de *E. coli* JM83. Los plásmidos que contienen insertos se identificaron como colonias lac- (blancas) en un medio X-gal que contiene ampicilina. Los plásmidos pUC18 que contenían el segmento de DNA 690:694 se identificaron mediante análisis con enzimas de restricción y la posterior electroforesis en geles de agarosa, y se designaron moléculas pUC18 690:694 rDNA.

(5) *Secuenciación de Clones de Hepatitis que codifican la Proteína Putativa de la Cápside*

Dos colonias independientes que se creía que contenían el vector pUC18 con el segmento de DNA 690:694 de la cepa Hutch del NANBV (pUC18 690:694) que codifica para el amino terminal de la proteína putativa de la cápside, se amplificaron y utilizaron para preparar el DNA plasmídico mediante un gradiente de densidad de CsCl por centrifugación mediante los procedimientos estándar (Sambrook *et al., supra*). Los plásmidos se secuenciaron utilizando procedimientos con didesoxi ^{35}S con cebadores específicos de pUC18. Los dos plásmidos se secuenciaron de forma independiente en ambas cadenas de DNA para asegurar la exactitud de la secuencia. La información de la secuencia resultante se presenta de la base 1 a la base 224 del Id de Sec Nº:1.

El plásmido pUC18 690:694 contiene un segmento de DNA del NANBV que tiene una longitud de 224 pb y cuando se compara con la secuencia del prototipo HCJ1, revela dos sustituciones de nucleótidos y una diferencia de un aminoácido en la región terminal de la proteína de la cápside putativa.

(6) *Preparación de los clones NANBV a partir del extremo 5' del Genoma*

Para obtener la secuencia del genoma de Hutch del NANBV que codifica el resto de la región de la cápside (Okamoto *et al., supra*), los oligonucleótidos 693 y 691 (descritos en la Tabla 1) se utilizaron en las reacciones de PCR. El cDNA se preparó como se describe en el Ejemplo 1A(1) del RNA viral de Hutch del NANBV y se utilizó en la amplificación por PCR como se describe en el Ejemplo 1A(3) con el par de oligonucleótidos 693:691. El DNA ds amplificado por PCR resultante, se clonó en vectores de clonaje pUC18 y se cribaron en búsqueda de insertos como se describe en el Ejemplo 1A(4) para formar pUC18 693:691. Los clones se secuenciaron entonces con cebadores específicos de pUC18 como se describe en el Ejemplo 1A(5).

El plásmido pUC18 693:691 contiene un segmento de DNA del NANBV que tiene 157 pb de longitud y abarca las bases de nucleótidos 203 a 360 del Id de Sec Nº:1. El segmento no se extiende hasta la secuencia del cebador 693 utilizada para generar el fragmento La secuencia de este fragmento revela tres nucleótidos diferentes cuando se compara con la secuencia conocida de HCJ1 y no presenta ningún cambio de aminoácido correspondiente a la secuencia de HCJ1.

Para obtener la secuencia del genoma de Hutch del NANBV que codifica la región putativa de la cubierta (Okamoto *et al., supra*), se utilizaron los cebadores de oligonucleótidos 14 al 18 (descritos en la Tabla 1) en varias combinaciones con muestras de RNA de Hutch de NANBV. Como fuente de RNA de NANBV, su utilizó una muestra de biopsia de hígado de un chimpancé inoculado con la cepa Hutch a las 4 semanas post-inoculación y que presentaba una infección aguda. La muestra biopsiada se congeló primero y luego se molió. El polvo resultante se trató entonces con isotiocianato de guanidina para la extracción de RNA. El RNA se extrajo a partir de muestras de hígado tratadas con guanidio con fenol en presencia de SDS a 65ºC. Las muestras de hígado se extrajeron una segunda vez, y luego se extrajeron con cloroformo. El RNA extraído se precipitó a -20ºC con isopropanol y acetato de sodio.

El RNA derivado de hígado purificado se utilizo como molde en las reacciones de extensión de cebadores con los oligonucleótidos 18 y 16 para generar cDNA específico NANBV. Para preparar el cDNA de las secuencias codificantes de la proteína amino terminal de la cepa Hutch, los oligonucleótidos 18 y 16 antisentido se hibridaron con el RNA derivado de hígado de la cepa Hutch en presencia de dNTP y transcriptasa reversa a 42ºC para formar los productos de extensión de cebadores. La primera ronda de amplificación de PCR de los dos cDNA, se realizó mezclando los productos de la reacción de extensión de cebadores con pares separados de oligonucleótidos 14:16 (cDNA cebado con 16) y 14:18 (cDNA cebado con 18) durante 30 ciclos a una temperatura de hibridación de 55ºC. Las reacciones de PCR se realizaron sobre la mezcla anterior como en 1A(3). Las alícuotas de las amplificaciones 14:16 y 14:18 se utilizaron como moldes para la segunda ronda de amplificación en que los pares de oligonucleótidos 15:17 y 15:18, respectivamente, se utilizaron como cebadores.

Los productos de reacción de PCR de cada reacción de par de cebadores, se analizaron mediante electroforesis sobre geles de agarosa de fusión baja. Tras la separación, las regiones del gel que contienen fragmentos de DNA correspondientes a los productos amplificados esperados 15:17 y 15:18 de aproximadamente 617 pb y 168 pb, respectivamente, se escindieron y eluyeron de los pedazos de gel a 65ºC. Los fragmentos eluídos resultantes se purificaron mediante extracciones con fenol y cloroformo. Para clonar los fragmentos 15:17 y 15:18, los fragmentos purificados se trataron de forma separada con el fragmento Klenow de la polimerasa de DNA y una quinasa para un posterior subclonaje en el sitio SmaI del vector plasmídico pBluescript (Stratagene Cloning Systems, La Jolla, CA). Las colonias transformadas de *E. coli* DH5 se analizaron para detectar el inserto plasmídico mediante análisis con enzima de restricción como se describe en el Ejemplo 1A (4).

El plásmido pBluescript que contiene los segmentos de DNA 15:17 o 15:18 se purificó utilizando los protocolos de preparación de plásmido a gran escala CsCl. Los segmentos de DNA presentes en los plásmidos amplificados y purificados se secuenciaron cada uno como se describe en el Ejemplo 1A(5).

La secuencia del segmento de DNA 15:17 está contenida en el Id. de Sec. Nº:1 del nucleótido 361 al 978.

La secuencia del segmento de DNA 15:18 también se presenta en el Id. de Sec. Nº:1 del nucleótido 361 al 529. Estos dos clones se solapan en 168 pb del segmento de DNA 15:18.

El resultado de la secuencia indica que el segmento de DNA 15:17 difiere en 30 nucleótidos cuando se compara la secuencia HCJ1 (Okamoto *et al., Supra*) y también difiere en diez residuos de aminoácidos. El segmento de DNA 15:18 difiere en siete nucleótidos y en tres residuos de aminoácidos cuando se compara con HCJ1. En la región de solapamiento, los dos segmentos de DNA difieren en dos bases de nucleótidos, llamadas, bases 510 y 511, en el que el segmento de DNA 15:18 contiene una C en lugar de una T y una A en lugar de una G, respectivamente, que resulta en un cambio de una serina en lugar de un residuo de aminoácido glicina, en el residuo 171 del Id. de Sec. Nº:1. La razón de estas diferencias se desconoce y puede ser debido a un artefacto de la PCR.

B. *Producción de DNA Recombinante (rDNA) que codifica una proteína de Fusión*

(1) *Aislamiento del Fragmento 690:694 del clon pUC18 e introducción del Fragmento en el Vector de Expresión pGEX-3X*

El vector pUC18 que contiene el segmento de DNA 690:694 se sometió a una digestión con las enzimas de restricción EcoRI y BamHI para liberar el segmento de DNA que incluye una secuencia contenida en el Id. de Sec. Nº:1 de la base 1 a la base 224 del vector pUC18. El segmento de DNA liberado va s se sometió a una electroforesis en gel de acrilamida y el segmento de DNA que contiene el inserto de 224 pb del NANBV además de las porciones del polienlazante pUC18, se escindió entonces y se eluyó del gel como se describe en el Ejemplo 1A(4). El segmento de DNA eluído se extrajo con una mezcla de fenol y cloroformo, y se precipitó.

El segmento de DNA precipitado se resuspendió a una concentración de 25 μg/ml en agua y se trató con el fragmento de la polimerasa de DNA I Klenow y dNTP para rellenar los extremos escalonados creados por la digestión de restricción. El segmento romo resultante 690:694 se mezcló con el vector de expresión bacteriano, pGEX-3X, (disponible de Pharmacia Inc., Piscataway, NJ) que se linearizó con el enzima de restricción en romo SmaI. Los DNA mezclados se unieron de forma covalente (ligaron) manteniendo la mezcla toda la noche a 16ºC en presencia de un tampón ligasa y 5 unidades de ligasa de DNA T4 para formar un plásmido del segmento de DNA 690:694 operativamente unido a pGEX-3X.

(2) *Selección y Verificación de Inserto Ligado Correctamente Insertado*

la mezcla de ligación que contiene el vector pGEX-3X y el segmento de DNA 690:694 se transformó en la cepa huésped de *E. coli* W3110. Los plásmidos que contienen insertos se identificaron mediante la selección de la bacteria huésped que contiene vector en medio de Luria (LB) que contiene ampicilina. Los cultivos bacterianos en fase estacionaria se sometieron a protocolos de lisis alcalina para formar una preparación de DNA bruta. El DNA se digirió con el enzima de restricción XhoI. El sitio único XhoI, que escinde el segmento de DNA 690:694 entre las posiciones de nucleótidos 173 a 178 del Id. de Sec. Nº:1, pero no dentro del vector pGEX-3X, se utilizó para cribar un vector que contiene el segmento de DNA 690:694.

Varios vectores que contienen el segmento de DNA 690:694 se amplificaron y el vector de DNA resultante amplificado se purificó mediante centrifugación en gradiente de densidad de CsCl. El DNA se secuenció a través de las juntas de ligación de DNA del segmento insertado mediante métodos de ^{35}S didesoxi, con un cebador que hibrida con la secuencia de pGEX-3X en las posiciones de nucleótidos 614 a 633 contenida en el Id. de Sec. Nº:2. Los vectores que contienen el segmento de DNA 690:694 con la secuencia codificante correcta para la traducción en marco de una proteína estructural de NANBV, se identificaron de esta manera y se seleccionaron para formar pGEX-3X-690:694.

(3) *Estructura de la Proteína de Fusión*

El vector pGEX-3X se construyó para permitir colocar insertos en el extremo C terminal de Sj26, una glutation S-transferasa de 26-kDa (GST; EC 2.5.1.l8) codificada por el parásito helminto *Schistosoma japonicum*. La inserción del fragmento en marco 690:694 NANBV tras el Sj26 permite la síntesis del polipéptido de fusión Sj26-NANBV. El polipéptido NANBV puede escindirse del transportador GST mediante la digestión con el factor de proteasa Xa específico de sitio (Smith *et al., Gene*, 67:31-40, 1988).

El nucleótido y la secuencia de aminoácidos predicha del tránscrito de fusión pGEX-3X-690:694 de la secuencia GST a través del inserto690:694, se presenta en el Id. de Sec. Nº:2. la molécula de rDNA resultante, pGEX-3X-690:694, se predice que codifique para una proteína de fusión NANBV con la secuencia de residuos aminoacídicos contenida en el Id. de Sec. Nº:2 del residuo de aminoácidos 1 al residuo 315. El producto de la proteína resultante generada de la expresión del plásmido se refiere como ambas proteínas de fusión GST:NANBV 690:694 y la proteína de fusión CAP-N.

C. *Producción deDNA Recombinante (rDNA) que Codifican las proteínas de fusión de la Cápside y de la Cubierta del NANBV*

pGEX-3X-693:691: El plásmido pGEX-3X-693:691 se formó sometiendo primero al plásmido pUC18 693:691 preparado en el Ejemplo 1A(6) a una digestión con enzimas de restricción con EcoRI y BamHI como se realizó en el Ejemplo 1B(1). El resultante segmento de DNA liberado, con una secuencia contenida en el Id. de Sec. Nº:1 de la base 205 a la base 360 se purificó como se muestra en el Ejemplo 1B(1). El segmento de DNA purificado se mezcló y se ligó con el vector pGEX-3X que se linearizó mediante la digestión con el enzima de restricción EcoRI y BamHI en presencia de ligasa T4 a 16ºC para formar el plásmido pGEX-3X-693:691.

Se identificó un plásmido pGEX-3X que contiene un segmento de DNA 693:691 mediante una selección realizada como en el Ejemplo 1B(2) con la excepción que las preparaciones de DNA brutas se digirieron con EcoRI y BamHI para liberar el inserto 693:691. Un vector pGEX-3X que contiene un segmento de DNA 693:691 con la secuencia codificante correcta para la traducción en marco de una proteína estructural de NANBV, se identificó mediante análisis de la secuencia como se realizó en el Ejemplo 1B(2) y se seleccionó para formar pGEX-3X-693:691.

El vector resultante codifica una proteína de fusión (GST:NANBV 693:691) que está comprendida por una porción de polipéptido amino-terminal que corresponde a los residuos 1 a 221 de GST como el contenido en el Id. de Sec. Nº:2, una porción de polipéptido intermedia que corresponde a los residuos 222 a 225 y que define un sitio de escisión para el Factor de proteasa Xa, una proteína enlazante que corresponde a los residuos 226 a 230 que consiste en la secuencia de residuos aminoacídicos (Id. de Sec. Nº:25):

Gly Ile Pro Asn Ser

codificada por la secuencia de bases nucleotídicas (Id. de Sec. Nº:24):

GGG ATC CCCAAT TCA, respectivamente;

una porción de polipéptido carboxi-terminal que corresponde a los residuos 231 a 282 que define un antígeno de la cápside del NANBV con la secuencia de residuos aminoacídicos 69 a 120 en el Id. de Sec. Nº:1, y una porción enlazante carboxi-terminal que corresponde a los residuos 283 a 287 que consiste en la secuencia de residuos aminoacídicos (Id. de Sec. Nº:27):

Asn Ser Ser FIN

codificada por la secuencia de bases nucleotídicas (Id. de Sec. Nº:26):

AAT TCA TCG TGA, respectivamente.

pGEX-3X-15:18: El plásmido pGEX-3X-15:18 se formó sometiendo primero al plásmido Bluescript 15:18 preparado en el Ejemplo 1A(6) a una digestión con enzimas de restricción con EcoRV y BamHI, y el extremo cohesivo de BamHI se rellenó como se realizó en el Ejemplo 1B(1). El resultante segmento de DNA liberado, con una secuencia contenida en el Id. de Sec. Nº:1 de la base 361 a la base 528 se purificó como se muestra en el Ejemplo 1B(1). El segmento de DNA purificado se mezcló y se ligó con el vector pGEX-3X que se linearizó mediante la digestión con el enzima de restricción SmaI como se realizó en el Ejemplo 1B(1) para formar el plásmido pGEX-3X-15:18.

Se identificó un plásmido pGEX-3X que contiene un segmento de DNA 15:18 mediante una selección realizada como en el Ejemplo 1B(2) y las preparaciones de DNA brutas se cortaron con EcoRI y BamHI para liberar los insertos 15:18. Un vector pGEX-3X que contiene un segmento de DNA 15:18 con la secuencia codificante correcta para la traducción en marco de una proteína estructural de NANBV, se identificó mediante análisis de la secuencia como se realizó en el Ejemplo 1B(2) y se seleccionó para formar pGEX-3X-15:18.

El vector resultante codifica una proteína de fusión (GST:NANBV 15:18) que está comprendida por una porción de polipéptido amino-terminal que corresponde a los residuos 1 a 221 de GST, una porción de polipéptido intermedia que corresponde a los residuos 222 a 225 y que define un sitio de escisión para el Factor de proteasa Xa, una proteína enlazante que corresponde a los residuos 226 a 234 que consiste en la secuencia de residuos aminoacídicos (Id. de Sec. Nº:29):

Gly Ile Pro Ile Glu Phe Leu Gln Pro,

codificada por la secuencia de bases nucleotídicas (Id. de Sec. Nº:28):

GGG ATC CCC ATC GAA TTC CTG CAG CCC, respectivamente; una porción de polipéptido carboxi-terminal que corresponde a los residuos 235 a 290 que define un antígeno de la cubierta del NANBV con la secuencia de residuos aminoacídicos 121 a 176 en el Id. de Sec. Nº:1, y una porción enlazante carboxi-terminal que corresponde a los residuos 291 a 298 que consiste en la secuencia de residuos aminoacídicos (Id. de Sec. Nº:31):

Trp Gly Ile Gly Asn Ser Ser FIN

codificada por la secuencia de bases nucleotídicas (Id. de Sec. Nº:30):

TGG GGG ATC GGG AAT TCA TCG TGA, respectivamente.

pGEX-3X-15-17: El plásmido pGEX-3X-15:17 se formó sometiendo primero al plásmido Bluescript 15:17 preparado en el Ejemplo 1A(6) a una digestión con enzimas de restricción con EcoRI y BamHI, y el extremo cohesivo se rellenó como se realizó en el Ejemplo 1B(1). El resultante segmento de DNA liberado, con una secuencia contenida en el Id. de Sec. Nº:1 de la base 361 a la base 978 se purificó como se muestra en el Ejemplo 1B(1). El segmento de DNA purificado se mezcló y se ligó con el vector pGEX-3X que se linearizó mediante la digestión con el enzima de restricción SmaI como se realizó en el Ejemplo 1B(1) para formar el plásmido pGEX-3X-15:17.

Se identificó un plásmido pGEX-3X que contiene un segmento de DNA 15:17 mediante una selección realizada como en el Ejemplo 1B(2) y las preparaciones de DNA brutas se digirieron con EcoRI y BamHI como se ha indicado antes. Un vector pGEX-3X que contiene un segmento de DNA 15:17 con la secuencia codificante correcta para la traducción en marco de una proteína estructural de NANBV, se identificó como se realizó en el Ejemplo 1B(2) y se seleccionó para formar pGEX-3X-15:17.

El vector resultante codifica una proteína de fusión (GST:NANBV 15:17) que está comprendida por una porción de polipéptido amino-terminal que corresponde a los residuos 1 a 221 de GST, una porción de polipéptido intermedia que corresponde a los residuos 222 a 225 y que define un sitio de escisión para el Factor de proteasa Xa, una proteína enlazante que corresponde a los residuos 226 a 233 que consiste en la secuencia de residuos aminoacídicos (Id. de Sec. Nº:33):

Gly Ile Pro Asn Ser Cys Ser Pro,

codificada por la secuencia de bases nucleotídicas (Id. de Sec. Nº:32):

GGG ATC CCC AAT TCC TGC AGC CCT, respectivamente; una porción de polipéptido carboxi-terminal que corresponde a los residuos 234 a 439 que define un antígeno de la cubierta del NANBV con la secuencia de residuos aminoacídicos 121 a 326 en el Id. de Sec. Nº:1, y una porción enlazante carboxi-terminal que corresponde a los residuos 440 a 446 que consiste en la secuencia de residuos aminoacídicos (Id. de Sec. Nº:35):

Gly Ile Gly Asn Ser Ser FIN

codificada por la secuencia de bases nucleotídicas (Id. de Sec. Nº:34):

GGG ATC GGG AAT TCA TCG TGA, respectivamente.

pGEX-2T-15:17: El plásmido pGEX-2T-15:17 se formó sometiendo primero al plásmido Bluescript 15:17 preparado en el Ejemplo 1A(6) a una digestión con enzimas de restricción con EcoRV y BamHI, y el extremo cohesivo de BamHI se rellenó como se realizó en el Ejemplo 1B(1). El resultante segmento de DNA liberado, con una secuencia contenida en el Id. de Sec. Nº:1 de la base 361 a la base 978 se purificó como se muestra en el Ejemplo 1B(1). El segmento de DNA purificado se mezcló y se ligó con el vector pGEX-2T (Pharmacia Inc.) que se linearizó mediante la digestión con el enzima de restricción SmaI como se realizó en el Ejemplo 1B(1) para formar el plásmido pGEX-2T-15:17.

Se identificó un plásmido pGEX-2T que contiene un segmento de DNA 15:17 mediante una selección realizada como en el Ejemplo 1B(2) y mediante la digestión de las preparaciones de DNA brutas con EcoRI y BamHI. Un vector pGEX-2T que contiene un segmento de DNA 15:17 con la secuencia codificante correcta para la traducción en marco de una proteína estructural de NANBV, se identificó como se realizó en el Ejemplo 1B(2) y se seleccionó para formar pGEX-2T-15:17.

El vector resultante codifica una proteína de fusión (GST:NANBV 15:17) que está comprendida por una porción de polipéptido amino-terminal que corresponde a los residuos 1 a 221 de GST, una porción de polipéptido intermedia que corresponde a los residuos 222 a 226 y que define un sitio de escisión para la proteasa Trombina que consiste en la secuencia de residuos aminoacídicos (Id. de Sec. Nº:37):

Val Pro Arg Gly Ser

codificada por la secuencia de bases nucleotídicas (Id. de Sec. Nº:36)

GTT CCG CGT GGA TCC, respectivamente; una proteína enlazante que corresponde a los residuos 227 a 233 que consiste en la secuencia de residuos aminoacídicos (Id. de Sec. Nº39):

Pro Ser Asn Ser Cys Ser Pro

codificada por la secuencia de bases nucleotídicas (Id. de Sec. Nº:38):

CCA TCG AAT TCC TGC AGC CCT, respectivamente; una porción de polipéptido carboxi-terminal que corresponde a los residuos 234 a 439 que define un antígeno de la cubierta de NANBV, y una porción enlazante carboxi-terminal que corresponde a los residuos 440 a 446 que consiste en la secuencia de residuos aminoacídicos (Id. de Sec. Nº:41):

Gly Ile His Arg Asp FIN

codificada por la secuencia de bases nucleotídicas (Id. de Sec. Nº:40):

GGA ATT CAT CGT GAC TGA, respectivamente.

pGEX-3X-690:691: Para obtener un segmento de DNA correspondiente a la secuencia Hutch del NANBV mostrada en el Id. de Sec. Nº:1 de la base 1 a la base 360, se utilizaron los oligonucleótidos 690:691 en las reacciones de PCR como se muestra en el Ejemplo 1A(6). El DNA ds amplificado resultante de la PCR se clonó entonces en los vectores de clonación pUC18 como se describe en el Ejemplo 1A(4) para formar pUC18 690:691. Los clones se secuenciaron entonces con cebadores pUC18 como se describe en el Ejemplo 1A(5) para identificar un plásmido que contiene la secuencia completa. El plásmido resultante identificado se seleccionó y se designó pUC18 690:691, y contiene un segmento de DNA NANBV que tiene361 pb de longitud y abarca los nucleótidos 1 a 360 del Id. de Sec. Nº:1.

El plásmido pGEX-3X-690:691 se formó sometiendo primero al plásmido pUC18 690:691 a una digestión con enzimas de restricción con EcoRI y BamHI como se realizó en el Ejemplo 1B(1). El resultante segmento de DNA liberado, con una secuencia contenida en el Id. de Sec. Nº:1 de la base 1 a la base 360 con una secuencia polienlazante pUC18, se purificó como se muestra en el Ejemplo 1B(1). El segmento de DNA purificado se mezcló y se ligó con el vector pGEX-3X que se linearizó mediante la digestión con el enzima de restricción SmaI como se realizó en el Ejemplo 1B(1) para formar el plásmido pGEX-3X-690:691.

Se identificó un plásmido pGEX-3X que contiene un segmento de DNA 690:691 mediante una selección realizada como en el Ejemplo 1B(2). El vector pGEX-3X que contiene un segmento de DNA 690:691 con la secuencia codificante correcta para la traducción en marco de una proteína estructural de NANBV, se identificó como se realizó en el Ejemplo 1B(2) y se seleccionó para formar pGEX-3X-690:691.

