

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 387 321**

51 Int. Cl.:
C12N 7/00 (2006.01)
C07K 14/18 (2006.01)
A61K 39/29 (2006.01)
C12N 7/04 (2006.01)
A61K 48/00 (2006.01)
C12N 5/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **07723067 .0**
96 Fecha de presentación: **06.03.2007**
97 Número de publicación de la solicitud: **1993603**
97 Fecha de publicación de la solicitud: **26.11.2008**

54 Título: **Proteína de fusión no estructural del virus de la hepatitis C**

30 Prioridad:
09.03.2006 EP 06360006

45 Fecha de publicación de la mención BOPI:
20.09.2012

45 Fecha de la publicación del folleto de la patente:
20.09.2012

73 Titular/es:
**TRANSGENE SA
PARC D'INNOVATION BOULEVARD GONTHIER
D'ANDERNACH
67400 ILLKIRCH GRAFFENSTADEN, FR**

72 Inventor/es:
**FOURNILLIER, Anne;
INCHAUSPE, Geneviève;
CHATEL, Laurence y
PENIN, François**

74 Agente/Representante:
Curell Aguilá, Mireia

ES 2 387 321 T3

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

DESCRIPCIÓN

Proteína de fusión no estructural del virus de la hepatitis C.

5 El VHC pertenece a la familia *Flaviviridae*, y es la causa principal de hepatopatías crónicas y hepatitis aguda. La infección vírica por el VHC está asociada con una tasa elevada (54-86%) de cronicidad, la cual, en el 5 a 24% de los casos, evoluciona a cirrosis y subsiguientemente a carcinoma hepatocelular durante un período de 20 a 30 años (McHutchinson, 2004, *The American Journal of Managed Care* 10, S21-29). Actualmente, en Europa y en los Estados Unidos de América, el 20-30% de los trasplantes hepáticos y el 15-33% de los cánceres hepáticos son atribuibles a infección por el VHC. La Organización Mundial de la Salud estimó en 1999 que alrededor e 170 millones de personas están crónicamente infectadas a nivel mundial por el VHC, y cada año se infectan 3 a 4 millones de personas.

15 El VHC es un virus de cubierta con un genoma de ácido ribonucleico (ARN) monocatenario positivo de aproximadamente 9.600 nucleótidos, cuya organización es habitual para todas las cepas y aislados del VHC. La traducción del genoma del VHC en la fase 0 genera una gran poliproteína de 3010 a 3030 aminoácidos según el genotipo, que es escindida proteolíticamente por proteasas víricas y celulares para producir 10 proteínas víricas. La tercera parte aminoterminal de la poliproteína codifica las proteínas estructurales del virión: la proteína Core (C), y las glucoproteínas de cubierta E1 y E2. Después de la región estructural viene una pequeña proteína de membrana integral, p7, que parece funcionar como un canal de iones. El resto del genoma codifica las proteínas no estructurales (NS) NS2, NS3, NS4A, NS4B, NS5A y NS5B, que median los procesos intracelulares del ciclo vital del virus (Penin, 2004, *Hepatology* 39, 5-19). El VHC también codifica proteínas pequeñas, denominadas F (desplazamiento del marco) o ARFP (proteína del marco de lectura alternativo), que se pueden producir mediante desplazamiento del marco ribosómico o iniciación interna en un marco de lectura +1 alternativo dentro del gen core (véanse Branch *et al.*, 2005, *Semin. Liver Dis.* 25:105-117; y el documento WO2004/069864). Las secuencias que codifican la poliproteína están flanqueadas en ambos extremos por dos regiones no traducidas muy conservadas (UTR), 5'UTR y 3'UTR, respectivamente. Un sitio de entrada al ribosoma interno, situado en 5'UTR, permite la fijación al ribosoma para la iniciación de la traducción, mientras que se piensa que la 3'UTR desempeña un papel importante en la iniciación de la replicación vírica.

30 La terapia estándar actual para pacientes infectados crónicamente con VHC es una combinación de interferón alfa pegilado (PEG-IFN- α) y ribavirina (Fried *et al.*, 2002, *N. Engl. J. Med.* 347, 975-982). Sin embargo, esta terapia es costosa, está asociada con efectos secundarios significativos que conducen a la terminación prematura del tratamiento en 10% de los casos, y es inadecuada para un número significativo de pacientes (por ejemplo aquellos con cirrosis hepática descompensada, enfermedades autoinmunitarias, antecedentes de depresión y embarazo). De forma más importante, sólo el 50% de los pacientes tratados responden al tratamiento (Falck-Ytter, *et al.*, 2002, *Ann. Intern. Med.* 136, 288-292), y la tasa de respuesta es todavía más baja para pacientes infectados con el genotipo 1 (27%-35%).

40 En conjunto, se acumula un número de pruebas experimentales para enfatizar el papel crítico desarrollado por la inmunidad de células T en el control de la infección por VHC, particularmente respuestas mediadas por células T CD4+ y CD8+ dirigidas a las proteínas no estructurales (NS) (por ejemplo Francavilla *et al.*, 2004, *Eur. J. Immunol.* 34, 427-37; Grakoui *et al.*, 2003, *Science* 302, 659-662; Shoukry *et al.*, 2003, *J. Exp. Med.* 197,1645-1655; Thimme *et al.*, 2001, *J. Exp. Med.* 194,1395-1406; Lechner *et al.*, 2000, *J. Exp. Med.* 191,1499-1512; Gerlach, 1999, *Gastroenterology* 117, 933-941; Folgori *et al.*, 2006, *Nat Med.* 12, 190-197). De hecho, se observan respuestas vigorosas de células T específicas del VHC en pacientes que se recuperan espontáneamente de hepatitis C aguda autolimitada, mientras que las respuestas de las células T son débiles y están centradas de forma muy particular en pacientes infectados persistentemente. Por otro lado, el papel desempeñado por los anticuerpos neutralizantes (inmunidad mediada por células B) es mucho menos claro (Logvinoff, 2004, *Proc. Natl. Acad. Sci. USA* 101, 10149-54).

55 Desde el descubrimiento del virus del VHC hace 15 años, todavía no hay ninguna vacuna contra el VHC debido a la dificultad de lograr un crecimiento eficiente del virus en cultivo celular y la falta de modelos de laboratorio fácilmente disponibles de infección vírica. Sin embargo, ahora ha aparecido un gran número de vacunas candidatas (Barth y Baumert, 2004, *Novel vaccine strategies*, p. 214-227, In *State of the Art of Hepatology: Molecular and Cell Biology* eds Kluwer Academic Publishers; Inchauspe and Feinstone, 2003, *Clin. Liver Dis.* 7, 243-259), que se basan, por ejemplo, en péptidos derivados de VHC (Cerny *et al.*, 1995, *J. Clin. Invest.* 95, 521-30), antígenos de VHC producidos recombinantemente, vacunas a base de ADN y de virus (Brinster *et al.*, 2001, *Hepatology* 34,1206-1217; Foms *et al.*, 2000, *Hepatology* 32, 618-625) y partículas semejantes al virus de VHC (Baumert *et al.*, 1999, *Gastroenterology* 117,1397-1407; Jeong *et al.*, 2004, *J. Virol.* 78,6995-7003). Actualmente se evalúan varios candidatos en modelos preclínicos, y la primera ola está entrando en ensayos clínicos. Las vacunas más avanzadas actualmente en fase II usan proteína de cubierta E1 con adyuvantes (Innogenetics) y epitopos de células CD4+ y CD8+ con adyuvantes (Intercell).

65 Un enfoque que se ha explorado en forma activa durante esta última década se basa en el uso de poliproteínas que asocian diversos dominios antigénicos como inmunógenos. Una amplia mayoría de las poliproteínas de la técnica

anterior resultan de la fusión a otro polipéptido antigénico, fragmentos o epítomos que se originan a partir de diversas proteínas NS de VHC, a fin de formar un único polipéptido esperanzadoramente inmunógeno. Por ejemplo, los documentos WO01/30812 y WO03/031588 describen la fusión de polipéptidos de VHC de NS3 a NS5B para formar un único polipéptido que engloba las proteínas NS principales. El documento WO03/097677 describe la fusión de fragmentos antigénicos de 30 a 70 aminoácidos que se originan a partir de polipéptidos NS3, NS4B y NS5B. El documento WO2004/0461 describe una poliproteína que comprende Core, NS3, NS4B y NS5B. El documento WO2004/1110 describe vectores adenovíricos y de MVA para coexpresar NS3, NS4 y NS5B, en el que NS3 y NS4 son expresadas como una proteína de fusión, y NS5B es expresada independientemente.

Se ha de señalar que la configuración de las poliproteínas de la técnica anterior es como en el contexto nativo, apareciendo los diversos componentes en el orden en el que aparecen de forma natural en el precursor de VHC nativo. En particular, NS3 es seguida de NS4, que es seguida de NS5. Sin embargo, la configuración "nativa" no es óptima en términos de expresión e inmunogenicidad de las poliproteínas resultantes.

Se puede esperar que el VHC continuará siendo una amenaza grave para la salud mundial durante muchos años, debido a la naturaleza crónica y persistente de infección, su elevada prevalencia y la morbilidad significativa de las enfermedades asociadas. De este modo, existe una necesidad importante de desarrollar polipéptidos más inmunógenos, vectores de expresión, métodos para la preparación de los mismos, y sus usos, para mejorar la prevención y tratamiento de infecciones por VHC o enfermedades o trastornos asociados con VHC.

La presente invención se refiere a proteínas de fusión que implican polipéptidos no estructurales (NS) dispuestos en una configuración no nativa según la reivindicación 1. Se seleccionaron fragmentos específicos para los polipéptidos MS de VHC (por ejemplo NS3, NS4A, NS5B y opcionalmente NS4B), y se configuraron específicamente a fin de optimizar la inmunogenicidad y/o el procedimiento de producción recombinante de las proteínas de fusión resultantes. En comparación con los polipéptidos NS nativos, las nuevas proteínas de fusión de VHC proporcionadas por la presente invención, o sus secuencias nucleotídicas codificantes, pueden permitir una mejora de una respuesta inmunitaria anti-VHC y/o una mejora de la respuesta citotóxica global con la introducción en un organismo hospedante, y/o una mejora de la producción de lotes clínicos. De este modo, la invención representa un avance significativo en el tratamiento y prevención actuales de infecciones por VHC o enfermedades o trastornos asociados con VHC. En particular, se puede usar para reforzar terapias existentes o proporcionar un tratamiento alternativo a pacientes infectados crónicamente, especialmente aquellos que no responden a la terapia convencional.

Este problema técnico se resuelve mediante la provisión de las realizaciones como se definen en las reivindicaciones.

Otros aspectos, características y ventajas adicionales de la invención serán manifiestos a partir de la descripción siguiente de las realizaciones actualmente preferidas de la invención. Estas realizaciones serán para los fines de la descripción.

En consecuencia, en un primer aspecto, la presente invención proporciona una proteína de fusión aislada que comprende al menos tres polipéptidos NS, entre ellos los polipéptidos NS3, NS4A y NS5B, que se originan a partir de un virus de la hepatitis C (VHC), en la que dichos polipéptidos NS están configurados en dicha proteína de fusión en un orden que es distinto del orden en el que aparecen en la configuración nativa, en la que el polipéptido NS4A está situado en el término N y NS5B está situado en el término C de dicha proteína de fusión.

Como se utiliza en la presente memoria a lo largo de toda la solicitud, los términos "un" y "una" se usan en el sentido de que significan "al menos un", "al menos un primer", "uno o más" o "una pluralidad" de los compuestos o etapas citados, excepto que el contexto dicte otra cosa. Por ejemplo, la expresión "una célula" incluye una pluralidad de células, incluyendo una mezcla de las mismas.

El término "y/o", siempre que se use en la presente memoria, incluye el significado de "y", "o" y "todos o cualquier otra combinación de los elementos conectados mediante dicha expresión".

El término "alrededor de" o "aproximadamente", como se utiliza en la presente memoria, significa en el 20%, preferentemente en el 10%, y más preferentemente en el 5% de un valor o intervalo dado.

Los términos "aminoácidos" y "restos" son sinónimos, y engloban aminoácidos naturales así como análogos de aminoácidos (por ejemplo aminoácidos no naturales, sintéticos y modificados, incluyendo los isómeros ópticos D o L).

Los términos "polipéptido", "péptido" y "proteína" se usan en la presente memoria de forma intercambiable para referirse a polímeros de restos de aminoácidos que comprenden diez o más aminoácidos enlazados vía enlaces peptídicos. El polímero puede ser lineal, ramificado o cíclico, y puede comprender aminoácidos de origen natural y/o análogos de aminoácidos, y puede estar interrumpido por no aminoácidos. Como indicación general, si el polímero de aminoácidos es largo (por ejemplo más de 50 restos de aminoácidos), se denomina preferentemente como polipéptido o proteína.

En el contexto de la presente invención, las expresiones “ácido nucleico”, “molécula de ácido nucleico”, “polinucleótido” y “secuencia nucleotídica” se usan de forma intercambiable y definen un polímero de cualquier longitud de polidesoxirribonucleótidos (ARN) (por ejemplo, ADNc, ADN genómico, plásmidos, vectores, genomas víricos, ADN aislados, sondas, cebadores, y cualquier mezcla de los mismos) C o polirribonucleótidos (ARN) (por ejemplo ARNm, ARN antisentido), o polirribo-polidesoxirribonucleótidos mixtos. Engloban polinucleótidos monocatenarios o bicatenarios, lineales o circulares, naturales o sintéticos. Además, un polinucleótido puede comprender nucleótidos de origen no natural, tales como nucleótidos metilados y análogos nucleotídicos (véanse los documentos US 5.525.711, US 4.711.955 o EPA 302 175 como ejemplos de modificaciones), y pueden estar interrumpidos por componentes no nucleotídicos. Si están presentes, las modificaciones al nucleótido se pueden impartir antes o después de la polimerización.

Como se utiliza en la presente memoria, cuando se use para definir productos, composiciones y métodos, la expresión “que comprende” quiere decir que los productos, composiciones y métodos incluyen los componentes o etapas citados, pero sin excluir otros. “Que consiste esencialmente en” debería significar que excluye otros componentes o etapas de cualquier significancia esencial. De este modo, una composición que consiste esencialmente en los componentes citados no excluiría contaminantes en trazas y vehículos farmacéuticamente aceptables. “Que consiste en” debe significar que excluye más de elementos en trazas de otros componentes o etapas. Por ejemplo, un polipéptido “consiste en” una secuencia de aminoácidos cuando el polipéptido no contiene ningún aminoácido sino la secuencia de aminoácidos citada. Un polipéptido “consiste esencialmente en” una secuencia de aminoácidos cuando tal secuencia de aminoácidos está presente junto con sólo unos pocos restos de aminoácidos adicionales, típicamente de alrededor de 1 a alrededor de 50 restos adicionales más o menos. Un polipéptido “comprende” una secuencia de aminoácidos cuando la secuencia de aminoácidos es al menos parte de la secuencia de aminoácidos final del polipéptido. Tal polipéptido puede tener unos pocos hasta varios cientos de restos de aminoácidos adicionales. Tales restos de aminoácidos adicionales pueden desempeñar un papel en el tráfico de polipéptidos, facilitar la producción o purificación de polipéptidos, prolongar la semivida, entre otros. Lo mismo se puede aplicar para las secuencias nucleotídicas.

Como se utiliza en la presente memoria, el término “aislado” se refiere a una proteína, polipéptido, péptido o ácido nucleico que está purificado o eliminado de su entorno natural. El término “purificado” se refiere a una proteína, polipéptido, péptido o ácido nucleico que se separa de al menos algún otro componente o componentes con los que está asociado naturalmente.

“VHC” significa “virus de la hepatitis C”. Los análisis filogenéticos amplios han conducido a la clasificación de aislados de VHC en 6 genotipos principales (1 a 6) que contienen diferentes subtipos (a, b, c, etc.) (Simmons *et al.*, 2005, *Hepatology* 42, 962-973). Los aislados de VHC ejemplares del genotipo 1a incluyen, sin limitación, HCV-1 (Choo *et al.*, 1991, *Proc. Natl. Acad. Sci. USA* 88, 2451-2455), -J I (Okamoto *et al.*, 1992, *Nucleic Acids Res.* 20, 6410-6410) y-H (Inchauspe *et al.*, 1991, *Proc. Natl. Acad. Sci.* 88, 10292-10296). Los aislados de VHC ejemplares de genotipo 1b incluyen, sin limitación, HCV-JA (Kato *et al.*, 1990, *Proc. Natl. Acad., Sci.* 87, 9524-9528) y BK (Takamizawa *et al.*, 1991, *J. Virol.* 65, 1105-1113). Los aislados de VHC ejemplares del genotipo 1c incluyen, sin limitación, HCV-G9 (Okamoto *et al.*, 1994, *J. Gen. Virol.* 45, 629-635). Los aislados de VHC ejemplares de genotipo 2a incluyen, sin limitación, HCV-J6 (Okamoto *et al.*, 1991, *J. Gen. Virol.* 72,2697-2704). Los aislados de VHC ejemplares del genotipo 2b incluyen, sin limitación, HCV-J8 (Okamoto *et al.*, 1992, *Virology* 188, 331-341). Los aislados de VHC ejemplares del genotipo 2c incluyen, sin limitación, HCV-BEBE1 (Nako *et al.*, 1996, *J. Gen. Virol.* 141, 701-704). Los aislados de VHC ejemplares del genotipo 3a incluyen, sin limitación, HCV-NZL1 (Sakamoto *et al.*, 1994, *J. Gen. Virol.* 75,1761- 1768). Los aislados de VHC ejemplares del genotipo 3b incluyen, sin limitación, HCV-Tr (Chayama *et al.*, 1994, *J. Gen. Virol.* 75, 3623-3628). Los aislados de VHC ejemplares del genotipo 4a incluyen, sin limitación, HCV-ED43 (Chamberlain *et al.*, 1997, *J. Gen. Virol.* 78,1341-1347). Los aislados de VHC ejemplares del genotipo 5a incluyen, sin limitación, HCV-EUH1480 (Chamberlain *et al.*, 1997, *Biochem. Biophys. Res. Commun.* 236, 44-49). Los aislados de VCH ejemplares del genotipo 6a incluyen, sin limitación, HCV-EUHK2 (Adams *et al.*, 1997, *Biochem. Biophys. Res. Commun.* 234, 393-396).

El término “fusión” o “proteína de fusión”, como se utiliza en la presente memoria, se refiere a la combinación con algún otro de al menos tres péptidos NS (o fragmento o fragmentos o variante o variantes de los mismos) en una cadena polipeptídica. Preferentemente, la fusión entre los diversos polipéptidos NS se lleva a cabo por medios genéticos, es decir, fusionando en el marco las secuencias nucleotídicas que codifican cada uno de los polipéptidos SN. Por “fusionado en el marco” se quiere decir que la expresión de las secuencias codificantes fusionadas da como resultado una única proteína sin ningún terminador transcripcional entre cada uno de los polipéptidos NS.

La fusión entre cada uno de los polipéptidos NS puede ser directa o a través de un ligador. Como se utiliza en la presente memoria, “directa” se refiere a una fusión entre dos polipéptidos NS sin ningún resto de aminoácido adicional entre medias (por ejemplo los codones que codifican un polipéptido NS son contiguos a los codones que codifican al siguiente polipéptido NS). Como alternativa, se puede usar un péptido ligador en unión de al menos dos polipéptidos NS. La presencia de un ligador puede facilitar la formación completa, el plegamiento y/o el funcionamiento de la proteína de fusión. La presente invención no está limitada por la forma, tamaño o número de secuencias ligadoras empleadas, y se pueden insertar en la unión entre dos polipéptidos NS múltiples copias de una

secuencia ligadora. Los ligadores adecuados según la invención tienen 3 a 30 aminoácidos de longitud, están compuestos de repeticiones de restos de aminoácidos tales como glicina, serina, treonina, asparagina, alanina y/o prolina (véase, por ejemplo, Wiederrecht *et al.*, 1988, Cell 54, 841; Aumailly *et al.*, 1990 FEBS Lett. 262, 82; y Dekker *et al.*, 1993, Nature 362, 852). Los ejemplos representativos de ligadores adecuados incluyen un ligador Ser-Gly-Ser para conectar el polipéptido NS del término N al segundo polipéptido NS, y un ligador Gly-Ser-Gly-Ser-Gly para conectar el segundo polipéptido NS al tercer polipéptido NS. Como alternativa, también se pueden usar secuencias derivadas del VHC para conectar entre sí dos polipéptidos NS (por ejemplo la porción C-terminal de un polipéptido NS4A nativo que se extiende desde aproximadamente la posición 1691 hasta aproximadamente la posición 1711, numeradas con relación a la poliproteína HCV-1 de longitud completa.

“Al menos tres” significa un número que es tres o mayor que tres, con preferencia especial por tres o cuatro.

Como se utiliza en la presente memoria, la expresión “polipéptido NS” se refiere a una proteína no estructural reconocida en la técnica, seleccionada preferentemente del grupo constituido por polipéptidos NS2, NS3, NS4A, NS4B y NS5B. No obstante, se prefiere que la proteína de fusión no contenga ningún polipéptido NS5A. En el contexto de la invención, los polipéptidos NS incluidos en la proteína de fusión de la invención se pueden originar independientemente a partir de cualquier cepa o aislado de VHC identificado actualmente, tal como los descritos anteriormente en relación con el término “VHC”. El término “originar” significa que se puede aislar, clonar, derivar o que está relacionado. De este modo, según la presente invención, cada uno de los polipéptidos NS incluidos en la proteína de fusión pueden originarse de un polipéptido NS nativo o a partir de un polipéptido NS modificado, como se define posteriormente más abajo.

Un polipéptido NS “nativo” se refiere a una proteína, polipéptido o péptido NS que se puede encontrar o aislar a partir de una fuente en la naturaleza, distinto de ser modificado o alterado artificialmente por el hombre en el laboratorio. De este modo, la expresión “polipéptido NS nativo” incluiría polipéptidos NS de origen natural y sus fragmentos. Tales fuentes en la naturaleza incluyen muestras biológicas (por ejemplo sangre, plasma o suero procedente de pacientes infectados por VHC, o que han sido infectados en el pasado por un VHC), células cultivadas, así como materiales recombinantes (por ejemplo virus o genoma de VHC, librerías de ADN genómico o ADNc, plásmidos que contienen fragmentos de genoma de VHC, precursor procesado previamente recombinante o polipéptido NS procesado maduro, y similares). Las secuencias nucleotídicas y de aminoácidos de un número de polipéptidos NS/genes nativos se han descrito en la bibliografía, y están disponibles en bancos de datos especializados. Los ejemplos representativos de polipéptidos NS nativos se exponen en SEC ID n°: 1-4 (SEC ID n°: 1-4 proporciona las secuencias de aminoácidos de los polipéptidos nativos NS3, NS4A, NS4B y NS5B de la cepa JA del VHC de genotipo 1b). Sin embargo, los polipéptidos NS nativos no están limitados a estas secuencias ejemplares. De hecho, las secuencias de aminoácidos pueden variar entre diferentes genotipos, subtipos, y aislados de VHC, y este alcance natural de variación genética está incluido en el alcance de la invención.

Uno o más de los polipéptidos NS incluidos en la proteína de fusión de la invención pueden incluir, independientemente entre sí, una o más modificaciones de aminoácidos de las secuencias ejemplares u otros polipéptidos NS nativos. La modificación o modificaciones se pueden generar por medio de mutación y/o adición de restos químicos (por ejemplo alquilación, acetilación, amidación, fosforilación y similar) o restos marcadores. La mutación o mutaciones incluyen supresión, sustitución o adición de uno o más restos de aminoácidos, o cualquier combinación de estas posibilidades. Cuando se contemplan varias modificaciones, pueden implicar restos consecutivos y/o restos no consecutivos. La modificación o modificaciones se pueden realizar de muchas maneras conocidas por los expertos en la técnica. Por ejemplo, la secuencia nucleotídica que codifica el polipéptido NS se puede modificar usando técnicas recombinantes habituales, tales como corte enzimático seguido de modificación y ligación de fragmento definido, mutagénesis dirigida al sitio (por ejemplo usando el sistema de mutagénesis *in vitro* Sculptor™ de Amersham, Les Ullis, Francia), mutagénesis de PCR o barajado de ADN.

Un polipéptido NS modificado preferido retiene un grado elevado de identidad de secuencia de aminoácidos con el polipéptido NS nativo correspondiente, por ejemplo al menos 75% de restos de aminoácidos idénticos a lo largo de la secuencia de aminoácidos de longitud completa o un fragmento más corto de la misma (por ejemplo de al menos 10, 15, 20, 25, 30, 40, 50, 100 aminoácidos de longitud). Más específicamente, en el contexto de la invención, en el contexto de la invención, el polipéptido NS modificado en uso en la invención comparte un grado de identidad con el polipéptido NS nativo correspondiente que es mayor que 75%, ventajosamente mayor que 80%, deseablemente mayor que 85%, preferentemente mayor que 90%, más preferentemente mayor que 95%, todavía más preferentemente mayor que 97%. El porcentaje de identidad entre los dos polipéptidos es una función del número de posiciones idénticas compartidas por la secuencia, teniendo en cuenta el número de espacios que es necesario introducir para el alineamiento óptimo y la longitud de cada espacio. En la técnica existen diversos programas de ordenador y algoritmos matemáticos para determinar el porcentaje de identidades entre secuencias de aminoácidos tales como, por ejemplo, el programa Blast (por ejemplo Altschul *et al.*, 1997, Nucleic Acids Res. 25, 3389-3402; Altschul *et al.*, 2005, FEBS J. 272, 51-01-5109) disponible en NCBI.

Por ejemplo, cualquiera o todos los polipéptidos NS incluidos en la proteína de fusión de la invención se puede modificar para ser representativo de un genotipo o subtipo específico, y de este modo comprende una secuencia de aminoácidos que corresponde a una secuencia de consenso o de casi consenso que se determina típicamente

después del alineamiento de secuencias de diversos NS de un genotipo o subtipo particular. Cualquiera o todos los polipéptidos NS también se pueden modificar para proporcionar un polipéptido NS con propiedades funcionales modificadas, tales como una función enzimática reducida, y/o una capacidad reducida para ser anclado en una membrana celular y/o una capacidad reducida para ser procesado post-traduccionalmente, como se describe más abajo. Los aminoácidos que son críticos para tales propiedades funcionales se pueden identificar por métodos habituales, tales como análisis estructural y funcional. El experto en la técnica puede determinar fácilmente el tipo de modificación o modificaciones que es capaz de reducir o interrumpir una actividad enzimática dada. Por ejemplo, se puede obrar mediante técnicas de PCR o mutagénesis dirigida al sitio a fin de suprimir o sustituir uno o más aminoácidos en una región activa de la actividad enzimática (por ejemplo en el sitio catalítico), de manera que la actividad enzimática nativa se reduzca significativamente o se anule. La reducción o falta de una actividad biológica se puede determinar fácilmente en ensayos apropiados según la actividad enzimática a ensayar usando métodos conocidos por los expertos en la técnica. Los dominios de anclaje a la membrana se predicen habitualmente basándose en su naturaleza hidrófoba.

Según la presente invención, los al menos tres polipéptidos MS incluidos en la proteína de fusión de la invención se configuran en un orden que es distinto del orden en el que aparecen en la configuración nativa. La configuración nativa es conocida en la técnica (véase el repaso dado en la sección de introducción de la presente solicitud), y el orden en el que aparecen los polipéptidos MS es como se encuentra en un precursor poliproteico de VHC, es decir, desde el término N al término C NS2-NS3-NS4A-NS4B-NS5A-NS5B. Según la configuración nativa:

- o un polipéptido NS2 siempre antecede a un polipéptido NS3 o un polipéptido NS4A o un polipéptido NS4B o un polipéptido NS5A o un polipéptido NS5B; y
- o un polipéptido NS3 siempre antecede a un polipéptido NS4A o un polipéptido NS4B o un polipéptido NS5A o un polipéptido NS5B; y
- o un polipéptido NS4A siempre antecede a un polipéptido NS4B o un polipéptido NS5A o un polipéptido NS5B; y
- o un polipéptido NS4B siempre antecede a un polipéptido NS5A o un polipéptido NS5B; y
- o un polipéptido NS5A siempre antecede a un polipéptido NS5B.

En la proteína de fusión de la invención, la configuración no es nativa en el sentido de que al menos uno de los polipéptidos NS aparece en un orden que es distinto de aquel de la configuración nativa. De este modo, si la proteína de fusión comprende un polipéptido NS3, un polipéptido NS4A y un polipéptido NS5B, la configuración nativa sería NS3-NS4A-NS5B, con NS3 en el término N, y NS5B en el término C. Por el contrario, una configuración no nativa puede ser NS5B-NS3- NS4A, NS5B-NS4A-NS3, NS4A-NS3-NS5B, NS4A-NS5B-NS3 o NS3-NS5B-NS4A.

En particular, la proteína de fusión de la invención comprende al menos uno de los siguientes:

- o un polipéptido NS4A fusionado directamente o a través de un ligador al término N de un polipéptido NS3;
- o un polipéptido NS3 fusionado directamente o a través de un ligador al término N de un polipéptido NS5B;
- o un polipéptido NS4B fusionado directamente o a través de un ligador al término N de un polipéptido NS5B;
- o un polipéptido NS4A fusionado directamente o a través de un ligador al término N de un polipéptido NS3, que está fusionado directamente o a través de un ligador al término N de un polipéptido NS4B; y/o
- o un polipéptido NS3 fusionado directamente o a través de un ligador al término N de un polipéptido NS4B que está fusionado directamente o a través de un ligador al término N de un polipéptido NS5B.

En tales porciones específicas de la proteína de fusión de la invención, cada uno de los polipéptidos NS puede ser independientemente nativo o puede estar modificado. Por ejemplo, el polipéptido NS4A incluido en la porción NS4A-NS3 puede ser nativo, mientras que el polipéptido NS3 comprende al menos una de las modificaciones descritas más abajo.

En una forma de realización, todos los polipéptidos NS incluidos en la proteína de fusión de la invención se originan de la misma cepa o aislado de VHC. Como alternativa, al menos dos de los polipéptidos NS se originan a partir de cepas o aislados de VHC diferentes, a fin de proporcionar protección frente a un intervalo más amplio de genotipos de VHC. También podría ser interesante adaptar la proteína de fusión a la región geográfica específica donde se usará, incluyendo al menos un polipéptido NS de un aislado de VHC que es endémico en esta región (por ejemplo, los genotipos 1a, 1b, 2 y 3 son los más prevalentes en América del Norte, Europa y Asia; el genotipo 4 es predominante en África del Norte y Central; el genotipo 5 hasta ahora se ha identificado mayoritariamente en África del Sur, y los aislados del genotipo 6 se han encontrado principalmente en Vietnam y en Hong Kong. Por ejemplo, si

la proteína de fusión comprende un polipéptido NS3, un polipéptido NS4A y un polipéptido NS5B, el polipéptido NS3 se puede originar a partir de una primera cepa de VHC, y los polipéptidos NS4A y NS5B a partir de una segunda cepa de VHC. Como alternativa, el polipéptido NS4A se puede originar a partir de la primera cepa de VHC, y los polipéptidos NS3 y NS5B a partir de una segunda cepa de VHC. Como alternativa, el polipéptido NS5B se puede originar a partir de una primera cepa de VHC, y los polipéptidos NS3 y NS4A a partir de una segunda cepa de VHC. Como alternativa, los polipéptidos NS3, NS4A y NS5B se originan cada uno de diferentes cepas de VHC. Cuando la proteína de fusión de la invención también comprende un polipéptido NS4B, se puede originar a partir de la misma cepa de VHC o a partir de una cepa de VHC diferente de los otros polipéptidos.

Las realizaciones preferidas de la invención están dirigidas a una fusión que comprende, o consiste esencialmente en, o consiste en un polipéptido NS4A, un polipéptido NS3 y un polipéptido NS5B con NS4A en el término N y NS5B en el término C (fusión NS4A-3-5B). Opcionalmente, la fusión puede incluir un polipéptido NS4B (o un fragmento o fragmentos o variante o variantes del mismo) para mejorar adicionalmente la actividad inmunógena de la proteína de fusión resultante. El polipéptido NS4B está incluido preferentemente entre los polipéptidos NS3 y NS5B. Por lo tanto, la presente invención también se refiere a una proteína de fusión que comprende, o consiste esencialmente en, o consiste en un polipéptido NS4A, un polipéptido NS3, un polipéptido NS4B y un polipéptido NS5B, con NS4A en el término N y NS5B en el término C (fusión NS4A-3-4B-5B). Aparte de los polipéptidos NS3, NS4A, NS5B y del NS4B opcional, se prefiere no obstante que la proteína de fusión de la invención no contenga otros polipéptidos de VHC, y especialmente ningún core y según se explica anteriormente en relación con los polipéptidos NS, ningún polipéptido NS5A.

Más específicamente, la expresión "polipéptido NS3" se refiere a una proteína, polipéptido o péptido NS3 nativo, obtenido de cualquier cepa de VHC, o a una proteína, polipéptido o péptido NS3 modificado como se define anteriormente. El polipéptido NS3 en uso en la invención tiene una longitud de al menos 200 aminoácidos, ventajosamente al menos 300 aminoácidos, preferentemente al menos 400 aminoácidos, más preferentemente al menos 500 aminoácidos, e incluso más preferentemente al menos 600 aminoácidos. Con el fin de ilustrar, la proteína NS3 nativa de HCV-1 tiene una longitud de 631 aminoácidos, y está situada aproximadamente desde las posiciones 1027 a 1657 en un precursor poliproteico. Un polipéptido NS3 nativo comprende dos dominios activos distintos, un dominio de serina proteasa en el término N y un dominio de helicasa en el término C. La NS3 proteasa es responsable del procesamiento del precursor de la poliproteína de VHC en las uniones NS3/NS4A, NS4A/NS4B, NS4B/NS5A y NS5A/NS5B, y requiere NS4A como cofactor para la actividad proteolítica (Tomei *et al.*, 1993, J. Virol. 67, 4017-4026) y para el plegamiento apropiado (Yao *et al.*, 1999, Structure (London) 7 p. 1353). Un motivo que consiste en un ácido aspártico o glutámico en la posición 6 y una cisteína o treonina en la posición 1 en dirección 5' del sitio de escisión, y una serina o alanina en la posición 1 en dirección 3' del sitio de escisión, es reconocido específicamente por la NS3 proteasa de VHC. Se piensa que la NS3 helicasa es importante para desenrollar el intermedio de ARN bicatenario durante la replicación vírica (Tai *et al.*, 1996, J. Virol. 70, 8477-8484).

Mediante un "polipéptido NS4A" se quiere decir una proteína, polipéptido o péptido NS4A nativo procedente de cualquier cepa de VHC, o una proteína, polipéptido o péptido modificado como se define anteriormente. El polipéptido NS4A en uso en la invención tiene una longitud de al menos 10 aminoácidos, ventajosamente al menos 11 aminoácidos, preferentemente al menos 12 aminoácidos, más preferentemente al menos 13 aminoácidos. Mediante un "polipéptido NS4B" se quiere decir una proteína, polipéptido o péptido NS4B nativo procedente de cualquier cepa de VHC, o una proteína, polipéptido o péptido NS4B modificado como se define anteriormente. El polipéptido NS4B en uso en la invención tiene una longitud de al menos 20 aminoácidos, ventajosamente al menos 25 aminoácidos, preferentemente al menos 30 aminoácidos, más preferentemente al menos 31 aminoácidos, e incluso más preferentemente al menos 32 aminoácidos. Con fines ilustrativos, la proteína NS4 (A-B) nativa de HCV-1 que con la longitud de 315 aminoácidos y está situada aproximadamente desde las posiciones 1658 a 1972 en el precursor poliproteico. El polipéptido NS4A (posiciones 1658 a 1711) es un cofactor de NS3 proteasa, pero también se requiere para la estabilidad y localización de NS3 en la membrana del retículo endoplásmico (ER) (Failla *et al.*, 1994, J. Virol. 68, 3753-3760). El Polipéptido NS4B (posiciones 1712 a 1972) es una proteína de membrana integral cuya función todavía es desconocida. Los algoritmos predichos sugieren la presencia de cuatro a seis segmentos transmembránicos. Sin embargo, su expresión induce la formación de una banda membranosa aparentemente derivada de ER (Egger *et al.*, 2002, J. Virol. 76, 5974-5984).

Mediante un "polipéptido NS5B" se quiere decir una proteína, polipéptido o péptido NS5B nativo procedente de VHC, o una proteína, polipéptido o péptido NS5B modificado como se define anteriormente. El polipéptido NS5B en uso de la invención tiene una longitud de al menos 200 aminoácidos, ventajosamente al menos 300 aminoácidos, preferentemente al menos 400 aminoácidos, más preferentemente al menos 500 aminoácidos, e incluso más preferentemente al menos 550 aminoácidos. Con fines ilustrativos, la proteína NS5B nativa de HCV-1 tiene una longitud de 591 aminoácidos y está situada aproximadamente desde las posiciones 2421 a 3011 en el precursor poliproteico. La proteína NS5B nativa actúa como una ARN polimerasa dependiente de ARN, lo que presumiblemente permite la síntesis de un intermedio de ARN de cadena negativa y copias de progenie de cadena positiva (Lohman *et al.*, 1997, J. Virol. 71, 8416-8428). Se encuentra asociada a la membrana del retículo endoplásmico mediante su secuencia de 21 aminoácidos del término C (Schmidt-Mende *et al.*, 2001, J. Biol. Chem. 276, 44052-44063).

En aras de la claridad, los tramos de aminoácidos citados en la presente memoria en relación con los polipéptidos NS3, NS4A, NS4B y NS5B se dan con respecto a sus posiciones en el precursor poliproteico de HCV-1 (como se describe por Choo *et al.*, 1991, Proc. Natl. Acad. Sci. USA 88, 2451-2455 o en GenBank con el número de acceso M62321). Sin embargo, como se explica anteriormente, la presente invención también engloba polipéptidos NS de otras cepas y aislados de VHC, así como polipéptidos NS modificados. Por lo tanto, excepto que el contexto indique claramente otra cosa, cuando se refiere en la presente memoria a un polipéptido NS o un polipéptido del mismo mediante referencia al precursor poliproteico de HCV-1, esto significa el polipéptido NS o el péptido del mismo de HCV-1, un polipéptido NS o el polipéptido del mismo de cualquier otra cepa o aislado de HCV, o un polipéptido NS modificado o un péptido del mismo. Una realización preferida de la presente invención se refiere a una proteína de fusión que comprenden los polipéptidos NS3, NS4A, NS5B y el NS4B opcional que se originan a partir de la cepa JA de VHC de genotipo 1b (Kato *et al.*, 1990, Proc. Natl. Acad., Sci. 87, 9524-9528).

En una forma de realización, el polipéptido NS3 incluido en la proteína de fusión de la presente invención está modificado en comparación con el polipéptido NS3 nativo correspondiente, para mostrar una actividad de proteasa significativamente reducida, y de este modo ser incapaz de mediar la escisión de la proteína de fusión en polipéptidos individuales. La actividad de proteasa se ha localizado en la porción N-terminal del polipéptido NS3, desde alrededor de la posición 1027 hasta alrededor de la posición 1207 numeradas con respecto a la poliproteína de HCV-1 de longitud completa (desde alrededor de la posición 1 hasta alrededor de la posición 181 del polipéptido NS3 nativo de HCV-Ja mostrado en SEC ID n°: 1). Más específicamente, la actividad de NS3 proteasa se ha atribuido a los restos de la tríada catalíticos His (H) en la posición 1083, Asp (D) en la posición 1107, y Ser (S) en la posición 1165 (respectivamente las posiciones 57, 81 y 139 de SEC ID n°: 1). Los ejemplos representativos de polipéptidos NS3 adecuados deficientes en proteasas se describen en Bartenshlager *et al.*, 1993, J. Virol. 67, 3835-3844 y Tomei *et al.*, 1993, J. Virol. 67, 4017-4026. Preferentemente, puesto que los restos Asp y Ser catalíticos se encuentran en un dominio antigénico predicho, el polipéptido NS3 en uso de la presente invención comprende la sustitución del resto His en la posición 1083 (que corresponde a la posición 57 de SEC ID n°: 1) o del resto de aminoácido situado en una posición equivalente de un polipéptido NS3 nativo de otro serotipo de VHC, por cualquier resto de aminoácido distinto de His, con especial preferencia por una sustitución por un resto de Ala (H1083A). La interrupción de la actividad de proteasa se puede determinar usando ensayos bien conocidos en la técnica (Takeshita *et al.*, 1997, Anal. Biochem. 247, 242-246; Kakiuchi *et al.*, 1997 J. Biochem 122, 749-755; Sali *et al.*, 1998, Biochemistry 37, 3392-3401; Kakiuchi *et al.*, 1999, J. Virol. Meth. 80, 77-84).

Como alternativa o en combinación, el polipéptido NS3 incluido la proteína de fusión de la invención está modificado en comparación con el polipéptido NS3 nativo correspondiente, para mostrar una actividad de helicasa significativamente reducida. Se han indicado cuatro restos en la actividad de helicasa asociada a NS3, especialmente Thr en posición 1295, Thr en posición 1437, Arg en posición 1490, y Arg en posición 1493, numeradas con relación a la poliproteína HCV-1 de longitud completa (que corresponden a los restos de Thr en las posiciones 269 y 411 y los restos de Arg en las posiciones 464 y 467 del polipéptido NS3 nativo de HCV-JA mostrado en SEC ID n°: 1). Los ejemplos representativos de polipéptidos NS3 adecuado deficientes en helicasa se describen en Kim *et al.* (1997, J. Virol. 71, 9400) y Lin y Kim (1999, J. Virol. 73, 8798-8807). Preferentemente, el polipéptido NS3 en uso de la presente invención comprende la sustitución del resto Arg en la posición 1490 (que corresponde a la posición 464 de SEC ID n°: 1) o del resto de aminoácido situado en una posición equivalente de un polipéptido NS3 nativo de otro serotipo de VHC por cualquier resto de aminoácido distinto de Arg, y la sustitución del resto Thr en la posición 1295 (que corresponde a la posición 269 de SEC ID n°: 1) o del resto de aminoácido situado en una posición equivalente de un polipéptido NS3 nativo de otro serotipo de VHC por cualquier resto de aminoácido distinto de Thr, con especial preferencia por una sustitución por un resto Ala en ambos casos (T1295A y R1490A). La interrupción de la actividad de helicasa se puede determinar usando ensayos bien conocidos en la técnica, por ejemplo evaluando la actividad desenrolladora.

Lo más preferible, el polipéptido NS2 en el uso de la presente invención está mutado para reducir o interrumpir tanto las actividades de proteasa como de helicasa del polipéptido NS3 nativo, y comprende las modificaciones explicadas anteriormente en relación con estas funciones enzimáticas, con preferencia especial por las mutaciones H1083A, T1295A y R1490A.

Como alternativa o en combinación con las modificaciones propuestas anteriormente en relación con el polipéptido NS3, el polipéptido NS5B incluido en la proteína de fusión de la invención está modificado en comparación con el polipéptido NS5B nativo correspondiente, para mostrar una actividad de ARN polimerasa dependiente de ARN. En esta función enzimática se han implicado dos restos Asp, respectivamente las posiciones 2640 y 2738, numeradas con relación a la poliproteína de HCV-1 de longitud completa (que corresponden a los restos Asp en las posiciones 220 y 318 de polipéptido NS5B nativo de HCV-JA mostrado en SEC ID n°: 4). Los ejemplos representativos de polipéptidos NS5B adecuados deficientes en polimerasas se describen en Lohmann *et al.* (1997, J. Virol. 71, 8416-8428). Preferentemente, el polipéptido NS5B en uso en la presente invención comprende al menos la sustitución del resto Asp en la posición 2640 (que corresponde a la posición 220 de SEC ID n°: 4) o del resto de aminoácido situado en una posición equivalente de un polipéptido NS5B nativo de otro serotipo de VHC por cualquier resto de aminoácido distinto de Asp, y/la sustitución del resto Asp en la posición 2738 (que corresponde a la posición 318 de SEC ID n°: 4) o del resto de aminoácido situado en una posición equivalente de un polipéptido NS5B nativo de otro serotipo de VHC por cualquier resto de aminoácido distinto de Asp, con referencia especial por la sustitución por un

resto Asn en ambos casos. La interrupción de la actividad de ARN polimerasa se puede determinar usando ensayos bien conocidos en la técnica, por ejemplo ensayos de replicasas convencionales basados en la incorporación de sustratos nucleotídicos radioactivos en un producto de ARN naciente (véanse, por ejemplo, Behrens *et al.*, 1996, EMBO J. 15, 12-22; Ferrari *et al.*, 1999, J. Virol. 73, 1649-1654).

En otra forma de realización, los polipéptidos NS incluidos en la proteína de fusión de la invención pueden comprender además modificaciones adicionales en comparación con los polipéptidos NS nativos correspondientes. Las modificaciones adecuadas son aquellas que son beneficiosas al procesamiento, estabilidad y solubilidad de la proteína de fusión resultante, por ejemplo aquellos dirigidos a inactivar sitios de escisión potenciales, el anclaje de la membrana y/o la glucosilación como se describe más abajo.

Ventajosamente, la proteína de fusión de la invención no comprende uno o más de los sitios de escisión reconocidos por NS3 normalmente presentes en el precursor de la poliproteína nativa de VHC en las uniones NS4A, NS4A/NS4B, NS4B/NS5A y NS5A/NS5B, a fin de evitar el procesamiento en polipéptidos NS individuales. Aunque el polipéptido NS3 preferido en uso en la presente invención muestra una actividad de proteasa significativamente reducida, la inactivación de los sitios de escisión reconocidos por NS3 introduce un grado adicional de seguridad. La secuencia de consenso del sitio de escisión es Asp/Glu-X4-Cys/Thr-Ser/Ala, en la que X es cualquier resto de aminoácido. Se demostró que la escisión mediante el polipéptido NS3 nativo se produce después del resto Cys/Thr. Preferentemente, el polipéptido NS3 en uso en la invención no comprende el resto Thr o Cys normalmente presente en el término C de un polipéptido NS3 nativo (posición 1657 numerada con respecto a la poliproteína de HCV-1, que corresponde a la posición 361 de SEC ID n°: 1 o una posición equivalente de un polipéptido nativo de otro serotipo de VHC). Como alternativa o en combinación, el polipéptido NS4A no comprende el resto Cys normalmente presente en el término C de un polipéptido NS4A nativo (posición 1711 numerada con relación a la poliproteína de HCV-1, que corresponde a la posición 54 de SEC ID n°: 2 o una posición equivalente de un polipéptido NS4A nativo de otro serotipo de VHC). Como alternativa o en combinación, el polipéptido NS4B opcional no comprende el resto Cys normalmente presente en el término C de un polipéptido NS4B nativo (posición 1972 numerada con relación a la poliproteína de HCV-1, que corresponde a la posición 261 de SEC ID n°: 3 o una posición equivalente de un polipéptido NS4B nativo de otro serotipo de VHC).

En todavía otra forma de realización, la proteína de fusión de la invención puede estar modificada adicionalmente para suprimir uno más dominios hidrófobos que están implicados normalmente en el anclaje a la membrana de los polipéptidos nativos NS4A, NS4B y/o NS5B. Se han identificado en el polipéptido NS4B nativo seis dominios hidrófobos de anclaje a la membrana, mientras que uno está presente en la porción del término N del polipéptido NS4A nativo y uno está en la porción C-terminal del polipéptido nativo NS5B. Su supresión al menos parcial puede permitir mejorar la solubilidad de la proteína de fusión y facilitar de este modo su producción por medios recombinantes. Esto también puede permitir limitar la citotoxicidad global de la proteína de fusión en comparación con la administración de los polipéptidos de VHC individuales en un organismo hospedante dado.