El vector resultante codifica una proteína de fusión (GST:NANBV 690:691) que está comprendida por una porción de polipéptido amino-terminal que corresponde a los residuos 1 a 221 de GST, una porción de polipéptido intermedia que corresponde a los residuos 222 a 225 y que define un sitio de escisión para el Factor de proteasa Xa, una proteína enlazante que corresponde a los residuos 226 a 234 que consiste en la secuencia de residuos aminoacídicos (Id. de Sec. Nº:43):

Gly IlePro Asn Ser Ser Ser Val Pro

codificada por la secuencia de bases nucleotídicas (Id. de Sec. Nº:42):

GGG ATC CCC AAT TCG AGC TCG GTA CCC respectivamente; una porción de polipéptido carboxi-terminal que corresponde a los residuos 235 a 355 que define un antígeno de la cápside de NANBV, y una porción enlazante carboxi-terminal que corresponde a los residuos 356 a 363 que consiste en la secuencia de residuos aminoacídicos (Id. de Sec. Nº:45):

Thr Gly Ile Gly Asn Ser Ser FIN

codificada por la secuencia de bases nucleotídicas (Id. de Sec. Nº:44):

ACG GGG ATC GGG AAT TCA TCG TGA, respectivamente.

pGEX-2T-CAP-A: Los oligonucleótidos 1-20(+) y 1-20(-) para construir el vector pGEX-2T-CAP-A para expresar la proteína de fusión CAP-A, se prepararon como se describe en el Ejemplo 1A(2) que tienen las secuencias de bases de nucleótidos correspondientes a los Id. de Sec. Nº:7 e Id. de Sec. Nº:8, respectivamente.

Los oligonucleótidos 1-20 (+) y 1-20 (-) se mezclaron en cantidades iguales con el vector de expresión pGEX-2T (Pharmacia) que había sido predigerido con EcoRI y BamHI y mantenidos bajo condiciones de hibridación para permitir la hibridación de los oligonucleótidos complementarios y para permitir hibridar los extremos cohesivos del producto oligonucleótido resultante de doble cadena (ds) con el pGEX-2T en los extremos cohesivos EcoRI y BamHI. Tras la ligación, el plásmido resultante designado como pGEX-2T-CAP-A contenía una sola copia del producto oligonucleótido ds y un gen estructural que codifica para una proteína de fusión designada CAP-A con la secuencia de residuos aminoacídicos mostrada en el Id. de Sec. Nº:3 del residuo 1 al residuo 252.

El vector pGEX-2T es similar al vector pGEX-3X descrito antes, excepto que la proteína de fusión resultante es escindible mediante digestión con la proteasa trombina específica de sitio.

PGEX-2T-CAP-B: Los oligonucleótidos 21-40(+) y 21-40 (-) para construir el vector pGEX-2T-CAP-B para expresar la proteína de fusión CAP-B, se prepararon como se describe en el Ejemplo 1A(2) que tienen las secuencias de bases de nucleótidos correspondientes a los Id. de Sec. Nº:9 e Id. de Sec. Nº:10, respectivamente.

Los oligonucleótidos 21-40 (+) y 21-40 (-) se mezclaron en cantidades iguales con el vector de expresión pGEX-2T que había sido predigerido con EcoRI y BamHI y mantenidos bajo condiciones de hibridación para permitir la hibridación de los oligonucleótidos complementarios y para permitir hibridar los extremos cohesivos del producto oligonucleótido resultante de doble cadena con el pGEX-2T en los extremos cohesivos EcoRI y BamHI. Tras la ligación, el plásmido resultante designado como pGEX-2T-CAP-B contenía una sola copia del producto oligonucleótido ds y un gen estructural que codifica para una proteína de fusión designada CAP-B con la secuencia de residuos aminoacídicos mostrada en el Id. de Sec. Nº:4 del residuo 1 al residuo 252.

pGEX-2T-CAP C: Los oligonucleótidos 41-60 (+) y 41-60 (-) para construir el vector pGEX-2T-CAP-C para expresar la proteína de fusión CAP-C, se prepararon como se describe en el Ejemplo 1A(2) que tienen las secuencias de bases de nucleótidos correspondientes a los Id. de Sec. Nº:11 e Id. de Sec. Nº:12, respectivamente.

Los oligonucleótidos 41-60 (+) y 41-60 (-) se mezclaron en cantidades iguales con el vector de expresión pGEX-2T que había sido predigerido con EcoRI y BamHI y mantenidos bajo condiciones de hibridación para permitir la hibridación de los oligonucleótidos complementarios y para permitir hibridar los extremos cohesivos del producto oligonucleótido resultante de doble cadena con el pGEX-2T en los extremos cohesivos EcoRI y BamHI. Tras la ligación, el plásmido resultante designado como pGEX-2T-CAP-C contenía una sola copia del producto oligonucleótido ds y un gen estructural que codifica para una proteína de fusión designada CAP-C con la secuencia de residuos aminoacídicos mostrada en el Id. de Sec. Nº:5 del residuo 1 al residuo 252.

pGEX-2T-CAP-A-B: Los oligonucleótidos para construir el vector pGEX-2T-CAP-A-B para expresar la proteína de fusión CAP-A-B, se prepararon como se describe en el Ejemplo 1A(2) que tienen las secuencias de bases de nucleótidos correspondientes a los Id. de Sec. Nº:13 e Id. de Sec. Nº:14, respectivamente.

Los oligonucleótidos de acuerdo con el Id. de Sec. Nº:14 se mezclaron en cantidades equimolares con el plásmido pGEX-3X-690:694 descrito en el Ejemplo 1B(2). La mezcla se combinó con los reactivos para una reacción en cadena de la polimerasa (PCR) y los dos oligonucleótidos mezclados se utilizaron como cebadores en la mezcla de pGEX-3X-690:694 como moldes en la reacción de PCR para formar un producto de extensión de PCR que consiste en una molécula de ácido nucleico de doble cadena que codifica la secuencia de residuos aminoacídicos contenida en el Id. de Sec. Nº:1 del residuo 2 al 40 y también incluye los sitios de restricción añadidos por PCR de BamHI en el extremo 5' y EcoRI en el extremo 3'. El producto de extensión de PCR se escindió entonces con las enzimas de restricción BamHI y EcoRI para formar extremos cohesivos en el producto de extensión de PCR. El producto resultante con extremos cohesivos se mezcló en las mismas cantidades con el vector de expresión pGEX-2T que había sido predigerido con EcoRI y BamHI y mantenidos bajo condiciones de hibridación para permitir la hibridación de los oligonucleótidos complementarios y para permitir hibridar los extremos cohesivos del producto oligonucleótido resultante de doble cadena con el pGEX-2T en los extremos cohesivos EcoRI y BamHI. Tras la ligación, el plásmido resultante designado como pGEX-2T-CAP-A-B contenía una sola copia del producto de extensión de PCR y un gen estructural que codifica para una proteína de fusión designada CAP-A-B con la secuencia de residuos aminoacídicos mostrada en el Id. de Sec. Nº:6 del residuo 1 al residuo 271.

Ejemplo 2

Expresión de la proteína de fusión NANBV 690:694 utilizando rDNA

Las colonias bacterianas que contienen el plásmido pGEX-3X-690:694 en la orientación correcta se seleccionaron para examinar las propiedades de la proteína de fusión. Los cultivos bacterianos de pGEX-3X-690:694 se hicieron crecer hasta alcanzar la fase estacionaria en presencia de ampicilina (concentración final 50 μg/ml) a 37ºC. Este cultivo se inoculó a una dilución 1:50 en medio fresco LB a 37ºC en presencia de ampicilina y mantenido a 37ºC con agitación a 250 rpm hasta que las bacterias alcanzaron una densidad óptica de 0,5 al medir con un espectrómetro con un detector de fuente de luz de longitud de onda igual a 550 nm. Se mezcló entonces isopropiltio-beta-D-galactósido (IPTG) con el cultivo bacteriano a una concentración final de1 mM para iniciar (inducir) la síntesis de la proteína de fusión bajo el control del promotor tac en el vector pGEX-3X.

Comenzando en el tiempo cero y al cabo de una hora y después a las tres horas tras la mezcla con IPTG (es decir, la fase de inducción), el cultivo bacteriano se mantuvo como antes para permitir la expresión de proteína recombinante. Durante esta fase de mantenimiento, se midió la densidad óptica del cultivo bacteriano y se recogieron alícuotas de 1 ml para su centrifugación. Cada botón celular resultante que contiene lisado de proteína bruta, se resuspendió en una mezcla de colorante de Laemmli que contiene 1% de beta-mercaptoetanol en un volumen final de 50 μl por cada 0,5 unidades de DO550. Las muestras se hirvieron durante 15 minutos y 10 μl de cada muestra se sometió a electroforesis en un gel al 10% de SDS-PAGE Laemmli.

Otras proteínas de fusión GST:NANBV se expresaron en bacterias mediante transformación con el vector de expresión apropiado y se indujeron como se ha descrito antes.

Ejemplo 3

Detección de Proteínas de Fusión Expresadas

Para visualizar las proteínas de fusión inducidas por IPTG, los geles de Laemmli se tiñeron con Azul de Coomassie y se destiñeron en ácido acético y metanol. Se examinaron proteínas inducidas de clones separados y se compararon en base al aumento de una banda de proteína en el rango de tamaño predicho a partir del tiempo cero hasta las tres horas de tratamiento post-IPTG. La expresión de proteína de fusión se observó en los clones que presentaron un aumento desde el tiempo cero en la intensidad de una banda de proteína que corresponde con la proteína de fusión.

Las proteínas de fusión GST:NANBV CAP-A, CAP-B, y CAP-C, cuando se analizaron en un gel de Laemmli PAGE al 12.5% como se describe en el Ejemplo 2, presentaron un peso molecular aparente de alrededor de 30.000 daltons.

Ejemplo 4

Análisis Western Blot

Las muestras de las inducciones con IPTG que contienen una proteína de fusión GST:NANBV de esta invención se separaron mediante electroforesis en gel y se transfirieron a nitrocelulosa para análisis posterior de inmunoblotting. El filtro de nitrocelulosa se mezcló con tampón de bloqueo de anticuerpo (fosfato sódico 20 mM, pH 7,5, cloruro sódico 0,5 M, suero de albúmina bovina al 1%, y Tween 40 al 0,05%) durante 3 a 12 horas a temperatura ambiente. El suero de humanos y chimpancés con hepatitis NANB que se creía que contenía anticuerpo inmunoreactivo contra la proteína estructural NANBV se diluyó 1:500 en tampón de bloqueo de anticuerpo y se mezcló con la nitrocelulosa y se mantuvo durante 12 horas a temperatura ambiente para permitir la formación de un producto de inmunoreacción en la fase sólida. La nitrocelulosa se lavó entonces tres veces con volúmenes en exceso de tampón de bloqueo de anticuerpo. Los lavados fueron seguidos por una mezcla de la nitrocelulosa con 50 μl deproteína A ^{125}I (New England Nuclear, Boston, MA) a una dilución 1:500 en tampón de bloqueo de anticuerpo durante una hora a temperatura ambiente para permitir a la proteína A marcada unirse a cualquier producto de inmunoreacción presente en la fase sólida sobre la nitrocelulosa. La nitrocelulosa se lavó entonces como se ha descrito aquí, se secó y se expuso a una película de rayos X entre 1 y 3 horas a -70ºC para visualizar la marca y por lo tanto cualquier producto de inmunoreacción en la nitrocelulosa.

Los resultados del inmunoensayo Western blot se muestra en la Tablas 2 a 7. Las muestras preparadas utilizando el vector pGEX-3X que produce el GST control se prepararon también como antes y se probaron utilizando el procedimiento Western blot como control. La proteína GST expresada no fue detectable al medirla por inmunoreactividad utilizando el suero que mostraba inmunoreacción con una proteína de fusión de esta invención (por ejemplo, proteína de fusión CST:NANBV 690:694).

Ejemplo 5

Purificación de Proteínas de Fusión Expresadas GST:NANBV

Los cultivos de la cepa de *E. coli* W3110 transformados con plásmidos pGEX-3X-690:694 recombinantes preparados en el Ejemplo 2, se cultivaron durante 3 horas tras el tratamiento con inducción de IPTG. Las células se centrifugaron entonces para formar 8 botones celulares bacterianos, las células se resuspendieron en un volumen de cultivo 1/200 en tampón de lisis (MTPBS: NaCl 150 mM, Na_2HPO_4 16 mM, NaH_2PO_4 4 mM, pH 7,3), y se lisó la suspensión celular con una celda de presión French. Se mezcló Triton X-100 con el lisado celular para producir una concentración final del 1%. La mezcla se centrifugó a 50,000 X g durante 30 minutos a 4ºC. El sobrenadante resultante se recogió y se mezcló con 2 ml de cuentas de glutatión agarosa 50% (p/v) (Sigma, St. Louis, MO) preinfladas enMTPBS. Tras mantener la mezcla durante 5 minutos a 25ºC para permitir la unión específica por afinidad entre el GST y el glutatión en la fasesólida, las cuentas se recogieron por centrifugación a 1000 X g y se lavaron en MTPBS tres veces.

La proteína de fusión GST:NANBV 690:694 se eluyó de las cuentas de glutatión lavadas mediante la mezcla e incubación de las cuentas de glutatión con 2 ml de Tris HCl 50 mM, pH 8,0, que contienen glutatión reducido 5 mM durante 2 minutos a 25ºC para formar proteína de fusión purificada GST:NANBV 690:694.

El procedimiento de purificación por afinidad anterior produjo más de un 95% de proteína de fusión pura como se determinó mediante SDS PAGE. Esto es, la proteína purificada estaba esencialmente libre de antígeno procariota y antígenos NANBV no estructurales como se ha definido aquí.

Alternativamente, la proteína de fusión GST:NANBV 690:694 se purificó mediante cromatografía de intercambio aniónico.

Los cultivos se prepararon como se ha descrito antes y los botones celulares se resuspendieron en guanidina 8M y se mantuvieron toda la noche a 4ºC para solubilizar la proteína de fusión. La suspensión celular se aplicó a una columna de cromatografía sepharose S-300 y se recogieron las fracciones con pico que contienen la proteína de fusión

GST:NANBV 690:694, se juntaron y dializaron en urea 4M y se sometieron a una cromatografía de intercambio aniónico para formar la proteína de fusión purificada.

Otras proteínas de fusión GST:NANBV descritas aquí se expresaron también en cultivos de la cepa de *E. coli* W3110 como se ha descrito antes utilizando los vectores de la proteína de fusión GST producidos en el Ejemplo 1 tras su introducción mediante transformación en el huésped de *E. coli*. Tras la inducción y lisis de los cultivos, las proteínas de fusión GST se purificaron como se ha descrito antes utilizando cromatografía de afinidad glutatión agarosa para proporcionar más del 95% de proteína de fusión pura como se determinó mediante SDS-PAGE. Así las proteínas de fusión, CAP-A, CAP-B y CAP-C se expresaron todas y purificaron como antes utilizando el vector pGEX-2T-CAP-A, el vector pGEX-2T-CAP-B o el vector pGEX-2T-CAP-C, respectivamente, y la proteína de fusión CAP-A-B se expresó y purificó utilizando el vector PGEX-2T-CAP-A-B.

Ejemplo 6

Escisión mediante Proteasa de la Proteína de fusión GST:NANBV 690:694 Purificada

La proteína de fusión GST:NANBV 690:694 purificada, preparada en el Ejemplo 5 se sometió a tratamiento con el Factor (Xa) activado (Sigma) para escindir el transportador GST de la proteína de fusión NANBV 690:694 (Smith *et al., supra*). Siete μg del Factor X se activaron antes de mezclar con las proteínas de fusión purificadas mediante la mezcla y mantenimiento con 75 nanogramos (ng) de enzima de activación, Tris-HCl 8 mM (pH 8,0), NaCl 70 mM y $CaCl_2$ 8 mM a 37ºC durante 5 minutos. Cincuenta μg de proteína de fusión purificada se mezcló entonces con 500 ng de factor Xa humano activado en el tampón de elución descrito en el Ejemplo 5 que contiene Tris HCl 50 mM, glutatión reducido 5 mM, NaCl 100 mM, y $CaCl_2$ 1 mM, y se mantuvo a 25ºC durante 30 minutos. Los productos de la reacción de escisión resultante se adsorbieron entonces sobre cuentas de glutatión-agarosa preparadas en el Ejemplo 5 mediante unión por afinidad y se separó el GST libre de cualquier proteína escindida que contiene antígeno estructural de NANBV. Tras esto, se recoge la fase líquida para formar una solución que contiene proteína estructural NANBV purificadacon una secuencia de residuos aminoacídicos contenida en el Id. de Sec. Nº:2 del residuo 226 al residuo 315.

Ejemplo 7

Detección Inmunológica de Anticuerpos de Anti-Proteína estructural del NANBV

La cepa Hutch del virus NANBV se inyectó en chimpancés y se recogieron muestras de sangre en varios intervalos semanales tras la inoculación (INOC para analizar la respuesta inmunológica a NANBV mediante cinco ensayos diferentes de diagnóstico. Los chimpancés se categorizaron como cualquier ser en la fase aguda o crónica de la infección. Los ensayos utilizados en la evaluación de la respuesta inmune incluye: 1) detección de la enzima alanina aminotransferasa (ALT) (Alter *et al., JAMA*, 246:630-634, 1981; y Aach *et al., N. Engl. J. Med*., 304:989-994, 1981); 2) evaluación histológica de viriones del NANBV mediante microscopía electrónica (ME); 3) detección de anticuerpos anti-HCV utilizando el equipo disponible comercialmente que contienen el antígeno C100-3 (Ortho Diagnostics, Inc.); 4) detección de anticuerpos anti-CAP-N mediante análisis inmunoblot como se describe en el Ejemplo 4 utilizando la proteína de fusión CAP-N; y 5) Detección de virus mediante amplificación por PCR como se describe en el Ejemplo 1.

En la Tabla 2, los resultados se presentan de ensayos ALT, ME, anti-HCV (anti-C100-3), anti-CAP-N, y PCR sobre sueros de un chimpancé con Hepatitis NANB aguda.

(Tabla pasa a página siguiente)

TABLA 2

Chimp 59 - hepatitis aguda NANB

SEMANA POST INOC[1]	ALT	ME	ANTI HCV	ANTI CAP-N[2]	PCR 690-691
8	26	++	-	-	-
10	26	+	-	+	-
12	107	+	-	+	-
14	115	+	+	+	-
16	26	+	+	+	+
18	17	ND	+	+	(+)
20	11	ND	+	+	-

1 Semana tras la inoculación.

2 un signo más (+) indica que se observó inmunoreacción entre el suero mezclado y la proteína de fusión, designada "CAP-N" ya que corresponde al extremo amino de la proteína de la cápside putativa NANBV, utilizando el inmunoensayo Western blot descrito en el Ejemplo 4.

Los resultados de la Tabla 2 muestran inmunoreacción entre la proteína de fusión y los anticuerpos anti- proteína estructural NANBV en el suero probado. Además, la seroconversión es detectable antes mediante el inmunoensayo utilizando la proteína de fusión que contiene antígeno de la cápside que si el mismo suero se prueba en el inmunoensayo basado en C100-3.

En la Tabla 3, se presentan los resultados de ensayos ALT, anti-HCV (anti-C100-3) y anti-CAP-N en suero recogido de un humano con hepatitis NANB definitiva.

TABLA 3

NYU - 169 - hepatitis NANB definitiva

SEMANA POST INFEC	ALT	ANTI HCV	ANTI CAP-N
2	34	-	-
6	8	-	-
10	150	-	-
12	118	-	-
14	183	-	+
16	317	-	+
19	213	-	+
23	53	-	+

Los resultados de la Tabla 3 muestran que en las series delhumano, 169 muestras de suero de seroconversión, el antígeno presente CAP-N en la proteína de fusión detecta anticuerpos específicos NANBV tan pronto como 14 semanas post inoculación, mientras que el inmunoensayo basado en C100-3 no detecta ningún anticuerpo anti-NANBV en los tiempos estudiados.

En la Tabla 4, los resultados se presentan de ensayos ALT, ME, anti-HCV, y anti-CAP-N en suero de un chimpancé con una infección auto limitada presentado.

TABLA 4

Chimp 213 - infección auto limitada

SEMANA POST INOC	ALT	ME	ANTI HCV	ANTI CAP-N
4	24	+	-	+
6	34	+	-	+
8	38	+	-	+
13	28	ND	-	+
16	25	ND	-	+
18	232	ND	+	+
20	25	-	+	+

Los resultados de la Tabla 4 muestran que el antígeno CAP-N detecta anticuerpos anti-NANBV antes que el antígeno C100 cuando se utiliza suero recogido durante el curso de una infección NANBV auto limitada.

En la Tabla 5 los resultados se presentan de ensayos ALT, anti-HCV y anti-CAP-N en suero de un chimpancé que pasó de un perfil de infección aguda a una crónica.

TABLA 5

Chimp 10 - hepatitis nanb crónica/aguda

SÍNTOMAS	SEMANA POST INOC	PICO ALT	ANTI HCV	ANTI CAP-N
aguda	2	223	-	+
crónica	40	223	+	+
crónica	42	223	+	+
crónica	44	223	+	+
crónica	51	223	+	-

Los resultados de la Tabla 5 indican que el antígeno CAP-N detecta preferentemente anticuerpos anti-NANBV en estadíos agudos de infección por NANBV.

En la Tabla 6 los resultados se presentan de ensayos ALT, EM, anti-HCV (anti-C100-3) y anti-CAP-N en suero recogido en varios intervalos de varios chimpancés con Hepatitis NANB aguda o crónica.

TABLA 6

Suero agudo adicional

SEMANA POST INOC	SEMANA POST ALT ELEVADO	PICO ALT	ANTI HCV	ANTI CAP-N
2	+1	73	-	+
14	+2	66	-	+
6	+2	197	-	+
11	+1	151	-	-
8	+4	125	-	+
15	+1	82	-	+
12	-4	73	ND	+
	SUERO CRÓNICO ADICIONAL			
156	+131	110	+	+
156	-	89	+	+
160	-	89	+	+

Los resultados de la Tabla 6 indican que el antígeno CAP-N se detecta más a menudo con anticuerpos anti-NANBV en suero de individuos infectados de forma aguda que con el antígeno C100-3.

Los resultados de las Tablas 2-6 muestran que la proteína estructural NANBV de la invención, en la forma de una proteína de fusión que contiene el antígeno CAP-N y producida por el vector pGEX-3X-690:694, detecta anticuerpos en las series de seroconversión definidas en los tiempos, en un paciente infectado o chimpancé antes que los métodos actuales en la materia que utilizan el antígeno C100-3. Además, los resultados muestran que el antígeno CAP-N es particularmente útil para detectar infección aguda del NANBV de forma temprana en la infección.

En conjunto, los resultados indican que los pacientes infectados con NANBV, contienen anticuerpos circulando en su sangre que son inmunoespecíficos para el antígeno NANBV designado aquí como antígenos estructurales, y particularmente han mostrado que inmunoreaccionan con el antígeno de la cápside putativo definido mediante CAP-N. Estos anticuerpos se refieren por lo tanto como anticuerpos anti proteína estructural del NANBV y se deben distinguir de la clase de anticuerpos detectados previamente utilizando el antígeno C100-3 de proteína no estructural de NANBV.

En la Tabla 7, se presentan resultados comparativos de ensayos anti proteína de fusión de la cápside de HCV de acuerdo con el ensayo básico de inmunoblot descrito en el Ejemplo 4 utilizando varios sueros de chimpancé y humano sobre las siguientes proteínas de fusión de la cápside de HCV: CAP-N, CAP-A, CAP-B y CAP-C.