A este respecto, el polipéptido NS4A está modificado ventajosamente mediante supresión o sustitución de uno o más restos de aminoácidos hidrófobos normalmente presentes en la porción N-terminal del polipéptido NS4A nativo. Un polipéptido NS4A preferido tiene una supresión de hasta 20 aminoácidos primeros en el término N (por ejemplo, supresión preferida desde aproximadamente la posición 1658 hasta aproximadamente la posición 1677, numeradas con relación a la poliproteína de HCV-1 de longitud completa, que corresponden a la supresión desde aproximadamente la posición 1 hasta aproximadamente la posición 20 del polipéptido nativo NS4A de HCV-JA mostrado en SEC ID n°: 2), mientras que no obstante retiene la porción del polipéptido NS4A nativo que se extiende desde aproximadamente la posición 1678 hasta aproximadamente la posición 1690, numeradas con relación a la poliproteína de HCV-1 de longitud completa (desde aproximadamente la posición 21 hasta aproximadamente la posición 33 del polipéptido nativo NS4A de HCV-JA mostrado en SEC ID n°: 2). Puede retener o no la porción C-terminal del polipéptido NS4A nativo, por ejemplo la porción que se extiende desde aproximadamente la posición 1691 hasta aproximadamente la posición 1711, numerada con relación a la poliproteína de HCV-1 de longitud completa (que corresponde desde aproximadamente la posición 34 hasta aproximadamente la posición 54 del polipéptido NS4A nativo de HCV-JA mostrado en SEC ID n°: 2). Lo más preferible, el polipéptido NS4A incluido en la proteína de fusión de la invención consiste en la porción de un polipéptido NS4A nativo que se extiende desde la posición 1678 hasta la posición 1690, numeradas con relación a la poliproteína de HCV-1 de longitud completa (que corresponde a la porción NS4A que se extiende desde la posición 21 hasta la posición 33 del polipéptido NS4A nativo de HCV-JA mostrado en SEC ID n°: 2).

Como alternativa o en combinación, un polipéptido NS5B preferido tiene suprimidos 10 a 30 restos de aminoácidos normalmente presentes en el término C de un polipéptido NS5B nativo, con especial preferencia por la supresión de al menos 21 restos de aminoácidos en el término C (por ejemplo desde la posición 2991 hasta la posición 3011, numeradas con relación a la polipéptido de HCV-1, que corresponden desde la posición 571 hasta la posición 591 de SEC ID n°: 4).

Como alternativa o en combinación, a un polipéptido NS4B opcional preferido se le suprime al menos uno de los seis dominios hidrófobos de anclaje a la membrana. Más preferentemente, el polipéptido NS4B opcional está truncado por uno o más aminoácidos tanto en el término N como C del polipéptido NS4B nativo correspondiente para retener

esencialmente el dominio antigénico interno. Preferentemente el polipéptido NS4B opcional en uso en la invención comprende o consiste esencialmente como alternativa en el tramo de aminoácidos que se extiende desde aproximadamente la posición 1789 hasta aproximadamente la posición 1890, numeradas con relación a la poliproteína de HCV-1 de longitud completa (que corresponde a aproximadamente la posición 78 a aproximadamente la posición 109 del polipéptido NS4B nativo de HCV-J mostrado en SEC ID n°: 3).

La proteína de fusión de la invención se puede modificar adicionalmente para suprimir la N-glucosilación potencial con la expresión en una célula hospedante dada. A este respecto, se ha identificado un sitio de glucosilación de consenso Asn-Val-Ser-Val en el polipéptido NS5B nativo desde la posición 2789 a la posición 2792 numeradas con relación a la poliproteína de HCV-1 de longitud completa (posición 369 a posición 372 de la proteína NS5B nativa de HCV-JA mostrada en SEC ID n°: 4). A este respecto, una mutación ejemplar consiste en la sustitución del resto Ser en la posición 2791, numerada con relación a la poliproteína de HCV-1 de longitud completa (posición 371 de SEC ID n°: 4), por un resto diferente de un resto Ser, tal como un resto Gly.

De forma deseable, la proteína de fusión de la invención es inmunógena, en el sentido de que es capaz de inducir o estimular una respuesta inmunitaria específica de antígeno, ya sea humoral o celular, o ambas, con la introducción en un organismo hospedante. La presente invención también engloba proteínas de fusión que se han manipulado mediante ingeniería para potenciar la inmunogenicidad (por ejemplo mediante oxidación del enlace de disulfuro). De forma deseable, el plegamiento de los diversos polipéptidos NS incluidos en la proteína de fusión de la invención es como el plegamiento de los polipéptidos NS nativos. En el contexto de la invención, se prefiere que la proteína de fusión retenga uno o más dominios antigénicos reconocidos por un receptor de células T. En la bibliografía se describe un número de dominios antigénicos de VHC que se pueden retener adecuadamente en la proteína de fusión (véase, por ejemplo, Chien *et al.*, 1992, Proc. Natl. Acad. Sci. USA 89, 10011-10015; Chien *et al.*, 1993, J. Gastroent Hepatol. 8, S33-39 y el documento WO03/097677).

Preferentemente, el polipéptido NS3 incluido en la proteína de fusión de la invención retiene las porciones del polipéptido NS3 nativo que se extienden desde aproximadamente las posiciones 1038 hasta aproximadamente 1082, desde aproximadamente la posición 1096 hasta aproximadamente la posición 1181, y desde aproximadamente la posición 1244 hasta aproximadamente la posición 1274, numeradas con relación a la poliproteína de HCV-1 de longitud completa (por ejemplo desde aproximadamente la posición 12 hasta aproximadamente la posición 56, desde aproximadamente la posición 70 hasta aproximadamente la posición 155, y/o desde aproximadamente la posición 218 hasta aproximadamente la posición 248 del polipéptido NS3 nativo de HCV-JA descrito en SEC ID n°: 1). Preferentemente, el polipéptido NS5B incluido en la proteína de fusión de la invención retiene la porción del polipéptido NS5B nativo que se extiende desde aproximadamente la posición 2573 hasta aproximadamente la posición 2601, numeradas con relación a la poliproteína de HCV-1 de longitud completa (por ejemplo desde aproximadamente la posición 155 hasta aproximadamente la posición 182 del polipéptido nativo NS5B de HCV-J descrito en SEC ID n°: 4). Cuando la proteína de fusión de la invención comprende un polipéptido NS4B, preferentemente retiene la porción del polipéptido NS4B nativo desde aproximadamente la posición 1789 hasta aproximadamente la posición 1820, numeradas con relación a la poliproteína de HCV-1 de longitud completa (por ejemplo desde aproximadamente la posición 78 hasta aproximadamente la posición 109 del polipéptido NS4B nativo de HCV-J descrito en SEC ID n°: 3).

La actividad inmunógena de la proteína de fusión de la invención se puede evaluar mediante un número de técnicas que son habituales en la técnica, tales como las descritas en lo sucesivo en relación con el método de la invención, o como se ilustra en los ejemplos. En el contexto de la invención, se prefiere que la actividad inmunógena de la proteína de fusión sea mayor en extensión y/o en naturaleza que la actividad inmunógena habitual proporcionada por un polipéptido NS nativo. Por ejemplo, con la introducción de la proteína de fusión en un organismo hospedante, la respuesta inmunitaria puede ser de mayor potencia, de naturaleza más amplia (por ejemplo que implica más células inmunitarias, tales como células CD4+ y CD8+), o su alcance puede ser diferente (por ejemplo dirigido a un epítipo críptico no dominante) que con la introducción de un polipéptido NS nativo. Esto proporcionaría un efecto terapéutico potenciado, permitiendo así reducir los regímenes de dosificación y mejorando la calidad de vida.

Un polipéptido NS3 preferido se origina a partir de la proteína NS3 nativa de HCV-JA mostrada en SEC ID n°: 1, que se modifica al menos de tal manera que:

- o retiene la porción que se extiende desde la posición 12 hasta la posición 56, desde la posición 70 hasta la posición 155, y desde la posición 218 hasta la posición 248;
- o comprende la sustitución del resto His en la posición 57 por un resto de aminoácido diferente de una His, tal como un resto Ala;
- o comprende la sustitución del resto Thr en la posición 269 por un resto de aminoácido diferente de una Thr, tal como un resto Ala; y
- o comprende la sustitución del resto Arg en la posición 464 por un resto de aminoácido diferente de una Arg, tal como un resto Ala; y

- o no comprende el resto Thr en la posición 631.

Incluso más preferentemente, el polipéptido NS3 en uso en la invención comprende, consiste esencialmente en, o
5 consiste como alternativa en la secuencia de aminoácidos mostrada en SEC ID n°: 5.

Un polipéptido NS4A preferido se origina a partir de la proteína NS4A de HCV-JA mostrada en SEC ID n°: 2, que se
modifica al menos de tal manera que:

- 10 o contiene la porción de SEC ID n°: 2 desde la posición 21 hasta la posición 33;
- o no contiene la porción de SEC ID n°: 2 desde la posición 1 hasta la posición 20.

Incluso más preferentemente, el polipéptido NS4A tiene una secuencia de aminoácidos que consiste esencialmente
15 en la porción de SEC ID n°: 2 desde la posición 21 hasta la posición 33, precedida por un resto Met iniciador. Lo más
preferible, el polipéptido NS4A consiste esencialmente en la secuencia de aminoácidos mostrada en SEC ID n°: 6,
precedida por un resto Met iniciador.

Un polipéptido NS4B opcional preferido se origina a partir de la proteína NS4B nativa de HCV-JA mostrada en SEC
20 ID n°: 3, que se modifica al menos de tal manera que:

- o comprende la porción que se extiende desde el resto Ser en la posición 78 hasta el resto Leu en la posición
109;
- 25 o no comprende el resto Cys en la posición 261.

Incluso más preferentemente, el polipéptido NS4B opcional consiste esencialmente o como alternativa consiste en la
secuencia de aminoácidos como se muestra en SEC ID n°: 7.

Un polipéptido NS5B preferido se origina a partir de la proteína NS5B nativa de HCV-JA mostrada en SEC ID n°: 4,
30 que se modifica al menos de tal manera que:

- o comprende la porción que se extiende desde el resto Arg en la posición 154 hasta el resto Leu en la posición
183;
- 35 o comprende la sustitución del resto Asp en la posición 220 por un resto de aminoácido diferente de una Asp,
tal como un resto Asn;
- o comprende la sustitución del resto Asp en la posición 318 por un resto de aminoácido diferente de una Asp,
tal como un resto Asn; y
- 40 o no comprende la porción que se extiende desde el resto Trp en la posición 571 hasta el resto Arg en la
posición 591.

Incluso más preferentemente, el polipéptido NS5B comprende, consiste esencialmente, o como alternativa consiste
45 en la secuencia de aminoácidos como se muestra en SEC ID n°: 8.

Las proteínas de fusión de la invención más preferidas comprenden, o como alternativa consisten esencialmente en,
o como alternativa consisten en una secuencia de aminoácidos que es homóloga o idéntica a la secuencia de
50 aminoácidos citada en SEC ID n°: 9 o 10. La secuencia citada en SEC ID n°: 9 (figura 1) corresponde a la fusión de
los polipéptidos NS4A, NS3 y NS5B descritos anteriormente, con una Met iniciadora N-terminal en la posición 1,
extendiéndose el polipéptido NS4A desde el resto de aminoácido 2 hasta el resto de aminoácido 14, extendiéndose
el péptido ligador desde el resto de aminoácido 15 hasta el resto de aminoácido 17, extendiéndose el polipéptido
NS3 desde el resto de aminoácido 18 hasta el resto de aminoácido 647, extendiéndose el ligador desde el resto de
55 aminoácidos 648 hasta el resto de aminoácido 652, y extendiéndose el polipéptido NS5B desde el resto de
aminoácidos 653 hasta el resto de aminoácido 1222. La secuencia citada en SEC ID n°: 10 (figura 2) corresponde a
la fusión de los polipéptidos más preferidos NS4A, NS3, NS4B y NS5B descritos anteriormente, con una Met
iniciadora N-terminal en la posición 1, extendiéndose el polipéptido NS4A desde el resto de aminoácido 2 hasta el
resto de aminoácido 14, extendiéndose el péptido ligador desde el resto de aminoácido 15 hasta el resto de
60 aminoácido 17, extendiéndose el polipéptido NS3 desde el resto de aminoácido 18 hasta el resto de aminoácido
647, extendiéndose el polipéptido NS4B desde el resto de aminoácidos 648 hasta el resto de aminoácidos 679, y
extendiéndose el polipéptido NS5B desde el resto de aminoácidos 680 hasta el resto de aminoácidos 1249.

Se describen en la presente memoria adicionalmente nuevos fragmentos peptídicos de las proteínas de fusión de la
65 invención, y especialmente aquellos citados en SEC ID n°: 9 o SEC ID n°: 10. Como se utiliza en la presente
memoria, un fragmento comprende al menos 10, 15, 20, 50 o más restos de aminoácidos contiguos de las proteínas
de fusión descritas en la presente memoria. Tales fragmentos se pueden escoger basándose en su capacidad para

llevar a cabo una función, por ejemplo unirse a un sustrato o actuar como un inmunógeno. Los fragmentos peptídicos adecuados son típicamente aquellos que comprenden un dominio o motivo de la proteína de fusión que contiene nuevas estructuras inmunógenas. Los dominios inmunógenos predichos son fácilmente identificables mediante programas de ordenador bien conocidos y fácilmente disponibles para los expertos en la técnica. Los fragmentos peptídicos se pueden sintetizar usando métodos de síntesis proteica conocidos.

La proteína de fusión de la presente invención se puede producir mediante cualquier método adecuado, por ejemplo mediante técnicas de síntesis peptídica directa estándar (por ejemplo Bodanszky, 1984, en Principles of peptide synthesis, Springer-Verlag) y mediante tecnología de ADN recombinante como se describe más abajo en relación con los vectores de la invención.

De este modo, la presente invención también proporciona una molécula de ácido nucleico aislada que codifica la proteína de fusión de la invención, con preferencia especial por moléculas de ácido nucleico que codifican una proteína de fusión que comprende o como alternativa consiste esencialmente en, o como alternativa consiste en una secuencia de aminoácidos como se muestra en SEC ID n°: 9 o SEC ID n°: 10 o una secuencia de aminoácidos homóloga a SEC ID n°: 9 o SEC ID n°: 10.

De forma deseable, las moléculas de ácido nucleico de la invención se optimizan para proporcionar una expresión de nivel elevado en una célula hospedante particular, por ejemplo células hospedantes de mamífero, de levadura (por ejemplo *Saccharomyces cerevisiae*, *Saccharomyces pombe* o *Pichia pastoris*) o procariontes (por ejemplo *E. coli*). De hecho se ha observado que, cuando existe más de un codón para codificar un aminoácido dado, los patrones de uso de codones de los organismos son muy no aleatorios (véase, por ejemplo, Wada *et al.*, 1992, Nucleic Acids Res. 20, 2111-2118), y la utilización de los codones puede ser notablemente diferente entre diferentes hospedantes (véase, por ejemplo, Nakamura *et al.*, 1996, Nucleic Acids Res. 24, 214-215). Puesto que las secuencias nucleotídicas que codifican NS usadas en la invención son de origen vírico (VHC), pueden tener un patrón de uso de codones inapropiado para la expresión eficiente en células hospedantes, tales como células bacterianas o eucariotas inferiores.

Típicamente, la optimización de codones se lleva a cabo sustituyendo uno o más codones "nativos" (por ejemplo de VHC), que corresponden a un codón usado de forma infrecuente en esta célula hospedante particular, por uno o más codones que codifican el mismo aminoácido que se usa más frecuentemente. Esto se puede lograr mediante mutagénesis convencional o mediante técnicas sintéticas químicas (por ejemplo, dando como resultado una molécula de ácido nucleico sintética). No es necesario sustituir todos los codones nativos que corresponden a codones usados infrecuentemente, puesto que se puede lograr una mayor expresión incluso con una sustitución parcial. Además, se pueden realizar algunas desviaciones de la adhesión estricta al uso optimizado de codones para adecuar la introducción del sitio o sitios de restricción en la molécula de ácido nucleico resultante.

Además de la optimización del uso de codones, la expresión en la célula hospedante se puede mejorar adicionalmente a través de modificaciones adicionales de la secuencia nucleotídica. Por ejemplo, la molécula de ácido nucleico de la invención se puede modificar para prevenir el agrupamiento de codones raros no óptimos que están presentes en áreas concentradas, y/o para suprimir o modificar elementos de secuencias al menos parcialmente negativos que se espera que influyan negativamente en los niveles de expresión. Tales elementos de secuencias negativos incluyen, sin limitación, las regiones que tienen un contenido muy elevado (>80%) o muy bajo (<30%) de GC; tramos de secuencias ricas en AT o ricas en GC; secuencias de repeticiones directas o invertidas inestables; estructuras secundarias de ARN; y/o elementos reguladores crípticos internos tales como cajas TATA internas, sitios chi, sitios de entrada al ribosoma, y/o sitios donantes/aceptores de ayuste.

La molécula de ácido nucleico optimizada de la invención es capaz preferentemente de expresar la proteína de fusión de la invención en una célula hospedante dada a un mayor nivel, es decir, al menos 110%, ventajosamente al menos 150%, y preferentemente al menos 200%, en comparación con el nivel expresado por una molécula de ácido nucleico que implica los genes nativos de VHC correspondientes en idénticas condiciones (por ejemplo mismo tipo celular, mismas condiciones de cultivo, mismo vectores de expresión, etc.). Los niveles de expresión se pueden evaluar mediante técnicas convencionales tales como transferencia Western, usando un anticuerpo específico para uno de los polipéptidos de VHC incluidos en la proteína de fusión de la invención.

Según una forma de realización preferida, la presente invención proporciona moléculas de ácido nucleico que comprende, o como alternativa consisten esencialmente en, o como alternativa consisten en una secuencia nucleotídica que es homóloga o, incluso más preferentemente, idéntica a cualquiera de las secuencias nucleotídicas mostradas en SEC ID n°: 11-16. La molécula de ácido nucleico de SEC ID n°: 11 codifica una proteína de fusión NS4A-3-5B cuya secuencia nucleotídica se ha optimizado para la expresión en células de mamíferos (por ejemplo humanas). La molécula de ácido nucleico de SEC ID n°: 12 codifica una fusión NS4A-3-4B-5B cuya secuencia nucleotídica se ha optimizado para la expresión en células de mamífero (por ejemplo células humanas). La molécula de ácido nucleico de SEC ID n°: 13 codifica una fusión NS4A-3-5B cuya secuencia nucleotídica se ha optimizado para la expresión en levadura (por ejemplo *P. pastoris*). La molécula de ácido nucleico de SEC ID n°: 14 codifica una fusión NS4A-3-4B-5B cuya secuencia nucleotídica se ha optimizado para la expresión en levadura (por ejemplo *P. pastoris*). La molécula de ácido nucleico de SEC ID n°: 15 codifica una fusión NS4A-3-5B cuya secuencia

nucleotídica se ha optimizado para la expresión en procariota (por ejemplo *E. coli*). La molécula de ácido nucleico de SEC ID n°: 16 codifica una fusión NS4A-3-4B-5B cuya secuencia nucleotídica se ha optimizado para la expresión en procariota (por ejemplo *E. coli*). Por supuesto, las secuencias se puede equipar con secuencias no codificantes en 5' y 3' que proporcionan si es necesario una Met iniciadora en el extremo 5', uno o más codones STOP en el extremo 3', y sitios de restricción adecuados en ambos extremos para facilitar las etapas de clonación. Las moléculas de ácido nucleico ejemplares también se proporcionan en las figuras 3-8.

También se describen moléculas de ácido nucleico que son capaces de hibridarse en condiciones restrictivas a las moléculas de ácido nucleico como se definen anteriormente. Las condiciones de hibridación restrictivas son conocidas por los expertos en la técnica. Típicamente, la hibridación se lleva a cabo en 6 veces de cloruro de sodio/citrato de sodio (SSC) a alrededor de 45°C, y a esto le sigue uno o más lavados en 0,2 veces SSC, 0,1% de SDS a 50-65°C. Preferentemente, tales ácidos nucleicos de hibridación comparten un grado de identidad de secuencias con las moléculas de ácido nucleico definidas anteriormente mayor que 70% a lo largo de la secuencia nucleotídica de longitud completa o un fragmento más corto de la misma (por ejemplo de al menos 30, 45, 50, 60, 80, 100, 150, 200 nucleótidos de longitud. Más preferentemente, el grado de identidad entre ellos es mayor que 70%, ventajosamente mayor que 80%, deseablemente mayor que 85%, preferentemente mayor que 90%, más preferentemente mayor que 95%, todavía más preferentemente mayor que 97%.

Los ejemplos representativos de tales moléculas de ácido nucleico de hibridación incluyen, sin limitación, moléculas de ácido nucleico que son complementarias en al menos alrededor de 20, 30, 40, 50, 100, o 150 nucleótidos contiguos incluidos en cualquiera de las SEC ID n°s: 11-16 (por ejemplo moléculas de ácido nucleico antisentido). También se describen variaciones de las secuencias ejemplares, tales como codones degenerados, o polimorfismo entre diferentes aislados/cepas de VHC, o variaciones que resultan de técnicas tales como el barajado de ADN.

También se describen fragmentos de la molécula de ácido nucleico de la invención, por ejemplo fragmentos generados mediante endonucleasas de restricción o mediante PCR. Tales fragmentos se pueden usar como sondas, cebadores o fragmentos que codifican una porción inmunógena de la proteína de fusión.

La molécula de ácido nucleico de la presente invención se puede generar usando datos de secuencias accesibles en la técnica y la información de secuencias proporcionada en la presente memoria. La secuencia de ADN que codifica cada uno de los polipéptidos de VHC incluidos en la proteína de fusión de la presente invención se puede aislar directamente de células que contienen VHC, ADNc y librerías genómicas, genomas víricos o cualquier vector de la técnica anterior que se sabe que los incluye, mediante técnicas de biología molecular o de PCR convencional, y, si es necesario, se puede modificar adicionalmente mediante técnicas de mutagénesis habituales (por ejemplo para optimizar la expresión en una célula hospedante particular, para cumplir con una secuencia específica del genotipo o del subtipo, como se describe anteriormente). La fusión de las secuencias que codifican cada uno de los polipéptidos de VHC se puede lograr, por ejemplo, mediante ligación en el marco ya sea directamente o a través de una secuencia que codifica un ligador peptídico. Como alternativa, la molécula de ácido nucleico de la invención también se puede generar mediante síntesis química en un procedimiento automatizado (por ejemplo ensamblado a partir de oligonucleótidos sintéticos solapantes como se describe, por ejemplo, en Edge, 1981, Nature 292, 756; Nambair *et al.*, 1984, Science 223, 1299; Jay *et al.*, 1984, J. Biol. Chem. 259, 6311).

También se proporciona mediante la presente invención un vector que comprende una o más copias de la molécula o moléculas de ácido nucleico de la invención. Por ejemplo, puede ser ventajoso incluir en el mismo vector moléculas de ácido nucleico que codifican proteínas de fusión que se originan a partir de diferentes genotipos o subtipos.

El término "vector", como se utiliza en la presente memoria, se refiere tanto a vectores de expresión como no de expresión, e incluye vectores víricos así como no víricos, incluyendo vectores extracromosómicos (por ejemplo plásmidos de múltiples copias) y vectores integrantes diseñados para ser incorporados en el cromosoma o cromosomas del hospedante. En el contexto de la invención son particularmente importantes los vectores para uso en terapia génica (es decir, que son capaces de suministrar la molécula de ácido nucleico a un organismo hospedante), así como vectores de expresión para uso en diversos sistemas de expresión.

Se puede usar una variedad de sistemas de hospedantes-vectores para expresar la proteína de fusión de la invención, incluyendo organismos procariotas tales como bacterias (por ejemplo *E. coli* o *Bacillus subtilis*) transformadas con vectores bacteriófagos, plasmídicos o cosmídicos que contienen la molécula de ácido nucleico que codifica la fusión; levaduras (por ejemplo *Saccharomyces cerevisiae*, *Saccharomyces pombe*, *Pichia pastoris*) transformadas con vectores de expresión de levadura que contienen la molécula de ácido nucleico que codifica la fusión; sistemas celulares de insectos (por ejemplo células Suficiente 9) infectados con vectores de expresión víricos (por ejemplo baculovirus) que contienen la molécula de ácido nucleico que codifica la fusión; sistemas de células vegetales infectados con vectores de expresión víricos (por ejemplo virus del mosaico de la coliflor CaMV; virus del mosaico del tabaco TMV) o transformados con vectores de expresión plasmídicos (por ejemplo plásmido Ti) que contienen la molécula de ácido nucleico que codifica la fusión; o sistemas de células de mamíferos (por ejemplo células cultivadas) transfectados o infectados con vectores de expresión plasmídicos o víricos que contienen la molécula de ácido nucleico que codifica la fusión.

- Los vectores adecuados para uso en sistemas procariotas incluyen sin limitación pBR322 (Gibco BRL), pUC (Gibco BRL), pBluescript (Stratagene), p Poly (Lathe *et al.*, 1987, Gene 57, 193-201), pTrc (Amann *et al.*, 1988, Gene 69, 301-315); pET 11d (Studier *et al.*, 1990, Gene Expression Technology: Methods in Enzymology 185, 60-89); pIN (Inouye *et al.*, 1985, Nucleic Acids Res. 13, 3101-3109; Van Heeke *et al.*, 1989, J. Biol. Chem. 264, 5503-5509); y vectores pGEX, en los que la molécula de ácido nucleico de la invención se puede expresar en fusión con glutatona S-transferasa (GST). El plásmido pGEX-2T (código de producto de Amersham Biosciences: 27-4801-01, número de acceso de GenBank U13850) es particularmente adecuado en el contexto de la invención.
- Los vectores adecuados para la expresión en levaduras (por ejemplo *S. cerevisiae*) incluyen, pero no se limitan a, pYepSecl (Baldari *et al.*, 1987, EMBO J. 6, 229-234), pMFa (Kujan *et al.*, 1982, Cell 30, 933-943), pJRY88 (Schultz *et al.*, 1987, Gene 54, 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.) y pTEF-MF (código de producto de Dualsystems Biotech: P03303).
- Los vectores adecuados para la expresión de células hospedantes de mamíferos pueden ser de origen vírico o no vírico (por ejemplo ADN plasmídico). Los vectores plasmídicos adecuados incluyen, sin limitación, pREP4, pCEP4 (Invitrogene), pCI (Promega), pCDM8 (Seed, 1987, Nature 329, 840) y pMT2PC (Kaufman *et al.*, 1987, EMBO J. 6, 187-195), pVAX y pgWiz (Gene Therapy System Inc; Himoudi *et al.*, 2002, J. Virol. 76, 12735-12746).
- Además, los sistemas de hospedantes-vectores usados en el contexto de la presente invención también pueden comprender uno o más elementos adicionales que permiten el mantenimiento, propagación o expresión de la molécula de ácido nucleico de la presente invención de la célula hospedante. Tales elementos adicionales incluyen sin limitación un gen o genes marcadores a fin de facilitar la identificación y aislamiento de las células hospedantes recombinantes (por ejemplo mediante complementación de una auxotrofia celular, o mediante resistencia a antibióticos), elementos estabilizantes (por ejemplo secuencia *cer* como se describe en Summers y Sherrat, 1984, Cell 36, 1097-1103 y el sistema DAP como se describe en el documento US 5.198.343), y elementos integradores (por ejemplo secuencias víricas LTR y transposones).
- Los genes marcadores adecuados para la expresión en células hospedantes procariotas incluyen genes de resistencia a tetraciclina y ampicilina. También, se pueden usar genes de resistencia para la expresión en células hospedantes de mamíferos, tales como dihidrofolato reductasa (*dhfr*), que confiere resistencia a metotrexato (Wigler *et al.*, 1980, Proc. Natl. Acad. Sci. USA 77, 3567; O'Hare *et al.*, 1981, Proc. Natl. Acad. Sci. USA 78, 1527); *gpt*, que confiere resistencia a ácido micofenólico (Mulligan y Berg, 1981, Proc. Natl. Acad. Sci. USA 78, 2072); *neo*, que confiere resistencia al aminoglucósido G-418 (Colberre-Garapin *et al.*, 1981, J. Mol. Biol. 150, 1); *zeo*, que confiere resistencia a zeomicina; *kana*, que confiere resistencia a kanamicina; e *hygro*, que confiere resistencia a higromicina (Santerre *et al.*, 1984, Gene 30, 147). Para la expresión en sistemas de levaduras, se pueden usar los genes *URA3* y *LEU2*, que proporcionan complementación de mutantes de levadura *ura3* o *leu2*. En particular, se puede usar ventajosamente el gen *URA3*, a partir del que se ha suprimido su promotor.
- También se puede utilizar en el contexto de la invención un número de sistemas de expresión a base de virus, derivados de una variedad de diferentes virus (por ejemplo retrovirus, adenovirus, AAV, poxvirus, virus del herpes, virus del sarampión, espumavirus y similares). Como se utiliza en la presente memoria, la expresión "vector vírico" engloba ADN vectorial así como partículas víricas generadas del mismo. Los vectores víricos pueden ser competentes en replicación, o se pueden inactivar genéticamente para ser defectuosos en la replicación o que tengan una replicación alterada. La expresión "competente en la replicación", como se utiliza en la presente memoria, engloba vectores víricos selectivos para la replicación y condicionalmente replicativos que se manipulan mediante ingeniería para replicarse mejor o selectivamente en células hospedantes específicas (por ejemplo células tumorales).
- Tales vectores incluyen, por ejemplo, vectores adenovíricos que tienen un número de ventajas bien documentadas para la transferencia génica o para la reproducción recombinante (para un repaso, véase "Adenoviral vectors for gene therapy", 2002, Ed D. Curiel and J. Douglas, Academic Press). Los vectores adenovíricos para uso según la presente invención pueden derivar de una variedad de fuentes humanas o animales. Se puede emplear cualquier serotipo desde los serotipos adenovíricos 1 a 51, con especial preferencia por los adenovirus humanos 2 (Ad2), 5 (Ad5), 6 (Ad6), 11 (Ad11), 24 (Ad24) y 35 (Ad35). Los adenovirus citados están disponibles de la American Type Culture Collection (ATCC, Rockville, Md.), y han sido el objeto de numerosas publicaciones que describen su secuencia, organización y métodos de producción, permitiendo al experto aplicarlos (véanse, por ejemplo, los documentos US 6.133.028; US 6.110.735; WO 02/40665; WO 00/50573; EP 1016711; Vogels *et al.*, 2003, J. Virol. 77, 8263-8271).
- El vector adenovírico de la presente descripción es competente en la replicación. Los ejemplos de tales vectores adenovíricos competentes en la replicación son bien conocidos en la técnica y están fácilmente disponibles para los expertos en la técnica (véanse, por ejemplo, Hernandez-Alcoceba *et al.*, 2000, Human Gene Ther. 11, 2009-2024; Nemunaitis *et al.*, 2001, Gene Ther. 8, 746-759; Alemany *et al.*, 2000, Nature Biotechnology 18, 723-727). Los vectores adenovíricos competentes para la replicación adecuados para uso en la invención se pueden manipular mediante ingeniería a partir de un genoma adenovírico de tipo salvaje mediante supresión en el dominio E1A CR2

para anular la unión al Rb (véase, por ejemplo, el documento WO 00/24408) y/o mediante sustitución de los promotores E1 y/o E4 nativos por promotores específicos del estado tisular, tumoral o celular (véanse, por ejemplo, los documentos US 5.998.205, WO 99/25860, US 5.698.443, WO 00/46355, WO 00/15820 y WO 01/36650).

5 El vector adenovirico de la invención tiene replicación defectuosa (véase, por ejemplo, el documento WO 94/28152; Lusky *et al.*, 1998, J. Virol 72, 2022-2032). Los vectores adenoviricos de replicación defectuosa preferidos son defectuosos en E1 (véanse, por ejemplo, los documentos 6.136.594 y US 6.013.638), extendiéndose una supresión E1 desde aproximadamente las posiciones 459 a 3328, o desde aproximadamente las posiciones 459 a 3510 (con referencia a la secuencia del adenovirus humano tipo 5 descrito en el GeneBank con el número de acceso M 73260 y en Chroboczek *et al.*, 1992, Virol. 186, 280-285). La capacidad de clonación se puede mejorar adicionalmente
10 suprimiendo una porción o porciones adicionales del genoma adenovirico (todo o parte de la región E3 no esencial o de otras regiones E2, E4 esenciales).

15 La molécula de ácido nucleico de la presente invención se puede insertar en cualquier localización del genoma adenovirico. Preferentemente, se inserta en sustitución de la región E1. Se puede colocar en orientación sentido o antisentido, con relación a la dirección transcripcional natural de la región en cuestión.

Otros vectores víricos adecuados derivan de poxvirus (véase, por ejemplo, Cox en "Viruses in Human Gene Therapy" Ed J. M. Hos, Carolina Academic Press). En el contexto de la presente invención, un vector poxvirico se puede obtener a partir de cualquier miembro de poxviridae, en particular la viruela del canario, la difteria aviar y el virus de la vacuna, siendo este último preferido. Los virus de la vacuna adecuados incluyen, sin limitación, la cepa de Copenhague (Goebel *et al.*, 1990, Virol. 179, 247-266 y 517-563; Johnson *et al.*, 1993, Virol. 196, 381-401), la cepa Wyeth y la cepa Ankara modificada (MVA) (Antoine *et al.*, 1998, Virol. 244, 365-396). Las condiciones generales para construir poxvirus recombinantes son bien conocidas en la técnica (véanse, por ejemplo, el documento EP 206 920; Mayr *et al.*, 1975, Infection 3, 6-14; Sutter y Moss, 1992, Proc. Natl. Acad. Sci. USA 89, 10847-10851; documento US 6.440.422). La molécula de ácido nucleico de la presente invención se inserta preferentemente en el genoma poxvirico en un locus no esencial. El gen de timidina cinasa es particularmente apropiado para la inserción en vectores de la vacuna de Copenhague (Hruby *et al.*, 1983, Proc. Natl. Acad. Sci USA 80, 3411-3415; Weir *et al.*, 1983, J. Virol. 46, 530-537), y la supresión II o III para la inserción en el vector de MVA (Meyer *et al.*, 1991, J. Gen. Virol. 72, 1031-1038; Sutter *et al.*, 1994, Vaccine 12, 1032-1040).
20
25
30

Los vectores de la invención comprenden la molécula de ácido nucleico de la invención en una forma adecuada para su expresión en una célula u organismo hospedante, lo que significa que la molécula de ácido nucleico se coloca bajo el control de una o más secuencias reguladoras, apropiadas para el vector y/o la célula hospedante. Como se utiliza en la presente memoria, la expresión "secuencia reguladora" se refiere a cualquier secuencia que permite, contribuye o modula la expresión de una molécula de ácido nucleico en una célula hospedante dada, incluyendo la replicación, duplicación, transducción, ajuste, traducción, estabilidad y/o transporte del ácido nucleico o uno de sus derivados (es decir, ARNm) en la célula hospedante. Los expertos en la técnica apreciarán que la elección de las secuencias reguladoras puede depender de factores tales como la célula hospedante, el nivel de expresión deseado, etc.
35
40

El promotor es de especial importancia, y los promotores adecuados útiles en el contexto de la presente invención incluyen promotores constitutivos que dirigen la expresión de la molécula de ácido nucleico en muchos tipos de célula hospedante y aquellos que dirigen la expresión de la secuencia nucleotídica sólo en ciertas células hospedantes (por ejemplo, secuencias reguladoras específicas del hígado) o en respuesta a sucesos específicos o factores exógenos (por ejemplo por temperatura, aditivo de nutrientes, hormona u otro ligando).
45

Los promotores adecuados para la expresión en una célula hospedante de *in E. coli* incluyen, pero no se limitan a, el promotor lambda pL de bacteriófago, los promotores lac, TRP y pTAC inducible por IPTG. Los promotores adecuados para la expresión en levaduras incluyen los promotores TF (Mumberg *et al.*, 1995, Gene 156, 119-122), PGK (Hitzeman *et al.*, 1983, Science 219, 620-625), MF alfa (Inokuchi *et al.*, 1987, Mol. Cell. Biol. 7, 3185-3193), CYC-1 (Guarente *et al.*, 1981, Proc. Natl. Acad. Sci. USA 78, 2199), GAL-1, GAL4, GAL10, PHO5, gliceraldehído-3-fosfato deshidrogenasa (GAP o GAPDH), y alcohol deshidrogenasa (ADH) (Denis *et al.*, 1983, J. Biol. Chem. 25, 1165). Para la expresión en vectores poxviricos, se puede usar el promotor de la vacuna 7.5K, H5R, TK, p28, p11 o K1L, los promotores sintéticos descritos en Chakrabarti *et al.* (1997, Biotechniques 23, 1094-1097), Hammond *et al.* (1997, J. Virological Methods 66, 135-138) y Kumar y Boyle (1990, Virology 179, 151-158), así como los promotores quiméricos temprano/tardío. Los promotores adecuados para la expresión constitutiva en células de mamíferos incluyen el promotor temprano inmediato de citomegalovirus (CMV) (Boshart *et al.*, 1985, Cell 41, 521-530), el promotor tardío principal de adenovirus, el promotor de fosfoglicerocinasa (PGK) (Adra *et al.*, 1987, Gene 60, 65-74), y el promotor de timidina cinasa (TK) del virus del herpes simple (HSV)-1. Los promotores eucariotas inducibles regulados por compuestos suministrados exógenamente incluyen, sin limitación, el promotor de metalotioneína (MT) inducible por zinc (Mc Ivor *et al.*, 1987, Mol. Cell Biol. 7, 838-848), el promotor del virus de tumor mamario de ratón (MMTV) inducible por dexametasona (Dex), el sistema del promotor de T7 polimerasa (documento WO 98/10088), el promotor de insecto de ecdisona (No *et al.*, 1996, Proc. Natl. Acad. Sci. USA 93, 3346-3351), el promotor represible mediante tetraciclina (Gossen *et al.*, 1992, Proc. Natl. Acad. Sci. USA 89, 5547-5551), el promotor inducible por tetraciclina (Kim *et al.*, 1995, J. Virol. 69, 2565-2573), el promotor inducible por RU486 (Wang *et al.*, 1997, Nat.
50
55
60
65

Biotech. 15, 239-243 y Wang *et al.*, 1997, Gene Ther. 4, 432-441), y el promotor inducible por rapamicina (Magari *et al.*, 1997, J. Clin. Invest. 100, 2865-2872).

También se pueden usar promotores específicos de tejidos, especialmente aquellos que permiten seleccionar como diana las células infectadas por VHC cuando se desea beneficio terapéutico. Los promotores adecuados incluyen promotores específicos del hígado, tales como HMG-CoA reductasa (Luskey, 1987, Mol. Cell. Biol. 7,1881-1893); el elemento 1 regulador de esterol (SRE-1; Smith *et al.*, 1990, J. Biol. Chem. 265, 2306-2310); albúmina (Pinkert *et al.*, 1987, Genes Dev. 1, 268-277); fosfoenol piruvato carboxi cinasa (PEPCK) (Eisenberger *et al.*, 1992, Mol. Cell Biol. 12,1396-1403); proteína reactiva C humana (CRP) (Li *et al.*, 1990, J. Biol. Chem. 265, 4136-4142); glucocinasa humana (Tanizawa *et al.*, 1992, Mol. Endocrinology 6, 1070-1081); colesterol 7-alfa hidrilasa (CYP-7) (Lee *et al.*, 1994, J. Biol. Chem. 269, 14681-14689); alfa-1 antitripsina (Ciliberto *et al.*, 1985, Cell 41, 531-540); proteína de unión al factor de crecimiento similar a insulina (IGFBP-1) (Babajko *et al.*, 1993, Biochem Biophys. Res. Comm. 196,480-486); transferrina humana (Mendelzon *et al.*, 1990, Nucleic Acids Res. 18, 5717-5721); colágeno tipo I (Houglum *et al.*, 1994, J. Clin. Invest. 94, 808-814) y los genes FIX (documento US 5.814.716).

Los promotores adicionales adecuados para uso en esta invención se pueden obtener a partir de genes que son expresados preferentemente en células tumorales proliferativas, tales como los promotores del gen de alfa-fetoproteína sobreexpresado en cánceres hepáticos (Kanai *et al.*, 1997, Cancer Res. 57, 461-465), la transcriptasa inversa de telomerasa (TERT) (documentos WO99/27113, WO 02/053760 y Horikawa *et al.*, 1999, Cancer Res. 59, 826), y elemento sensible a hipoxia (HRE).

Los expertos en la materia apreciarán que los elementos reguladores que controlan la expresión de la molécula de ácido nucleico de la invención pueden comprender además elementos adicionales para el inicio, regulación y/o terminación apropiados de la transcripción (por ejemplo secuencias de terminación de la transcripción poliA), transporte de ARNm (por ejemplo secuencias de señal de localización nuclear), procesamiento (por ejemplo señales de ajuste), y estabilidad (por ejemplo intrones y secuencias de 5' y 3' no codificantes), traducción (por ejemplo señal peptídica, propéptido, secuencias líder tripartitas, sitios de unión a ribosoma, secuencias de Shine-Dalgarno, etc.) en la célula hospedante u organismo y las etapas de purificación (por ejemplo un marcador).

Por ejemplo, se puede usar un péptido señal, eventualmente en combinación con un propéptido, para facilitar la secreción de la proteína de fusión en el medio de cultivo. El péptido señal se inserta típicamente en el término N de la proteína de fusión inmediatamente después del iniciador Met, y, si es necesario, el propéptido se inserta en dirección 3' del péptido señal. La elección de la señal y/o de los propéptidos es amplia y es accesible a las personas expertas en la técnica. Los ejemplos de señal/propéptidos apropiados para la presente invención incluyen, pero no se limitan a, las secuencias de péptido señal del factor de coincidencia (MF) alfa (Kurjan y Herskowitz, 1982, Cell 30, 933-934 y el documento US 5.879.926); invertasa (documento WO84/01153); PHOS (documento DK 3614/83); YAP3 (proteasa aspártica 3 de levadura; documento WO95/02059); y BAR1 (documento WO87/02670). Durante la entrada en el ER, el péptido señal se separa por escisión del polipéptido precursor en un sitio de procesamiento. El sitio de procesamiento puede comprender cualquier secuencia peptídica que es reconocida por una enzima proteolítica de la célula hospedante. Los ejemplos de sitios de procesamiento preferidos incluyen, pero no se limitan a, cualquier combinación de los dos restos básicos Lys y Arg (con preferencia especial por Lys-Arg) que son escindidos por la endopeptidasa codificada por el gen KEX2 de *Saccharomyces cerevisiae* (véase Fuller *et al.*, 1986, Microbiology, 273-278) o la proteasa equivalente de otra especie de levadura (véase Julius *et al.*, 1983, Cell 32, 839-852).

Para facilitar la purificación de la proteína de fusión recombinante a partir de la célula hospedante o del sobrenadante de cultivo, se puede usar un marcador peptídico (por ejemplo una secuencia peptídica corta capaz de ser reconocida por antisueros disponibles). Preferentemente, el marcador se inserta en el término C de la proteína de fusión. Los marcadores ejemplares incluyen, sin limitación, el marcador His (por ejemplo, compuesto de 6 o más restos de histidina), secuencia peptídica de glutatona-S-transferasa (GST) de *S mansoni*, proteína de unión a maltosa (MPB) de *E. coli*, fosfatasa alcalina humana, el octapéptido FLAG, y el marcador e (documento US 6.686.152).

También se considera un plásmido lanzadera para la expresión en levadura, que comprende la molécula de ácido nucleico de la invención (por ejemplo la secuencia nucleotídica de SEC ID n°: 13 y 14) colocada bajo el control del promotor TEF y una secuencia terminadora transcripcional apropiada (por ejemplo terminador del citocromo *cyc*) que comprende además orígenes de replicación para células hospedantes de *E. coli* y levadura (por ejemplo ORI y 2 μ respectivamente), marcadores de selección adecuados para la expresión en *E. coli* (por ejemplo gen de resistencia a ampicilina) y levadura (por ejemplo auxotrofia URA3 y leu2-3).

Además, se describe un plásmido pGEX adecuado para la expresión en *E. coli* que comprende la molécula de ácido nucleico de la invención (por ejemplo la secuencia nucleotídica de SEC ID n°: 15 y 16) colocada bajo el control del promotor pTAC inducible por IPTG, y una secuencia terminadora transcripcional apropiada que comprende además un origen adecuado de replicación para *E. coli* (por ejemplo ORI), y un marcador de selección adecuado (por ejemplo gen de resistencia a ampicilina).

Todavía otra forma de realización preferida se refiere a un vector adenovirico defectuoso en la replicación para la expresión en células de mamíferos, que comprende la molécula de ácido nucleico de la invención (por ejemplo la secuencia nucleotídica de SEC ID n°: 11 y 12) colocada bajo elementos reguladores apropiados. Preferentemente, el vector adenovirico defectuoso en la replicación es un genoma Ad5 al que se le ha suprimido E1 y E3, extendiéndose la supresión E1 desde aproximadamente el nucleótido 459 hasta aproximadamente el nucleótido 3511, y la supresión E3 desde aproximadamente el nucleótido 28592 hasta aproximadamente el nucleótido 30470. Un vector adenovirico preferido comprende las secuencias Ad5 desde aproximadamente el nucleótido 1 hasta aproximadamente el nucleótido 458, conteniendo el casete de expresión, desde 5' hasta 3', el potenciador/promotor temprano inmediato de CMV, un intrón de β -globina/IgG humano quimérico (como se encuentra en el vector pCI disponible en Promega), la secuencia que codifica cualquiera de las dos formas poliproteicas (SEC ID n°: 11 o 12) y la señal de poliadenilación de SV40 seguido de una secuencia de Ad5 desde aproximadamente el nucleótido 3512 hasta aproximadamente el nucleótido 28592, y desde aproximadamente el nucleótido 30470 hasta aproximadamente el nucleótido 35935.

Si es necesario, el vector de la invención puede comprender además uno o más transgenes, por ejemplo un gen de interés a expresar junto con la molécula de ácido nucleico de la invención en una célula hospedante u organismo. De forma deseable, la expresión del transgén tiene una actividad terapéutica o protectora frente a una enfermedad o afección asociada a VHC. El transgén adecuado incluye, sin limitación, genes que codifican citocinas (por ejemplo IL-2, IL-7, IL-15, IL-18, IL-21, IFN γ), genes que codifican toxinas (por ejemplo ricina, toxina de la difteria, toxina del cólera), genes suicidas, y especialmente los genes que codifican K HSV-1 (Caruso *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90, 7024-7028) que se va a usar en combinación con el profármaco aciclovir o ganciclovir, citoxina desaminasa (CDasa), uracilo fosforribosilo transferasa (UPRTasa) que se va a usar en combinación con el profármaco 5-fluorocitosina (5-FC). El producto génico *FCU-1* (descrito en el documento WO 99/54481) es particularmente útil en este contexto. Si se usa un transgén, se coloca bajo el control de elementos reguladores apropiados, y se puede insertar en cualquier localización del vector de la invención o en un vector independiente que se usa en combinación con el vector de la invención.

La inserción de la molécula de ácido nucleico de la invención en una estructura principal vectorial se puede llevar a cabo mediante biología molecular habitual, por ejemplo como se describe en Sambrook *et al.* (2001, Molecular Cloning-A Laboratory Manual, Cold Spring Harbor Laboratory). La inserción en el vector adenovirico o en un vector poxvirico se puede llevar a cabo a través de recombinación homóloga como se describe respectivamente en Chartier *et al.* (1996, J. Virol. 70, 4805-4810) y Paul *et al.* (2002, Cancer gene Ther. 9, 470-477).

La presente invención también engloba vectores (por ejemplo ADN plasmídico) complejados a lípidos o polímeros para formar estructuras en partículas tales como liposomas, lipocomplejos o nanopartículas. Tales tecnologías están disponibles en la técnica (véase, por ejemplo, Arangoa *et al.*, 2003, Gene Ther. 10: 5-14; Eliaz *et al.*, 2002, Gene Ther. 9, 1230-1237 y Betageri *et al.*, 1993, "Liposome drug delivery systems", Technomic Publishing Company, Inc).

Se describen partículas víricas infecciosas que comprenden las moléculas de ácido nucleico o vectores de la presente invención descritos anteriormente.

Típicamente, tales partículas víricas se producen mediante un procedimiento que comprende las etapas de:

- (a) introducir el vector vírico de la invención en una extirpe celular adecuada,
- (b) cultivar dicha extirpe celular en condiciones adecuadas para permitir la producción de dicha partícula vírica infecciosa,
- (c) recuperar la partícula vírica infecciosa producida a partir del cultivo de dicha extirpe celular, y
- (d) opcionalmente purificar dicha partícula vírica infecciosa recuperada.