TABLA 7

SUERO	TIPO[a]	CAP-N[b]	CAP-A[c]	CAP-B[d]	CAP-C[e]
C18	Chimp 10 (A)	+++	+	+	-
C10	Chimp 194 (A)	+++	+++	+++	-
59-16	Chimp 59 (A)	+++	+	+++	ND
59-12	Chimp 59 (A)	ND[f]	++	+++	-
C9	Chimp 181 (A)	+++	-	+++	-
213-18	Chimp 213 (A)	ND	+	+	-
C2	Chimp 10 (C)	++	-	-	-
C1	Chimp 10 (C)	+++	-	-	-
C19	Chimp 10 (C)	+++	-	-	-
C4	Chimp 68 (C)	+++	+++	+++	ND
169-16	Humano	ND	+++	+++	-
169-23	Humano	ND	+++	+++	-
191-1	Humano	+	+	+	ND
191-2	Humano	+	+	++	ND
191-3	Humano	+	+	+	ND
216-1	Humano	-	+/-	+/-	ND
216-2	Humano	+	+	+	ND
216-3	Humano	+	+	+	ND

a El tipo de suero probado está indicado por las especies (chimpancé o humano), un número de identificación de chimpancé si la muestra es de un chimpancé, y una designación (entre paréntesis) si el donante de suero presenta infección de HCV aguda (A) o crónica (C) en el momento en que la muestra se tomó.

b CAP-N indica la proteína de fusión GST:NANBV 690:694 producida en el Ejemplo 5 que incluye los residuos 1 a 74 de la proteína de la cápside de HCV.

c CAP-A indica la proteína de fusión GST:NANBV producida en el Ejemplo 5 que incluye los residuos 1 a 20 de la proteína de la cápside de HCV.

d CAP-B indica la proteína de fusión GST:NANBV producida en el Ejemplo 5 que incluye los residuos 21 a 40 de la proteína de la cápside de HCV.

e CAP-C indica la proteína de fusión GST:NANBV producida en el Ejemplo 5 que incluye los residuos 41 a 60 de la proteína de la cápside de HCV.

f +, ++ y +++ indica cantidades relativas de anticuerpo anti cápside de HCV producto de la inmunización detectado mediante el ensayo Western blot, en el que + indica una banda débil tras la exposición durante la noche de la película de rayos x, ++ indica una banda fuerte tras la exposición durante la noche de la película de rayos x, +++ indica una banda fuerte tras 1 o 2 horas de exposición de la película de rayos x, y +/- o - indica una banda muy débil o que no hay banda, respectivamente, tras la exposición durante la noche de la película de rayos x

g "ND" indica no probado.

Los resultados mostrados en la Tabla 7 indica que las proteínas de fusión que contienen el antígeno CAP-A o CAP-B son inmunoreactivos con anticuerpos presentes en suero de chimpancé o humano infectados con HCV. Además, el antígeno CAP-C no inmunoreacciona significativamente con suero de chimpancé o humano infectados con HCV.

Ejemplo 8

Caracterización de la Secuencia de RNA Genómico del NANBV

A. *Caracterización de los Clones de cDNA y Estructura primaria del NANBV*

(1) *Aislamiento del RNA Viral del NANBV*

NANBV, también referido como virus de la hepatitis C (HCV), se aisló de dos tejidos fuente, de un chimpancé infectado con HCV, número 59 (c59), que había sido inoculado con la cepa Hutch (H) del HCV (designado HCV-Hc59) como se describe en el Ejemplo 1A(1). El hígado del chimpancé se biopsió durante la fase aguda de la infección (4 semanas después de la inoculación) y se tomó plasma de chimpancé 13 semanas después de la inoculación.

La extracción de ácidos nucleicos del hígado se realizó como se describe en Ogata *et al., Proc. Natl. Acad. Sci., USA*, 88:3392-3396 (1991) y en el Ejemplo 1A(6). Los viriones del HCV se aislaron del plasma con títulos virales de $10^{5,5}$ a $10^{6,5}$ CID_{50}/ml. El RNA de HCV se purificó de las muestras de plasma mediante cromatografía de inmunoafinidad como se describe en el Ejemplo 1A(1) o mediante precipitación con isopropanol.

En breve, 50 μl de plasma se diluyó con una solución tampón enfriada en hielo que contenía isotiocianato de guanidinio 4,2 M, sarcosil 0,5% y Tris-HCl 0,025 M a pH 8,0. El plasma diluido se mezcló entonces con 50 μl de tampón de extracción que contiene Tris-HCl 100 mM a pH 8,0, EDTA 10 mM y SDS 1% para formar una mezcla de extracción. La mezcla se vorteó y se mantuvo 5 minutos a 65ºC para iniciar la extracción. Las proteínas del suero se eliminaron de la mezcla con fenol/cloroformo a 65ºC seguido por una extracción con cloroformo solo. El RNA de HCV se precipitó entonces de la mezcla libre de proteína añadiendo dos volúmenes de isopropanol en hielo frío y un décimo de volumen de acetato sódico 3 M y se mantuvo la mezcla toda la noche a -20ºC. Tras precipitar por centrifugación en una centrífuga Eppendorf a 1400 rpm durante 30 minutos a 4ºC, el RNA de HCV se lavó una vez con etanol al 70%, se secó al vacío y luego se resuspendió en 9 μl de agua libre de RNAsa. Las muestras de RNA de HCV purificadas se calentaron durante 5 minutos a 65ºC antes de realizar la síntesis de cDNA como se describe posteriormente y en el Ejemplo 1A(1).

(2) *Clonaje del cDNA HCV-Hc59*

Cinco μg de RNA de HCV purificado de hígado o plasma se utilizaron para la reacción de cebado de cDNA. Para cebar la reacción, se utilizaron los cebadores nucleótidos específicos derivados de las secuencias de HCV publicadas y que abarcan las secuencias genómicas enteras descritas. Véase Okamoto *et al., Japan. J. Exp. Med*., 60:167-177 (1990); Kato *et al., Proc. Natl. Acad. Sci., USA*, 87:9524-9528 (1990); Han *et al., Proc. Natl. Acad. Sci.,USA*. 88:1711-1715 (1991); y Houghton *et al*., Solicitud de Patente Europea Número 88310922.5 y Pub1icación Número 318216. Las secuencias diana seleccionadas se amplificaron utilizando una aproximación basada en la PCR utilizando una variedad de cebadores de nucleótidos como se describe en el Ejemplo 1A(3). Las secuencias de nucleótido de los cebadores se resumen en la Tabla 8 posterior y han sido identificados mediante el número de cebador y el correspondiente Nº de Id. de Sec.

TABLA 8

Cebadores de nucleótidos utilizados en el clonaje de cDNA HCv-HC59

CEBADOR (N°)	ID DE SEC N°	SECUENCIA DE NUCLEÓTIDOS (5'-3')	POLARIDAD[a]
1	47	CAGCCCCCTGATGGGGGCGAC	+
22	48	ACTCGCAAGCACCCTATCA	-
21	49	CTGTGAGGAACTACTGTCT	+
690	50	ATGAGCACGAATCCTCAAACCT	-
694	51	GTCCTGCCCTCGGGCC	+
693	52	CGAGGAAGACTTCCGAGC	-
691	53	ACCCAAATTGCGCGACCTAC	+
15	54	TAAGGTCATCGATACCCT	-
17	55	CAGTTCATCATCATATCCCA	+
18	56	AGATAGAGAAAGAGCAAC	-
23	57	AGACTTCCGAGCGGTCGCAA	+
717	58	GACCTGTGCGGGTCTGTC	+
567	59	GGGTCGGCAGCTGGCTAGCCTCTCA	-
801	60	TCCTGGCGGGCATAGCGT	+
8	61	CCCCAGCCCTGGTCAAAATCGGTAA	-
568	62	TGAGAGGCTAGCCAGCTGCCGACCC	+
745	63	CTGTCGGTCGTTCCCACCA	-
626	64	CCGCGAAGAGTGTGTGTGGT	+
627	65	CAATGTTCTGGTGGAGGTG	-
617	66	GCCATTAAGTGGGAGTACGTCGTTCTCC	+

652	67	CGAGGAAGGATACAAGACC	-
628	68	TGCTTGTGGATGATGCTACT	+
629	69	CACACGTGCAGTTGCGCT	-
701	70	CTGCTGACCACTACACAG	+
654	71	GACCAGAGTGGAAGCGCAA	+
653	72	TACCAGAGTCGGGTGTACAG	-
500	73	CTAGGAGGCCCCTTGTCTGC	-
688	74	CTCGGGCCAGCCGATGGA	+
633	75	GGGGACCTCATGGTTGTCT	-
846	76	CCCGTGGAGTGGCTAAGG	+
831	77	CTCCTCGATGTTGGGATGG	-
830	78	CAGAGCTTCCAGGTGGCTC	+
795	79	CGGGCTCCGTCACTGTG	+
794	80	GTATTGCAGTCTATCACCGAG	-
464	81	GGCTATACCGGCGACTTCGA	+
40	82	CGTTGAGTGCGGGAGACAG	-
463	83	TCACCATTGAGACAATCACG	+
788	84	GTAAGGAAGGTTCTCCCCACTC	-
571	85	ATGCCCACTTTCTATCCCAGACAAAGC	+
623	86	TGCATGTCATGATGTAT	-
841	87	GGACAAGACGACCCTGCC	-
625	8	CGTATTGCCTGTCAACAGGC	+
631	89	AGCGCCCACAAAGGCAGTAG	-
842	90	CCTCTTCAACATATTGGGG	+
843	91	CCAGGAACCGGAGCATGG	-
859	92	ACCAGTGGATAAGCTCGG	+
904	93	CGTGGTGTAGGCATTAATG	-
862	94	ATGTGGAGTGGGACCTTCC	+

861	95	CTCTGCTGTTATATGGGAGG	-
F4	96	GTTGACGTCCATGCTCACTG	+
A4	97	TTTCCACGTCTCCACTAGCG	-
849	98	GTGAGGACCACCGTCCGC	-
F1	99	TTCCACCTCCAAAGTCCCCT	+
2_1	100	AGAACTTGCAGTCTGTCAAATGTGA	-
621	101	GGAAGAACAGAAACTGCCCATCAATGCACTA AGC	+
2_0	102	TGACGCCGCTGCTTTAACCT	-
2_2	103	TGCAAGCTTCCTCTACGGAT	-
51	104	AGGTTAAAGCAGCGGCGTCA	+
50	105	AGCTTCCCATCACGGCCAA	-
502	106	GATGGCTTTGTACGACGTG	+
55	107	GCACCTGCGATAGCCGCAGT	-
632	108	GTCCCTGACCGAGACGCT	+
853	109	GATTGGAGGTAGATCAAGTG	-
4	110	TACGACTTGGAGCTCATAAC	+
62	111	AGCAAGACACACTCCAGTCA	+
61	112	GCCTATTGGCCTGGAGTGGTTAGC	-

a (+) indica cadena sentido
(-) indica cadena anti sentido

Las secuencias amplificadas se aislaron a continuación, se dejaron con los extremos romos y se insertaron en un vector de clonación pUC o pBluescript (Stratagene) mediante procedimientos estándar como se describe en el Ejemplo 1A(4).

(3) *Análisis de la Secuencia de los cDNA HCV-Hc59 clonados*

Los clones se secuenciaron utilizando el método didesoxi de terminación de cadena utilizando un secuenciador automático duPont Genesis 2000. Para poder minimizar los errores de secuenciación debido a los artefactos de PCR (error de lectura de la polimerasa Taq), se aislaron tres clones independientes de cada secuencia diana y luego se secuenciaron. Las secuencias resultantes se compararon para derivar la secuencia final consenso representativa del genoma de la cepa HCV Hutch (HCV-H). En algunos casos, varios clones derivados de estudios independientes abarcan el mismo dominio genómico. Las secuencies de estos clones proporcionaron más datos confirmatorios.

(4) *Caracterización de los Clones de cDNA y Estructura Primaria de HCV-Hc59*

Se seleccionaron pares de cebadores tal como se ha descrito antes y en el Ejemplo 1A(3) para amplificar las regiones específicas del genoma de HCV-Hc59 para generar clones solapantes, las secuencias de los cuales puede comprender el genoma completo. Los pares de cebadores utilizados en reacciones específicas de PCR se listan en la Tabla 9 a continuación. Los cuarenta clones de cDNA resultantes, generados de los pares de cebadores seleccionados se detallan numéricamente comenzando con cero y acabando con 39 en la misma tabla y corresponde a la localización putativa en el mapa mostrado en la Figura 1. El tamaño deducido en el número de pares de bases de cada cDNA aislado también se detalla en la Tabla 9.

TABLA 9

Clones de HCV-HC59 derivados de PCR

Clon Nº[a]	Par de cebadores[b]	Tamaño de Inserto (pb)[c]
0	1:22	309
1	21:22	268
2	690:694	224
3	693:691	216
4	15:18	170
5	23:18	378
6	15:17	618
7	717:567	548
8	801:8	346
9	568:745	205
10	626:627	597
11	617:652	173
12	628:652	119
13	628:629	390
14	701:652	314
15	654:653	106
16	654:500	572
17	688:633	590
18	846:831	537
19	830:831	432
20	795:794	313
21	464:40	134
22	463:788	347
23	571:623	241
24	571:841	362

25	625:631	482
26	842:843	568
27	859:904	320
28	862:861	390
29	F4:A4	397
30	F4:849	498
31	F1: 2_1	493
32	621: 2_1	132
33	621:2_0	181
34	621: 2_2	221
35	51:50	360
36	502:55	322
37	852:853	625
38	4:853	315
39	62:61	611

[a] Localización relativa en el genoma de HCV-Hc59 mostrado en la Figura 1.

[b] Pares de cebadores sentido (+) y anti-sentido (-) con las secuencias de nucleótidos mostrada en la Tabla 1 y en los listados de Secuencia.

[c] Tamaño deducido en pares de bases (pb) del inserto clonado producido por PCR utilizando el par de cebadores indicado como se describe en el Ejemplo 1A(3) y 8A(3).

La comparación de las secuencias de tres clones de cDNA aislados de forma independiente del mismo dominio genómico, revelaron muy pocas diferencias de nucleótido indicando que la reserva de virus era homogénea. La secuencia del genoma completo de HCV-H se dedujo, representando 9416 nucleótidos, que es similar en longitud a la de los genomas de HCV previamente aislados, HCV-1, HCV-J y HCV-BK Kato *et al., supra*; Choo *et al., Proc. Natl. Acad. Sci., USA*, 88:2451-2455 (1991); y Takamizawa *et al., J. Virol.*, 65:1105-1113 (1991). La secuencia posee un alto contenido de GC (58.8%), y contiene un gran marco de lectura abierto que comienza en la base de nucleótido número 342 y termina en la base de nucleótido número 9374 (Id. de Sec. Nº:46) correspondiente a la proteína de 3011 residuos de aminoácido (Id. de Sec. Nº:46). Las secuencias de nucleótidos deducidas de HCV-Hc59 han sido depositadas en el GenBank con el número de acceso M67463.

Se identificaron las secuencias de HCV-Hc59 de los dominios de los extremos 5' y 3' no codificantes (NC), que abarcan 341 y 42 nucleótidos respectivamente. Los primeros 12 nucleótidos y los últimos 20 nucleótidos (Id. de Sec. Nº:46, ver características) corresponden a los cebadores de nucleótidos utilizados en el proceso de amplificación y, por lo tanto no están confirmadas como secuencias de HCV-H. No obstante, las secuencias 5' no codificantes de los genomas de HCV anteriormente descritos están extremadamente conservados (>98%), haciendo que probablemente la secuencia del extremo 5' del HCV-H descrito aquí sea muy similar sino idéntica a la otra indicada. No obstante, debido a la mayor divergencia entre las secuencias no codificantes 3' de HCV, la secuencia del extremo 3' del HCV-Hc59 continúa sujeta a confirmación. Cuando un cebador oligo (dT) se utiliza para la síntesis de cDNA seguida por una amplificación de PCR utilizando diferentes combinaciones de cebadores, no se obtuvieron secuencias virales. Este resultado indica que el genoma viral carece de tramos internos ricos en A en el extremo terminal 3' o una secuencia terminal 3' poli (rA). De forma similar, no se amplificaron secuencias cuando se utilizaron los cebadores ricos en A complementarios a la secuencia de nucleótidos del extremo 3' (ricos en U) de los dos aislamientosJaponeses descritos, HCV-J y HCV-BK, en la reacción de cebamiento de la TR, sugiriendo así la ausencia de una secuencia terminal rica en U en el genoma de HCV-Hc59.

El gran marco de lectura abierto del genoma de RNA de HCV-Hc59 está precedido por cinco codones AUG (cDNA = ATG - número de bases de nucleótido 13, 32, 85, 96 y 214 como se muestra en el Id. de Sec. Nº:46) confirmando la existencia de pequeños marcos de lectura abiertos hipotéticos en la región 5' NC de los genomas de HCV. Varias secuencias repetidas como las mostradas en Id. de Sec. Nº:46 numeradas de R_1 a R_5 en la parte de características del listado, se identifican en las regiones 5' y 3' NC, y en el C terminal del dominio putativo NS5. Estas secuencias pueden corresponder a importantes elementos que actúan en cis involucrados en la regulación de la replicación viral.

Las secuencias repetidas, R_2 y R_3, aparecen conservadas entre todos los aislamientos de HCV. Aunque no se han encontrado otras secuencias repetidas en los extremos terminales de los genomas de HCV, es posible que las secuencias con una función reguladora sean secuencias conservadas entre todos los virus HCV, como la R_2 y R_3.

La secuencia repetida R_2 es particularmente significativa ya que está representada por el número de copias más alto, cuatro. Se encuentra entre ambos extremos terminales 5' y 3', y está localizada corriente arriba de un lazo en horquilla del extremo 3' que puede estar involucrado en la ciclación de los virus. Nada se conoce todavía sobre la ciclación putativa de los virus HCV. También es posible que estas secuencias muy conservadas auto-complementarias puedan representar sitios de reconocimiento de replicasas, utilizadas posiblemente por ambas cadenas, la más y la menos del genoma viral.

Tal como se ha descrito en publicaciones previas para otros aislamientos de HCV, (Kato *et al., Proc. Natl. Acad. Sci. USA*, 87:9524-9528 (1990); Choo *et al., Proc. Natl. Acad. Sci. USA*, 88:2451-2455 (1991); y Takamizawa *et al., J. Virol*., 65:1105-1113 (1991)) el genoma de HCV-Hc59 o la proteína comparte solo una similitud limitada con otras secuencias virales conocidas, excepto para tres dominios: (1) unos pocos tramos de nucleótidos en la secuencia 5' NC están conservados con pestivirus idénticos a aquellos descritos por Choo *et al., Supra*, para el prototipo Americano de HCV-1 (Id. de Sec. Nº:46), (2) los bloques de aminoácidos encontrados en el dominio putativo NS3 (números de bases de nucleótido 3693 a 5198; Id. de Sec. Nº:46) correspondiente a la helicasa de unión a NTP putativa y serin proteasas similares a tripsina, están conservados entre los flavivirus y los pestivirus y (3) la secuencia consenso GDD conservada entre todas las polimerasas de RNA dependientes de RNA codificadas por el virus (residuos de aminoácidos 2737 a 2739; Id. de Sec. Nº:46). Además, se localizaron un total de diecinueve sitios de N-glicosilación putativos, esencialmente agrupados entre los residuos de aminoácidos 196 y 647 de forma similar a la organización observada para las proteínas de la cubierta de los pestivirus como se ha descrito por Meyers *et al., Virol*., 171:555-567 (1989); y Collett *et al., Virol*., 162:167-180 (1988).

B. *Comparación de Secuencias de Nucleótidos y Proteína de HCV-Hc59 y Aislamientos Heterólogos de HCV*

Un resumen de la comparación entre diferentes dominios genómicos de HCV-Hc59 y las secuencias previamente repetidas para los aislamientos Americano (HCV-1) o similar al Americano (HC-J1), y para los aislamientos Japoneses HC-J4, HCV-JH, HCV-J y HCV-BK se muestran en la Tabla 10. La comparación de secuencias está limitada en el caso de HC-J1, HC-J4 y HC-JH ya que la secuencia completa del genoma de estos aislamientos no ha sido aún descrita. Las asignaciones hipotéticas al mapa de las proteínas codificadas por HCV deducidas de la secuencia y de la similitud del perfil de hidrofobicidad entre los genomas de HCV y los flavivirus y/o pestivirus se utilizaron para realizar la comparación. Las referencias para las secuencias comparadas se resumen al final de la Tabla 10. Basado en las comparaciones de secuencias de los virus relatados, el genoma de HCV se cree que codifica al menos 8 dominios como se indica en la Tabla 10: el dominio estructural consistente en la nucleocápside (C) y dos proteínas de la cubierta (El y E2), y la región no estructural que consta de cinco proteínas, NS2, NS3, NS4a, NS4b, y NS5. Las designaciones del dominio se basan en la organización de las cepas de HCV relatadas con propósitos comparativos, y no necesariamente reflejan los dominios de HCV-Hc59 debido al presente estado de la materia en la caracterización de los dominios de HCV-Hc59.

TABLA 10

Homología de la secuencia de nucleótidos y aminoácidos deducida entre HCV-Hc59 y aislamientos heterólogos

Dominio[1]	Aislamiento[2]					
5' NC	HCV-1	HC-J1	HC-J4	HCV-JH	HCV-J	HCV-BK
-326-1						
%pb[3]	99,7	99,1	99,1	98,9	98,2	98,8
C						
1-570						
pb	98,4	98,9	90,0	90,3	91,0	90,3
aa	98,9	98,9	97,9	98,4	98,9	98,4
E1						
571-1140						
pb	93,5	93,1	74,1	73,7	73,9	73,8
aa	94,1	93,2	78,9	79,4	78,8	77,9
E2/NS2						
1141-2197						
pb	93,6	91,7	67,7	65,4	73,5	71,2
aa	92,9	88,7	70,7	65,6	79,3	80,4
NS2						
2198-3350						

pb	93,8	--	--	--	72,4	72,7
aa	95,1	--	--	--	80,0	78,2
NS3						
3351-4856						
pb	95,4	--	--	--	80,1	78,9
aa	97,2	--	--	--	92,2	92,6
NS4a						
4857-5596						
pb	95,8	--	--	--	80,4	80,0
aa	95,5	--	--	--	87,0	86,2
NS4b						
5597-6049						
pb	95,4	--	--	--	76,9	77,7
aa	96,7	--	--	--	84,8	85,4
NS5						
6050-9036						
pb	95,9	--	--	--	78,3	79,3
aa	96,7	--	--	--	83,2	83,7
3' NC						
9037-9055						
pb	83,3	--	--	--	73,6	63,1

[1] Posición de nucleótido para C y E1 deducida de Weiner *et al.*, Virol., 180:842-848 (1991) y para E_2 y NS2-NS5 de Takamizawa *et al.*, J. Virol., 65:1105-1113 (1991);

[2] Las posiciones de nucleótido están calculadas a partir del codón de inicio AUG en el que la A es la base número 1.

[3] Se resume el porcentaje de homología en los pares de bases (pb) y aminoácidos (aa).

Los datos indican un muy alto grado de identidad encontrado en dos dominios genómicos (5' NC y C) para todos los aislamientos a pesar de la separación geográfica (90,0-98,9% de homología de nucleótidos y 97,9 a 98,9% de homología de aminoácidos). Se ha realizado una observación similar en flavivirus que son miembros del mismo subgrupo sero-relacionado por Brinton *et al., Virol*., 162:290-299 (1988), mientras que los miembros de los diferentes subgrupos antigénicos comparten solo bajos niveles de homología en esta región. Se encontraron en el dominio 5' NC dos grupos de secuencias repetidas, R_2 y R_3 (Id. de Sec. Nº:46), que están conservadas entre todos los aislamientos descritos. Dos copias de la secuencia repetida R_1 también están conservadas entre los dos aislamientos Americanos HCV-Hc59 y HCV-1 pero solo una copia se encuentra en ambos aislamientos Japoneses HCV-J y HCV-BK. La secuencia 5' NC de estos genomas no se extiende muy lejos de la segunda copia. Esta secuencia de nucleótido descrita para los otros aislamientos de HCV no se extiende muy lejos del 5' NC para permitir la comparación.