Cuando el vector vírico es defectuoso, las partículas infecciosas se producen habitualmente en una extirpe celular de complementación o vía el uso de un virus auxiliar, que suministra *en trans* los genes víricos no funcionales. Por ejemplo, las extirpes celulares adecuadas para complementar vectores adenoviricos a los que se les ha suprimido E1 incluyen las células 293 (Graham *et al.*, 1997, J. Gen. Virol. 36, 59-72) así como las células PER-C6 (Fallaux *et al.*, 1998, Human Gene Ther. 9, 1909-1917). Las células apropiadas para propagar vectores poxviricos son células aviares, y los más preferible fibroblastos de embriones de pollo primarios (CEF) preparados a partir de embriones de pollo obtenidos de huevos fertilizados.

Las partículas víricas infecciosas se pueden recuperar del sobrenadante de cultivo o a partir de las células tras la lisis. Se pueden purificar además según técnicas estándar (cromatografía, ultracentrifugación en un gradiente de cloruro de cesio como se describe, por ejemplo, en los documentos WO96/27677, WO98/00524, WO98/22588, WO98/26048, WO00/40702, EP1016700 y WO00/50573).

La presente invención también engloba vectores o partículas víricas que se han modificado para permitir la selección preferente a una célula diana particular (véase, por ejemplo Wickam *et al.*, 1997, J. Virol. 71, 8221-8229; Arnberg *et al.*, 1997, Virol. 227, 239-244; Michael *et al.*, 1995, Gene Therapy 2, 660-668; documento WO94/10323; documento WO02/96939 y documento EP 1 146 125). Un rasgo característico de los vectores seleccionados como dianas y partículas víricas de la invención es la presencia en su superficie de un ligando capaz de reconocer y unirse a un componente celular y expuesto a la superficie, tal como un marcador específico de las células (por ejemplo una célula infectada con VHC), un marcador específico de tejidos (por ejemplo un marcador específico de hígado), así como un antígeno vírico (por ejemplo HVC). Los ejemplos de ligandos adecuados incluyen anticuerpos o sus fragmentos dirigidos contra un dominio antigénico de VHC. La selección de las dianas se puede llevar a cabo insertando genéticamente el ligando en un polipéptido presente sobre la superficie del virus (por ejemplo fibra adenovírica, penton, pIX o producto génico p14 vacunal).

La invención también se refiere a células hospedantes que comprenden las moléculas de ácido nucleico, vectores o partículas víricas infecciosas de la invención descritas en la presente memoria. En el contexto de la presente invención, la expresión "célula hospedante" debería ser entendida ampliamente sin ninguna limitación referida en particular a la organización en tejido, órgano, o células aisladas. Tales células pueden ser de un único tipo de células o un grupo de tipos diferentes de células, y englobar extirpes celulares cultivadas, células primarias y células proliferativas.

En el contexto de la invención, las células hospedantes incluyen células procariotas, células eucariotas inferiores tales como levadura, y otras células eucariotas tales como células de insectos, células vegetales y de mamíferos (por ejemplo humanas o no humanas). Las células hospedantes de *E. coli* preferidas incluyen sin limitación BL21 (disponibles en Amersham Biosciences). Las células hospedantes de levadura preferidas incluyen, sin limitación, *S. cerevisiae*, *S. pombe*, *Pichia pastoris*, con preferencia especial por células de levadura que expresan KEX-2, tales como la cepa TGY47.1 (o su GY73.4 conversor Leu⁺ isogénico) descrito en el documento EP 607 080, y aquellas descritas en los documentos US 5.879.926, US 5.162.208 y US 6.103.515. Las células hospedantes de mamíferos preferidas, sin limitación, células BHK (riñón de hámster neonato), CV-1 (extirpe celular de riñón de mono africano), COS (por ejemplo COS-7), células de ovario de hámster chino (CHO), células NIH/3T3 de ratón, células HeLa, y células Vero. Las células hospedantes también engloban células de complementación capaces de complementar al menos una función defectuosa de un vector defectuoso en la replicación de la invención (por ejemplo vector adenovírico), tal como células 293 y PERC.6.

La célula hospedante se puede encapsular adicionalmente. La tecnología de encapsulamiento celular se ha descrito previamente (Tresco *et al.*, 1992, ASAIO J. 38, 17-23; Aebischer *et al.*, 1996, Human Gene Ther. 7, 851-860).

Todavía un aspecto adicional de la presente invención es un método para producir recombinantemente la proteína de fusión, que emplea los vectores, partículas víricas infecciosas y/o células hospedantes de la invención. El método para producir la proteína de fusión comprende (a) introducir un vector o una partícula vírica infecciosa de la invención en una célula hospedante adecuada para producir una célula hospedante transfectada o infectada, (b) cultivar *in vitro* dicha célula hospedante transfectada o infectada bajo condiciones adecuadas para el crecimiento de la célula hospedante, (c) recuperar la proteína de fusión del cultivo celular, y (d) opcionalmente purificar la proteína de fusión recuperada.

Se espera que aquellos expertos en la materia sepan los numerosos sistemas de expresión disponibles para producir las proteínas de fusión de la invención en células hospedantes apropiadas y los métodos para producir un vector o una partícula vírica infecciosa en una célula hospedante. Tales métodos incluyen, pero no se limitan a, microinyección (Capechi *et al.*, 1980, Cell 22, 479-488), transfección mediada por CaPO₄ (Chen y Okayama, 1987, Mol. Cell Biol. 7, 2745-2752), transfección mediada por DEAE-dextrano, electroporación (Chu *et al.*, 1987, Nucleic Acid Res. 15,1311-1326), fusión de lipofección/liposoma (Felgner *et al.*, 1987, Proc. Natl. Acad. Sci. USA84, 7413-7417), bombardeo de partículas (Yang *et al.*, 1990, Proc. Natl. Acad. Sci. USA 87, 9568-9572), pistolas génicas, transducción, infección vírica, así como la administración directa en un organismo hospedante vía diversos medios.

Los vectores de la invención se pueden usar en asociación con reactivos de transfección, a fin de facilitar la introducción del vector en la célula hospedante, tal como polímeros policationicos (por ejemplo citosano, polimetacrilato, PEI, etc.) y lípidos catiónicos (por ejemplo DC-Chol/DOPE, lipofectina transfectam disponible ahora de Promega). Además, como se explica en la presente memoria, se pueden usar tecnologías de ADN recombinante para mejorar la expresión de la molécula de ácido nucleico en la célula hospedante, por ejemplo usando vectores de número elevado de copias, sustituyendo o modificando una o más secuencias reguladoras transcripcionales (por ejemplo promotor, potenciador y similar), optimizando el uso de codones de la molécula de ácido nucleico que codifica la fusión a la célula hospedante, y suprimiendo secuencias negativas que pueden desestabilizar el transcrito.

Las células hospedantes de la presente invención se pueden cultivar en biorreactores de fermentación convencionales, matraces, y cápsulas de petri. El cultivo se puede llevar a cabo a una temperatura, pH y contenido de oxígeno apropiados para una célula hospedante dada. No se harán en la presente memoria intentos para describir con detalle los diversos métodos conocidos para la producción de proteínas en células procariotas y

eucariotas. La producción de la proteína de fusión puede ser periplásmica, intracelular o segregada fuera de la célula hospedante (por ejemplo en el medio de cultivo).

5 Cuando la proteína de fusión no se segrega fuera de la célula hospedante traductora, o cuando no se segrega completamente, se puede recuperar mediante procedimientos de lisis estándar, incluyendo congelación descongela- ción, ultrasonidos, destrucción mecánica, uso de agentes de lisis, y similares. Si se segrega, se puede recuperar directamente a partir del medio de cultivo.

10 La proteína de fusión se puede purificar entonces por métodos de purificación bien conocidos, incluyendo precipitación con sulfato de amonio, extracción ácida, electroforesis en gel, filtración y métodos cromatográficos (por ejemplo cromatografía de fase inversa, exclusión molecular, intercambio iónico, afinidad, fosfocelulosa, interacción hidrófoba, hidroxilapatita, o cromatografía de líquidos de altas prestaciones). Las condiciones y tecnología usadas para purificar una proteína de fusión particular de la invención dependerán de factores tales como la carga neta, el peso molecular, la hidrofobia, la hidrofilia, y serán manifiestas para aquello que tienen pericia en la técnica. Además, el nivel de purificación dependerá del uso pretendido. También se entiende que, dependiendo de la célula hospedante productora, las proteínas de fusión pueden tener diversos patrones de glucosilación, o puede estar no glucosiladas (por ejemplo cuando se producen en bacterias).

20 En otro aspecto, esta invención proporciona una composición que comprende la proteína de fusión, la molécula de ácido nucleico, el vector, la partícula vírica infecciosa, la célula hospedante de la invención (también denominada en la presente memoria "agente activo") o cualquier combinación de los mismos (por ejemplo combinación de proteínas de fusión o vectores/partículas víricas que codifican proteínas de fusión de diferentes genotipos o subtipos). Preferentemente, la composición es una composición farmacéutica que comprende además de una cantidad terapéuticamente eficaz del agente o agentes activos, un vehículo farmacéuticamente aceptable. Como se utiliza en la presente memoria, una "cantidad terapéuticamente eficaz" es una dosis suficiente para el alivio de uno o más síntomas asociados normalmente con la enfermedad o afección que se desea tratar. Cuando está implicado el uso profiláctico, esta expresión significa una dosis suficiente para prevenir o retrasar el establecimiento de una enfermedad o afección. Por ejemplo, una cantidad terapéuticamente eficaz para inducir una respuesta inmunitaria sería aquella cantidad necesaria para provocar la activación del sistema inmunitario (por ejemplo que da como resultado el desarrollo de una respuesta anti-VHC).

30 Como se utiliza en la presente memoria, un "vehículo farmacéuticamente aceptable" pretende incluir todos y cualesquiera vehículos, disolventes, diluyentes, excipientes, adyuvantes, medios de dispersión, revestimientos, agentes antibacterianos y antifúngicos, y agentes que retrasan la absorción y similares, compatibles con la administración farmacéutica.

40 De forma adecuada, la composición farmacéutica de la invención comprende un diluyente apropiado para uso humano o animal. Es preferentemente isotónica, hipotónica o débilmente hipertónica, y tiene una fuerza iónica relativamente baja. Los ejemplos representativos incluyen agua estéril, disolución salina fisiológica (por ejemplo cloruro de sodio), disolución de Ringer, glucosa, trealosa o disoluciones de sacarosa, disolución de Hank, u otras disoluciones salinas balanceadas fisiológicamente acuosas (véase, por ejemplo, la edición más actual de Remington: The Science and Practice of Pharmacy, A. Gennaro, Lippincott, Williams & Wilkins). La composición de la invención está tamponada adecuadamente a fin de que sea apropiada para uso humano a un pH fisiológico o ligeramente básico (por ejemplo entre alrededor de pH 7 y alrededor de pH 9). Los tampones adecuados incluyen, sin limitación, tampón de fosfato (por ejemplo PBS), tampón de bicarbonato y/o tampón Tris.

50 La composición también contiene otros excipientes farmacéuticamente aceptables para proporcionar propiedades farmacéuticas o farmacodinámicas deseables, incluyendo, por ejemplo, modificar o mantener el pH, osmolaridad, viscosidad, claridad, color, esterilidad, estabilidad, velocidad de disolución de la formulación, modificar o mantener la liberación o absorción en el organismo humano o animal, promover el transporte a través de la barrera de la sangre o la penetración en un órgano particular (por ejemplo el hígado). Los excipientes adecuados incluyen aminoácidos.

55 Los vehículos farmacéuticamente aceptables incluidos en la composición de la invención deben permitir también conservar su estabilidad en las condiciones de fabricación y almacenamiento a largo plazo (es decir, al menos un mes) a temperaturas de congelación (por ejemplo -70°C, -20°C), temperaturas refrigeradas (por ejemplo 4°C) o temperatura ambiente. A este respecto, las formulaciones que están particularmente adaptadas a la composición de la invención incluyen:

- 60 o 1M de sacarosa, 150 mM de NaCl, 1 mM de MgCl₂, 54 mg/l de Tween 80, 10 mM de Tris pH 8,5 (especialmente cuando el agente activo es un vector adenovírico) y
- o 10 mg/ml de manitol, 1 mg/ml de HSA, 20 mM de Tris, pH 7,2, y 150 mM de NaCl
- o disolución salina fisiológica.

65

Además, la composición de la invención puede comprender uno o más adyuvantes adecuados para la aplicación sistémica o mucosal en seres humanos. Preferentemente, el adyuvante es capaz de estimular la inmunidad a la composición de la invención, especialmente una inmunidad mediada por células T, por ejemplo a través de receptores de tipo toll (TLR), tales como TLR-7, TLR-8 y TLR-9. Los ejemplos representativos de adyuvantes útiles incluyen, sin limitación, alum, emulsión de aceite mineral tal como completa de Freund e incompleta i(IFA), lipopolisacárido o un derivado del mismo (Ribi *et al.*, 1986, *Immunology and Immunopharmacology of Bacterial Endotoxins*, Plenum Publ. Corp., NY, p407-419), saponinas tales como QS21 (Sumino *et al.*, 1998, *J. Virol.* 72, 4931-4939; documento WO 98/56415); compuestos de imidazoquinolina tales como Imiquimod (Suader, 2000, *J. Am Acad Dermatol.* 43, S6-S11) y un compuesto relacionado S-27609 (Smorlesi, 2005, *Gene Ther.* 12,1324-1332), oligodesoxinucleótidos de citosina fosfato guanosa tales CpG (Chu *et al.*, 1997, *J. Exp. Med.* 186:1623; Tritel *et al.*, 2003, *J. Immunol.* 171: 2358-2547) y péptidos catiónicos tales como IC-31 (Kritsch *et al.*, 2005, *J. Chromatogr Anal. Technol Biomed Life Sci* 822, 263-270).

La composición de la presente invención se puede administrar mediante una variedad de modos de administración, incluyendo la administración sistémica, tópica y localizada. La inyección se puede llevar cabo por cualquier medio, por ejemplo mediante inyección subcutánea, intradérmica, intramuscular, intraperitoneal, intratumoral, intravascular, intraarterial o mediante inyección directa en una arteria (por ejemplo mediante infusión en la arteria hepática) o una vena que alimenta al hígado (por ejemplo inyección en la vena porta). Las inyecciones se pueden realizar con jeringuillas y agujas convencionales, o cualesquiera otros dispositivos apropiados disponibles en la técnica. Como alternativa, la composición de la presente invención se puede administrar vía una ruta mucosal, tal como la vía oral/alimentaria, nasal, intratraqueal, intrapulmonar, intravaginal o intrarrectal. La administración en el aparato respiratorio se puede realizar a través de nebulización o en aerosol de composiciones en gotitas, de pulverización, o de polvo seco, usando un recipiente a presión (por ejemplo con un propelente adecuado tal como diclorodifluorometano, propano, nitrógeno y similar), o en un dispensador que no está a presión. La administración tópica también se puede llevar a cabo usando medios transdérmicos (por ejemplo parche y similar). En el contexto de la invención, las administraciones intramuscular y subcutánea constituyen las vías preferidas.

La composición de la invención puede estar en diversas formas, por ejemplo sólida, líquida o congelada. Las composiciones sólidas (por ejemplo en polvo seco o liofilizadas) se pueden obtener mediante un procedimiento que implica el secado a vacío y la liofilización. Para la administración mucosal, las composiciones se pueden formular como cápsulas y gránulos gastroresistentes para la administración oral, supositorios para administración rectal o vaginal, eventualmente combinación con potenciadores de la absorción útiles para incrementar el tamaño de poros de las membranas mucosales. Tales potenciadores de la absorción son típicamente sustancias que tienen similitudes estructurales a los dominios fosfolípidicos de las membranas mucosales, tales como desoxicolato de sodio, glucocolato de sodio, dimetil-beta-ciclodextrina, lauroil-1-lisofosfatidilcolina).

La dosis apropiada se puede adaptar en función de diversos parámetros, en particular el modo de administración; la composición empleada; la edad, salud y peso del organismo hospedante; la naturaleza y grado de síntomas; el tipo de tratamiento concurrente; la frecuencia de tratamiento; y/o la necesidad de prevención o terapia. El médico realiza habitualmente otro refinamiento de los cálculos necesarios para determinar la dosificación apropiada para el tratamiento, a la luz de las circunstancias relevantes. Para una guía general, la dosis adecuada para una composición que comprende virus (por ejemplo partículas adenovíricas) varía desde alrededor de 10^5 a 10^{13} ui (unidades infecciosas), deseablemente de alrededor de 10^7 a 10^{12} ui. Una composición a base de plásmidos vectoriales se puede administrar en dosis de entre 10 μ g y 20 mg, ventajosamente entre 100 μ g y 2 mg. Una composición proteica se puede administrar en una o más dosis de entre 10 ng y 20 mg, con una especial preferencia por una dosis de alrededor de 0,1 μ g a alrededor de 2 mg de la proteína terapéutica por kg de peso corporal. La administración puede tener lugar en una única dosis o una dosis repetida una o varias veces después de un cierto intervalo de tiempo.

La composición farmacéutica de la invención se puede emplear en métodos para tratar una variedad de enfermedades y estados patológicos, especialmente aquellos asociados con una infección por VHC. Como se utiliza en la presente memoria, el término "tratamiento" o "tratar" engloba la profilaxis y/o terapia. Es especialmente útil para tratar infección persistente por VHC y cáncer hepático en pacientes infectados con VHC. El término "cáncer" engloba cualesquiera afecciones cancerosas, incluyendo tumores difusos o localizados, metástasis, pólipos cancerosos así como lesiones preneoplásicas (por ejemplo cirrosis). Preferentemente, con la introducción en un organismo hospedante según las modalidades descritas en la presente memoria, la composición de la invención proporciona un beneficio terapéutico al hospedante tratado. La expresión "organismo hospedante" está destinada a englobar a cualquier animal, particularmente mamífero, tal como cualquier sujeto murino, rata, bovino, porcino, canino, felino, equino, simio o humano, por ejemplo un ser humano infectado con VHC. El beneficio terapéutico se puede evidenciar de muchas maneras, por ejemplo una disminución de la viremia de VHC detectada en sangre, plasma o sueros del individuo infectado en comparación con antes del tratamiento, y/o mediante la detección de una respuesta inmunitaria anti-VHC (por ejemplo producción de anticuerpos anti-VHC y/o inmunidad mediada por células T) o mediante el retraso de los síntomas asociados con una infección por VHC (por ejemplo retraso en el desarrollo de cirrosis hepática o cáncer), o mediante una disminución de afecciones de inflamación/esteatosis/fibrosis hepáticas, típicamente asociadas con infección por VHC, o mediante una respuesta mejorada del individuo a terapias convencionales.

En consecuencia, la presente invención también engloba el uso de la proteína de fusión, molécula de ácido nucleico, vector, partícula vírica infecciosa, célula hospedante o composición de la invención para la preparación de un fármaco destinado a tratar o prevenir infecciones por VHC, enfermedades y estados patológicos asociados a VHC, según las modalidades descritas anteriormente.

En otro aspecto, la presente invención se refiere a la proteína de fusión, molécula de ácido nucleico, vector, partícula vírica infecciosa, célula hospedante o composición de la invención para su uso para tratar o prevenir infecciones por VHC, enfermedades asociadas con VHC y estados patológicos, según las modalidades descritas anteriormente.

Si se desea, el uso de la invención se puede llevar a cabo conjuntamente con una o más modalidades terapéuticas convencionales (por ejemplo radiación, quimioterapia y/o cirugía). El uso de enfoques terapéuticos múltiples proporciona al paciente con una intervención más amplia. En una realización, el método de la invención se puede anteceder o puede estar seguido de una intervención quirúrgica. En otra realización, puede estar precedido o seguido por radioterapia (por ejemplo radiación gamma). Los expertos en la técnica pueden formular fácilmente protocolos apropiados de terapia de radiación y parámetros que se pueden usar (véanse, por ejemplo, Pérez y Brady, 1992, Principles and Practice of Radiation Oncology, segunda edición JB Lippincott Co; usando adaptaciones y modificaciones apropiadas como será fácilmente manifiesto para los expertos en el campo.

En todavía otra forma de realización, el uso de la composición de la invención está asociado a quimioterapia con uno o más fármacos contra VIH que se usan convencionalmente para tratar o prevenir infecciones por VHC, enfermedades y estados patológicos asociados a VHC. Los ejemplos representativos de fármacos contra VHC incluyen, sin limitación, inhibidores de proteasas, por ejemplo inhibidores de serina proteasas tales como VX950 de Vertex), inhibidores de polimerasas, inhibidores de helicasas, antifibróticos, análogos nucleosídicos, agonistas de TLR, inhibidores de N-glucosilación, ARNip, oligonucleótidos antisentido, anticuerpos anti-VHC, moduladores inmunitarios, vacunas terapéuticas y agentes antitumorales usados habitualmente en el tratamiento de hepatocarcinomas asociados a VHC (por ejemplo adriamicina o una mezcla de adriamicina, lipiodol y spongel, administrada habitualmente mediante quimioembolización en la arteria hepática). Por ejemplo, las vacunas terapéuticas incluyen antígenos recombinantes, VLPs, vectores o péptidos sintéticos basados en o que codifican proteínas estructurales de VHC (Core, cubierta E1 y/o E2) que son particularmente adecuados para disparar una respuesta humoral anti-VHC. Tales fármacos contra el VHC se pueden proporcionar en una única dosis, o, como alternativa, en múltiples dosis según protocolos estándar, dosificaciones y regímenes durante varias horas, días y/o semanas. Su administración puede anteceder, ser concomitante, o subsiguiente a la administración de la composición de la invención. Una combinación particularmente adecuada incluye el tratamiento con IFN- α -2a pegilado (por ejemplo una dosis de 10 μ g/semana) y/o ribavirina (por ejemplo a 800 a 1200 mg/día) durante 24 a 48 semanas, antes, en paralelo o subsiguientemente al uso de la composición de la invención.

En otra forma de realización, el uso de la invención se lleva a cabo según una modalidad terapéutica de vacunación-revacuación, que comprende la administración secuencial de una o más composiciones de sensibilización y una o más composiciones de revacuación. Típicamente, las composiciones de sensibilización y de revacuación usan diferentes vehículos que comprenden o codifican al menos un dominio antigénico en común. La composición de sensibilización se administra inicialmente al organismo hospedante, y la composición de revacuación se administra subsiguientemente al mismo organismo hospedante después de un período que varía desde un día hasta doce meses. El uso de la invención puede comprender una a diez administraciones secuenciales de la composición de sensibilización, seguido de una a diez administraciones secuenciales de la composición de revacuación. De forma deseable, los intervalos de inyección son cuestión de una semana a seis meses. Además, las composiciones de sensibilización y revacuación se pueden administrar en el mismo sitio o en sitios alternativos, mediante la misma vía o mediante vías diferentes de administración. Por ejemplo, las composiciones basadas en polipéptido se puede administrar mediante una vía mucosal, mientras que las composiciones basadas en vectores se inyectan preferentemente, por ejemplo inyección subcutánea para un vector de MVA, inyección intramuscular para un plásmido de ADN y para un vector adenovírico.

En el contexto de la invención, una composición se usa para sensibilizar o revacunar, o tanto sensibilizar como revacunar, una respuesta inmunitaria anti-VHC. Las composiciones de sensibilización preferidas de la invención son preferentemente aquellas que comprenden un vector adenovírico/virus que codifica la fusión y un ADN plasmídico que codifica la fusión. Los vectores de MVA/virus que codifican la fusión se usan preferentemente para la revacuación. Por otro lado, la proteína de fusión producida de forma recombinantemente se puede usar independientemente para la sensibilización o la revacuación. Por ejemplo, se puede sensibilizar con una proteína de fusión producida recombinantemente que tiene la secuencia de aminoácidos mostrada en SEC ID n°: 9 o SEC ID n°: 10, y se puede revacunar con un vector adenovírico que codifica tal proteína de fusión. Como otro ejemplo, se puede sensibilizar con un plásmido de ADN que codifica una proteína de fusión que tiene la secuencia de aminoácidos mostrada en SEC ID n°: 9 o SEC ID n°: 10, y se puede revacunar con partículas de virus de MVA que codifican tal proteína de fusión.

También es posible usar la composición de la invención en combinación con cualquier material de la técnica anterior que codifica o que comprende un dominio antigénico en común con la composición de la invención (por ejemplo un

dominio antigénico del polipéptido NS3, NS4B y/o NS5B). La fuente de tal material es amplia, e incluye, sin limitación, péptidos, proteínas, vector vírico de una variedad de virus, ADN plasmídico, partículas proteinosas tales como partículas semejantes a virus (Bums *et al.*, 1994, Mol. Biol. Techno. 1, 137-145), materiales celulares tales como células irradiadas, partículas víricas, etc. Por ejemplo, los vectores particularmente apropiados útiles en este contexto son los vectores adenovírico y MVA descritos en el documento WO2004/111082 que codifican NS3, NS4 y NS5B. Como otro ejemplo, también son útiles en el contexto de la invención los péptidos descritos en el documento WO03/097677 que comprende dominios antigénicos de NS3, NS4B y NS5B. Con fines ilustrativos, se puede sensibilizar con las partículas adenovíricas del documento WO2004/111082 y revacunar subsiguientemente con la proteína de fusión producida recombinantemente de SEC ID n°: 9 o SEC ID n°: 10 o con el vector/virus MVA que codifica la proteína de fusión de SEC ID n°: 9 o SEC ID n°: 10 como se define anteriormente. Como otro ejemplo, se puede sensibilizar con un plásmido de ADN que codifica la proteína de fusión de SEC ID n°: 9 o SEC ID n°: 10, y revacunar con el vector de MVA del documento WO2004/111082.

Sin embargo, también son posibles en el contexto de la invención otras combinaciones de sensibilización revacunación.

También se describen anticuerpos que se unen selectivamente a la proteína de fusión de la presente invención o sus fragmentos peptídicos. Como se utiliza en la presente memoria, un anticuerpo se une selectivamente a un péptido diana cuando se une al péptido diana y no se une significativamente a proteínas no relacionadas. En ciertos casos, se entendería que el anticuerpo que se une al péptido todavía es selectivo a pesar de cierto grado de reactividad cruzada. No obstante, se prefiere que el anticuerpo no se una con afinidad elevada o selectividad elevada a polipéptidos NS nativos individuales y a fusiones de polipéptidos NS configurados en una configuración nativa.

Como se utiliza en la presente memoria, un anticuerpo se define en términos consistentes con los reconocidos en la técnica. Los anticuerpos incluyen anticuerpos policlonales y anticuerpos monoclonales, así como fragmentos de tales anticuerpos, incluyendo, pero sin limitarse a, Fab o F(ab').sub.2, y fragmentos Fv. Los anticuerpos se pueden producir usando técnicas convencionales en la técnica, por ejemplo administrando a un animal una cantidad eficaz de una proteína de fusión de la presente invención y/o un fragmento peptídico del mismo. Los anticuerpos se preparan preferentemente a partir de regiones o fragmentos discretos de la proteína de fusión de la invención que comprenden secuencias únicas, tales como aquellas que solapan el sitio de fusión entre los polipéptidos NS4A y el NS3 y los polipéptidos NS3 y NS5B.

Los anticuerpos descritos en la presente memoria tienen una variedad de usos potenciales que están dentro del alcance de la presente invención. Por ejemplo, tales anticuerpos se pueden usar (a) como reactivos en ensayos para detectar una proteína de fusión de la presente invención, (b) como reactivos en ensayos para detectar la presencia de un virus de VHC en una muestra biológica, y/o (c) como herramientas para recuperar la proteína de fusión producida recombinantemente de la presente invención a partir de una mezcla de proteínas y otros contaminantes (por ejemplo permitiendo la purificación mediante cromatografía de afinidad o inmunoprecipitación a partir de células hospedantes cultivadas).

También se describe un método para la detección y/o cuantificación de un virus de VHC o un anticuerpo anti-VHC en una muestra biológica (por ejemplo plasma, suero, tejido) tomada de un individuo susceptible de ser infectado por dicho virus de VHC usando la proteína de fusión, molécula de ácido nucleico, vector, partícula vírica infecciosa, célula hospedante, composición o anticuerpo de la invención.

El método es más particularmente adecuado para la detección y/o cuantificación de un virus de VHC en una muestra biológica, y comprende al menos las etapas de poner en contacto dicha muestra biológica con al menos uno de los anticuerpos de la invención en condiciones que permitan la formación de un complejo entre el virus y el anticuerpo, y detectar y/o cuantificar la formación de dicho complejo por cualquier medio apropiado.

El método también es más particularmente adecuado para la detección y/o cuantificación de un anticuerpo anti-VHC en una muestra biológica, y comprende al menos las etapas de poner en contacto dicha muestra biológica con al menos una de la proteína de fusión, molécula de ácido nucleico, vector, partícula vírica infecciosa, célula hospedante, composición de la invención en condiciones que permiten la formación de un complejo entre el anticuerpo anti-VHC y la proteína de fusión, molécula de ácido nucleico, vector, partícula vírica infecciosa, célula hospedante, composición de la invención, y detectar y/o cuantificar la formación de dicho complejo por cualquier medio apropiado.

Un experto en la materia determinará fácilmente la cantidad de anticuerpo, proteína de fusión, molécula de ácido nucleico, vector, partícula vírica infecciosa, célula hospedante, composición a usar en los métodos de la invención. Los medios de detección y/o cuantificación del virus son rutinarios y bien conocidos por una persona experta en la técnica. A título de ilustración, se puede mencionar transferencias, Elisa, técnicas denominadas de sándwich, técnicas de competición, y técnicas de PCR. En particular las técnicas denominadas "de tiempo real". El uso de un anticuerpo, proteína de fusión, molécula de ácido nucleico, vector, partícula vírica infecciosa, célula hospedante, o composición de la presente invención como reactivo se puede facilitar mediante acoplamiento (es decir, enlazando

físicamente a una sustancia detectable). Los ejemplos de sustancias detectables incluyen diversas enzimas (por ejemplo peroxidasa de rábano picante, fosfatasa alcalina, beta-galactosidasa o acetilcolinesterasa), grupos prostéticos (por ejemplo estreptavidina/biotina, o avidina/biotina), materiales fluorescentes (por ejemplo umbeliferona, fluoresceína, o derivados de fluoresceína), materiales luminiscentes, materiales luminiscentes, materiales bioluminiscentes (por ejemplo luciferasa, luciferina, o acurina), y materiales radioactivos (por ejemplo ^{125}I , ^{131}I , ^{35}S o ^3H).

Finalmente, también se describe el uso de la proteína de fusión, molécula de ácido nucleico, vector, partícula vírica infecciosa, célula hospedante, composición, o anticuerpo de la invención para el diagnóstico *in vitro* de una infección por VHC en una muestra biológica.

La invención se ha descrito de una manera ilustrada, y se entenderá que la terminología que se ha estado usando pretende ser la naturaleza de las palabras de la descripción en lugar de una limitación. Obviamente, son posibles muchas modificaciones y variaciones de la presente invención a la luz de las enseñanzas anteriores.

Leyendas de las figuras

La figura 1 ilustra la secuencia de aminoácidos de una proteína de fusión NS4A-3-5B. Los restos de aminoácidos añadidos o modificados están subrayados, mientras que los restos nativos sustituidos se muestran debajo de la secuencia.

La figura 2 ilustra la secuencia de aminoácidos de una proteína de fusión NS4A-3-4B-5B. Los restos de aminoácidos añadidos o modificados están subrayados, mientras que los restos nativos sustituidos se muestran debajo de la secuencia.

La figura 3 ilustra la secuencia nucleotídica que codifica la proteína de fusión NS4A-3-5B optimizada para la expresión en células de mamíferos.

La figura 4 ilustra la secuencia nucleotídica que codifica la proteína de fusión NS4A-3-4B-5B optimizada para la expresión en células de mamíferos.

La figura 5 ilustra la secuencia nucleotídica que codifica la proteína de fusión NS4A-3-5B optimizada para la expresión en células de levadura.

La figura 6 ilustra la secuencia nucleotídica que codifica la proteína de fusión NS4A-3-4B-5B optimizada para la expresión en células de levadura.

La figura 7 ilustra la secuencia nucleotídica que codifica la proteína de fusión NS4A-3-5B optimizada para la expresión en células procariotas.

La figura 8 ilustra la secuencia nucleotídica que codifica la proteína de fusión NS4A-3-4B-5B optimizada para la expresión en células procariotas.

La figura 9 ilustra el diseño esquemático de las proteínas de fusión NS4A-3-5B y NS4A-3-4B-5B ilustradas en los siguientes ejemplos. A, B, C, D y E representan dominios antigénicos que se pueden predecir en los polipéptidos NS3, NS4B y NS5B (véase el documento WO03/097677).

La figura 10 ilustra la construcción esquemática de los vectores de expresión gWiz NS4A-3-5B Hs y gWiz NS4A-3-4B-5B Hs.

La figura 11 ilustra la expresión *in vitro* mediante transferencia Western tras la infección de células Huh-7 con los plásmidos gWiz NS4A-3-5B (A línea 1 y C línea 3), gWiz NS4A-3-4B-5B (B línea 1 y C línea 2) y gWiz (A línea 2, B línea 2 y C línea 1). Las células Huh-7 transfectadas se lisaron a 48 h después de la transfección, y las proteínas de fusión se detectaron mediante análisis de transferencia Western usando anticuerpos monoclonal de ratón anti-NS3 (A y B) o monoclonal de ratón anti-NS5B (C).

La figura 12 ilustra la evaluación de las respuestas de células T CD8+ específicas de NS3 mediante ensayo de ELISPOT de IFN γ tras la inyección intramuscular de gWiz NS4A-3-5B Hs. Los esplenocitos de ratones individuales se cultivaron en presencia de péptido GLL específico de NS3 o en presencia de un péptido irrelevante, durante 40 h. Los resultados se muestran como el valor medio del número de manchas detectadas para 10^6 esplenocitos obtenidos para pocillos triplicados. M1, M2, M3 y M4 representan 4 ratones inmunizados con el gWiz NS4A-3-5B Hs. Mc es un ratón de control.

La figura 13 ilustra la expresión de las proteínas de fusión a partir de células infectadas con adenovirus. Se infectaron células A549 ($1,5 \times 10^6$ células) a una MOI de 10 con diferentes adenovirus, respectivamente los adenovirus AdTG17476 y AdTG17477, que expresan la fusión, así como un adenovirus vacío y un adenovirus

recombinante y relevante como controles negativos. Las células se cultivaron durante 48 h con 10% de suero antes de lisarlas con tampón de Laemmli. Las muestras proteicas (volumen 1/20) se cargaron en un gel de electroforesis SDS PAGE. Las proteínas se detectaron mediante azul de Coomassie. Línea 1: marcadores de peso molecular; Línea 2: células A549 no infectadas; Línea 3: células A549 infectadas con adenovirus vacío; Línea 4: células A549 infectadas con adenovirus irrelevante; Línea 5: células A549 infectadas con AdTG17476; y Línea 6: células A549 infectadas con AdTG17477.

La figura 14 ilustra respuestas de Elispot IFN γ inducidas con proteínas de fusión de Ad de VHC en ratones transgénicos HLA-A2.1. Se inyectaron AdTG17476, AdTG 17477 o Ad vacío intramuscularmente en el músculo anterior de la tibia una vez a una dosis de 10⁹ UI/ratón. Las respuestas inmunitarias celulares se investigaron 2 semanas después de la inyección mediante ELISPOT de IFN γ para cada grupo de animal (incluyendo 4 ratones para grupos de ensayo y 2 ratones para grupo de control). Se presentan a respuestas de ratones individuales (M1 a 4) frente a los epítomos GLL (figura 14A) y KLT (figura 14B) restringidos a HLA-A2 específicos de NS3 y el epítomo KLQ (figura 14C) restringido a HLA-A2 específico de NS5B. Como control negativo, se usó el péptido irrelevante DLM. Una respuesta ELISPOT se considera positiva si el número de manchas para 10⁶ esplenocitos es mayor que 50 (línea de corte).

Los siguientes ejemplos sirven para ilustrar la presente invención.

20 Ejemplos

Materiales y Métodos

Los constructos descritos más abajo se preparan según las técnicas generales de ingeniería genética y de clonación molecular, como se detalla en Sambrook *et al.* (2001, Molecular Cloning; A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY) o según las recomendaciones del fabricante cuando se usa un kit comercial. Las técnicas de amplificación mediante PCR son conocidas por los expertos en la técnica (véase, por ejemplo PCR Protocols - A guide to Methods and applications, 1990; Ed Innis, Gelfand, Sninsky and White, Academic Press Inc). La extirpe celular de hepatoma Huh-7 humana se usó para evaluar la expresión correcta de las proteínas de fusión a partir de los constructos descritos en los ejemplos que siguen. Las condiciones de cultivo son convencionales en la técnica. Con fines ilustrativos, las células se hicieron crecer a 37°C en atmósfera de CO₂ al 5% en medio completo DMEM que contiene medio de Eagle modificado de Dulbecco suplementado con 10% de suero fetal de ternera, 2 mM de L-glutamina y 100 UI/ml de penicilina/estreptomicina (Sigma). Las células se transfectaron usando el reactivo Lipofectamine Plus (Invitrogen) y según las instrucciones del fabricante. Los anticuerpos monoclonales murinos dirigidos contra NS3 y NS5B se pueden adquirir comercialmente o se pueden preparar según técnicas convencionales en la técnica.

Construcciones plasmídicas

Se usaron oligonucleótidos solapantes para generar genes sintéticos (GeneART, Regensburg, Alemania) que codifican, respectivamente, las proteínas de fusión NS4A-NS3-NS5B (3666 nucleótidos) y NS4A-NS3-NS4B-NS5B (3747 nucleótidos). El diseño de las fusiones se ilustra en la figura 9. Las secuencias nucleotídicas se utilizaron para la expresión elevada en homo sapiens (SEC ID n°: 11 y 12 y figuras 3 y 4), *Pichia* (SEC ID n°: 13 y 14 y figuras 5 y 6) y *E. coli* (SEC ID n°: 15 y 16 y figuras 7 y 8). La secuencia de aminoácidos de las fusiones resultantes se ilustra en SEC ID n°: 9 para NS4A-NS3-NS5B y en SEC ID n°: 10 para NS4A-NS3-NS4B-NS5B. Cada una de las secuencias nucleotídicas sintéticas se clonó en los sitios de restricción *SacI* y *KpnI* del plásmido pPCR-Script (Stratagene).

Para la expresión transitoria en células de mamíferos, los genes sintéticos que codifican la fusión, optimizados para la expresión en homo sapiens (SEC ID n°: 11-12 y figuras 3 y 4), se insertaron en los sitios de restricción *SacI* y *NotI* del vector de expresión gWiz (Gene Therapy Systems) bajo el control transcripcional del promotor de citomegalovirus, para generar los vectores gWiz NS4A-3-5B Hs y gWiz NS4A-3-4B-5B Hs (figura 10). La secuencia se modificó entonces para mejorar la secuencia de consenso de Kozak (subrayada) usando el oligonucleótido mostrado en SEC ID n°: 17 (5'-ATG TAC GTC GAC CACC ATG GGC AGC GTG GTG ATT GTG GGC CGG ATC 3') y para crear codones de parada (en negrita) usando el oligonucleótido mostrado en SEC ID n°: 18 (5'-CAT GTA GC GGC CGC TCA TCA TCT AGG CCT GGC CCT GGA CAG-3'). Todos los plásmidos se verificaron mediante secuenciación.

Estudios de expresión *in vitro*

Análisis de transferencia Western

Un día antes de la transfección, se colocaron células Huh-7 (5x 10⁵ por pocillo) en placas de 6 pocillos. Las células se transfectaron con el vector gWiz NS4A-3-5B Hs, gWiz NS4A-3-4B-5B Hs, o el plásmido gWiz como control negativo. A las 48 h después de la transfección, las células se lisaron en 50mM de Tris-HCl, 150mM de NaCl, 0,1% de SDS, 1% de NP40 y 0,5% de desoxicolato de sodio. Las muestras de los lisados se calentaron a 95°C durante 5

minutos, y las proteínas se separaron mediante electroforesis en SD-8% de poliacrilamida. La inmunotransferencia se llevó a cabo usando anticuerpos monoclonales murinos específicos de S3 (por ejemplo 8G4C3, bioMerieux) o NS5B (por ejemplo 11F6C8, bioMerieux) como anticuerpo primario, seguido de un anticuerpo anti-IgG de ratón de cabra conjugado con peroxidasa (Sigma) como anticuerpo secundario. Las señales se revelaron usando el kit ECL comercial (Amersham Biosciences).

Análisis de inmunofluorescencia

Un día antes de la transfección, se colocaron cubreobjetos de vidrio en placas de 6 pocillos y se trataron con gelatina al 0,2% durante 10 minutos, y después 5×10^5 células Huh-7 se colocaron en placas en cada pocillo. Las monocapas celulares se transfectaron con el vector gWiz NS4A-3-5B Hs, gWiz NS4A-3-4B-5B Hs o con el plásmido gWiz como control negativo. A las 48 h después de la transfección, los cubreobjetos se lavaron dos veces en PBS y las células se fijaron durante 10 minutos con 4% de PFA, y después se lavaron y se permeabilizaron con 0,1% de Triton X-100 en PBS durante 10 minutos. Los cubreobjetos se lavaron y se trataron con anticuerpos monoclonales murinos específicos de NS3 (por ejemplo 8D8E1, bioMerieux) o NS5B (por ejemplo 5B-12B7, proporcionado por D. Moradpour) como anticuerpo primario durante 1h a temperatura ambiente. Después de lavar se añadió anti-IgG de ratón TRITC (Dako) durante 20 minutos. Después de lavar, los cubreobjetos se montaron en glicerol al 80%. Se observaron hilos con el microscopio Carl Zeiss Axioplan, y las imágenes se tomaron con una cámara digital AxioCam Color.

Estudios de expresión *in vivo*

Los ratones transgénicos HLA-A2.1 producidos como se describe Pascolo *et al.* (1997, J. of Experimental Medicine 185, 2043-2051) se hicieron reproducir. A estos ratones se les suprimieron los genes H-2D^b y β_2 -microglobulina murina, y expresaron una molécula de histocompatibilidad clase I monocatenaria transgénica en la que el término C de la β_2m humana se enlazó covalentemente al término N de una cadena pesada quimérica (dominios intracitoplásmicos y transmembránicos HLA-A2.1 $\alpha 1$ - $\alpha 2$, H-2D^b a3).

Protocolos de inmunización

Los animales transgénicos HLA-A2 recibieron un pretratamiento con 10 μ M de cardiotoxina (Latoxan) cinco días antes de la inyección de los plásmidos. Se inyectaron intramuscularmente plásmidos gWiz NS4A-3-5B, gWIZNS4A-3-4B-5B y gWiz (control negativo) en 2 veces a intervalos de 2 semanas a una dosis de 100 μ g (en 100 μ l 1X PBS). Se investigaron las respuestas de las células T CD8+ restringidas a HLA-A2 2 semanas después de la segunda inmunización, tal como se describe en Himoudi *et al.* (2002, J. Virol. 76, 12735-46).

Monitorización de la respuesta inmunitaria mediante ensayos ELISPOT

Se cultivaron esplenocitos (2×10^5) en pocillos triplicados durante 40 h en placas de Multiscreen (Millipore) revestidas con anticuerpo monoclonal anti-IFN γ de ratón (Pharmlingen; 10 μ g/ml) en medio de cultivo completo α MEM (Invitrogen) en presencia de 10 unidades/ml de IL-2 recombinante (PeproTech Inc, Inglaterra) sólo como control negativo, o 5 μ g/ml de concavalina A como control positivo, o con 10 μ g de péptido (péptido irrelevante DLM, péptido GLL situado en la proteína NS3 descrita en Martin *et al.*, 2004, J. Med. Virol. 74, 397-405 y Himoudi *et al.*, 2002, J. Virol. 76, 12735-46). Las células productoras de IFN γ se cuantificaron mediante ensayo de inmunomancha enlazada encima específico de citocinas (ELISPOT) como se describe previamente (Himoudi *et al.*, 2002, J. Virol. 76, 12735-46). El número de manchas (que representa células de IFN γ individuales) en los pocillos de control negativo se restó del número en los pocillos de ensayo que contienen péptidos. Los resultados se muestran como el valor medio obtenido para pocillos triplicados.

Ejemplo 1: Expresión *in vitro* de las poliproteínas NS4A-3-5B y NS4A-3-4B-5B

Las secuencias que codifican las proteínas de fusión NS4A-3-5B y NS4A-3-4B-5B se construyeron como se esquematiza en la figura 9. Las secuencias se optimizaron para la expresión en tres hospedantes diferentes (GeneART, Regensburg, Alemania), ser humano (Hs), *Pichia pastoris* (Pp) y *E. coli* (Ec), respectivamente, como se da a continuación:

- NS4A-3-5B Hs optimizada para la expresión en homo sapiens (SEC ID n°: 11 y figura 3)
- NS4A-3-5B Pp optimizada para la expresión en Pichia (SEC ID n°: 13 y figura 5)
- NS4A-3-5B Ec optimizada para la expresión en *E. coli* (SEC ID n°: 15 y figura 7)
- NS4A-3-4B-5B Hs optimizada para la expresión en homo sapiens (SEC ID n°: 12 y figura 4)
- NS4A-3-4B-SB Pp optimizada para la expresión en Pichia (SEC ID n°: 14 y figura 6)
- NS4A-3-4B-5B Ec optimizada para la expresión en *E. coli* (SEC ID n°: 16 y figura 8)

Las secuencias optimizadas para la expresión en homo sapiens que codifican las fusiones NS4A-NS3-NS4B-NS5B y NS4A-NS3-NS5B se clonaron en los vectores de expresión gWiz (GeneThecap Systems) bajo el control

transcripcional del promotor de citomegalovirus (figura 10). La expresión de las proteínas de fusión NS se evaluó mediante transferencia Western e inmunofluorescencia en células Huh-7 transfectadas con los vectores gWiz NS4A-3-5B Hs, gWiz NS4A-3-4B-5B Hs, o el plásmido gWiz como control negativo.

5 Como se muestra en la figura 11, el análisis de transferencia Western reveló una única banda que tiene el peso molecular esperado de alrededor de 135 kDa para cada fusión individual (figura 11A línea 1, y figura 11C línea 3 para la fusión NS4A-3-5B corta, y figura 11B línea 1, y figura 11C línea 2 para la fusión larga NS4A-3-4B-5B) tras la detección con los anticuerpos específicos anti-NS3 (figuras 11A y 7B) o anti-NS5B (figura 11C). Como se esperaba, no se detectaron proteínas en las muestras obtenidas a partir de células Huh-7 transfectadas con el plásmido gWiz (figura 11A línea 2, figura 11B línea 2 y figura 11C línea 1). Estos resultados confirman que la fusión corta NS4A-NS3-NS5B y la fusión larga que incorpora adicionalmente NS4B son expresadas en células Huh-7.

15 Los análisis de inmunofluorescencia confirmaron la expresión de las dos proteínas de fusión en células transfectadas Huh-7. De forma interesante, ambas proteínas de fusión mostraron una localización citoplásmica que sugiere un procesamiento apropiado como resultado de la supresión de la mayoría de los dominios de anclaje a membrana hidrófobos. Las células transfectadas con el plásmido gWiz no presentaron ninguna señal.

Ejemplo 2: Evaluación *in vivo* de la inmunogenicidad de gWIZ NS4A-3-5B Hs y NS4A-3-4B-5B Hs

20 Para evaluar la capacidad de los plásmidos gWiz NS4A-3-5B Hs and NS4A-3-4B-5B Hs para inducir respuestas de células T CD8+ en ratones transgénicos HLA-A2, se llevó a cabo un estudio de inmunogenicidad usando ratones transgénicos HLA-A2 como se describe en Material y Métodos. Se investigaron las respuestas de células T CD8+ restringidas a HLA-A2 mediante ensayo ELISPOT de IFN γ 2 semanas después de la segunda inmunización. Como se ilustra en la figura 12, el pgWiz NS4A-3-5B fue capaz de inducir células T productoras de IFN γ específicas del péptido GLL bien conocido e inmunodominante restringido a HLA-A2 (proteína NS3, Martin *et al.*, 2004, J. Med. Virol. 74, 397-405), sugiriendo una presentación correcta de estos epítopos.