Las regiones con identidad moderada se encontraron a lo largo de los dominio no estructurales, en los que se puede apreciar una clara separación entre los dos grupos de aislamientos (Americano/Japonés). Mientras que el 93,8 al 95,9% de identidad de nucleótidos se observó cuando el HCV-Hc59 se comparó con el primer grupo, solo se encontró entre un 72,7 y un 80,0% de identidad con el segundo grupo (95,1 al 97,2% y 78,2-92,6% de identidad de aminoácido, respectivamente). Una región, encontrada en la posición putativa NS5 (posición de residuo de aminoácidos 2356 a 2379 del Id. de Sec. Nº:46) y llamada Región V_3 como se muestra en la Tabla 11 a continuación, reflejó aún una mayor divergencia entre los dos subgrupos de aislamientos de HCV. Esta región mostró un 100% de identidad entre los dos aislamientos Americanos (datos no mostrados) pero solo un 12,5% con cualquiera de las cepas Japonesa. Aunque la mayoría de los cambios aparecen como cambios conservadores y pueden, por lo tanto, no resultar en una modificación funcional de la proteína, podría ser interesante evaluar si esta región genómica es inmunológicamente activa y si la variación antigénica también existe entre los dos subgrupos de aislamientos de HCV.

TABLA 11'

REGIÓN V

(Residuos 386 a 411 del Id. de Sec. N°:46)

Aislamientos[2]	
HCV-Hc59:	
	His Val Thr Gly Gly Asn Ala Gly Arg Thr Thr Ala
	Gly Leu Val Gly Leu Leu Thr Pro Gly Ala Lys Gln
	Asn Ile (Id. de Sec. N°:113)
HCV-1:	
	His Val Thr Gly Gly Ser Ala Gly His Thr Val Ser
	Gly Phe Val Ser Leu Leu Ala Pro Gly Ala Lys Gln
	Asn Val (Id. de Sec. N°:114)
HC-J1:	

His Val Thr Gly Gly Gln Ala Ala Arg Ala Met Ser
Gly Leu Val Ser Leu Phe Thr Pro Gly Ala Lys Gln
Asn Ile (Id. de Sec. N°:115)

HCV-J:

His Val Thr Gly Gly Arg Val Ala Ser Ser Thr Gln
Ser Leu Val Ser Trp Leu Ser Gln Gly Pro Ser Gln
Lys Ile (Id. de Sec. N°:116)

HCV-BK:

His Val Thr Gly Gly Ala Gln Ala Lys Thr Thr Asn
Arg Leu Val Ser Met Phe Ala Ser Gly Pro Ser Gln
Lys Ile (Id. de Sec. N°:117)

HC-J4:

Tyr Thr Ser Gly Gly Ala Ala Ser His Thr Thr Ser
Thr Leu Ala Ser Leu Phe Ser Pro Gly Ala Ser Arg
Asn Ile (Id. de Sec. N°:118)

HCV-JH:

His Val Thr Gly Gly Val Gln Gly His Val Thr Ser
Thr Leu Thr Ser Leu Phe Arg Pro Gly Ala Ser Gln
Lys Ile (Id. de Sec. N°: 11.9)

HCV-Hh-H77 :

His Val Thr Gly Gly Ser Ala Gly Arg Thr Thr Ala
Gly Leu Val Gly Leu Leu Thr Pro Gly Ala Lys Gln
Asn Ile (Id. de Sec. N°:;120)

HCV-Hh-H90:

His Val Thr Gly Gly Ser Ala Gly Arg Ser Val Leu
Gly Ile Ala Ser Phe Leu Thr Arg Gly Pro Lys Gln
Asn Ile (Id. de Sec. N°:121)

REGIÓN V_1

(Residuo 246 a 275 Id. de Sec. N°: 46)

HCV-Hc59:

Val Ala Thr Arg Asp Gly Lys Leu Pro Thr Thr Gln
Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser Ala
Thr Leu Cys Ser Ala Leu (Id. de Sec. N°:122)

HCV-1:

Val Ala Thr Arg Asp Gly Lys Leu Pro Ala Thr Gln
Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser Ala
Thr Leu Cys Ser Ala Leu (Id. de Sec. N°:123)

HC-J1:

Val Ala Thr Arg Asp Gly Lys Leu Pro Ala Thr Gln
Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser Ala
Thr Leu Cys Ser Ala Leu (Id. de Sec. N°:123)

HCV-J:

Leu Ala Ala Arg Asn Ser Ser Ile Pro Thr Thr Thr
Ile Arg Arg His Val Asp Leu Leu Val Gly Ala Ala
Ala Leu Cys Ser Ala Met (Id. de Sec. N°:124)

HCV-BK:

Leu Ala Ala Arg Asn Val Thr Ile Pro Thr Thr Thr
Ile Arg Arg His Val Asp Leu Leu Val Gly Ala Ala
Ala Phe Cys Ser Ala Met (Id. de Sec. N°: 125)

HC-J4:

Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr
Ile Arg Arg Bis Val Asp Leu Leu Val Gly Ala Ala
Ala Phe Cys Ser Ala Met (Id. de Sec. N°:126)

HCV-JH:

Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr
Leu Arg Arg His Val Asp Leu Leu Val Gly Thr Ala

Ala Phe Cys Ser Ala Met (Id. de Sec. N°:127)

HCV-Hh-H77:

Val Ala Thr Arg Asp Gly Lys Leu Pro Thr Thr Gln
Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser Ala
Thr Leu Cys Ser Ala Leu (Id. de Sec. N°: 122)

HCV-Hh-H90:

Val Ala Thr Arg Asp Gly Lys Leu Pro Thr Thr Gln
Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser Ala
Thr Leu Cys Ser Ala Leu (Id. de Sec. N°: 122)

REGIÓN V_2

(Residuo 456 a 482 del Id. de Sec. N°:46)

HCV-Hc59:

Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Ala Gln
Gly Trp Gly Pro Ile Ser Tyr Ala Asn Gly Ser Gly
Leu Asp Glu (Id. de Sec. N°:128)

HCV-1:

Leu Ala Ser Cys Arg Pro Leu Thr Asp Phe Asp Gln
Gly Trp Gly Pro Ile Ser Tyr Ala Asn Gly Ser Gly
Pro Asp Gln (Id. de Sec. N°:129)

HC-J1:

Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Asp Gln
Gly Trp Gly Pro Ile Ser His Ala Asn Gly Ser Gly
Pro Asp Gln (Id. de Sec. N°: 130)

HCV-J:

Met *ala* Ser Cys Arg Pro Ile Asp Glu Phe Ala Gln
Gly Trp Gly Pro Ile Thr His Asp Met Pro Glu Ser
Ser Asp Gln (Id. de Sec. N°:131)

HCV-BK:

Met *ala* Gln Cys Arg Thr Ile Asp Lys Phe Asp Gln Gly Trp Gly PrO Ile Thr Tyr Ala Glu Ser Ser Arg Ser Asp Gln (Id. de Sec. N°:132)

HC-J4:

Met *ala* Ser Cys Arg Pro Ile Gln Trp Phe Ala Gln Gly Trp Gly Pro Ile Thr Tyr Thr Glu Pro Asp Ser Pro Asp Gln (Id. de Sec. N°:133)

HCV-Hh-H77:

Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Ala Gln Gly Trp Gly Pro Ile Ser Tyr Ala Asn Gly Ser Gly Leu Asp Glu (Id. de Sec. N°:128)

HCV-Hh-H90:

Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Asp Gln Gly Trp Gly Pro Ile Ser Tyr Ala Asn Gly Ser Gly Pro Asp Glu (Id. de Sec. N°:134)

REGIÓN V_3

(Residuo 2356 a 2379 del Id. de Sec. N°:46)

HCV-Hc59 :

Ser Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro Ala Pro Ser Gly Cys Pro Pro Asp (Id. de Sec. N°:135)

HCV-J

Gly Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Gly Pro Pro Asp Gln Ala Ser Asp Asp Gly Asp Lys Gly (Id. de Sec. N°:136)

HCV-BK:

Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro Asp Gln Ala Ser Asp Asp Gly Asp Lys Gly (Id. de Sec. N°:137)

[1] Alineamiento de la secuencia de residuos aminoacídicos deducida de las Regiones V, V_1, V_2, y V_3 de HCV-Hc59 con otros aislamientos Americanos y Japoneses.

2 Aislamientos:

HCV-Hc59:	Americano/Chimpancé 59; Inschauspe *et al., Proc. Natl. Acad. Sci. USA*, 1991; Número de Acceso al GenBank M67463;
HCV-1:	Americano: Choo *et al., Proc. Natl. Acad. Sci. USA*, 88:2451-2455 (1991); Número de Acceso al GenBank M62321;
HCV-1:	5' terminal - Han *et al., Proc. Natl. Acad. Sci. USA*, 88: 1711-1715 (1991); Número de Acceso al GenBank M58407;
HCV-1:	3' terminal - Han *et al., supra*; Número de Acceso al GenBank M58406;
HC-J1:	Americano; Okamoto *et al., Japan J. Exp. Med*., 60:167-177 (1990);
HCV-J:	Japonés; Kato *et al., Proc. Natl. Acad. Sci., USA*, 87:9524-9528 (1990); Número de Acceso al GenBank D90208;
HCV-BK:	Japonés; Takamizawa *et al., J. Virol*., 65:1105-1113 (1991); Número de Acceso al GenBank M58335;
HC-J4:	Japonés; Okamoto *et al., supra*.,
HCV-JH:	Japonés; Takeuchi *et al., Nucl. Acids Res*., 18:4626 (1990);
HCV-Hh-H77 y H90:	Americano/humano; Ogata *et al., Proc. Natl. Acad. Sci., USA*, 88:3392-3396 (1991).

Las regiones de mayor divergencia se encontraron en la cubierta putativa E1 (número de bases de nucleótidos 571 a 1140} y E2 (número de bases de nucleótidos 1141 a 2197 calculados a partir del inicio de codón AUG), en el que se observó una identidad de aminoácidos del 77,9 al 94,1% y del 65,6 al 92,9%, respectivamente, entre HCV-Hc59 y el resto de aislamientos. Además de las regiones moderadas e hipervariables identificadas por Weiner *et al., Virol*., 180:842-848 (1991) en El y E2 (residuos de aminoácidos 214 a 254 y 386 a 411, respectivamente) en la que la heterogeneidad de dicha proteína entre HCV-Hc59 y el resto de aislamientos de HCV está entre un 70,7 y un 97,6% para la región moderada (datos no mostrados) y del 51,7 al 72,4% para la Región V como se muestra en la Tabla 11, se identificaron dos regiones de alta variabilidad. Ambas regiones, Región V, y Región V_2 (residuos de aminoácidos 246 a 275 y 456-482, respectivamente) aparecieron muy conservadas entre los tipos Americano o Japonés de HCV (96% de identidad) pero mostraron una marcada heterogeneicidad cuando se compararon ambos grupos (55-58% de identidad de proteína, Tabla 11). En contraste con la observación realizada por Weiner *et al., supra*, quien describió que aproximadamente el 50% de los cambios de aminoácidos observados en la Región V entre cuatro aislamientos Americanos y uno Italiano, no son cambios conservadores, más del 85% de los cambios observados en cualquiera de la Región V, V_1 o V_2 consistieron en cambios conservadores. Aunque la función de estas regiones permanece siendo desconocida, estos datos sugieren que están bajo presión inmunológica y pueden ser buenos candidatos para buscar dominios de epítopos protectores diana que pueden ser específicos del subtipo en el caso de las Regiones V_1 y V_2.

Así, el genoma de HCV-Hc59 muestra una homología total de aminoácidos del 96% con el prototipo Americano HCV-1 y del 84,9% con ambos aislamientos HCV-J y HCV-BK. Se identificaron tres nuevas regiones de alta variabilidad en E1, E2 y NS5 (Regiones V_1, V_2 y V_3 respectivamente). En las tres regiones, la heterogeneidad de secuencia aparece como específica de subgrupo (es decir, aislamientos Americanos frente a Japoneses), en particular para la Región V_3 en la que se encontró una divergencia de hasta el 87,5% entre los dos subgrupos tal como se muestra en las Tablas 10 y 11. La heterogeneidad de secuencia se observó en las regiones de la cubierta/NS1 de los flavivirus (véase, Meyers *et al., Virol*., 171:555-567 (1989); Collett *et al., Virol*., 165:191-199 (1988); y Hahn *et al., Virol*., 162:167-180 (1988) pero no para la extensión aquí descrita de las Regiones V, y V_2, sugiriendo así que la estructura de HCV es significativamente divergente de esta familia de virus. El hecho de que tres de las cuatro regiones variables del genoma de HCV están localizadas en los dominios putativos de la cubierta confirma que estos dominios están bajo una gran presión inmunológica, asociada posiblemente con la divergencia molecular relacionada con la evolución. Una tasa alta de cambios de nucleótido (28,2%) en el dominio putativo de E2/NS1 de HCV-H durante un intervalo de trece años sugiere una evolución significativa del genoma de HCV en este dominio. Véase Ogata *et al., supra*.

La secuencia de cDNA de la cepa del prototipo humano H de HCV (9416 nucleótidos) es el sujeto de esta invención. Hasta la fecha, esta es la segunda secuencia de nucleótidos de un genoma de HCV determinado para una cepa prototipo, ya que las dos secuencias Japonesas RCV-J y RCV-BK derivan de clones aislados de una mezcla de plasma, representando de esta manera secuencia genómicas probables de múltiples aislamientos. Los datos confirman que el

HCV presenta una estructura única y una organización más cercanamente relacionada con los pestivirus que con los flavivirus por la presencia de tramos de nucleótidos altamente conservados en el dominio 5' NC, pequeños marcos de lectura abiertos putativos precedentes al codón inicial AUG, y helicasas putativas de unión a NTP o serin proteasas similares a tripsina.

Descripción de los Id. de Sec. Nº:1-6 en los Listados de Secuencias

El Id. de Sec. Nº:1 contiene la secuencia de bases nucleotídicas lineal de cadena sencilla de un segmento preferidos de DNA de la presente invención que codifica porciones de las proteínas estructurales de la cepa Hutch de NANBV. La secuencia de las bases se muestran convencionalmente de izquierda a derecha y en la dirección de 5' terminal a 3' terminal utilizando el código de bases nucleótidos de una sola letra (A=adenina, T=timina, C=citosina y G=guanina) con el número de posición del primer residuo de base en cada línea, indicado a la izquierda de la línea la secuencia de bases nucleotídicas.

El marco de lectura de la secuencia de nucleótidos del Id. de Sec. Nº:1 está indicado por la colocación de la secuencia de residuos aminoacídicos de la proteína para la que codifica, debajo de la secuencia de nucleótidos, de tal forma que el código de tres letras de cada residuo de aminoácido (Tabla de Correspondencia) está localizado directamente debajo de las tres bases (codón) que codifican para cada residuo. El Id. de Sec. Nº:1 también contiene la secuencia lineal de residuos de aminoácidos codificadas por la secuencia de nucleótidos del Id. de Sec. Nº:1 y se muestra convencionalmente de izquierda a derecha y en la dirección del amino terminal al carboxi terminal. El número de posición para cada quinto residuo de aminoácido se indica debajo de cada secuencia de residuo de aminoácido.

Id. de Sec. Nº:2 contiene la secuencia de residuos aminoacídicos lineal de una proteína de fusión preferida designada CAP-N y está compuesta por una porción amino-terminal de un polipéptido que corresponde a los residuos 1 a 221 de la glutatión-S-transferasa, una porción de un polipéptido intermedio que corresponde a los residuos 222 a 225 y que define un sitio de escisión para el Factor Xa de la proteasa, una porción enlazante que corresponde a los residuos 226 a 234, una porción de un polipéptido que corresponde a los residuos 235 a 308 que definen un antígeno de la cápside del NANBV que posee la secuencia de residuos aminoacídicos 1 a 74 en el Id. de Sec. Nº:1, y una porción enlazante carboxi-terminal que corresponde a los residuos 309 a 315. El Id. de Sec. Nº:2 también contiene la secuencia de bases nucleotídicas de un segmento lineal de DNA de cadena sencilla que codifica la proteína de fusión descrita aquí. La nomenclatura y la presentación de la información de la secuencia son como se ha descrito para el Id. de Sec. Nº:1.

El Id. de Sec. Nº:3 contiene la secuencia de residuos aminoacídicos lineal de una proteína de fusión preferida designada CAP-A y que comprende de una porción amino-terminal de un polipéptido que corresponde a los residuos 1 a 220 de la glutatión-S-transferasa, una porción de un polipéptido intermedio que corresponde a los residuos 221 a 226 y que define un sitio de escisión para la proteasa Trombina, una porción de un polipéptido que corresponde a los residuos 227 a 246 que definen una porción del antígeno de la cápside del NANBV que posee la secuencia de residuos aminoacídicos 1 a 20 en el Id. de Sec. Nº:1, y una porción enlazante carboxi-terminal que corresponde a los residuos 247 a 252. El Id. de Sec. Nº:3 también contiene la secuencia de bases nucleotídicas de un segmento lineal de DNA de cadena sencilla que codifica la proteína de fusión descrita allí. La nomenclatura y la presentación de la información de la secuencia son como se ha descrito para el Id. de Sec. Nº:1.

El Id. de Sec. Nº:4 contiene la secuencia de residuos aminoacídicos lineal de una proteína de fusión preferida designada CAP-B y que comprende de una porción amino-terminal de un polipéptido que corresponde a los residuos 1 a 220 de la glutatión-S-transferasa, una porción de un polipéptido intermedio que corresponde a los residuos 221 a 226 y que define un sitio de escisión para la proteasa Trombina una porción de un polipéptido que corresponde a los residuos 227 a 246 que definen una porción del antígeno de la cápside del NANBV que posee la secuencia de residuos aminoacídicos 21 a 40 en el Id. de Sec. Nº:1, y una porción enlazante carboxi-terminal que corresponde a los residuos 247 a 252. El Id. de Sec. Nº:4 también contiene la secuencia de bases nucleotídicas de un segmento lineal de DNA de cadena sencilla que codifica la proteína de fusión descrita allí. La nomenclatura y la presentación de la información de la secuencia son como se ha descrito para el Id. de Sec. Nº:1.

El Id. de Sec. Nº:5 contiene la secuencia de residuos aminoacídicos lineal de una proteína de fusión preferida designada CAP-C y que comprende de una porción amino-terminal de un polipéptido que corresponde a los residuos 1 a 220 de la glutatión-S-transferasa, una porción de un polipéptido intermedio que corresponde a los residuos 221 a 226 y que define un sitio de escisión para la proteasa Trombina una porción de un polipéptido que corresponde a los residuos 227 a 246 que definen una porción del antígeno de la cápside del NANBV que posee la secuencia de residuos aminoacídicos 41 a 60 en el Id. de Sec. Nº:1, y una porción enlazante carboxi-terminal que corresponde a los residuos 247 a 252. El Id. de Sec. Nº:5 también contiene la secuencia de bases nucleotídicas de un segmento lineal de DNA de cadena sencilla que codifica la proteína de fusión descrita allí. La nomenclatura y la presentación de la información de la secuencia son como se ha descrito para el Id. de Sec. Nº:1.

El Id. de Sec. Nº:6 contiene la secuencia de residuos aminoacídicos lineal de una proteína de fusión preferida designada CAP-A-B y que comprende de una porción amino-terminal de un polipéptido que corresponde a los residuos 1 a 220 de la glutatión-S-transferasa, una porción de un polipéptido intermedio que corresponde a los residuos 221 a 226 y que define un sitio de escisión para la proteasa Trombina una porción de un polipéptido que corresponde a los residuos 227 a 265 que definen una porción del antígeno de la cápside del NANBV que posee la secuencia de residuos

aminoacídicos 2 a 40 en el Id. de Sec. Nº:1, y una porción enlazante carboxi-terminal que corresponde a los residuos 266 a 271. El Id. de Sec. Nº:6 también contiene la secuencia de bases nucleotídicas de un segmento lineal de DNA de cadena sencilla que codifica la proteína de fusión descrita allí. La nomenclatura y la presentación de la información de la secuencia son como se ha descrito para el Id. de Sec. Nº:1.

REIVINDICACIONES

1. Un método para detectar la seroconversión asociada con la infección por NANBV en etapas tempranas después de la infección que comprende:

(a) formar una mezcla de inmunoreacción acuosa mezclando una muestra de fluido corporal con un antígeno de la cápside NANBV.

(b) mantener dicha mezcla de inmunoreacción acuosa durante un periodo de tiempo suficiente que permita que los anticuerpos contra el antígeno de la cápside del NANBV presentes en la muestra de fluido corporal inmunoreaccionen con dicho antígeno de la cápside del NANBV para formar un producto de inmunoreacción; y

(c) detectar la presencia de cualquier producto de inmunoreacción formada y por lo tanto detectar una seroconversión temprana.

2. El método de la reivindicación 1, en el que dicho paso de detección (c) comprende los pasos de

(i) mezclar dicho producto de inmunoreacción formado en el paso (b) con un agente específico de unión marcado para formar una mezcla de marcaje, dicho agente específico de unión marcado comprende un agente específico de unión y un marcaje.

(ii) mantener dicha mezcla de marcaje durante un periodo de tiempo suficiente para que cualquier producto de inmunoreacción presente se una a dicho producto marcado; y

(iii) detectar la presencia de cualquier producto marcado formado, y por lo tanto la presencia de dicho producto de inmunoreacción.

3. El método de la reivindicación 2, en el que dicho agente específico de unión se selecciona del grupo que consiste en la Proteína A y al menos uno de los anticuerpos anti-IgG humana y anti-IgM humana.

4. El método de la reivindicación 2 o 3, en el que dicho marcaje se selecciona del grupo que consiste en quelato lantánido, biotina, enzima e isótopo radioactivo.

5. El uso del antígeno de la cápside del NANBV para la preparación de una composición diagnóstica para la detección de la seroconversión asociada con la infección por NANBV en las etapas tempranas tras la infección.

6. Un equipo para diagnosticar la seroconversión asociada con la infección por NANBV en las etapas tempranas tras la infección en una muestra de fluido corporal que comprende un antígeno de la cápside de NANBV, seleccionado del grupo que consiste en:

(a) un antígeno CAP-A de la cápside del NANBV con la secuencia de aminoácidos desde el residuo 1 al 20 del Id. de Sec. Nº:1;

(b) un antígeno CAP-B de la cápside del NANBV con la secuencia de aminoácidos desde el residuo 21 al 40 del Id. de Sec. Nº:1;

(c) un antígeno CAP-N de la cápside del NANBV con la secuencia de aminoácidos desde el residuo 1 al 74 del Id. de Sec. Nº:1;

7. El Equipo de la reivindicación 6, que comprende además un marcaje o medio indicativo capaz de señalizar la formación de un complejo que contenga un anticuerpo anti-NANBV.

8. El método de cualquiera de las reivindicaciones 1 a 4, el uso de la reivindicación 5 o el Equipo de la reivindicación 6 o 7, en que dicho antígeno de la cápside del NANBV está fijado a una matriz sólida.

9. El método de cualquiera de las reivindicaciones 1 a 4, el uso de la reivindicación 5 o el Equipo de la reivindicación 6 o 7, o el método, uso o equipo de la reivindicación 8, en que dicho antígeno de la cápside del NANBV está compuesto por una proteína de fusión.