30 También es posible usar ensayos de exposición murino preclínicos a fin de evaluar las proteínas de fusión para determinar la inmunidad protectora. Puesto que los ratones no son capaces de ser infectados directamente por virus de VHC, estos ensayos emplean virus de la vacuna recombinantes (cepa WR) o Listeria que expresa antígenos NS de VHC. Tras la inyección del virus de la vacuna o de Listeria, los ratones desarrollan una infección que alcanza un pico 2 a 6 días después de la inyección, dando como resultado que se detecten virus o bacterias en los ovarios (modelo de la vacuna) o hígados (modelo de Listeria) de ratones inyectados. La evaluación preclínica implica la vacunación de ratones con la vacuna candidata (administración de la proteína de fusión recombinante o de partículas adenovíricas que expresan la proteína de fusión), controles (por ejemplo administración de fusión de polipéptidos NS en configuración nativa) y constructos negativos (por ejemplo antígenos no de VHC, o adenovirus vacío). Los animales se expusieron subsiguientemente al virus de la vacuna recombinante que expresa todos los antígenos NS de VHC (NS2 a NS5) o a Listeria que expresa uno de los antígenos NS incluidos en la vacuna candidata. Si la vacuna sensibiliza con éxito a células T específicas, los ratones expuestos desarrollan respuesta inmunitaria a la vacuna o al vector de Listeria vía los antígenos NS. La actividad protectora de la vacuna candidata da como resultado títulos más bajos de vacuna o de Listeria, según se mire en los ovarios o hígados de los ratones vacunados, en comparación con los títulos de vacuna o Listeria medidos en los ovarios o hígados de los ratones no vacunados.

45 **Ejemplo 3: Expresión de la proteína de fusión NS4A-3-5B y NS4A-3-4B-5B en células de mamíferos.**

Los genes sintéticos que codifican la fusión, optimizados para la expresión en homo sapiens (SEC ID n°: 11-12), que corresponden a NS4A-NS3-NS5B (3666 nucleótidos) y NS4A-NS3-NS4B-NS5B (3747 nucleótidos), se amplificaron mediante PCR a partir de los plásmidos correspondientes pPCR-Script, usando los siguientes cebadores OTG18033 (GGGGGGCTAGCGCCACCATGGGCAGCGTGATTG; SEC ID n°: 19) y OTG 18036 (GGGGGGGTACCCTCATCTAGGCCTGGCCCTG; SEC ID n°: 20). Ambos fragmentos se insertaron en los sitios de restricción NheI y KpnI de un plásmido lanzadera adenovírico que contiene un casete de expresión conducido por CMV rodeado por secuencias adenovíricas (nucleótidos adenovíricos 1-458 y nucleótidos 3328-5788, respectivamente) para permitir la generación del genoma del vector mediante recombinación homologada (Chartier *et al.*, 1996, J. Virol. 70,4805-4810). Los vectores adenovíricos resultantes están suprimidos en E3 (nucleótidos 28592-30470) y E1 (nucleótidos 459-3511), sustituyéndose la región E1 por el casete de expresión que contiene, desde 5' hasta 3', el potenciador/promotor temprano inmediato de CMV, un intrón β -globina humana/IgG quimérico (como se muestra en el vector pCI disponible en Promega), la secuencia que codifica una de las dos formas poliproteicas y la señal de poliadenilación tardía de SV40. Los vectores adenovíricos resultantes se denominaron pTG17476 (NS4A-3-4B-5B) y pTG17477 (NS4A-3-5B). Los adenovirus recombinantes, AdTG17476 y AdTG1.7477, se generaron transfectando los genomas víricos linealizados con *PacI* en una extirpe celular de complementación E1 convencional. La propagación, purificación y titulación del virus se realizaron como se describe previamente (Erbs, 2000, Cancer Res. 60, 3813-3822).

65 La expresión de las proteínas de fusión se evaluó en células infectadas con adenovirus mediante electroforesis en SDS PAGE. Se infectaron células A549 (1,5 x 10⁶ células) (ATCC CCL-185) a una MOI de 10 o 100 con los

adenovirus AdTG17476 y AdTG17477 que expresan la fusión, así como con un adenovirus vacío y un adenovirus recombinante irrelevante como controles negativos. Las células se cultivaron durante 48 h con 10% de suero antes de ser lisado con tampón de Laemmli. Se cargaron muestras de proteína (volumen 1/20) en un gel de electroforesis de SDS PAGE, y las proteínas se tiñeron mediante azul de Coomassie. Como se ilustra en la figura 13, se observó claramente una única banda que tiene el peso molecular esperado de aproximadamente 135 kDa en células infectadas con AdTG17476 y AdTG17477 (Líneas 5 y 6), que refleja niveles de expresión elevada en células infectadas con los vectores adenovíricos que codifican la fusión. Por otro lado, esta banda está ausente en las muestras recogidas de las células infectadas con los controles negativos (Líneas 3 y 4) así como células A549 no infectadas (Línea 2).

Los niveles de expresión se analizaron mediante transferencia Western en las células A549 infectadas con los constructos víricos a MOI de 10 durante 48 horas. Los peletes celulares se recogieron y se sondaron con un anticuerpo monoclonal murino anti-NS3 (por ejemplo 1B6, bioMerieux) y un anticuerpo policlonal de conejo anti-NS4 (por ejemplo RB, bioMerieux). Se reveló una banda principal que tiene el peso molecular esperado en las muestras recogidas de células infectas con AdTG17476 y AdTG17477 tras la detección con los anticuerpos específicos anti-NS3 o anti-NS4, confirmando los niveles de expresión elevados detectados tras la tinción con azul de Coomassie.

Ejemplo 4: Inmunogenicidad de Adenovirus que expresan las poliproteínas de VHC en ratones transgénicos HLA A2

Modelo de animal

Los ratones transgénicos HLA-A2.1 se produjeron como se describe por Pascolo *et al.* (1997, J Ex Med). Estos ratones tienen los genes de H-2D^b y β_2 -microglobulina murina suprimidos y expresan una molécula de histocompatibilidad de clase I monocatenaria transgénica en la que el término C de la β m humana está enlazado covalentemente al término N de una cadena pesada quimérica (dominios transmembránicos e intracitoplásmicos (HLA-A2.1 α 1- α 2, H-2D^b a3).

Protocolo de inmunización

Se incluyeron tres grupos de ratones en el protocolo, respectivamente cuatro ratones inyectados con AdTG17476 (Ad NS4A- 3-4B-5B), cuatro ratones inyectados con AdTG17477 (Ad NS4A-3-5B) y dos ratones de control inyectados con un Ad vacío (Ad514). Los adenovirus se inyectaron intramuscularmente en el músculo de la tibia anterior una vez a una dosis de 10^9 UI en 10 μ l de 1X PBS. Las respuestas inmunitarias celulares se investigaron dos semanas después de la inyección.

Monitorización de las respuestas inmunitarias

Ensayos ELISPOT

Se recogieron esplenocitos de ratones de cada grupo y los glóbulos rojos se lisaron. Se cultivaron $2 \cdot 10^5$ células en pocillos triplicados durante 40 h en placas Multiscreen (Millipore) revestidas con anticuerpo monoclonal IFN γ de ratón (PharMingen; 10 mg/ml) en medio de cultivo completo α MEM (GIBCO BRL) en presencia de 10 unidades/ml de IL-2 murina recombinante (PeproTech Inc, Inglaterra), sola como control negativo, o con 10 μ m de péptido (péptido irrelevante DLM, péptidos GLL y KLT situados en la proteína NS3, péptido KLQ situado en la proteína NS5B (Martin *et al.*, 2004, J Med Virol. 74, 397-405) y (Himoudi *et al.*, 2002, J Virol., 76, 12735-12746), o 5 μ g/ml de concavalina A como control positivo. Las células T productoras de IFN γ se cuantificaron mediante ensayo de inmunomancha enlazado a enzima específico de citocinas (ELISPOT) como se describe previamente (Himoudi *et al.*, J Virol 2002). El número de manchas (que representa las células T productoras de IFN γ individuales) en pocillos de control negativo se restaron del número en pocillos de ensayo que contienen péptidos. Los resultados se muestran como el valor medio obtenido para pocillos triplicados.

Resultados

Se investigaron las respuestas de células T restringidas a HLA-A2 mediante ensayo de ELISPOT de IFN γ dos semanas después de la inmunización con adenovirus. Como se muestra en la figura 14, tanto AdTG17476 como AdTG 17477 fueron capaces de inducir en animales todos los animales vacunados, frecuencias elevadas de células T productoras de IFN γ específicas del epítipo GLL restringido a HLA-A2 (figura 14A; valor de la mediana: 1210 y 868 manchas respectivamente) con una eficiencia mayor estadísticamente significativa para AdTG17476 (ensayo de Mann-Whitney: P=0,0209). Ambos virus también indujeron células T productoras de IFN γ específicas del epítipo KLT subdominante (figura 14B) con menores frecuencias (valor de la mediana: 187 y 90 manchas). Además, 2:4 ratones vacunados con AdTG17476 y 1:4 ratones vacunados con AdTG17477 desarrollaron respuestas dirigidas al epítipo KLQ de la proteína NS5B (figura 14C).

Esto demuestra que las proteínas de fusión según la invención se pueden producir en niveles elevados en sistemas eucariotas, y que las proteínas resultantes son inmunógenas en modelos de animales convencionales.

Listado de secuencias

- 5 <110> TRANSGENE S.A.
- <120> Proteína de fusión no estructural del virus de la hepatitis C
- 10 <130> TG173
- <160> 20
- <170> PatentIn versión 3.1
- 15 <210> 1
- <211> 631
- <212> PRT
- <213> Virus de la hepatitis C
- 20 <400> 1

```

Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys
1           5           10           15

Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Asp Gly Glu
          20           25           30

Val Gln Val Leu Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val
          35           40           45

Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu
50           55           60

Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp Gln
65           70           75           80

Asp Leu Val Gly Trp Pro Ala Pro Pro Gly Ala Arg Ser Met Thr Pro
          85           90           95

Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp
100          105          110

Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser
115          120          125

Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu
130          135          140

Cys Pro Ser Gly His Val Val Gly Ile Phe Arg Ala Ala Val Cys Thr
145          150          155          160
    
```

Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Ser Met Glu
 165 170 175
 Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala
 180 185 190
 Val Pro Gln Thr Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser
 195 200 205
 Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys
 210 215 220
 Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala
 225 230 235 240
 Tyr Met Ser Lys Ala His Gly Ile Glu Pro Asn Ile Arg Thr Gly Val
 245 250 255
 Arg Thr Ile Thr Thr Gly Gly Pro Ile Thr Tyr Ser Thr Tyr Gly Lys
 260 265 270
 Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile
 275 280 285
 Cys Asp Glu Cys His Ser Thr Asp Trp Thr Thr Ile Leu Gly Ile Gly
 290 295 300
 Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu
 305 310 315 320
 Ala Thr Ala Thr Pro Pro Gly Ser Ile Thr Val Pro His Pro Asn Ile
 325 330 335
 Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys
 340 345 350
 Ala Ile Pro Ile Glu Ala Ile Lys Gly Gly Arg His Leu Ile Phe Cys
 355 360 365
 His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Thr Gly Leu
 370 375 380

ES 2 387 321 T3

Gly Leu Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile
 385 390 395 400

Pro Thr Ser Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr
 405 410 415

Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val
 420 425 430

Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr
 435 440 445

Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly Arg
 450 455 460

Thr Gly Arg Gly Arg Ser Gly Ile Tyr Arg Phe Val Thr Pro Gly Glu
 465 470 475 480

Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp
 485 490 495

Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg
 500 505 510

Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His
 515 520 525

Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala
 530 535 540

His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu
 545 550 555 560

Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro
 565 570 575

Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu
 580 585 590

His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu
 595 600 605

Ile Thr Leu Thr His Pro Ile Thr Lys Phe Val Met Ala Cys Met Ser
 610 615 620

Ala Asp Leu Glu Val Val Thr
 625 630

ES 2 387 321 T3

<211> 54
 <212> PRT
 <213> Virus de la hepatitis C

5 <400> 2

Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr
 1 5 10 15
 Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser
 20 25 30
 Gly Arg Pro Ala Val Val Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe
 35 40 45
 Asp Glu Met Glu Glu Cys
 50

10 <210> 3
 <211> 261
 <212> PRT
 <213> Virus de la hepatitis C

<400> 3

Ala Ser His Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln
 1 5 10 15
 Phe Lys Gln Lys Ala Leu Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala
 20 25 30
 Glu Ala Ala Ala Pro Val Val Glu Ser Arg Trp Arg Ala Leu Glu Ala
 35 40 45
 Phe Trp Ala Lys His Met Trp Asn Phe Ile Ser Gly Ile Gln Tyr Leu
 50 55 60
 Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met
 65 70 75 80

Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln Asn Thr Leu
 85 90 95

Leu Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu Ala Pro Pro
 100 105 110

Ser Ala Ala Ser Ala Phe Val Gly Ala Gly Ile Ala Gly Ala Ala Ile
 115 120 125

Gly Ser Ile Gly Leu Gly Lys Val Leu Val Asp Ile Leu Ala Gly Tyr
 130 135 140

Gly Ala Gly Val Ala Gly Ala Leu Val Ala Phe Lys Val Met Ser Gly
 145 150 155 160

Glu Ala Pro Ser Ala Glu Asp Leu Val Asn Leu Leu Pro Ala Ile Leu
 165 170 175

Ser Pro Gly Ala Leu Val Val Gly Ile Val Cys Ala Ala Ile Leu Arg
 180 185 190

Arg His Val Gly Pro Gly Glu Gly Ala Val Gln Trp Met Asn Arg Leu
 195 200 205

Ile Ala Phe Ala Ser Arg Gly Asn His Val Ser Pro Thr His Tyr Val
 210 215 220

Pro Glu Ser Asp Ala Ala Ala Arg Val Thr Gln Ile Leu Ser Ser Leu
 225 230 235 240

Thr Ile Thr Gln Leu Leu Lys Arg Leu His Gln Trp Ile Asn Glu Asp
 245 250 255

Cys Ser Thr Pro Cys
 260

- <210> 4
- <211> 591
- 5 <212> PRT
- <213> Virus de la hepatitis C

<400> 4

Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala

Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys Ser Leu
 245 250 255

Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln
 260 265 270

Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser
 275 280 285

Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Thr Ala Ala Cys Arg
 290 295 300

Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Asn Gly Asp Asp Leu
 305 310 315 320

Val Val Ile Cys Glu Ser Ala Gly Thr Gln Glu Asp Ala Ala Ser Leu
 325 330 335

Arg Val Phe Thr Glu Ala Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp
 340 345 350

Pro Pro Gln Pro Glu Tyr Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser
 355 360 365

Asn Val Ser Val Ala His Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu
 370 375 380

Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Val
 385 390 395 400

Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala
 405 410 415

Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Ile
 420 425 430

Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr
 435 440 445

Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile Ile Glu
 450 455 460

ES 2 387 321 T3

Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly
465 470 475 480

Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val Pro Pro
485 490 495

Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Lys Leu Leu
500 505 510

Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp
515 520 525

Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln
530 535 540

Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Asn Gly Gly Asp Ile
545 550 555 560

Tyr His Ser Leu Ser Arg Ala Arg Pro Arg Trp Phe Met Leu Cys Leu
565 570 575

Leu Leu Leu Ser Val Gly Val Gly Ile Tyr Leu Leu Pro Asn Arg
580 585 590

<210> 5

<211> 630

5 <212> PRT

<213> secuencia artificial

<220>

10 <223> resto NS3 derivado de polipéptido NS3 de HCV-JA al que se le inactivó las actividades de proteasa y helicasa

<400>5

Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys
1 5 10 15

Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Asp Gly Glu
20 25 30

Val Gln Val Leu Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val
35 40 45

Asn Gly Val Cys Trp Thr Val Tyr Ala Gly Ala Gly Ser Lys Thr Leu

ES 2 387 321 T3

50		55		60											
Ala 65	Gly	Pro	Lys	Gly	Pro	Ile	Thr	Gln	Met	Tyr	Thr	Asn	Val	Asp	Gln 80
Asp	Leu	Val	Gly	Trp	Pro	Ala	Pro	Pro	Gly	Ala	Arg	Ser	Met	Thr	Pro 95
Cys	Thr	Cys	Gly	Ser	Ser	Asp	Leu	Tyr	Leu	Val	Thr	Arg	His	Ala	Asp 110
			100					105							
Val	Ile	Pro	Val	Arg	Arg	Arg	Gly	Asp	Ser	Arg	Gly	Ser	Leu	Leu	Ser 125
			115					120							
Pro	Arg	Pro	Val	Ser	Tyr	Leu	Lys	Gly	Ser	Ser	Gly	Gly	Pro	Leu	Leu 140
Cys	Pro	Ser	Gly	His	Val	Val	Gly	Ile	Phe	Arg	Ala	Ala	Val	Cys	Thr 160
Arg	Gly	Val	Ala	Lys	Ala	Val	Asp	Phe	Ile	Pro	Val	Glu	Ser	Met	Glu 175
Thr	Thr	Met	Arg	Ser	Pro	Val	Phe	Thr	Asp	Asn	Ser	Ser	Pro	Pro	Ala 190
Val	Pro	Gln	Thr	Phe	Gln	Val	Ala	His	Leu	His	Ala	Pro	Thr	Gly	Ser 205
Gly	Lys	Ser	Thr	Lys	Val	Pro	Ala	Ala	Tyr	Ala	Ala	Gln	Gly	Tyr	Lys 220
Val	Leu	Val	Leu	Asn	Pro	Ser	Val	Ala	Ala	Thr	Leu	Gly	Phe	Gly	Ala 240
Tyr	Met	Ser	Lys	Ala	His	Gly	Ile	Glu	Pro	Asn	Ile	Arg	Thr	Gly	Val 255
Arg	Thr	Ile	Thr	Thr	Gly	Gly	Pro	Ile	Thr	Tyr	Ser	Ala	Tyr	Gly	Lys 270
Phe	Leu	Ala	Asp	Gly	Gly	Cys	Ser	Gly	Gly	Ala	Tyr	Asp	Ile	Ile	Ile 285

ES 2 387 321 T3

Cys Asp Glu Cys His Ser Thr Asp Trp Thr Thr Ile Leu Gly Ile Gly
 290 295 300

Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu
 305 310 315 320

Ala Thr Ala Thr Pro Pro Gly Ser Ile Thr Val Pro His Pro Asn Ile
 325 330 335

Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly Lys
 340 345 350

Ala Ile Pro Ile Glu Ala Ile Lys Gly Gly Arg His Leu Ile Phe Cys
 355 360 365

His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Thr Gly Leu
 370 375 380

Gly Leu Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile
 385 390 395 400

Pro Thr Ser Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met Thr
 405 410 415

Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val
 420 425 430

Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr
 435 440 445

Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly Ala
 450 455 460

Thr Gly Arg Gly Arg Ser Gly Ile Tyr Arg Phe Val Thr Pro Gly Glu
 465 470 475 480

Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp
 485 490 495

Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val Arg
 500 505 510

ES 2 387 321 T3

Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His
515 520 525

Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala
530 535 540

His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu
545 550 555 560

Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro
565 570 575

Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu
580 585 590

His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu
595 600 605

Ile Thr Leu Thr His Pro Ile Thr Lys Phe Val Met Ala Cys Met Ser
610 615 620

Ala Asp Leu Glu Val Val
625 630

5 <210> 6
<211> 13
<212> PRT
<213> secuencia artificial

10 <220>
<223> resto NS4A derivado de polipéptido JA NS4A de HCV-JA mediante truncamiento del C y N-terminal
<400> 6

Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser Gly
1 5 10

15 <210> 7
<211> 32
<212> PRT
<213> secuencia artificial

20 <220>
<223> resto NS4B derivado de polipéptido NS4B de HCV-JA mediante truncamiento de N y C
<400> 7

Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser Pro Leu Thr Thr Gln
1 5 10 15

Asn Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala Ala Gln Leu
20 25 30

25 <210> 8
<211> 570
<212> PRT

ES 2 387 321 T3

<213> secuencia artificial

<220>

5 <223> resto NS5B derivado de polipéptido NS5B de HCV-JA al que se le inactivó la actividad de polimerasa y truncamiento del término C

<400> 8

```

Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala
1          5          10          15

Glu Glu Ser Lys Leu Pro Ile Asn Pro Leu Ser Asn Ser Leu Leu Arg
20          25          30

His His Ser Met Val Tyr Ser Thr Thr Ser Arg Ser Ala Ser Leu Arg
35          40          45

Gln Lys Lys Val Thr Phe Asp Arg Leu Gln Val Leu Asp Asp His Tyr
50          55          60

Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val Lys Ala
65          70          75          80

Arg Leu Leu Ser Ile Glu Glu Ala Cys Lys Leu Thr Pro Pro His Ser
85          90          95

Ala Lys Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Ser Leu Ser
100         105         110

Ser Arg Ala Val Asn His Ile Arg Ser Val Trp Glu Asp Leu Leu Glu
115        120        125

Asp Thr Glu Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val
130        135        140

Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile

```


ES 2 387 321 T3

Thr Arg Asp Pro Thr Thr Pro Leu Ala Arg Ala Ala Trp Glu Thr Val
 385 390 395 400

Arg His Thr Pro Val Asn Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala
 405 410 415

Pro Thr Leu Trp Ala Arg Met Ile Leu Met Thr His Phe Phe Ser Ile
 420 425 430

Leu Leu Ala Gln Glu Gln Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr
 435 440 445

Gly Ala Cys Tyr Ser Ile Glu Pro Leu Asp Leu Pro Gln Ile Ile Glu
 450 455 460

Arg Leu His Gly Leu Ser Ala Phe Ser Leu His Ser Tyr Ser Pro Gly
 465 470 475 480

Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys Leu Gly Val Pro Pro
 485 490 495

Leu Arg Val Trp Arg His Arg Ala Arg Ser Val Arg Ala Lys Leu Leu
 500 505 510

Ser Gln Gly Gly Arg Ala Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp
 515 520 525

Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln
 530 535 540

Leu Asp Leu Ser Gly Trp Phe Val Ala Gly Tyr Asn Gly Gly Asp Ile
 545 550 555 560

Tyr His Ser Leu Ser Arg Ala Arg Pro Arg
 565 570

<210> 9
 <211> 1222
 <212> PRT
 <213> secuencia artificial

5

<220>
 <223> fusión de los restos de NS4A, NS3 y NS5B de HCV JA

10

<400> 9

ES 2 387 321 T3

Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser Gly Ser Gly
 1 5 10 15

Ser Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly
 20 25 30

Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Asp Gly
 35 40 45

Glu Val Gln Val Leu Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys
 50 55 60

Val Asn Gly Val Cys Trp Thr Val Tyr Ala Gly Ala Gly Ser Lys Thr
 65 70 75 80

Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp
 85 90 95

Gln Asp Leu Val Gly Trp Pro Ala Pro Pro Gly Ala Arg Ser Met Thr
 100 105 110

Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala
 115 120 125

Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu
 130 135 140

Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu
 145 150 155 160

Leu Cys Pro Ser Gly His Val Val Gly Ile Phe Arg Ala Ala Val Cys
 165 170 175

Thr Arg Gly Val Ala Lys Ala Val Asp Phe Ile Pro Val Glu Ser Met
 180 185 190

Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro
 195 200 205

Ala Val Pro Gln Thr Phe Gln Val Ala His Leu His Ala Pro Thr Gly
 210 215 220

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

ES 2 387 321 T3

Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu
 450 455 460

Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly
 465 470 475 480

Ala Thr Gly Arg Gly Arg Ser Gly Ile Tyr Arg Phe Val Thr Pro Gly
 485 490 495

Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr
 500 505 510

Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val
 515 520 525

Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp
 530 535 540

His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp
 545 550 555 560

Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr
 565 570 575

Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro
 580 585 590

Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr
 595 600 605

Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn
 610 615 620

Glu Ile Thr Leu Thr His Pro Ile Thr Lys Phe Val Met Ala Cys Met
 625 630 635 640

Ser Ala Asp Leu Glu Val Val Gly Ser Gly Ser Gly Ser Met Ser Tyr
 645 650 655

Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Ser Lys
 660 665 670

ES 2 387 321 T3

Leu Pro Ile Asn Pro Leu Ser Asn Ser Leu Leu Arg His His Ser Met
 675 680 685
 Val Tyr Ser Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val
 690 695 700
 Thr Phe Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu
 705 710 715 720
 Lys Glu Met Lys Ala Lys Ala Ser Thr Val Lys Ala Arg Leu Leu Ser
 725 730 735
 Ile Glu Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Lys Ser Lys
 740 745 750
 Phe Gly Tyr Gly Ala Lys Asp Val Arg Ser Leu Ser Ser Arg Ala Val
 755 760 765
 Asn His Ile Arg Ser Val Trp Glu Asp Leu Leu Glu Asp Thr Glu Thr
 770 775 780
 Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln
 785 790 795 800
 Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp
 805 810 815
 Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser
 820 825 830
 Thr Leu Pro Gln Ala Val Met Gly Pro Ser Tyr Gly Phe Gln Tyr Ser
 835 840 845
 Pro Gly Gln Arg Val Glu Phe Leu Val Asn Thr Trp Lys Ser Lys Lys
 850 855 860
 Cys Pro Met Gly Phe Ser Tyr Asn Thr Arg Cys Phe Asp Ser Thr Val
 865 870 875 880
 Thr Glu Asn Asp Ile Arg Thr Glu Glu Ser Ile Tyr Gln Cys Cys Asp
 885 890 895
 Leu Ala Pro Glu Ala Arg Gln Ala Ile Lys Ser Leu Thr Glu Arg Leu