10. El método de cualquiera de las reivindicaciones 1 a 4, el uso de la reivindicación 5, el método o uso de la reivindicación 8 o 9, en que dicho antígeno de la cápside del NANBV está seleccionado del grupo que consiste en:

(a) un antígeno de la cápside del NANBV con la secuencia de aminoácidos desde el residuo 1 al 120 del Id. de Sec. Nº:1;

(b) un antígeno de la cápside del NANBV con la secuencia de aminoácidos desde el residuo 1 al 20 del Id. de Sec. Nº:1;

(c) un antígeno de la cápside del NANBV con la secuencia de aminoácidos desde el residuo 1 al 74 del Id. de Sec. Nº:1;

(d) un antígeno de la cápside del NANBV con la secuencia de aminoácidos desde el residuo 2 al 40 del Id. de Sec. Nº:1; y

(e) un antígeno de la cápside del NANBV con la secuencia de aminoácidos desde el residuo 69 al 120 del Id. de Sec. Nº:1;

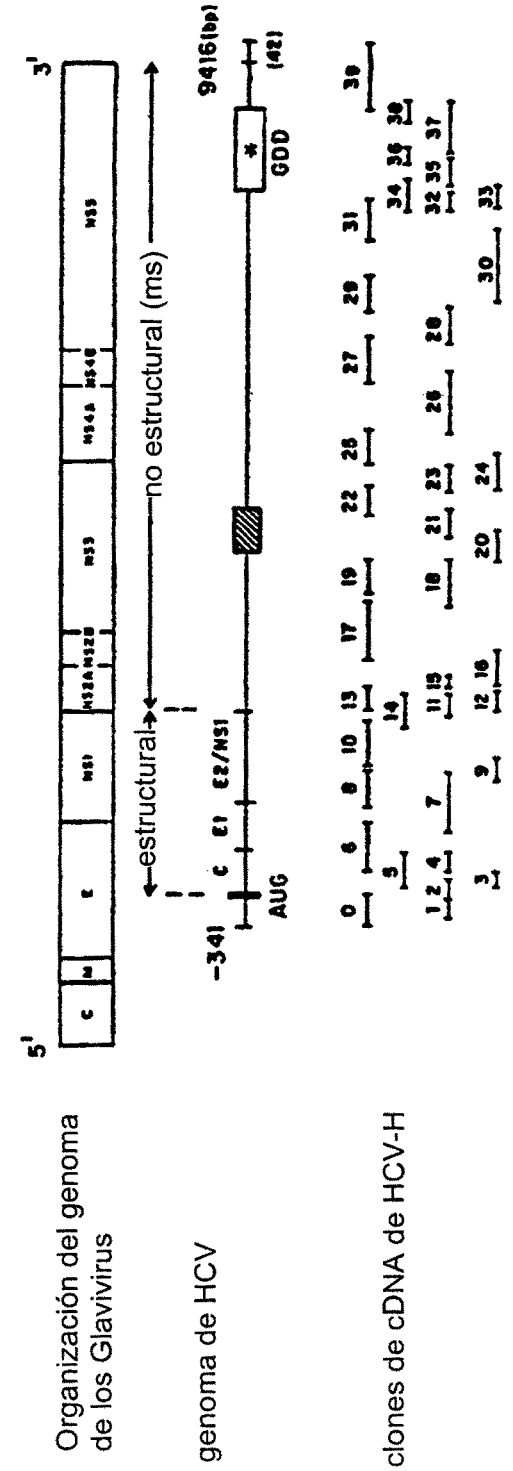

Organización del genoma de los Glavivirus
genoma de HCV
clones de cDNA de HCV-H
estructural
no estructural (ms)
AUG
GDD
9416(bp)
= NS3 del dengue tipo 2
= repticasa del CARMV

FIGURA 1

LISTA DE SECUENCIAS

(1) INFORMACIÓN GENERAL:

(i) SOLICITANTE: Zebedee, Suzanne
Inchauspe, Genevieve
Nasoff, Marc
Prince, Alfred

(ii) TÍTULO DE LA INVENCIÓN: ANTÍGENO DEL VIRUS DE LA HEPATITIS NO-A, NO-B, MÉTODOS DE DIAGNÓSTICO Y VACUNAS

(iii) NÚMERO DE SECUENCIAS: 137

(iv) DIRECCIÓN DE CORRESPONDENCIA:
(A) DESTINATARIO: DRESSLER, GOLDSMITH, SHORE, SUTKER & MILNAMOW, LTD.
(B) CALLE: 180 N. Stetson, Suite 4700
(C) CIUDAD: Chicago
(D) ESTADO: MIL
(E) PAÍS: EEUU
(F) CP: 60601

(v) FORMA DE LECTURA DEL COMPUTADOR:
(A) TIPO DE MEDIO: Disco flexible
(B) COMPUTADOR: IBM PC compatible
(C) SISTEMA OPERATIVO: PC-DOS/MS-DOS
(O) PROGRAMA: PatentIn Release #1.0, Versión 1.25

(vi) DATOS DE LA PRESENTE SOLICITUD:
(A) NÚMERO DE SOLICITUD: US
(B) FECHA DE DEPÓSITO:
(C) CLASIFICACIÓN:

(vii) DATOS ANTERIORES DE LA SOLICITUD:
(A) NÚMERO DE SOLICITUD: US 07/616369
(B) FECHA DE DEPÓSITO: 21 NOV 1990

(vii) DATOS ANTERIORES DE LA SOLICITUD:
(A) NÚMERO DE SOLICITUD: US 07/573643
(B) FECHA DE DEPÓSITO: 25 AGO 1990

(viii) INFORMACIÓN DEL ABOGADO/AGENTE:
(A) NOMBRE: Gamson, Edward P.
(B) NÚMERO DE REGISTRO: 29.381
(C) NÚMERO DE REFERENCIA/SUMARIO: PHA0029P

(ix) INFORMACIÓN DE TELECOMUNICACIÓN:
(A) TELÉFONO: 312-616-5400
(B) TELEFAX: 312-616-54060

(2) INFORMACIÓN PARA ID de SEC Nº:1

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 978 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 978

(D) OTRA INFORMACIÓN: /inicio de codón=1
/producto= "Antígeno Estructural NANBV"
/número=1

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:1:

```
ATG AGC ACG ATT CCC AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC          48
Met Ser Thr Ile Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
  1               5                  10                  15

CGT CGC CCA CAG GAC GTC AAG TTC CCG GGT GGC GGT CAG ATC GTT GGT          96
Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly
            20                  25                  30

GGA GTT TAC TTG TTG CCG CGC AGG GGC CCT AGA TTG GGT GTG CGC GCG         144
Gly Val Tyr Leu Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala
        35                  40                  45

ACG AGG AAG ACT TCC GAG CGG TCG CAA CCT CGA GGT AGA CGT CAG CCT         192
Thr Arg Lys Thr Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro
    50                  55                  60

ATC CCC AAG GCA CGT CGG CCC GAG GGC AGG ACC TGG GCT CAG CCC GGG         240
Ile Pro Lys Ala Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly
65                  70                  75                  80

TAC CCT TGG CCC CTC TAT GGC AAT GAG GGT TGC GGG TGG GCG GGA TGG         288
Tyr Pro Trp Pro Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp
                85                  90                  95

CTC CTG TCT CCC CGT GGC TCT CGG CCT AGC TGG GGC CCC ACA GAC CCC         336
Leu Leu Ser Pro Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro
            100                 105                 110

CGG CGT AGG TCG CGC AAT TTG GGT AAG GTC ATC GAT ACC CTT ACG TGC         384
Arg Arg Arg Ser Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys
        115                 120                 125

GGC TTC GCC GAC CTC ATG GGG TAC ATA CCG CTC GTC GGC GCC CCT CTT         432
Gly Phe Ala Asp Leu Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu
    130                 135                 140

GGA GGC GCT GCC AGG GCC CTG GCG CAT GGC GTC CGG GTT CTG GAA GAC         480
Gly Gly Ala Ala Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp
145                 150                 155                 160

GGC GTG AAC TAT GCA ACA GGG AAC CTT CCT GGT TGC TCT TTC TCT ATC         528
Gly Val Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile
                165                 170                 175

TTC CTT CTG GCC CTG CTC TCT TGC CTG ACT GTG CCC GCT TCA GCC TAC         576
Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr
            180                 185                 190

CAA GTG CGC AAT TCC TCG GGG CTT TAC CAT GTC ACC AAT GAT TGC CCT         624
Gln Val Arg Asn Ser Ser Gly Leu Tyr His Val Thr Asn Asp Cys Pro
        195                 200                 205

AAC TCG AGT GTT GTG TAC GAG GCG GCC GAT GCC ATC CTG CAC ACT CCG         672
Asn Ser Ser Val Val Tyr Glu Ala Ala Asp Ala Ile Leu His Thr Pro
    210                 215                 220

GGG TGT GTC CCT TGC GTT CGC GAG GGT AAC GCC TCG AGG TGT TGG GTG         720
Gly Cys Val Pro Cys Val Arg Glu Gly Asn Ala Ser Arg Cys Trp Val
225                 230                 235                 240
```

```
GCG GTG ACC CCC ACG GTG GCC ACC AGG GAC GGC AAA CTT CCC ACA ACG        768
Ala Val Thr Pro Thr Val Ala Thr Arg Asp Gly Lys Leu Pro Thr Thr
                245                 250                 255

CAG CTT CGA CGT CAT ATC GAT CTG CTT GTC GGG AGC GCC ACC CTC TGC        816
Gln Leu Arg Arg His Ile Asp Leu Leu Val Gly Ser Ala Thr Leu Cys
            260                 265                 270

TCG GCC CTC TAC GTG GGG GAC CTG TGC GGG TCT GTC TTT CTC GTT GGT        864
Ser Ala Leu Tyr Val Gly Asp Leu Cys Gly Ser Val Phe Leu Val Gly
        275                 280                 285

CAA CTG TTT ACC TTC TCT CCC AGG CGC CAC TGG ACG ACG CAA GAC TGC        912
Gln Leu Phe Thr Phe Ser Pro Arg Arg His Trp Thr Thr Gln Asp Cys
    290                 295                 300

AAT TGT TCT ATC TAT CCC GGC CAT ATA ACG GGT CAT CGC ATG GCA TGG        960


Asn Cys Ser Ile Tyr Pro Gly His Ile Thr Gly His Arg Met Ala Trp
305                 310                 315                 320

GAT ATG ATG ATG AAC TGG                                               978
Asp Met Met Met Asn Trp
                325
```

(2) INFORMACIÓN PARA ID de SEC Nº:2

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 948 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 945

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:2:

```
ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GGC CTT GTG CAA CCC          48
Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro
  1               5                  10                  15

ACT CGA CTT CTT TTG GAA TAT CTT GAA GAA AAA TAT GAA GAG CAT TTG          96
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu
            20                  25                  30

TAT GAG CGC GAT GAA GGT GAT AAA TGG CGA AAC AAA AAG TTT GAA TTG         144
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu
        35                  40                  45

GGT TTG GAG TTT CCC AAT CTT CCT TAT TAT ATT GAT GGT GAT GTT AAA         192
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
    50                  55                  60

TTA ACA CAG TCT ATG GCC ATC ATA CGT TAT ATA GCT GAC AAG CAC AAC         240
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn
65                  70                  75                  80

ATG TTG GGT GGT TGT CCA AAA GAG CGT GCA GAG ATT TCA ATG CTT GAA         288
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu

                85                  90                  95

GGA GCG GTT TTG GAT ATT AGA TAC GGT GTT TCG AGA ATT GCA TAT AGT         336
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser
            100                 105                 110

AAA GAC TTT GAA ACT CTC AAA GTT GAT TTT CTT AGC AAG CTA CCT GAA         384
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu
        115                 120                 125

ATG CTG AAA ATG TTC GAA GAT CGT TTA TGT CAT AAA ACA TAT TTA AAT         432
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn
    130                 135                 140

GGT GAT CAT GTA ACC CAT CCT GAC TTC ATG TTG TAT GAC GCT CTT GAT         480
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp
145                 150                 155                 160

GTT GTT TTA TAC ATG GAC CCA ATG TGC CTG GAT GCG TTC CCA AAA TTA          52
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu
                165                 170                 175

GTT TGT TTT AAA AAA CGT ATT GAA GCT ATC CCA CAA ATT GAT AAG TAC         576
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr
            180                 185                 190

TTG AAA TCC AGC AAG TAT ATA GCA TGG CCT TTG CAG GGC TGG CAA GCC         624
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala
        195                 200                 205
```

```
ACG TTT GGT GGT GGC GAC CAT CCT CCA AAA TCG GAT CTG ATC GAA GGT        672
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Ile Glu Gly
    210                 215                 220

CGT GGG ATC CCC AAT TCG AGC TCG GTA CCC ATG AGC ACG ATT CCC AAA        720
Arg Gly Ile Pro Asn Ser Ser Ser Val Pro Met Ser Thr Ile Pro Lys
225                 230                 235                 240

CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGT CGC CCA CAG GAC GTC        768
Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln Asp Val
                245                 250                 255

AAG TTC CCG GGT GGC GGT CAG ATC GTT GGT GGA GTT TAC TTG TTG CCG        816
Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu Leu Pro
            260                 265                 270

CGC AGG GGC CCT AGA TTG GGT GTG CGC GCG ACG AGG AAG ACT TCC GAG        864
Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu
        275                 280                 285

CGG TCG CAA CCT CGA GGT AGA CGT CAG CCT ATC CCC AAG GCA CGT CGG        912
Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala Arg Arg


    290                 295                 300

CCC GAG GGC AGG ACG GGG ATC GGG AAT TCA TCG TGA                       948
Pro Glu Gly Arg Thr Gly Ile Gly Asn Ser Ser
305                 310                 315
```

(2) INFORMACIÓN PARA ID de SEC Nº:3

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 759 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 756

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:3:

```
ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GGC CTT GTG CAA CCC          48
Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro
  1               5                  10                  15

ACT CGA CTT CTT TTG GAA TAT CTT GAA GAA AAA TAT GAA GAG CAT TTG          96
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu
            20                  25                  30

TAT GAG CGC GAT GAA GGT GAT AAA TGG CGA AAC AAA AAG TTT GAA TTG         144
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu
        35                  40                  45

GGT TTG GAG TTT CCC AAT CTT CCT TAT TAT ATT GAT GGT GAT GTT AAA         192
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
    50                  55                  60

TTA ACA CAG TCT ATG GCC ATC ATA CGT TAT ATA GCT GAC AAG CAC AAC         240
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn
65                  70                  75                  80

ATG TTG GGT GGT TGT CCA AAA GAG CGT GCA GAG ATT TCA ATG CTT GAA         288
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu
                85                  90                  95

GGA GCG GTT TTG GAT ATT AGA TAC GGT GTT TCG AGA ATT GCA TAT AGT         336
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser
            100                 105                 110

AAA GAC TTT GAA ACT CTC AAA GTT GAT TTT CTT AGC AAG CTA CCT GAA         384
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu
        115                 120                 125

ATG CTG AAA ATG TTC GAA GAT CGT TTA TGT CAT AAA ACA TAT TTA AAT         432
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn
    130                 135                 140

GGT GAT CAT GTA ACC CAT CCT GAC TTC ATG TTG TAT GAC GCT CTT GAT         480
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp
145                 150                 155                 160

GTT GTT TTA TAC ATG GAC CCA ATG TGC CTG GAT GCG TTC CCA AAA TTA         528
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu
                165                 170                 175

GTT TGT TTT AAA AAA CGT ATT GAA GCT ATC CCA CAA ATT GAT AAG TAC         576
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr
            180                 185                 190

TTG AAA TCC AGC AAG TAT ATA GCA TGG CCT TTG CAG GGC TGG CAA GCC         624
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala
        195                 200                 205

ACG TTT GGT GGT GGC GAC CAT CCT CCA AAA TCG GAT CTG GTT CCG CGT         672
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg
    210                 215                 220

GGA TCC ATG AGC ACG ATT CCC AAA CCT CAA AGA AAA ACC AAA CGT AAC         720
Gly Ser Met Ser Thr Ile Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn
225                 230                 235                 240

ACC AAC CGT CGC CCA CAG GAA TTC ATC GTG ACT GAC TGA                     759
Thr Asn Arg Arg Pro Gln Glu Phe Ile Val Thr Asp
                245                 250
```

(2) INFORMACIÓN PARA ID de SEC Nº:4

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 759 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 756

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

```
ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GGC CTT GTG CAA CCC          48
Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro
  1               5                  10                  15

ACT CGA CTT CTT TTG GAA TAT CTT GAA GAA AAA TAT GAA GAG CAT TTG          96
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu
            20                  25                  30

TAT GAG CGC GAT GAA GGT GAT AAA TGG CGA AAC AAA AAG TTT GAA TTG         144
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu
        35                  40                  45

GGT TTG GAG TTT CCC AAT CTT CCT TAT TAT ATT GAT GGT GAT GTT AAA         192
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
    50                  55                  60

TTA ACA CAG TCT ATG GCC ATC ATA CGT TAT ATA GCT GAC AAG CAC AAC         240
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn
65                  70                  75                  80

ATG TTG GGT GGT TGT CCA AAA GAG CGT GCA GAG ATT TCA ATG CTT GAA         288
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu
                85                  90                  95

GGA GCG GTT TTG GAT ATT AGA TAC GGT GTT TCG AGA ATT GCA TAT AGT         336
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser
            100                 105                 110

AAA GAC TTT GAA ACT CTC AAA GTT GAT TTT CTT AGC AAG CTA CCT GAA         384
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu
        115                 120                 125

ATG CTG AAA ATG TTC GAA GAT CGT TTA TGT CAT AAA ACA TAT TTA AAT         432
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn
    130                 135                 140

GGT GAT CAT GTA ACC CAT CCT GAC TTC ATG TTG TAT GAC GCT CTT GAT         480
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp
145                 150                 155                 160

GTT GTT TTA TAC ATG GAC CCA ATG TGC CTG GAT GCG TTC CCA AAA TTA         528
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu
                165                 170                 175

GTT TGT TTT AAA AAA CGT ATT GAA GCT ATC CCA CAA ATT GAT AAG TAC         576
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr
            180                 185                 190

TTG AAA TCC AGC AAG TAT ATA GCA TGG CCT TTG CAG GGC TGG CAA GCC         624
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala
        195                 200                 205

ACG TTT GGT GGT GGC GAC CAT CCT CCA AAA TCG GAT CTG GTT CCG CGT         672
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg
    210                 215                 220
```

```
GGA TCC GAC GTC AAG TTC CCG GGT GGC GGT CAG ATC GTT GGT GGA GTT        720
Gly Ser Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val
225                 230                 235                 240

TAC TTG TTG CCG CGC AGG GAA TTC ATC GTG ACT GAC TGA                  759
Tyr Leu Leu Pro Arg Arg Glu Phe Ile Val Thr Asp
                245                 250
```

(2) INFORMACIÓN PARA ID de SEC Nº:5

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 759 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 756

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:5:

```
ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GGC CTT GTG CAA CCC          48
Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro
  1               5                  10                  15

ACT CGA CTT CTT TTG GAA TAT CTT GAA GAA AAA TAT GAA GAG CAT TTG          96

 Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu
             20                  25                  30

TAT GAG CGC GAT GAA GGT GAT AAA TGG CGA AAC AAA AAG TTT GAA TTG         144
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu
        35                  40                  45

GGT TTG GAG TTT CCC AAT CTT CCT TAT TAT ATT GAT GGT GAT GTT AAA         192
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
    50                  55                  60

TTA ACA CAG TCT ATG GCC ATC ATA CGT TAT ATA GCT GAC AAG CAC AAC         240
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn
65                  70                  75                  80

ATG TTG GGT GGT TGT CCA AAA GAG CGT GCA GAG ATT TCA ATG CTT GAA         288
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu
                85                  90                  95

GGA GCG GTT TTG GAT ATT AGA TAC GGT GTT TCG AGA ATT GCA TAT AGT         336
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser
            100                 105                 110

AAA GAC TTT GAA ACT CTC AAA GTT GAT TTT CTT AGC AAG CTA CCT GAA         384
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu
        115                 120                 125

ATG CTG AAA ATG TTC GAA GAT CGT TTA TGT CAT AAA ACA TAT TTA AAT         432
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn
    130                 135                 140

GGT GAT CAT GTA ACC CAT CCT GAC TTC ATG TTG TAT GAC GCT CTT GAT         480
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp
145                 150                 155                 160

GTT GTT TTA TAC ATG GAC CCA ATG TGC CTG GAT GCG TTC CCA AAA TTA         528
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu
                165                 170                 175

GTT TGT TTT AAA AAA CGT ATT GAA GCT ATC CCA CAA ATT GAT AAG TAC         576
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr
            180                 185                 190

TTG AAA TCC AGC AAG TAT ATA GCA TGG CCT TTG CAG GGC TGG CAA GCC         624
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala
        195                 200                 205

ACG TTT GGT GGT GGC GAC CAT CCT CCA AAA TCG GAT CTG GTT CCG CGT         672
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg
    210                 215                 220

GGA TCC GGC CCT AGA TTG GGT GTG CGC GCG ACG AGG AAG ACT TCC GAG         720

Gly Ser Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr Ser Glu
225                 230                 235                 240

CGG TCG CAA CCT CGA GGT GAA TTC ATC GTG ACT GAC TGA                     759
Arg Ser Gln Pro Arg Gly Glu Phe Ile Val Thr Asp
                245                 250
```

(2) INFORMACIÓN PARA ID de SEC Nº:6

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 816 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 816

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:6:

```
ATG TCC CCT ATA CTA GGT TAT TGG AAA ATT AAG GGC CTT GTG CAA CCC        48
Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro
  1               5                  10                  15

ACT CGA CTT CTT TTG GAA TAT CTT GAA GAA AAA TAT GAA GAG CAT TTG        96
Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu
            20                  25                  30

TAT GAG CGC GAT GAA GGT GAT AAA TGG CGA AAC AAA AAG TTT GAA TTG       144
Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe Glu Leu
        35                  40                  45

GGT TTG GAG TTT CCC AAT CTT CCT TAT TAT ATT GAT GGT GAT GTT AAA       192
Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp Val Lys
    50                  55                  60

TTA ACA CAG TCT ATG GCC ATC ATA CGT TAT ATA GCT GAC AAG CAC AAC       240
Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn
65                  70                  75                  80

ATG TTG GGT GGT TGT CCA AAA GAG CGT GCA GAG ATT TCA ATG CTT GAA       288
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu
```

```
                    85                  90                  95

GGA GCG GTT TTG GAT ATT AGA TAC GGT GTT TCG AGA ATT GCA TAT AGT      336
Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser
            100                 105                 110

AAA GAC TTT GAA ACT CTC AAA GTT GAT TTT CTT AGC AAG CTA CCT GAA      384
Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu Pro Glu
        115                 120                 125

ATG CTG AAA ATG TTC GAA GAT CGT TTA TGT CAT AAA ACA TAT TTA AAT      432
Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr Leu Asn
    130                 135                 140

GGT GAT CAT GTA ACC CAT CCT GAC TTC ATG TTG TAT GAC GCT CTT GAT      480
Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp
145                 150                 155                 160

GTT GTT TTA TAC ATG GAC CCA ATG TGC CTG GAT GCG TTC CCA AAA TTA      528
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu
                165                 170                 175

GTT TGT TTT AAA AAA CGT ATT GAA GCT ATC CCA CAA ATT GAT AAG TAC      576
Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr
            180                 185                 190

TTG AAA TCC AGC AAG TAT ATA GCA TGG CCT TTG CAG GGC TGG CAA GCC      624
Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp Gln Ala
        195                 200                 205

ACG TTT GGT GGT GGC GAC CAT CCT CCA AAA TCG GAT CTG GTT CCG CGT      672
Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp Leu Val Pro Arg
    210                 215                 220

GGA TCC AGC ACG ATT CCC AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC      720
Gly Ser Ser Thr Ile Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr
225                 230                 235                 240

AAC CGT CGC CCA CAG GAC GTC AAG TTC CCG GGT GGC GGT CAG ATC GTT      768
Asn Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val
                245                 250                 255

GGT GGA GTT TAC TTG TTG CCG CGC AGG GAA TTC ATC GTG ACT GAC          813
Gly Gly Val Tyr Leu Leu Pro Arg Arg Glu Phe Ile Val Thr Asp
            260                 265                 270

TGA                                                                 816
```

(2) INFORMACIÓN PARA ID de SEC Nº:7

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 66 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:7:

```
GATCCATGAG CACGATTCCC AAACCTCAAA GAAAAACCAA ACGTAACACC AACCGTCGCC    60

CACAGG                                                               66
```

(2) INFORMACIÓN PARA ID de SEC Nº:8

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 66 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:8:

```
ATTCCTGTG GGCGACGGTT GGTGTTACGT TTGGTTTTTC TTTGAGGTTT GGGAATCGTG   60
CTCATG                                                             66
```

(2) INFORMACIÓN PARA ID de SEC Nº:9

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 66 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:9:

```
GATCCGACGT CAAGTTCCCG GGTGGCGGTC AGATCGTTGG TGGAGTTTAC TTGTTGCCGC   60
GCAGGG                                                             66
```

(2) INFORMACIÓN PARA ID de SEC Nº:10

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 66 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:10:

```
AATTCCCTGC GCGGCAACAA GTAAACTCCA CCAACGATCT GACCGCCACC CGGGAACTTG   60
ACGTCG                                                             66
```

(2) INFORMACIÓN PARA ID de SEC Nº:11

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 66 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:11:

GATCCGGCCC TAGATTGGGT GTGCGCGCGA CGAGGAAGAC TTCCGAGCGG TCGCAACCTC 60

GAGGTG 66

(2) INFORMACIÓN PARA ID de SEC Nº:12

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 66 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:12:

AATTCACCTC GAGGTTGCGA CCGCTCGGAA GTCTTCCTCG TCGCGCGCAC ACCCAATCTA 60

GGGCCG 66

(2) INFORMACIÓN PARA ID de SEC Nº:13

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 32 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:13:

GAATTCTTAC CTGCGCGGCA ACAAGTAAACTC 32

(2) INFORMACIÓN PARA ID de SEC Nº:14

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 32 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:14:

GCTGGATCCA GCACGATTCC CAAACCTCAA AG 32

(2) INFORMACIÓN PARA ID de SEC Nº:15

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 21 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:15:

ATGAGCACGA TTCCCAAACC T 21

(2) INFORMACIÓN PARA ID de SEC Nº:16

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 17 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:16:

GAGGAAGACT TCCGAGC 17

(2) INFORMACIÓN PARA ID de SEC Nº:17

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 17 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:17:

GTCCTGCCCT CGGGCCG 17

(2) INFORMACIÓN PARA ID de SEC Nº:18

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 21 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:18:

ACCCAAATTG CGCGACCTAC G 21

(2) INFORMACIÓN PARA ID de SEC Nº:19

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 19 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:19:

TGGGTAAGGT CATCGATAC 19

(2) INFORMACIÓN PARA ID de SEC Nº:20

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 17 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:20:

AAGGTCATCG ATACCCT 17

(2) INFORMACIÓN PARA ID de SEC Nº:21

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:21:

AGATAGAGAA AGAGCAAC 18

(2) INFORMACIÓN PARA ID de SEC Nº:22

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 22 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:22:

GGACCAGTTC ATCATCATAT AT 22

(2) INFORMACIÓN PARA ID de SEC Nº:23

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:23:

CAGTTCATCA TCATATCCCA 20

(2) INFORMACIÓN PARA ID de SEC Nº:24

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 15 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 15

(D) OTRA INFORMACIÓN: /producto= "Proteína Enlazante en GST-NANBV 693-691"

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:24:

15

```
GGG ATC CCC AAT TCA
Gly Ile Pro Asn Ser
  1               5
```

(2) INFORMACIÓN PARA ID de SEC Nº:25

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 5 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:25:

```
Gly Ile Pro Asn Ser
1               5
```

(2) INFORMACIÓN PARA ID de SEC Nº:26

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 12 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 9

(D) OTRA INFORMACIÓN: /producto= "Proteína Enlazante Carboxi-terminal en GST-NANBV 693-691"

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:26:

```
AAT TCA TCG TGA                                                    12
Asn Ser Ser
  1
```

(2) INFORMACIÓN PARA ID de SEC Nº:27

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 3 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:27:

Asn Ser Ser
1

(2) INFORMACIÓN PARA ID de SEC Nº:28

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 27 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 27

(D) OTRA INFORMACIÓN: /producto= “Proteína Enlazante en GST-NANBV 15-18”

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:28:

```
GGG ATC CCC ATC GAA TTC CTG CAG CCC                                27
Gly Ile Pro Ile Glu Phe Leu Gln Pro
  1               5
```

(2) INFORMACIÓN PARA ID de SEC Nº:29

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 9 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:29:

Gly Ile Pro Ile Glu Phe Leu Gln Pro
1   5

(2) INFORMACIÓN PARA ID de SEC Nº:30

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 24 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 21

(D) OTRA INFORMACIÓN: /producto= "Proteína Enlazante Carboxi-terminal en GST-NANBV 15-18"

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:30:

```
TGG GGG ATC GGG AAT TCA TCG TGA
Trp Gly Ile Gly Asn Ser Ser
  1               5
```

24

(2) INFORMACIÓN PARA ID de SEC Nº:31

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 7 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:31:

```
Trp Gly Ile Gly Asn Ser Ser
1               5
```

(2) INFORMACIÓN PARA ID de SEC Nº:32

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 24 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 24

(D) OTRA INFORMACIÓN: /producto= "Proteína Enlazante en GST-NANBV 15-17"

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:32:

```
GGG ATC CCC AAT TCC TGC AGC CCT
Gly Ile Pro Asn Ser Cys Ser Pro
  1               5
```

24

(2) INFORMACIÓN PARA ID de SEC Nº:33

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 8 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:33:

Gly Ile Pro Asn Ser Cys Ser Pro
1 5

(2) INFORMACIÓN PARA ID de SEC Nº:34

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 21 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 18

(D) OTRA INFORMACIÓN: /producto= “Proteína Enlazante Carboxi-terminal en GST-NANBV 15-17”

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:34:

GGG ATC GGG AAT TCA TCG TGA
Gly Ile Gly Asn Ser Ser
1 5

21

(2) INFORMACIÓN PARA ID de SEC Nº:35

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 6 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:35:

Gly Ile Gly Asn Ser Ser
1 5

(2) INFORMACIÓN PARA ID de SEC Nº:36

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 15 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 15

(D) OTRA INFORMACIÓN: /producto= “Sitio de Escisión de la Trombina en GST-NANBV 15-17”

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:36:

```
GTT CCG CGT GGA TCC
Val Pro Arg Gly Ser
 1               5
```

15

(2) INFORMACIÓN PARA ID de SEC Nº:37

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 5 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:37:

```
Val Pro Arg Gly Ser
1               5
```

(2) INFORMACIÓN PARA ID de SEC Nº:38

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 21 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 21

(D) OTRA INFORMACIÓN: /producto= “Proteína Enlazante en GST-NANBV 15-17”

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:38:

```
CCA TCG AAT TCC TGC AGC CCT
Pro Ser Asn Ser Cys Ser Pro
1               5
```

21

(2) INFORMACIÓN PARA ID de SEC Nº:39

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 7 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:39:

Pro Ser Asn Ser Cys Ser Pro
1 5

(2) INFORMACIÓN PARA ID de SEC Nº:40

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 18 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 15

(D) OTRA INFORMACIÓN: /producto= “Proteína Enlazante Carboxi-terminal en GST-NANBV 15-17”

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:40:

GGA ATT CAT CGT GAC TGA 18
Gly Ile His Arg Asp
1 5

(2) INFORMACIÓN PARA ID de SEC Nº:41

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 5 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:41:

Gly Ile His Arg Asp
1 5

(2) INFORMACIÓN PARA ID de SEC Nº:42

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 27 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 27

(D) OTRA INFORMACIÓN: /producto= “Proteína Enlazante en GST-NANBV 690-691”

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:42:

```
GGG ATC CCC AAT TCG AGC TCG GTA CCC                                    27
Gly Ile Pro Asn Ser Ser Ser Val Pro
  1               5
```

(2) INFORMACIÓN PARA ID de SEC Nº:43

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 9 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:43:

```
Gly Ile Pro Asn Ser Ser Ser Val Pro
1               5
```

(2) INFORMACIÓN PARA ID de SEC Nº:44

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 24 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 1.. 21

(D) OTRA INFORMACIÓN: /producto= “Proteína Enlazante Carboxi-terminal en GST-NANBV 690-691”

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:44:

```
ACG GGG ATC GGG AAT TCA TCG TGA                                        24
Thr Gly Ile Gly Asn Ser Ser
  1               5
```

(2) INFORMACIÓN PARA ID de SEC Nº:45

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 7 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:45:

Thr Gly Ile Gly Asn Ser Ser
1 5

(2) INFORMACIÓN PARA ID de SEC Nº:46

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 9416 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: cDNA

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 342.. 9374

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: característica miscelánea

(B) LOCALIZACIÓN: 1.. 12

(D) OTRA INFORMACIÓN: /nota= “No confirmada como secuencia de HCV-Hc59”

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: característica miscelánea

(B) LOCALIZACIÓN: 9397.. 9416

(D) OTRA INFORMACIÓN: /nota= “No confirmada como secuencia de HCV-Hc59”

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: unidad de repetición

(B) LOCALIZACIÓN: grupo (7...12, 42...47)

(D) OTRA INFORMACIÓN: /tipo de rpt= “otro”
/familia de rpt= “1”

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: unidad de repetición

(B) LOCALIZACIÓN: grupo (23...28, 38...43, 9209...9214, 9391...9396)

(D) OTRA INFORMACIÓN: /tipo de rpt= “otro”
/familia de rpt= “2”

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: unidad de repetición

(B) LOCALIZACIÓN: grupo (128...135, 315...322)

(D) OTRA INFORMACIÓN: /tipo de rpt= “otro”
/familia de rpt= “3”

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: unidad de repetición

(B) LOCALIZACIÓN: grupo (9231...9237, 9245...9251, 9256...9262)

(D) OTRA INFORMACIÓN: /tipo de rpt= “otro”
/familia de rpt= “4”

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: unidad de repetición

(B) LOCALIZACIÓN: grupo (9248...9253, 9221...9226, 9227...9232)

(D) OTRA INFORMACIÓN: /tipo de rpt= “otro”
/familia de rpt= “5”

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:46:

```
GCCAGCCCCC TGATGGGGGC GACACTCCAC CATGAATCAC TCCCCTGTGA GGAACTACTG      60

TCTTCACGCA GAAAGCGTCT AGCCATGGCG TTAGTATGAG TGTCGTGCAG CCTCCAGGAC     120

CCCCCCTCCC GGGAGAGCCA TAGTGGTCTG CGGAACCGGT GAGTACACCG GAATTGCCAG     180

GACGACCGGG TCCTTTCTTG GATAAACCCG CTCAATGCCT GGAGATTTGG GCGTGCCCCC     240

GCAAGACTGC TAGCCGAGTA GTGTTGGGTC GCGAAAGGCC TTGTGGTACT GCCTGATAGG     300

GTGCTTGCGA GTGCCCCGGG AGGTCTCGTA GACCGTGCAC C ATG AGC ACG AAT         353
                                               Met Ser Thr Asn
                                                 1

CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGT CGC CCA CAG       401
Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn Arg Arg Pro Gln
  5                  10                  15                  20

GAC GTC AAG TTC CCG GGT GGC GGT CAG ATC GTT GGT GGA GTT TAC TTG       449
Asp Val Lys Phe Pro Gly Gly Gly Gln Ile Val Gly Gly Val Tyr Leu
                  25                  30                  35

TTG CCG CGC AGG GGC CCT AGA TTG GGT GTG CGC GCG ACG AGG AAG ACT       497
Leu Pro Arg Arg Gly Pro Arg Leu Gly Val Arg Ala Thr Arg Lys Thr
              40                  45                  50

TCC GAG CGG TCG CAA CCT CGA GGT AGA CGT CAG CCT ATC CCC AAG GCA       545
Ser Glu Arg Ser Gln Pro Arg Gly Arg Arg Gln Pro Ile Pro Lys Ala
          55                  60                  65

CGT CGG CCC GAG GGC AGG ACC TGG GCT CAG CCC GGG TAC CCT TGG CCC       593
Arg Arg Pro Glu Gly Arg Thr Trp Ala Gln Pro Gly Tyr Pro Trp Pro
      70                  75                  80

CTC TAT GGC AAT GAG GGT TGC GGG TGG GCG GGA TGG CTC CTG TCT CCC       641
Leu Tyr Gly Asn Glu Gly Cys Gly Trp Ala Gly Trp Leu Leu Ser Pro
 85                  90                  95                  100

CGT GGC TCT CGG CCT AGC TGG GGC CCC ACA GAC CCC CGG CGT AGG TCG       689
Arg Gly Ser Arg Pro Ser Trp Gly Pro Thr Asp Pro Arg Arg Arg Ser
                 105                 110                 115

CGC AAT TTG GGT AAG GTC ATC GAT ACC CTT ACG TGC GGC TTC GCC GAC       737
Arg Asn Leu Gly Lys Val Ile Asp Thr Leu Thr Cys Gly Phe Ala Asp
             120                 125                 130

CTC ATG GGG TAC ATA CCG CTC GTC GGC GCC CCT CTT GGA GGC GCT GCC       785
Leu Met Gly Tyr Ile Pro Leu Val Gly Ala Pro Leu Gly Gly Ala Ala
         135                 140                 145
```

```
AGG GCC CTG GCG CAT GGC GTC CGG GTT CTG GAA GAC GGC GTG AAC TAT        833
Arg Ala Leu Ala His Gly Val Arg Val Leu Glu Asp Gly Val Asn Tyr
    150                 155                 160

GCA ACA GGG AAC CTT CCT GGT TGC TCT TTC TCT ATC TTC CTT CTG GCC        881
Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser Ile Phe Leu Leu Ala
165                 170                 175                 180

CTG CTC TCT TGC CTG ACT GTG CCC GCT TCA GCC TAC CAA GTG CGC AAT        929
Leu Leu Ser Cys Leu Thr Val Pro Ala Ser Ala Tyr Gln Val Arg Asn
                185                 190                 195

TCC TCG GGG CTT TAC CAT GTC ACC AAT GAT TGC CCT AAC TCG AGT GTT        977
Ser Ser Gly Leu Tyr His Val Thr Asn Asp Cys Pro Asn Ser Ser Val
            200                 205                 210

GTG TAC GAG GCG GCC GAT GCC ATC CTG CAC ACT CCG GGG TGT GTC CCT       1025
Val Tyr Glu Ala Ala Asp Ala Ile Leu His Thr Pro Gly Cys Val Pro
        215                 220                 225

TGC GTT CGC GAG GGT AAC GCC TCG AGG TGT TGG GTG GCG GTG ACC CCC       1073
Cys Val Arg Glu Gly Asn Ala Ser Arg Cys Trp Val Ala Val Thr Pro
    230                 235                 240

ACG GTG GCC ACC AGG GAC GGC AAA CTC CCC ACA ACG CAG CTT CGA CGT       1121
Thr Val Ala Thr Arg Asp Gly Lys Leu Pro Thr Thr Gln Leu Arg Arg
245                 250                 255                 260

CAT ATC GAT CTG CTT GTC GGG AGC GCC ACC CTC TGC TCG GCC CTC TAC       1169
His Ile Asp Leu Leu Val Gly Ser Ala Thr Leu Cys Ser Ala Leu Tyr
                265                 270                 275

GTG GGG GAC CTG TGC GGG TCT GTC TTT CTT GTT GGT CAA CTG TTT ACC       1217
Val Gly Asp Leu Cys Gly Ser Val Phe Leu Val Gly Gln Leu Phe Thr
            280                 285                 290

TTC TCT CCC AGG CAC CAC TGG ACG ACG CAA GAC TGC AAT TGT TCT ATC       1265
Phe Ser Pro Arg His His Trp Thr Thr Gln Asp Cys Asn Cys Ser Ile
        295                 300                 305

TAT CCC GGC CAT ATA ACG GGT CAT CGC ATG GCA TGG AAT ATG ATG ATG       1313
Tyr Pro Gly His Ile Thr Gly His Arg Met Ala Trp Asn Met Met Met
    310                 315                 320

AAC TGG TCC CCT ACG GCA GCG TTG GTG GTA GCT CAG CTG CTC CGA ATC       1361
Asn Trp Ser Pro Thr Ala Ala Leu Val Val Ala Gln Leu Leu Arg Ile
325                 330                 335                 340

CCA CAA GCC ATC ATG GAC ATG ATC GCT GGC GCC CAC TGG GGA GTC CTG       1409
Pro Gln Ala Ile Met Asp Met Ile Ala Gly Ala His Trp Gly Val Leu
                345                 350                 355
```

```
GCG GGC ATA AAG TAT TTC TCC ATG GTG GGG AAC TGG GCG AAG GTC CTG      1457
Ala Gly Ile Lys Tyr Phe Ser Met Val Gly Asn Trp Ala Lys Val Leu
            360                 365                 370

GTA GTG CTG CTG CTA TTT GCC GGC GTC GAC GCG GAA ACC CAC GTC ACC      1505
Val Val Leu Leu Leu Phe Ala Gly Val Asp Ala Glu Thr His Val Thr
        375                 380                 385

GGG GGA AAT GCC GGC CGC ACC ACG GCT GGG CTT GTT GGT CTC CTT ACA      1553
Gly Gly Asn Ala Gly Arg Thr Thr Ala Gly Leu Val Gly Leu Leu Thr
    390                 395                 400

CCA GGC GCC AAG CAG AAC ATC CAA CTG ATC AAC ACC AAC GGC AGT TGG      1601
Pro Gly Ala Lys Gln Asn Ile Gln Leu Ile Asn Thr Asn Gly Ser Trp
405                 410                 415                 420

CAC ATC AAT AGC ACG GCC TTG AAC TGC AAT GAA AGC CTT AAC ACC GGC      1649
His Ile Asn Ser Thr Ala Leu Asn Cys Asn Glu Ser Leu Asn Thr Gly
                425                 430                 435

TGG TTA GCA GGG CTC TTC TAT CAG CAC AAA TTC AAC TCT TCA GGC TGT      1697
Trp Leu Ala Gly Leu Phe Tyr Gln His Lys Phe Asn Ser Ser Gly Cys
            440                 445                 450

CCT GAG AGG TTG GCC AGC TGC CGA CGC CTT ACC GAT TTT GCC CAG GGC      1745
Pro Glu Arg Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Ala Gln Gly
        455                 460                 465

TGG GGT CCT ATC AGT TAT GCC AAC GGA AGC GGC CTC GAC GAA CGC CCC      1793
Trp Gly Pro Ile Ser Tyr Ala Asn Gly Ser Gly Leu Asp Glu Arg Pro
    470                 475                 480

TAC TGC TGG CAC TAC CCT CCA AGA CCT TGT GGC ATT GTG CCC GCA AAG      1841
Tyr Cys Trp His Tyr Pro Pro Arg Pro Cys Gly Ile Val Pro Ala Lys
485                 490                 495                 500

AGC GTG TGT GGC CCG GTA TAT TGC TTC ACT CCC AGC CCC GTG GTG GTG      1889
Ser Val Cys Gly Pro Val Tyr Cys Phe Thr Pro Ser Pro Val Val Val
                505                 510                 515

GGA ACG ACC GAC AGG TCG GGC GCG CCT ACC TAC AGC TGG GGT GCA AAT      1937
Gly Thr Thr Asp Arg Ser Gly Ala Pro Thr Tyr Ser Trp Gly Ala Asn
            520                 525                 530

GAT ACG GAT GTC TTC GTC CTT AAC AAC ACC AGG CCA CCG CTG GGC AAT      1985
Asp Thr Asp Val Phe Val Leu Asn Asn Thr Arg Pro Pro Leu Gly Asn
        535                 540                 545

TGG TTC GGT TGT ACC TGG ATG AAC TCA ACT GGA TTC ACC AAA GTG TGC      2033
Trp Phe Gly Cys Thr Trp Met Asn Ser Thr Gly Phe Thr Lys Val Cys
    550                 555                 560
```

```
GGA GCG CCC CCT TGT GTC ATC GGA GGG GTG GGC AAC AAC ACC TTG CTC      2081
Gly Ala Pro Pro Cys Val Ile Gly Gly Val Gly Asn Asn Thr Leu Leu
565                 570                 575                 580

TGC CCC ACT GAT TGC TTC CGC AAA TAT CCG GAA GCC ACA TAC TCT CGG      2129
Cys Pro Thr Asp Cys Phe Arg Lys Tyr Pro Glu Ala Thr Tyr Ser Arg
                585                 590                 595

TGC GGC TCC GGT CCC AGG ATT ACA CCC AGG TGC ATG GTC GAC TAC CCG      2177
Cys Gly Ser Gly Pro Arg Ile Thr Pro Arg Cys Met Val Asp Tyr Pro
            600                 605                 610

TAT AGG CTT TGG CAC TAT CCT TGT ACC ATC AAT TAC ACC ATA TTC AAA      2225
Tyr Arg Leu Trp His Tyr Pro Cys Thr Ile Asn Tyr Thr Ile Phe Lys
        615                 620                 625

GTC AGG ATG TAC GTG GGA GGG GTC GAG CAC AGG CTG GAA GCG GCC TGC      2273
Val Arg Met Tyr Val Gly Gly Val Glu His Arg Leu Glu Ala Ala Cys
    630                 635                 640

AAC TGG ACG CGG GGC GAA CGC TGT GAT CTG GAA GAC AGG GAC AGG TCC      2321
Asn Trp Thr Arg Gly Glu Arg Cys Asp Leu Glu Asp Arg Asp Arg Ser
645                 650                 655                 660

GAG CTC AGC CCG TTG CTG CTG TCC ACC ACA CAG TGG CAG GTC CTT CCG      2369
Glu Leu Ser Pro Leu Leu Leu Ser Thr Thr Gln Trp Gln Val Leu Pro
                665                 670                 675

TGT TCT TTC ACG ACC CTG CCA GCC TTG TCC ACC GGC CTC ATC CAC CTC      2417
Cys Ser Phe Thr Thr Leu Pro Ala Leu Ser Thr Gly Leu Ile His Leu
            680                 685                 690

CAC CAG AAC ATT GTG GAC GTG CAG TAC TTG TAC GGG GTA GGG TCA AGC      2465
His Gln Asn Ile Val Asp Val Gln Tyr Leu Tyr Gly Val Gly Ser Ser
        695                 700                 705

ATC GCG TCC TGG GCC ATT AAG TGG GAG TAC GTC GTT CTC CTG TTC CTT      2513
Ile Ala Ser Trp Ala Ile Lys Trp Glu Tyr Val Val Leu Leu Phe Leu
    710                 715                 720

CTG CTT GCA GAC GCG CGC GTC TGT TCC TGC TTG TGG ATG ATG TTA CTC      2561
Leu Leu Ala Asp Ala Arg Val Cys Ser Cys Leu Trp Met Met Leu Leu
725                 730                 735                 740

ATA TCC CAA GCG GAG GCG GCT TTG GAG AAC CTC GTA ATA CTC AAT GCA      2609
Ile Ser Gln Ala Glu Ala Ala Leu Glu Asn Leu Val Ile Leu Asn Ala
                745                 750                 755

GCA TCC CTG GCC GGG ACG CAT GGT CTT GTG TCC TTC CTC GTG TTC TTC      2657
Ala Ser Leu Ala Gly Thr His Gly Leu Val Ser Phe Leu Val Phe Phe
            760                 765                 770
```

```
TGC TTT GCG TGG TAT CTG AAG GGT AGG TGG GTG CCC GGA GCG GTC TAC      2705
Cys Phe Ala Trp Tyr Leu Lys Gly Arg Trp Val Pro Gly Ala Val Tyr
        775                 780                 785

GCC CTC TAC GGG ATG TGG CCT CTC CTC CTG CTC CTG CTG GCG TTG CCT      2753
Ala Leu Tyr Gly Met Trp Pro Leu Leu Leu Leu Leu Leu Ala Leu Pro
    790                 795                 800

CAG CGG GCA TAC GCA CTG GAC ACG GAG GTG GCC GCG TCG TGT GGC GGC      2801
Gln Arg Ala Tyr Ala Leu Asp Thr Glu Val Ala Ala Ser Cys Gly Gly
805                 810                 815                 820

GTT GTT CTT GTC GGG TTA ATG GCG CTG ACT CTG TCG CCA TAT TAC AAG      2849
Val Val Leu Val Gly Leu Met Ala Leu Thr Leu Ser Pro Tyr Tyr Lys
                825                 830                 835

CGC TAT ATC AGC TGG TGC ATG TGG TGG CTT CAG TAT TTT CTG ACC AGA      2897
Arg Tyr Ile Ser Trp Cys Met Trp Trp Leu Gln Tyr Phe Leu Thr Arg
            840                 845                 850

GTA GAA GCG CAA CTG CAC GTG TGG GTT CCC CCC CTC AAC GTC CGG GGG      2945
Val Glu Ala Gln Leu His Val Trp Val Pro Pro Leu Asn Val Arg Gly
        855                 860                 865

GGG CGC GAT GCC GTC ATC TTA CTC ACG TGT GTA GTA CAC CCG GCC CTG      2993
Gly Arg Asp Ala Val Ile Leu Leu Thr Cys Val Val His Pro Ala Leu
    870                 875                 880

GTA TTT GAC ATC ACC AAA CTA CTC CTG GCC ATC TTC GGA CCC CTT TGG      3041
Val Phe Asp Ile Thr Lys Leu Leu Leu Ala Ile Phe Gly Pro Leu Trp
885                 890                 895                 900

ATT CTT CAA GCC AGT TTG CTT AAA GTC CCC TAC TTC GTG CGC GTT CAA      3089
Ile Leu Gln Ala Ser Leu Leu Lys Val Pro Tyr Phe Val Arg Val Gln
                905                 910                 915

GGC CTT CTC CGG ATC TGC GCG CTA GCG CGG AAG ATA GCC GGA GGT CAT      3137
Gly Leu Leu Arg Ile Cys Ala Leu Ala Arg Lys Ile Ala Gly Gly His
            920                 925                 930

TAC GTG CAA ATG GCC ATC ATC AAG TTA GGG GCG CTT ACT GGC ACC TGT      3185
Tyr Val Gln Met Ala Ile Ile Lys Leu Gly Ala Leu Thr Gly Thr Cys
        935                 940                 945

GTG TAT AAC CAT CTC GCT CCT CTT CGA GAC TGG GCG CAC AAC GGC CTG      3233
Val Tyr Asn His Leu Ala Pro Leu Arg Asp Trp Ala His Asn Gly Leu
    950                 955                 960

CGA GAT CTG GCC GTG GCT GTG GAA CCA GTC GTC TTC TCC CGA ATG GAG      3281
Arg Asp Leu Ala Val Ala Val Glu Pro Val Val Phe Ser Arg Met Glu
965                 970                 975                 980
```

```
ACC AAG CTC ATC ACG TGG GGG GCA GAT ACC GCC GCG TGC GGT GAC ATC      3329
Thr Lys Leu Ile Thr Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile
                985                 990                 995

ATC AAC GGC TTG CCC GTC TCT GCC CGT AGG GGC CAG GAG ATA CTG CTT      3377
Ile Asn Gly Leu Pro Val Ser Ala Arg Arg Gly Gln Glu Ile Leu Leu
           1000                1005                1010

GGG CCA GCC GAC GGA ATG GTC TCC AAG GGG TGG AGG TTG CTG GCG CCC      3425
Gly Pro Ala Asp Gly Met Val Ser Lys Gly Trp Arg Leu Leu Ala Pro
      1015                1020                1025

ATC ACG GCG TAC GCC CAG CAG ACG AGA GGC CTC CTA GGG TGT ATA ATC      3473
Ile Thr Ala Tyr Ala Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile
 1030                1035                1040

ACC AGC CTG ACT GGC CGG GAC AAA AAC CAA GTG GAG GGT GAG GTC CAG      3521
Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Glu Gly Glu Val Gln
1045                1050                1055                1060

ATC GTG TCA ACT GCT ACC CAG ACC TTC CTG GCA ACG TGC ATC AAT GGG      3569
Ile Val Ser Thr Ala Thr Gln Thr Phe Leu Ala Thr Cys Ile Asn Gly
                1065                1070                1075

GTA TGC TGG ACT GTC TAC CAC GGG GCC GGA ACG AGG ACC ATC GCA TCA      3617
Val Cys Trp Thr Val Tyr His Gly Ala Gly Thr Arg Thr Ile Ala Ser
           1080                1085                1090

CCC AAG GGT CCT GTC ATC CAG ACG TAT ACC AAT GTG GAT CAA GAC CTC      3665
Pro Lys Gly Pro Val Ile Gln Thr Tyr Thr Asn Val Asp Gln Asp Leu
      1095                1100                1105

GTG GGC TGG CCC GCT CCT CAA GGT TCC CGC TCA TTG ACA CCC TGC ACC      3713
Val Gly Trp Pro Ala Pro Gln Gly Ser Arg Ser Leu Thr Pro Cys Thr
 1110                1115                1120

TGC GGC TCC TCG GAC CTT TAC CTG GTC ACG AGG CAC GCC GAT GTC ATT      3761
Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile
1125                1130                1135                1140

CCC GTG CGC CGG CGA GGT GAT AGC AGG GGT AGC CTG CTT TCG CCC CGG      3809
Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg
                1145                1150                1155

CCC ATT TCC TAC TTG AAA GGC TCC TCG GGG GGT CCG CTG TTG TGC CCC      3857
Pro Ile Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro
           1160                1165                1170

ACG GGA CAC GCC GTG GGC CTA TTC AGG GCC GCG GTG TGC ACC CGT GGA      3905
Thr Gly His Ala Val Gly Leu Phe Arg Ala Ala Val Cys Thr Arg Gly
      1175                1180                1185
```

```
GTG GCT AAG GCG GTG GAC TTT ATC CCT GTG GAG AAC CTA GAG ACA ACC      3953
Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Asn Leu Glu Thr Thr
    1190                1195                1200

ATG AGA TCC CCG GTG TTC ACG GAC AAC TCC TCT CCA CCA GCA GTG CCC      4001
Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro
1205                1210                1215                1220

CAG AGC TTC CAG GTG GCC CAC CTG CAT GCT CCC ACC GGC AGC GGT AAG      4049
Gln Ser Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser Gly Lys
                1225                1230                1235

AGC ACC AAG GTC CCG GCT GCG TAC GCA GCC AAG GGC TAC AAG GTG TTG      4097
Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Lys Gly Tyr Lys Val Leu
            1240                1245                1250

GTG CTC AAC CCC TCT GTT GCT GCA ACA CTG GGC TTT GGT GCT TAC ATG      4145
Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met
        1255                1260                1265

TCC AAG GCC CAT GGG GTT GAT CCT AAT ATC AGG ACC GGG GTG AGA ACA      4193
Ser Lys Ala His Gly Val Asp Pro Asn Ile Arg Thr Gly Val Arg Thr
    1270                1275                1280

ATT ACC ACT GGC AGC CCC ATC ACG TAC TCC ACC TAC GGC AAG TTC CTT      4241
Ile Thr Thr Gly Ser Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu
1285                1290                1295                1300

GCC GAC GCC GGG TGC TCA GGA GGT GCT TAT GAC ATA ATA ATT TGT GAC      4289
Ala Asp Ala Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp
                1305                1310                1315

GAG TGC CAC TCC ACG GAT GCC ACA TCC ATC TCG GGC ATC GGC ACT GTC      4337
Glu Cys His Ser Thr Asp Ala Thr Ser Ile Ser Gly Ile Gly Thr Val
            1320                1325                1330

CTT GAC CAA GCA GAG ACT GCG GGG GCG AGA CTG GTT GTG CTC GCC ACT      4385
Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr
        1335                1340                1345

GCT ACC CCT CCG GGC TCC GTC ACT GTG TCC CAT CCT AAC ATC GAG GAG      4433
Ala Thr Pro Pro Gly Ser Val Thr Val Ser His Pro Asn Ile Glu Glu
    1350                1355                1360

GTT GCT CTG TCC ACC ACC GGA GAG ATC CCC TTT TAC GGC AAG GCT ATC      4481
Val Ala Leu Ser Thr Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile
1365                1370                1375                1380

CCC CTC GAG GTG ATC AAG GGG GGA AGA CAT CTC ATC TTC TGC CAC TCA      4529
Pro Leu Glu Val Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser
                1385                1390                1395
```

```
AAG AAG AAG TGC GAC GAG CTC GCC GCG AAG CTG GTC GCA TTG GGC ATC      4577
Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Val Ala Leu Gly Ile
            1400                1405                1410

AAT GCC GTG GCC TAC TAC CGC GGT CTT GAC GTG TCT GTC ATC CCG ACC      4625
Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr
        1415                1420                1425

AGC GGC GAT GTT GTC GTC GTG TCG ACC GAT GCT CTC ATG ACT GGC TTT      4673
Ser Gly Asp Val Val Val Val Ser Thr Asp Ala Leu Met Thr Gly Phe
    1430                1435                1440

ACC GGC GAC TTC GAC TCT GTG ATA GAC TGC AAC ACG TGT GTC ACT CAG      4721
Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln
1445                1450                1455                1460

ACA GTC GAT TTT AGC CTT GAC CCT ACC TTT ACC ATT GAG ACA ACC ACG      4769
Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr
                1465                1470                1475

CTC CCC CAG GAT GCT GTC TCC AGG ACT CAA CGC CGG GGC AGG ACT GGC      4817
Leu Pro Gln Asp Ala Val Ser Arg Thr Gln Arg Arg Gly Arg Thr Gly
            1480                1485                1490

AGG GGG AAG CCA GGC ATC TAT AGA TTT GTG GCA CCG GGG GAG CGC CCC      4865
Arg Gly Lys Pro Gly Ile Tyr Arg Phe Val Ala Pro Gly Glu Arg Pro
        1495                1500                1505

TCC GGC ATG TTC GAC TCG TCC GTC CTC TGT GAG TGC TAT GAC GCG GGC      4913
Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly
    1510                1515                1520

TGT GCT TGG TAT GAG CTC ACG CCC GCC GAG ACT ACA GTT AGG CTA CGA      4961
Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg Leu Arg
1525                1530                1535                1540

GCG TAC ATG AAC ACC CCG GGG CTT CCC GTG TGC CAG GAC CAT CTT GGA      5009
Ala Tyr Met Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Gly
                1545                1550                1555

TTT TGG GAG GGC GTC TTT ACG GGC CTC ACT CAT ATA GAT GCC CAC TTT      5057
Phe Trp Glu Gly Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe
            1560                1565                1570

CTA TCC CAG ACA AAG CAG AGT GGG GAG AAC TTT CCT TAC CTG GTA GCG      5105
Leu Ser Gln Thr Lys Gln Ser Gly Glu Asn Phe Pro Tyr Leu Val Ala
        1575                1580                1585

TAC CAA GCC ACC GTG TGC GCT AGG GCT CAA GCC CCT CCC CCA TCG TGG      5153
Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp
    1590                1595                1600
```

```
GAC CAG ATG CGG AAG TGT TTG ATC CGC CTT AAA CCC ACC CTC CAT GGG      5201
Asp Gln Met Arg Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly
1605                1610                1615                1620

CCA ACA CCC CTG CTA TAC AGA CTG GGC GCT GTT CAG AAT GAA GTC ACC      5249
Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr
                    1625                1630                1635

CTG ACG CAC CCA ATC ACC AAA TAC ATC ATG ACA TGC ATG TCG GCC GAC      5297
Leu Thr His Pro Ile Thr Lys Tyr Ile Met Thr Cys Met Ser Ala Asp
            1640                1645                1650

CTG GAG GTC GTC ACG AGC ACC TGG GTG CTC GTT GGC GGC GTC CTG GCT      5345
Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala
        1655                1660                1665

GCT CTG GCC GCG TAT TGC CTG TCA ACA GGC TGC GTG GTC ATA GTG GGC      5393
Ala Leu Ala Ala Tyr Cys Leu Ser Thr Gly Cys Val Val Ile Val Gly
    1670                1675                1680

AGG ATC GTC TTG TCC GGG AAG CCG GCA ATT ATA CCT GAC AGG GAG GTT      5441
Arg Ile Val Leu Ser Gly Lys Pro Ala Ile Ile Pro Asp Arg Glu Val
1685                1690                1695                1700

CTC TAC CAG GAG TTC GAT GAG ATG GAA GAG TGC TCT CAG CAC TTA CCG      5489
Leu Tyr Gln Glu Phe Asp Glu Met Glu Glu Cys Ser Gln His Leu Pro
                1705                1710                1715

TAC ATC GAG CAA GGG ATG ATG CTC GCT GAG CAG TTC AAG CAG AAG GCC      5537
Tyr Ile Glu Gln Gly Met Met Leu Ala Glu Gln Phe Lys Gln Lys Ala
            1720                1725                1730

CTC GGC CTC CTG CAG ACC GCG TCC CGC CAT GCA GAG GTT ATC ACC CCT      5585
Leu Gly Leu Leu Gln Thr Ala Ser Arg His Ala Glu Val Ile Thr Pro
        1735                1740                1745

GCT GTC CAG ACC AAC TGG CAG AAA CTC GAG GTC TTT TGG GCG AAG CAC      5633
Ala Val Gln Thr Asn Trp Gln Lys Leu Glu Val Phe Trp Ala Lys His
    1750                1755                1760

ATG TGG AAT TTC ATC AGT GGG ATA CAA TAC TTG GCG GGC CTG TCA ACG      5681
Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr
1765                1770                1775                1780

CTG CCT GGT AAC CCC GCC ATT GCT TCA TTG ATG GCT TTT ACA GCT GCC      5729
Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ala
                1785                1790                1795

GTC ACC AGC CCA CTA ACC ACT GGC CAA ACC CTC CTC TTC AAC ATA TTG      5777
Val Thr Ser Pro Leu Thr Thr Gly Gln Thr Leu Leu Phe Asn Ile Leu
            1800                1805                1810
```

```
GGG GGG TGG GTG GCT GCC CAG CTC GCC GCC CCC GGT GCC GCT ACC GCC      5825
Gly Gly Trp Val Ala Ala Gln Leu Ala Ala Pro Gly Ala Ala Thr Ala
        1815                1820                1825

TTT GTG GGC GCT GGC TTA GCT GGC GCC GCA CTC GAC AGC GTT GGA CTG      5873
Phe Val Gly Ala Gly Leu Ala Gly Ala Ala Leu Asp Ser Val Gly Leu
    1830                1835                1840

GGG AAG GTC CTC GTG GAC ATT CTT GCA GGC TAT GGC GCG GGC GTG GCG      5921
Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr Gly Ala Gly Val Ala
1845                1850                1855                1860

GGA GCT CTT GTG GCA TTC AAG ATC ATG AGC GGT GAG GTC CCC TCC ACG      5969
Gly Ala Leu Val Ala Phe Lys Ile Met Ser Gly Glu Val Pro Ser Thr
                1865                1870                1875

GAG GAC CTG GTC AAT CTG CTG CCC GCC ATC CTC TCA CCT GGA GCC CTT      6017
Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu Ser Pro Gly Ala Leu
            1880                1885                1890

GCA GTC GGT GTG GTC TTT GCA TCA ATA CTG CGC CGG CGT GTT GGC CCG      6065
Ala Val Gly Val Val Phe Ala Ser Ile Leu Arg Arg Arg Val Gly Pro
        1895                1900                1905

GGC GAG GGG GCA GTG CAA TGG ATG AAC CGG CTA ATA GCC TTC GCC TCC      6113
Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu Ile Ala Phe Ala Ser
    1910                1915                1920

CGG GGG AAC CAT GTT TCC CCC ACA CAC TAC GTG CCG GAG AGC GAT GCA      6161
Arg Gly Asn His Val Ser Pro Thr His Tyr Val Pro Glu Ser Asp Ala
1925                1930                1935                1940

GCC GCC CGC GTC ACT GCC ATA CTC AGC AGC CTC ACT GTA ACC CAG CTC      6209
Ala Ala Arg Val Thr Ala Ile Leu Ser Ser Leu Thr Val Thr Gln Leu
                1945                1950                1955

CTG AGG CGA CTG CAT CAG TGG ATA AGC TCG GAG TGT ACC ACT CCA TGC      6257
Leu Arg Arg Leu His Gln Trp Ile Ser Ser Glu Cys Thr Thr Pro Cys
            1960                1965                1970

TCC GGT TCC TGG CTA AGG GAC ATC TGG GAC TGG ATA TGC GAG GTG CTG      6305
Ser Gly Ser Trp Leu Arg Asp Ile Trp Asp Trp Ile Cys Glu Val Leu
        1975                1980                1985

AGC GAC TTT AAG ACC TGG CTG AAA GCC AAG CTC ATG CCA CAA CTG CCT      6353
Ser Asp Phe Lys Thr Trp Leu Lys Ala Lys Leu Met Pro Gln Leu Pro
    1990                1995                2000

GGG ATT CCC TTT GTG TCC TGC CAG CGC GGG TAT AGG GGG GTC TGG CGA      6401
Gly Ile Pro Phe Val Ser Cys Gln Arg Gly Tyr Arg Gly Val Trp Arg
2005                2010                2015                2020
```

```
GGA GAC GGC ATT ATG CAC ACT CGC TGC CAC TGT GGA GCT GAG ATC ACT      6449
Gly Asp Gly Ile Met His Thr Arg Cys His Cys Gly Ala Glu Ile Thr
                2025                2030                2035

GGA CAT GTC AAA AAC GGG ACG ATG AGG ATC GTC GGT CCT AGG ACC TGC      6497
Gly His Val Lys Asn Gly Thr Met Arg Ile Val Gly Pro Arg Thr Cys
            2040                2045                2050

AAG AAC ATG TGG AGT GGG ACG TTC TTC ATT AAT GCC TAC ACC ACG GGC      6545
Lys Asn Met Trp Ser Gly Thr Phe Phe Ile Asn Ala Tyr Thr Thr Gly
        2055                2060                2065

CCC TGT ACT CCC CTT CCT GCG CCG AAC TAT AAG TTC GCG CTG TGG AGG      6593
Pro Cys Thr Pro Leu Pro Ala Pro Asn Tyr Lys Phe Ala Leu Trp Arg
    2070                2075                2080

GTG TCT GCA GAG GAA TAC GTG GAG ATA AGG CGG GTG GGG GAC TTC CAC      6641
Val Ser Ala Glu Glu Tyr Val Glu Ile Arg Arg Val Gly Asp Phe His
2085                2090                2095                2100

TAC GTA TCG GGC ATG ACT ACT GAC AAT CTC AAA TGC CCG TGC CAG ATC      6689
Tyr Val Ser Gly Met Thr Thr Asp Asn Leu Lys Cys Pro Cys Gln Ile
                2105                2110                2115

CCA TCG CCC GAA TTT TTC ACA GAA TTG GAC GGG GTG CGC CTA CAT AGG      6737
Pro Ser Pro Glu Phe Phe Thr Glu Leu Asp Gly Val Arg Leu His Arg
            2120                2125                2130

TTT GCG CCC CCT TGC AAG CCC TTG CTG CGG GAG GAG GTA TCT TTC AGA      6785
Phe Ala Pro Pro Cys Lys Pro Leu Leu Arg Glu Glu Val Ser Phe Arg
        2135                2140                2145

GTA GGA CTC CAC GAG TAC CCG GTG GGG TCG CAA TTA CCT TGC GAG CCC      6833
Val Gly Leu His Glu Tyr Pro Val Gly Ser Gln Leu Pro Cys Glu Pro
    2150                2155                2160

GAA CCG GAC GTA GCC GTG TTG ACG TCC ATG CTC ACT GAT CCC TCC CAT      6881
Glu Pro Asp Val Ala Val Leu Thr Ser Met Leu Thr Asp Pro Ser His
2165                2170                2175                2180

ATA ACA GCA GAG GCG GCC GGG AGA AGG TTG GCG AGA GGG TCA CCC CCT      6929
Ile Thr Ala Glu Ala Ala Gly Arg Arg Leu Ala Arg Gly Ser Pro Pro
                2185                2190                2195

TCT ATG GCC AGC TCC TCG GCT AGC CAG CTG TCC GCT CCA TCT CTC AAG      6977
Ser Met Ala Ser Ser Ser Ala Ser Gln Leu Ser Ala Pro Ser Leu Lys
            2200                2205                2210

GCA ACT TGC ACC GCC AAC CAT GAC TCC CCT GAC GCC GAG CTC ATA GAG      7025
Ala Thr Cys Thr Ala Asn His Asp Ser Pro Asp Ala Glu Leu Ile Glu
        2215                2220                2225
```

```
GCT AAC CTC CTG TGG AGG CAG GAG ATG GGC GGC AAC ATC ACC AGG GTT      7073
Ala Asn Leu Leu Trp Arg Gln Glu Met Gly Gly Asn Ile Thr Arg Val
    2230                2235                2240

GAG TCA GAG AAC AAA GTG GTG ATT CTG GAC TCC TTC GAT CCG CTT GTG      7121
Glu Ser Glu Asn Lys Val Val Ile Leu Asp Ser Phe Asp Pro Leu Val
2245                2250                2255                2260

GCA GAG GAG GAT GAG CGG GAG GTC TCC GTA CCC GCA GAA ATT CTG CGG      7169
Ala Glu Glu Asp Glu Arg Glu Val Ser Val Pro Ala Glu Ile Leu Arg
                2265                2270                2275

AAG TCT CGG AGA TTC GCC CCA GCC CTG CCC GTC TGG GCG CGG CCG GAC      7217
Lys Ser Arg Arg Phe Ala Pro Ala Leu Pro Val Trp Ala Arg Pro Asp
            2280                2285                2290

TAC AAC CCC CTG CTA GTA GAG ACG TGG AAA AAG CCT GAC TAC GAA CCA      7265
Tyr Asn Pro Leu Leu Val Glu Thr Trp Lys Lys Pro Asp Tyr Glu Pro
        2295                2300                2305

CCT GTG GTC CAT GGC TGC CCG CTA CCA CCT CCA CGG TCC CCT CCT GTG      7313
Pro Val Val His Gly Cys Pro Leu Pro Pro Pro Arg Ser Pro Pro Val
    2310                2315                2320

CCT CCG CCT CGG AAA AAG CGT ACG GTG GTC CTC ACC GAA TCA ACC CTA      7361
Pro Pro Pro Arg Lys Lys Arg Thr Val Val Leu Thr Glu Ser Thr Leu
2325                2330                2335                2340

CCT ACT GCC TTG GCC GAG CTT GCC ACC AAA AGT TTT GGC AGC TCC TCA      7409
Pro Thr Ala Leu Ala Glu Leu Ala Thr Lys Ser Phe Gly Ser Ser Ser
                2345                2350                2355

ACT TCC GGC ATT ACG GGC GAC AAT ACG ACA ACA TCC TCT GAG CCC GCC      7457
Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro Ala
            2360                2365                2370

CCT TCT GGC TGC CCC CCC GAC TCC GAC GTT GAG TCC TAT TCT TCC ATG      7505
Pro Ser Gly Cys Pro Pro Asp Ser Asp Val Glu Ser Tyr Ser Ser Met
        2375                2380                2385

CCC CCC CTG GAG GGG GAG CCT GGG GAT CCG GAT CTC AGC GAC GGG TCA      7553
Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser
    2390                2395                2400

TGG TCG ACG GTC AGT AGT GGG GCC GAC ACG GAA GAT GTC GTG TGC TGC      7601
Trp Ser Thr Val Ser Ser Gly Ala Asp Thr Glu Asp Val Val Cys Cys
2405                2410                2415                2420

TCA ATG TCT TAT TCC TGG ACA GGC GCA CTC GTC ACC CCG TGC GCT GCG      7649
Ser Met Ser Tyr Ser Trp Thr Gly Ala Leu Val Thr Pro Cys Ala Ala
                2425                2430                2435
```

```
GAG GAA CAA AAA CTG CCC ATC AAC GCA CTG AGC AAC TCG TTG CTA CGC      7697
Glu Glu Gln Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg
            2440                2445                2450

CAT CAC AAT CTG GTG TAT TCC ACC ACT TCA CGC AGT GCT TGC CAA AGG      7745
His His Asn Leu Val Tyr Ser Thr Thr Ser Arg Ser Ala Cys Gln Arg
        2455                2460                2465

AAG AAG AAA GTC ACA TTT GAC AGA CTG CAA GTT CTG GAC AGC CAT TAC      7793
Lys Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Ser His Tyr
    2470                2475                2480

CAG GAC GTG CTC AAG GAG GTC AAA GCA GCG GCG TCA AAA GTG AAG GCT      7841
Gln Asp Val Leu Lys Glu Val Lys Ala Ala Ala Ser Lys Val Lys Ala
2485                2490                2495                2500

AAC TTG CTA TCC GTA GAG GAA GCT TGC AGC CTG GCG CCC CCA CAT TCA      7889
Asn Leu Leu Ser Val Glu Glu Ala Cys Ser Leu Ala Pro Pro His Ser
                2505                2510                2515

GCC AAA TCC AAG TTT GGC TAT GGG GCA AAA GAC GTC CGT TGC CAT GCC      7937
Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Cys His Ala
            2520                2525                2530

AGA AAG GCC GTA GCC CAC ATC AAC TCC GTG TGG AAA GAC CTT CTG GAA      7985
Arg Lys Ala Val Ala His Ile Asn Ser Val Trp Lys Asp Leu Leu Glu
        2535                2540                2545

GAC AGT GTA ACA CCA ATA GAC ACT ACC ATC ATG GCC AAG AAC GAG GTT      8033
Asp Ser Val Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val
    2550                2555                2560

TTC TGC GTT CAG CCT GAG AAG GGG GGT CGT AAG CCA GCT CGT CTC ATC      8081
Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile
2565                2570                2575                2580

GTG TTC CCC GAC CTG GGC GTG CGC GTG TGC GAG AAG ATG GCC CTG TAC      8129
Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr
                2585                2590                2595

GAC GTG GTT AGC AAG CTC CCC TTG GCC GTG ATG GGA AGC TCC TAC GGA      8177
Asp Val Val Ser Lys Leu Pro Leu Ala Val Met Gly Ser Ser Tyr Gly
            2600                2605                2610

TTC CAA TAC TCA CCA GGA CAG CGG GTT GAA TTC CTC GTG CAA GCG TGG      8225
Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Gln Ala Trp
        2615                2620                2625

AAG TCC AAG AAG ACC CCG ATG GGG CTC TCG TAT GAT ACC CGC TGT TTT      8273
Lys Ser Lys Lys Thr Pro Met Gly Leu Ser Tyr Asp Thr Arg Cys Phe
    2630                2635                2640
```

```
GAC TCC ACA GTC ACT GAG AGC GAC ATC CGT ACG GAG GAG GCA ATT TAC      8321
Asp Ser Thr Val Thr Glu Ser Asp Ile Arg Thr Glu Glu Ala Ile Tyr
2645                 2650                 2655                 2660

CAA TGT TGT GAC CTG GAC CCC CAA GCC CGC GTG GCC ATC AAG TCC CTC      8369
Gln Cys Cys Asp Leu Asp Pro Gln Ala Arg Val Ala Ile Lys Ser Leu
                2665                 2670                 2675

ACT GAG AGG CTT TAT GTT GGG GGC CCT CTT ACT AAT TCA AGG GGG GAA      8417
Thr Glu Arg Leu Tyr Val Gly Gly Pro Leu Thr Asn Ser Arg Gly Glu
           2680                 2685                 2690

AAC TGC GGC TAC CGC AGG TGC CGC GCG AGC AGA GTA CTG ACA ACT AGC      8465
Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Arg Val Leu Thr Thr Ser
      2695                 2700                 2705

TGT GGT AAC ACC CTC ACT CGC TAC ATC AAG GCC CGG GCA GCC TGT CGA      8513
Cys Gly Asn Thr Leu Thr Arg Tyr Ile Lys Ala Arg Ala Ala Cys Arg
 2710                 2715                 2720

GCC GCA GGG CTC CAG GAC TGC ACC ATG CTC GTG TGT GGC GAC GAC TTA      8561
Ala Ala Gly Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu
2725                 2730                 2735                 2740

GTC GTT ATC TGT GAA AGT GCG GGG GTC CAG GAG GAC GCG GCG AGC CTG      8609
Val Val Ile Cys Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu
                2745                 2750                 2755

AGA GCC TTC ACG GAG GCT ATG ACC AGG TAC TCC GCC CCC CCC GGG GAC      8657
Arg Ala Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp
           2760                 2765                 2770

CCC CCA CAA CCA GAA TAC GAC TTG GAG CTT ATA ACA TCA TGC TCC TCC      8705
Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser
      2775                 2780                 2785

AAC GTG TCA GTC GCC CAC GAC GGC GCT GGA AAG AGG GTC TAC TAC CTT      8753
Asn Val Ser Val Ala His Asp Gly Ala Gly Lys Arg Val Tyr Tyr Leu
 2790                 2795                 2800

ACC CGT GAC CCT ACA ACC CCC CTC GCG AGA GCC GCG TGG GAG ACA GCA      8801
Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala
2805                 2810                 2815                 2820

AGA CAC ACT CCA GTC AAT TCC TGG CTA GGC AAC ATA ATC ATG TTT GCC      8849
Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Phe Ala
                2825                 2830                 2835

CCC ACA CTG TGG GCG AGG ATG ATA CTG ATG ACC CAC TTC TTT AGC GTC      8897
Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Val
           2840                 2845                 2850
```

```
CTC ATA GCC AGG GAT CAG CTT GAA CAG GCT CTC AAC TGC GAG ATC TAC      8945
Leu Ile Ala Arg Asp Gln Leu Glu Gln Ala Leu Asn Cys Glu Ile Tyr
        2855                2860                2865

GGA GCC TGC TAC TCC ATA GAA CCA CTG GAT CTA CCT CCA ATC ATT CAA      8993
Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Pro Ile Ile Gln
    2870                2875                2880

AGA CTC CAT GGC CTC AGC GCA TTT TCA CTC CAC AGT TAC TCT CCA GGT      9041
Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly
2885                2890                2895                2900

GAA ATT AAT AGG GTG GCC GCA TGC CTC AGA AAA CTT GGG GTC CCG CCC      9089
Glu Ile Asn Arg Val Ala Ala Cys Leu Arg Lys Leu Gly Val Pro Pro
                2905                2910                2915

TTG CGA GCT TGG AGA CAC CGG GCC TGG AGC GTC CGC GCT AGG CTT CTG      9137
Leu Arg Ala Trp Arg His Arg Ala Trp Ser Val Arg Ala Arg Leu Leu
            2920                2925                2930

GCC AGA GGA GGC AAG GCT GCC ATA TGT GGC AAG TAC CTC TTC AAC TGG      9185
Ala Arg Gly Gly Lys Ala Ala Ile Cys Gly Lys Tyr Leu Phe Asn Trp
        2935                2940                2945

GCA GTA AGA ACA AAG CTC AAA CTC ACT CCG ATA ACG GCC GCT GGC CGG      9233
Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Thr Ala Ala Gly Arg
    2950                2955                2960

CTG GAC TTG TCC GGC TGG TTC ACG GCT GGC TAC AGC GGG GGA GAC ATT      9281
Leu Asp Leu Ser Gly Trp Phe Thr Ala Gly Tyr Ser Gly Gly Asp Ile
2965                2970                2975                2980

TAT CAC AGC GTG TCT CAT GCC CGG CCC CGC TGG TTC TGG TTT TGC CTA      9329
Tyr His Ser Val Ser His Ala Arg Pro Arg Trp Phe Trp Phe Cys Leu
                2985                2990                2995

CTC CTG CTT GCT GCA GGG GTA GGC ATC TAC CTC CTC CCC AAC CGA          9374
Leu Leu Leu Ala Ala Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
            3000                3005                3010

TGAAGATTGG GCTAACCACT CCAGGCCAAT AGGCCATTCC CT                      9416

 ***
```

(2) INFORMACIÓN PARA ID de SEC Nº:47

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 21 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:47:

CAGCCCCCTG ATGGGGGA C 21

(2) INFORMACIÓN PARA ID de SEC Nº:48

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 19 pares de bases

(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:48:

ACTCGCAAGC ACCCTATCA 19

(2) INFORMACIÓN PARA ID de SEC Nº:49

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 19 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:49:

CTGTGAGGAA CTACTGTCT 19

(2) INFORMACIÓN PARA ID de SEC Nº:50

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 22 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:50:

ATGAGCACGA ATCCTCAAAC CT 22

(2) INFORMACIÓN PARA ID de SEC Nº:51

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 16 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:51:

GTCCTGCCCT CGGGCC 16

(2) INFORMACIÓN PARA ID de SEC Nº:52

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:52:

CGAGGAAGAC TTCCGAGC 18

(2) INFORMACIÓN PARA ID de SEC Nº:53

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:53:

ACCCAAATTG CGCGACCTAC 20

(2) INFORMACIÓN PARA ID de SEC Nº:54

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:54:

TAAGGTCATC GATACCCT 18

(2) INFORMACIÓN PARA ID de SEC Nº:55

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:55:

CAGTTCATCA TCATATCCCA 20

(2) INFORMACIÓN PARA ID de SEC Nº:56

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:56:

GATAGAGAA AGAGCAAC 18

(2) INFORMACIÓN PARA ID de SEC Nº:57

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:57:

AGACTTCCGA GCGGTCGCAA 20

(2) INFORMACIÓN PARA ID de SEC Nº:58

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:58:

GACCTGTGCG GGTCTGTC 18

(2) INFORMACIÓN PARA ID de SEC Nº:59

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:59:

GGGTCGGCAG CTGGCTAGCC TCTCA 25

(2) INFORMACIÓN PARA ID de SEC Nº:60

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 18 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:60:

TCCTGGCGGG CATAGCGT 18

(2) INFORMACIÓN PARA ID de SEC Nº:61

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:61:

CCCCAGCCCT GGTCAAAATC GGTAA 25

(2) INFORMACIÓN PARA ID de SEC Nº:62

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 25 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:62:

(2) INFORMACIÓN PARA ID de SEC Nº:63

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 19 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:63:

CTCGGTCG TTCCCACCA 19

(2) INFORMACIÓN PARA ID de SEC Nº:64

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 20 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:64:

CCGCGAAGAG TGTGTGTGGT 20

(2) INFORMACIÓN PARA ID de SEC Nº:65

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 19 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:65:

CAATGTTCTG GTGGAGGTG 19

(2) INFORMACIÓN PARA ID de SEC Nº:66

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 28 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:66:

GCCATTAAGT GGGAGTACGT CGTTCTCC 28

(2) INFORMACIÓN PARA ID de SEC Nº:67

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 19 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:67:

CGAGGAAGGA TACAAGACC 19

(2) INFORMACIÓN PARA ID de SEC Nº:68

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 20 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:68:

TGCTTGTGGA TGATGCTACT 20

(2) INFORMACIÓN PARA ID de SEC Nº:69

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 18 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:69:

CACACGTGCA GTTGCGCT 18

(2) INFORMACIÓN PARA ID de SEC Nº:70

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 18 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:70:

CTGCTGACCA CTACACAG 18

(2) INFORMACIÓN PARA ID de SEC Nº:71

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 19 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:71:

GACCAGAGTG GAAGCGCAA 19

(2) INFORMACIÓN PARA ID de SEC Nº:72

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:72:

TACCAGAGTC GGGTGTACAG 20

(2) INFORMACIÓN PARA ID de SEC Nº:73

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:73:

CTAGGAGGCC CCTTGTCTGC 20

(2) INFORMACIÓN PARA ID de SEC Nº:74

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:74:

CTCGGGCCAG CCGATGGA 18

(2) INFORMACIÓN PARA ID de SEC Nº:75

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 19 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:75:

GGGGACCTCA TGGTTGTCT 19

(2) INFORMACIÓN PARA ID de SEC Nº:76

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 18 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:76:

CCCGTGGAGT GGCTAAGG 18

(2) INFORMACIÓN PARA ID de SEC Nº:77

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 19 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:77:

CTCCTCGATG TTGGGATGG 19

(2) INFORMACIÓN PARA ID de SEC Nº:78

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 19 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:78:

CAGAGCTTCC AGGTGGCTC 19

(2) INFORMACIÓN PARA ID de SEC Nº:79

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 17 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:79:

CGGGCTCCGT CACTGTG 17

(2) INFORMACIÓN PARA ID de SEC Nº:80

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 21 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:80:

GTATTGCAGT CTATCACCGA G 21

(2) INFORMACIÓN PARA ID de SEC Nº:81

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:81:

GGCTATACCG GCGACTTCGA 20

(2) INFORMACIÓN PARA ID de SEC Nº:82

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 19 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:82:

CGTTGAGTGC GGGAGACAG 19

(2) INFORMACIÓN PARA ID de SEC Nº:83

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:83:

TCACCATTGA GACAATCACG 20

(2) INFORMACIÓN PARA ID de SEC Nº:84

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 22 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: si

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:84:

GTAAGGAAGG TTCTCCCCAC TC 22

(2) INFORMACIÓN PARA ID de SEC Nº:85

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 27 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:85:

ATGCCCACTT TCTATCCCAG ACAAAGC 27

(2) INFORMACIÓN PARA ID de SEC Nº:86

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 17 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: si

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:86:

TGCATGTCAT GATGTAT 17

(2) INFORMACIÓN PARA ID de SEC Nº:87

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 18 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: si

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:87:

GGACAAGACG ACCCTGCC 18

(2) INFORMACIÓN PARA ID de SEC Nº:88

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 20 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:88:

CGTATTGCCT GTCAACAGGC 20

(2) INFORMACIÓN PARA ID de SEC Nº:89

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:89:

AGCGCCCACA AAGGCAGTAG 20

(2) INFORMACIÓN PARA ID de SEC Nº:90

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 19 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:90:

[illegible]TCTTCAAC ATATTGGGG 19
***

(2) INFORMACIÓN PARA ID de SEC Nº:91

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:91:

CCAGGAACCG GAGCATGG 18

(2) INFORMACIÓN PARA ID de SEC Nº:92

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 18 pares de bases

(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:92:

ACCAGTGGAT AAGCTCGG 18

(2) INFORMACIÓN PARA ID de SEC Nº:93

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 19 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:93:

CGTGGTGTAG GCATTAATG 19

(2) INFORMACIÓN PARA ID de SEC Nº:94

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 19 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:94:

ATGTGGAGTG GGACCTTCC 19

(2) INFORMACIÓN PARA ID de SEC Nº:95

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 20 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:95:

CTCTGCTGTT ATATGGGAGG 20

(2) INFORMACIÓN PARA ID de SEC Nº:96

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:96:

GTTGACGTCC ATGCTCACTG 20

(2) INFORMACIÓN PARA ID de SEC Nº:97

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:97:

TTTCCACGTC TCCACTAGCG 20

(2) INFORMACIÓN PARA ID de SEC Nº:98

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:98:

GTGAGGACCA CCGTCCGC 18

(2) INFORMACIÓN PARA ID de SEC Nº:99

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:99:

TTCCACCTCC AAAGTCCCCT 20

(2) INFORMACIÓN PARA ID de SEC Nº:100

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 25 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:100:

AGAACTTGCA GTCTGTCAAA TGTGA 25

(2) INFORMACIÓN PARA ID de SEC Nº:101

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 34 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:101:

GGAAGAACAG AAACTGCCCA TCAATGCACT AAGC 34

(2) INFORMACIÓN PARA ID de SEC Nº:102

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:102:

TGACGCCGCT GCTTTAACCT 20

(2) INFORMACIÓN PARA ID de SEC Nº:103

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 20 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:103:

TGCAAGCTTC CTCTACGGAT 20

(2) INFORMACIÓN PARA ID de SEC Nº:104

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 20 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:104:

AGGTTAAAGC AGCGGCGGTCA 20

(2) INFORMACIÓN PARA ID de SEC Nº:105

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 19 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:105:

AGCTTCCCAT CACGGCCAA 19

(2) INFORMACIÓN PARA ID de SEC Nº:106

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 19 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:106:

GATGGCTTTG TACGCGTG 19

(2) INFORMACIÓN PARA ID de SEC Nº:107

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:107:

GCACCTGCGA TAGCCGCAGT 20

(2) INFORMACIÓN PARA ID de SEC Nº:108

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 18 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:108:

GTCCCTCACC GAGAGGCT 18

(2) INFORMACIÓN PARA ID de SEC Nº:109

(i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 20 pares de bases

(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:109:

GATTGGAGGT AGATCAAGTG 20

(2) INFORMACIÓN PARA ID de SEC Nº:110

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 20 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:110:

TACGACTTGG AGCTCATAAC 20

(2) INFORMACIÓN PARA ID de SEC Nº:111

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 20 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: NO

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:111:

AGCAAGACAC ACTCCAGTCA 20

(2) INFORMACIÓN PARA ID de SEC Nº:112

(i) CARACTERÍSTICAS DE LA SECUENCIA:
(A) LONGITUD: 24 pares de bases
(B) TIPO: ácido nucleico
(C) TIPO DE CADENA: sencilla
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HIPOTÉTICO: NO

(iv) ANTISENTIDO: SI

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:112:

GCCTATTGGC CTGGAGTGGT TAGC 24

(2) INFORMACIÓN PARA ID de SEC Nº:113

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 26 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:113:

His Val Thr Gly Gly Asn Ala Gly Arg Thr Thr Ala Gly Leu Val Gly
1 5 10 15
Leu Leu Thr Pro Gly Ala Lys Gln Asn Ile
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:114

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 26 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:114:

His Val Thr Gly Gly Ser Ala Gly His Thr Val Ser Gly Phe Val Ser
1 5 10 15
Leu Leu Ala Pro Gly Ala Lys Gln Asn Val
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:115

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 26 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:115:

His Val Thr Gly Gly Gln Ala Ala Arg Ala Met Ser Gly Leu Val Gly
1 5 10 15
Leu Phe Thr Pro Gly Ala Lys Gln Asn Ile
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:116

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 26 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:116:

His Val Thr Gly Gly Arg Val Ala Ser Ser Thr Gln Ser Leu Val Ser
1 5 10 15
Trp Leu Ser Gln Gly Pro Ser Gln Lys Ile
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:117

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 26 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:117:

His Val Thr Gly Gly Ala Gln Ala Lys Thr Thr Asn Arg Leu Val Ser
1 5 10 15
Met Phe Ala Ser Gly Pro Ser Gln Lys Ile
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:118

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 26 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:118:

Tyr Thr Ser Gly Gly Ala Ala Ser His Thr Thr Ser Thr Leu Ala Ser
1 5 10 15
Leu Phe Ser Pro Gly Ala Ser Arg Asn Ile
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:119

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 26 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:119:

His Val Thr Gly Gly Val Gln Gly His Val Thr Ser Thr Leu Thr Ser
1 5 10 15
Leu Phe Arg Pro Gly Ala Ser Gln Lys Ile
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:120

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 26 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:120:

His Val Thr Gly Gly Ser Ala Gly Arg Thr Thr Ala Gly Leu Val Gly
1 5 10 15
Leu Leu Thr Pro Gly Ala Lys Gln Asn nIle
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:121

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 26 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:121:

His Val Thr Gly Gly Ser Ala Gly Arg Ser Val Leu Gly Ile Ala Ser
1 5 10 15
Phe Leu Thr Arg Gly Pro Lys Gln Asn Ile
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:122

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 30 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:122:

Val Ala Thr Arg Asp Gly Lys Leu Pro Thr Thr Gln Leu Arg Arg His
1 5 10 15
Ile Asp Leu Leu Val Gly Ser Ala Thr Leu Cys Ser Ala Leu
20 25 30

(2) INFORMACIÓN PARA ID de SEC Nº:123

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 30 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:123:

Val Ala Thr Arg Asp Gly Lys Leu Pro Ala Thr Gln Leu Arg Arg His
1 5 10 15
Ile Asp Leu Leu Val Gly Ser Ala Thr Leu Cys Ser Ala Leu
20 25 30

(2) INFORMACIÓN PARA ID de SEC Nº:124

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 30 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:124:

Leu Ala Ala Arg Asn Ser Ser Ile Pro Thr Thr Thr Ile Arg Arg His
1 5 10 15
Val Asp Leu Leu Val Gly Ala Ala Ala Leu Cys Ser Ala Met
20 25 30

(2) INFORMACIÓN PARA ID de SEC Nº:125

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 30 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:125:

Leu Ala Ala Arg Asn Val Thr Ile Pro Thr Thr Thr Ile Arg Arg His
1 5 10 15
Val Asp Leu Leu Val Gly Ala Ala Ala Phe Cys Ser Ala Met
20 25 30

(2) INFORMACIÓN PARA ID de SEC Nº:126

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 30 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:126:

Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr Ile Arg Arg His
1 5 10 15
Val Asp Leu Leu Val Gly Ala Ala Ala Phe Cys Ser Ala Met
20 25 30

(2) INFORMACIÓN PARA ID de SEC Nº:127

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 30 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:127:

Leu Ala Ala Arg Asn Ala Ser Val Pro Thr Thr Thr Leu Arg Arg His
1 5 10 15
Val Asp Leu Leu Val Gly Thr Ala Ala Phe Cys Ser Ala Met
20 25 30

(2) INFORMACIÓN PARA ID de SEC Nº:128

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 27 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:128:

Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Ala Gln Gly Trp Gly Pro
1 5 10 15
Ile Ser Tyr Ala Asn Gly Ser Gly Leu Asp Glu
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:129

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 27 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:129:

Leu Ala Ser Cys Arg Pro Leu Thr Asp Phe Asp Gln Gly Trp Gly Pro
1 5 10 15
Ile Ser Tyr Ala Asn Gly Ser Gly Pro Asp Gln
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:130

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 27 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:130:

Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Asp Gln Gly Trp Gly Pro
1 5 10 15
Ile Ser His Ala Asn Gly Ser Gly Pro Asp Gln
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:131

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 27 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:131:

Met Ala Ser Cys Arg Pro Ile Asp Glu Phe Ala Gln Gly Trp Gly Pro
1 5 10 15
Ile Thr His Asp Met Pro Glu Ser Ser Asp Gln
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:132

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 27 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:132:

Met Ala Gln Cys Arg Thr Ile Asp Lys Phe Ala Gln Gly Trp Gly Pro
1 5 10 15
Ile Thr Tyr Ala Glu Ser Ser Arg Ser Asp Gln
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:133

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 27 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:133:

Met Ala Ser Cys Arg Pro Ile Gln Trp Phe Ala Gln Gly Trp Gly Pro
1 5 10 15
Ile Thr Tyr Thr Glu Pro Asp Ser Pro Asp Gln
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:134

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 27 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:134:

Leu Ala Ser Cys Arg Arg Leu Thr Asp Phe Asp Gln Gly Trp Gly Pro
1 5 10 15
Ile Ser Tyr Ala Asn Gly Ser Gly Pro Asp Glu
20 25

(2) INFORMACIÓN PARA ID de SEC Nº:135

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 24 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:135:

SEr Thr Ser Gly Ile Thr Gly Asp Asn Thr Thr Thr Ser Ser Glu Pro
1 5 10 15
Ala Pro Ser Gly Cys Pro Pro Asp
20

(2) INFORMACIÓN PARA ID de SEC Nº:136

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 24 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:136:

Gly Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Gly Pro Pro Asp Gln
1 5 10 15
Ala Ser Asp Asp Gly Asp Lys Gly
20

(2) INFORMACIÓN PARA ID de SEC Nº:137

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 24 aminoácidos

(B) TIPO: aminoácido

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(xi) DESCRIPCIÓN DE LA SECUENCIA: ID de SEC Nº:137:

Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Leu Pro Asp Gln
1 5 10 15
Ala Ser Asp Asp Gly Asp Lys Gly
20