ES 2 387 321 T3

Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala Ser Cys Leu Arg Lys
 1130 1135 1140

Leu Gly Val Pro Pro Leu Arg Val Trp Arg His Arg Ala Arg Ser
 1145 1150 1155

Val Arg Ala Lys Leu Leu Ser Gln Gly Gly Arg Ala Ala Thr Cys
 1160 1165 1170

Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu Lys Leu
 1175 1180 1185

Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Gly Trp Phe
 1190 1195 1200

Val Ala Gly Tyr Asn Gly Gly Asp Ile Tyr His Ser Leu Ser Arg
 1205 1210 1215

Ala Arg Pro Arg
 1220

<210> 10
 <211> 1249
 <212> PRT
 <213> secuencia artificial

<220>
 <223> fusión de los restos de NS4A, NS3, NS4B y NS5B de HCV JA

<400> 10

Met Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser Gly Ser Gly
 1 5 10 15

Ser Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly
 20 25 30

Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Lys Asn Gln Val Asp Gly
 35 40 45

Glu Val Gln Val Leu Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys
 50 55 60

Val Asn Gly Val Cys Trp Thr Val Tyr Ala Gly Ala Gly Ser Lys Thr

5

10

ES 2 387 321 T3

Ile Cys Asp Glu Cys His Ser Thr Asp Trp Thr Thr Ile Leu Gly Ile
305 310 315 320

Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val
325 330 335

Leu Ala Thr Ala Thr Pro Pro Gly Ser Ile Thr Val Pro His Pro Asn
340 345 350

Ile Glu Glu Val Ala Leu Ser Asn Thr Gly Glu Ile Pro Phe Tyr Gly
355 360 365

Lys Ala Ile Pro Ile Glu Ala Ile Lys Gly Gly Arg His Leu Ile Phe
370 375 380

Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Thr Gly
385 390 395 400

Leu Gly Leu Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val
405 410 415

Ile Pro Thr Ser Gly Asp Val Val Val Val Ala Thr Asp Ala Leu Met
420 425 430

Thr Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys
435 440 445

Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu
450 455 460

Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly
465 470 475 480

Ala Thr Gly Arg Gly Arg Ser Gly Ile Tyr Arg Phe Val Thr Pro Gly
485 490 495

Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr
500 505 510

Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Thr Val
515 520 525

ES 2 387 321 T3

Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp
 530 535 540

His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp
 545 550 555 560

Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr
 565 570 575

Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro
 580 585 590

Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr
 595 600 605

Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn
 610 615 620

Glu Ile Thr Leu Thr His Pro Ile Thr Lys Phe Val Met Ala Cys Met
 625 630 635 640

Ser Ala Asp Leu Glu Val Val Ser Leu Met Ala Phe Thr Ala Ser Ile
 645 650 655

Thr Ser Pro Leu Thr Thr Gln Asn Thr Leu Leu Phe Asn Ile Leu Gly
 660 665 670

Gly Trp Val Ala Ala Gln Leu Ser Met Ser Tyr Thr Trp Thr Gly Ala
 675 680 685

Leu Ile Thr Pro Cys Ala Ala Glu Glu Ser Lys Leu Pro Ile Asn Pro
 690 695 700

Leu Ser Asn Ser Leu Leu Arg His His Ser Met Val Tyr Ser Thr Thr
 705 710 715 720

Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe Asp Arg Leu
 725 730 735

Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu Met Lys Ala
 740 745 750

ES 2 387 321 T3

Lys Ala Ser Thr Val Lys Ala Arg Leu Leu Ser Ile Glu Glu Ala Cys
 755 760 765
 Lys Leu Thr Pro Pro His Ser Ala Lys Ser Lys Phe Gly Tyr Gly Ala
 770 775 780
 Lys Asp Val Arg Ser Leu Ser Ser Arg Ala Val Asn His Ile Arg Ser
 785 790 795 800
 Val Trp Glu Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile Asp Thr Thr
 805 810 815
 Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly
 820 825 830
 Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val
 835 840 845
 Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu Pro Gln Ala
 850 855 860
 Val Met Gly Pro Ser Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val
 865 870 875 880
 Glu Phe Leu Val Asn Thr Trp Lys Ser Lys Lys Cys Pro Met Gly Phe
 885 890 895
 Ser Tyr Asn Thr Arg Cys Phe Asp Ser Thr Val Thr Glu Asn Asp Ile
 900 905 910
 Arg Thr Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala
 915 920 925
 Arg Gln Ala Ile Lys Ser Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro
 930 935 940
 Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala
 945 950 955 960
 Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu
 965 970 975
 Lys Ala Thr Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp Cys Thr Met

ES 2 387 321 T3

		980						985						990	
Leu	Val	Asn	Gly	Asn	Asp	Leu	Val	Val	Ile	Cys	Glu	Ser	Ala	Gly	Thr
		995					1000							1005	
Gln	Glu	Asp	Ala	Ala	Ser	Leu	Arg	Val	Phe	Thr	Glu	Ala	Met	Thr	
	1010					1015					1020				
Arg	Tyr	Ser	Ala	Pro	Pro	Gly	Asp	Pro	Pro	Gln	Pro	Glu	Tyr	Asp	
	1025					1030					1035				
Leu	Glu	Leu	Ile	Thr	Ser	Cys	Ser	Ser	Asn	Val	Ser	Val	Ala	His	
	1040					1045					1050				
Asp	Ala	Ser	Gly	Lys	Arg	Val	Tyr	Tyr	Leu	Thr	Arg	Asp	Pro	Thr	
	1055					1060					1065				
Thr	Pro	Leu	Ala	Arg	Ala	Ala	Trp	Glu	Thr	Val	Arg	His	Thr	Pro	
	1070					1075					1080				
Val	Asn	Ser	Trp	Leu	Gly	Asn	Ile	Ile	Met	Tyr	Ala	Pro	Thr	Leu	
	1085					1090					1095				
Trp	Ala	Arg	Met	Ile	Leu	Met	Thr	His	Phe	Phe	Ser	Ile	Leu	Leu	
	1100					1105					1110				
Ala	Gln	Glu	Gln	Leu	Glu	Lys	Ala	Leu	Asp	Cys	Gln	Ile	Tyr	Gly	
	1115					1120					1125				
Ala	Cys	Tyr	Ser	Ile	Glu	Pro	Leu	Asp	Leu	Pro	Gln	Ile	Ile	Glu	
	1130					1135					1140				
Arg	Leu	His	Gly	Leu	Ser	Ala	Phe	Ser	Leu	His	Ser	Tyr	Ser	Pro	
	1145					1150					1155				
Gly	Glu	Ile	Asn	Arg	Val	Ala	Ser	Cys	Leu	Arg	Lys	Leu	Gly	Val	
	1160					1165					1170				
Pro	Pro	Leu	Arg	Val	Trp	Arg	His	Arg	Ala	Arg	Ser	Val	Arg	Ala	
	1175					1180					1185				
Lys	Leu	Leu	Ser	Gln	Gly	Gly	Arg	Ala	Ala	Thr	Cys	Gly	Lys	Tyr	
	1190					1195					1200				

ES 2 387 321 T3

Leu Phe Asn Trp Ala Val Arg Thr Lys Leu Lys Leu Thr Pro Ile
 1205 1210 1215

Pro Ala Ala Ser Gln Leu Asp Leu Ser Gly Trp Phe Val Ala Gly
 1220 1225 1230

Tyr Asn Gly Gly Asp Ile Tyr His Ser Leu Ser Arg Ala Arg Pro
 1235 1240 1245

Arg

- <210> 11
- <211> 3666
- 5 <212> ADN
- <213> secuencia artificial

- <220>
- 10 <223> secuencia nucleotídica que codifica la fusión NS4A-3-5B de HCV JA optimizada para la expresión en células de mamíferos

<400> 11

```

atggggcagcg tgggtgattgt gggccggatc atcctgagcg gcagcggcag cgcctccatc      60
accgcttaca gccagcagac cagaggcctg ctgggctgta tcatcaccag cctgaccggc      120
agagacaaga atcagggtga cggcagagtg cagggtgctgt ccaccgccac ccagagcttc      180
ctggccaact gtgtgaatgg cgtgtgttgg accgtgtacg ccggagccgg cagcaagacc      240
ctggccggac ccaagggcc catcaccag atgtacacca acgtggacca ggacctggtg      300
ggctggcctg cccctcctgg cgcagaagc atgaccctt gtacctgtgg cagcagcagc      360
ctgtaectgg tgaccagaca cgcgatgtg atcctgtga ggaggagagg cgatagcaga      420
ggcagcctgc tgtctectag acccgtgtc tacctgaggg gcagcagcgg cggaccctg      480
ctgtgcccca gcggccacgt ggtgggcac ttcagagccg ccgtgtgtac cagaggcgtg      540
gccaaaggcc tggatttcat cccctggag agcatggaga ccaccatgag gagccctg      600
ttcaccgaca atagcagccc cctgcccgtg cctcagacct tccaggtggc ccacctgac      660
gccccaccg gctccggcaa gaggaccaag gtgccagccg cctacgccgc ccagggctac      720
aaggtgctgg tgttgaatcc cagcgtggcc gccaccctgg gcttcggcgc ctacatgagc      780
aaggcccacg gcctcagacc caatatccgg accggagtga ggaccatcac cacaggcggc      840
cctatcacct acagcgccta cggcaagttc ctggccgacg gcggctgtag cggcggagcc      900
    
```

taaggacatca tcatctgtga cgagtgccac agcaccgatt ggaccaccat cctgggcatc 960
 ggcaccgtgc tggaccaggg ccgagaccgcc ggagccagac tgggtggtgct ggcacacagcc 1020
 acccccctg gcagcatcac cgtgccccac cccaacatcg aggaggtggc cctgagcaac 1080
 accggcgaga tccccctcta cggcaaggcc atccctatcg aggccatcaa gggcggcaga 1140
 cacctgatct tctgccacag caagaagaag tgtgaccgagc tggccgccaa gctgaccggc 1200
 ctgggcctga acgccgtggc ctactacaga ggccctggagc tgtccgtgat cctaccagc 1260
 ggcgatgtgg tgggtggtggc caccgacgcc ctgatgaccg gcttcaccgg cgatttcgac 1320
 agcgtgatcg actgtaatac ctgtgtgacc cagaccgtgg acttcagcct ggacccccacc 1380
 ttcaccatcg agaccaccac cgtgccacag gatgccgtgt ccagaagcca gagaagaggg 1440
 gccaccggca gaggcagaag cggcatctac agattcgtga cccctggcga gagaccacagc 1500
 ggcattgttc atagcagcgt gctgtgtgag tgcacgacg ccggtctgtc ctggtacgag 1560
 ctgacccccag ccgagaccac agtgaggctg agggcctacc tgaacacccc tggcctgcct 1620
 gtgtgccagg atcacctgga gttctgggag agcgtgttta ccggcctgac ccacatcgat 1680
 gcccacttct tgagccagac caagcaggcc ggcgacaact tccccacct ggtggcctac 1740
 caggccaccg tgtgtgccag agcccaggcc cctcccccca gctgggacca gatgtggaag 1800
 tgccctgatca ggctgaagcc caccctgcac ggcctacc cctgctgta cagactgggc 1860
 gccgtgcaga atgagatcac cctgaccac cctatcacca agttcgtgat ggcctgtatg 1920
 agcggcgacc tggaggtggt gggcagcggc tccggctcca tgagctacac ctggacaggg 1980
 gccctgatca caccctgtgc ccggaggag agcaagctgc ccataacccc cctgagcaat 2040
 agcctgctga ggcaccacag catggtgtac agcaccacct ccagaagcgc cagcctgagg 2100
 cagaagaagg tgaccttcca caggctgcag gtgctggacg accactacag ggacgtgctg 2160
 aaggagatga aggcccaagg cagcaccgtg aaggccagac tgctgtctat cgaggaggcc 2220
 tgtaagctga cccccctca cagcgccaag agcasgttcg gctacggcgc caaggatgtg 2280
 agaagcctga gcagcagagc cgtgaaccac atccggtctg tgtgggagga tctgctggag 2340
 qataccgaga cccccatcga caccacaatc atggccaaga acgaggtgtt ctgctgacg 2400
 cccgagaagg gcggaagaaa gcccgccagg ctgatactgt tccctgacct gggagtgaga 2460
 gtgtgtgaga agatggccct gtaagacgtg gtgtccaccc tgcgccaggg cgtgatgggc 2520
 cccagctacg gcttcacgta cagccctggc cagagagtgg agttcctggt gaacacctgg 2580

```

aagagcaaga aatgccccat gggcttcagc tacaacaccc ggtgcttcga cagcacagtg 2640
accgagaacg acatcaggac cgaggagtcc atctaccagt gctgtgacct ggcctccgag 2700
gccagacagg ccatacaaaag cctgaccgag cggctgtaca tcggcggacc tctgaccaac 2760
agcaagggcc aqaactgtgg ctacagcaga tgtaggcca gcggcgtgct gaccacctct 2820
tgtggcaaca cctgacctg ttacctgaag gccaccgccc cctgtagagc cgcacaactg 2880
caggactgta ccattgctgg gaacggcaac gaacctgggg tgatctgtga gacgcgggc 2940
accagggagg atgccgcctc cctgagagtg tttaccgagg ccattgaccag atacagccc 3000
cctcccgccg accctcccca gcccgagtac gatctggagc tgatcaccag ctgctccagc 3060
aatgtggggc tygctcagca cgcagcggc aagagagtgt actacctgac cagggacccc 3120
accacccctc tggccagagc cgcctgggag acagtgagac acacccccgt gaacagctgg 3180
ctgggcaaca tcattatgta cgcctacc ctgtggcca gaatgacct gatgaccac 3240
ttcttcagca tctgtctggc ccaggagcag ctggagaagg ccctggactg ccagatctac 3300
ggcgcctgct acagcctcga gctctggac ctgcctcaga tcattgagag actgcacggc 3360
ctgagcgcct tcagcctgca cagctacagc ccaggcgaga tcaatagagt ggcagctgc 3420
ctgagaaagc tggcgtgccc acctctgaga gtgtggcggc acagagccag atctgtgagg 3480
gccaaqctgc tgtcccaggc cggcagggcc gccacctgtg gcaagtacct gttcaactgg 3540
gccgtgagga caaagctgaa gctgacaccc atccctgccc ccagccagct ggacctgagc 3600
ggttggctcg tggccgcta caatgcccgc gacatctacc acagcctgtc cagggccagg 3660
cctaga 3666

```

<210> 12
 <211> 3747
 5 <212> ADN
 <213> secuencia artificial

<220>
 10 <223> secuencia nucleotídica que codifica la fusión NS4A-3-4B-5B de HCV JA optimizada para la expresión en células de mamíferos

```

<400> 12
atgggcagcg tggtgattgt gggccggatc atccctgagc gcagcggcag cgcctccatc 60
accgcctaca gccagcagac cagagccctg ctgggctgta tcattaccag cctgaccggc 120
agagacaaga atcaggtgga cggcgagggt cagggtgctgt ccaccgccac ccagagcttc 180
ctggccacct gtgtgaatgg cgtgtgttgg accgtgtacg ccggagccgg cagcaagacc 240

```

ctggccggac ccaagggcc catcaccag atgtacacca acgtggacca ggacctggtg 300
ggetgqctg cccctcctg cgcagaagc atgacccctt gtacctgtgg cagcagcgac 360
ctgtacctg tgaccagaca cgcgatgtg atccctgtga ggaggagagg cgtatgcaga 420
ggcagcctg tgtctcctag acccgtgtcc tacctgaagg gcagcagcgg cggaccctg 480
ctgtgcccc gcygccacgt ggtgggcac ttcagagccg ccgtgtgtac cagaggcgtg 540
gccaaggccg tggatttcat ccccggtggg agcatggaga ccaccatgag gagccccgtg 600
ttcacggaca atagcagccc ccctgcccgt cctcagacct tccaggtggc ccacctgcac 660
gcccccccg gctccggca gagcaccag gtgcacggc cctacggcc ccagggtac 720
aagggtctg tgcgtaacc cagcgtggc gccaccctg gcttcggcc ctacatgagc 780
aaggcccac gcatcgagcc caatatccg accggagtga ggacctcac cacaggcggc 840
cctatcacct acagcgccta cggcaagtc ctggccgac gggctgtag cggcggagcc 900
tcgacatca tcactctga caggtgccac agcaccgatt ggaccaccat cctgggcac 960
ggcaccgtg tggaccagg cagagccgc ggagccagac tgggtggtg ggcacagcc 1020
acccccctg gcagcatcac cgtgccccac cccaacatc aggaggtggc cctgagcaac 1080
accggcgaga tccccctc cggcaaggc atccctatc aggccatcaa gggcggcaga 1140
cacctgatct tctgccacg caagaagaag tgtgacgagc tggccgcaa gctgaccggc 1200
ctgggcctga acgcctggc ctactacaga ggcctggac tgtccgtgat cctaccagc 1260
ggcgtatgtg tggtygtgg caccgacgcc ctgatgacc gcttcaccg cgtttcagc 1320
agcgtgatc actgtaata ctgtgtgacc cagaccgtg acttcagcct ggacccacc 1380
ttcacatc agaccaccac cgtgccacg gatgcccgtt ccaqagcca gagaagggc 1440
gcbaccggca gaggcagaag cggcatctac agattcgtg cccctggcga gagaccagc 1500
ggcatgttc atagcagcgt gctgtgtgag tgcacgacg ccggtgtgc ctggtacgag 1560
ctgaccccag ccgagaccac agtgaggctg agggcctacc tgaacacccc tggcctgct 1620
gtgtgccagg atcacctgga gttctgggag agcgtgtta ccggcctgac ccacatcgt 1680
gcccacttc tgagccagac caagcagcc cgcgcaact tcccctacct ggtggcctac 1740
caggccaccg tgtgtgccag agcccaggcc cctccccca gctgggacca gatgtggaag 1800
tgctgatca ggtgaagcc caccctgcac ggcctacc cctgctgta cagactggc 1860
gccgtgcaga atgagatcac cctgacccac cctatcaca agttcgtgat ggcctgtatg 1920
agcggcagc tggaggtgt gtcctgatg gccctaccg ccagcatcac aagccccctg 1980

accaccaga ataccctgct gttcaacatc ctggggggct gggtgccgc ccagctgtcc 2040
 atgagctaca cctggacagg cgcctgctc acaccctgtg ccgcccagga gagcaagctg 2100
 cccatcaacc cctgagcaa tagcctgctg aggcaccaca gcatggtgta cagcaccacc 2160
 tccagaagcg ccagcctgag gcagaagaag gtgaccttcg acaggctgca ggtgctggac 2220
 gaccactaca gggacgtgct gaaggagatg aaggccaagg ccagcaccgt gaaggccaga 2280
 ctgctgtcta tegaggaggc ctgtaagctg acccccctc acagcgccaa gagcaagtcc 2340
 ggtacggcg ccaaggatgt gagaagcctg agcagcagag ccgtgaacca catccggtct 2400
 gtgtgggagg atctgctgga ggataccgag acccccctc acaccacaat catggccaag 2460
 aacgaggtgt tctgctgca gcccgagaag ggcggaaga agcccgcag gctgatctg 2520
 ttccctgacc tgggagtgag agtgtgtgag aagatggccc tgtacgact ggtgtccacc 2580
 ctgcccagg ccgtgatggg ccccagctac ggttccagt acagccctgg ccagagagtg 2640
 gagtccctgg tgaacacctg gaagagcaa aatgcccc tgggcttcag ctaccacc 2700
 ccgtgcttcg acagcacagt gaccgagaac gacatcagga ccgaggagt ccctaccag 2760
 tctgtgacc tggccccga gccagacag cccatcaaaa gctgaccga gcgctgtac 2820
 atcggggac ctctgaccaa cagcaaggc cagaactgtg gctacagaag atgtagggcc 2880
 agcggcgtg tgaaccctc ttgtggcaac accctgacct gttacctga gccaccgcc 2940
 gctgtagag ccgccaaact gcaggactgt accatgctgg tgaacggcaa cgacctggtg 3000
 gtgatctgt agagcgccgg ccccaggag gatgcccct ccctgagagt gttaccgag 3060
 gccatgacca gatacagcgc cctcccggc gaccctccc agcccagta cgatctggag 3120
 ctgatccca gctgctccag caatgtggc gtggctccg acgccagcg caagagagtg 3180
 tactacctga ccagggacc caccaccct ctggccagag ccgctggga gacagtgaga 3240
 cacaccccg tgaacagctg gctggcaac atcatcatgt acgcccctac cctgtgggcc 3300
 agaatgatcc tgatgacca ctcttcagc atcctgctgg ccaggagca gctggagaag 3360
 gccctggact gccagatcta cggcgcctg tacagcatcg agcctctgga cctgcccag 3420
 atcctcaga gactgcacgg cctgagccc ttccgctgc acagctacag cccaggcgag 3480
 atcaatagag tggccagctg cctgagaag ctgggctgc cccctctgag agtgtggcg 3540
 cacagagca gatctgtgag ggccaagctg ctgtcccag gggcagggc cggccctgt 3600
 ggcaagrac tgtcaactg gcccgtgagg acaaagctga agctgacacc catccctgcc 3660

 gccagccagc tggacctgag ccgctggctc gtggccgct acaatggcg gcacatctac 3720
 cacagcctgt ccagggccag gcctaga 3747

ES 2 387 321 T3

<210> 13
 <211> 3666
 <212> ADN
 <213> secuencia artificial

5

<220>
 <223> secuencia nucleotídica que codifica la fusión NS4A-3-5B de HCV JA optimizada para la expresión en Pichia

<400> 13

```

atgggggccg ttgttatcgt tggtagaato atcttgtctg gttctggttc cgtccaatt    60
actgcttact cccagcagac tagaggattg ttgggttcta tcatcacttc cttgactggt    120
agagacaaga accaagtga cggagaggtt cagggtttgt ccactgctac tcagtcttcc    180
ttggctactt gtgttaacgg tgtttgttgg actgtttacg ctggtgctgg ttctaaaact    240
ttggctgggc caaagggtcc aatcactcag atgtacacas acgtcgacca ggatttgggt    300
ggttggccag ctccaccagg tgctagatcc atgactccat gtacttgtgg ttectccgac    360
ttgtacttgg ttactagaca cgtcgacgtt atcccagtta gaagaagagg agactccaga    420
ggatctttgt tgtcccaag accagtttct tacttgaagg gatcttccgg tggttccattg    480
ttgtgtccat cgggtcacgt tgttggkatt ttcagagctg ctgtttgtac tagagggttt    540
gctaaaggctg ttgacttcat cccagttgag tccatggaga ctactatgag atccccagtt    600
ttcactgaca actcttcccc acctgctgtt ccacaaactt tccaagttgc tcacctgcat    660
gtccaaactg gttctggtaa gtccactaag gttccagctg cttacgctgc tcaaggttac    720
aagggttttg ttttgaacct atccgttget gctactttgg gtttcgggtgc ttacatgtct    780
aaggctcacg gtattgagcc aaacatcaga actgggtgta gaactatcac tactggtggt    840
cctattactt actccgctta cggaaagttt ttggetgacg gtggttgttc tggtygtget    900
tacgacatca tcactctgtg cgggtgtcac tctactgact ggactactat cttgggtatc    960
ggtactgttt tggaccaagc tgaacttctt ggtgctagat tggttgtttt ggtactgct    1020
actccaccag gttccattac tgttccacac ccaaacatcg aggaagttgc tttgtctaac    1080
actggagaga tcccattcta cggaaaggct atcccattg aggetatcaa gggtygtaga    1140
cacttgattt tctgtcactc caagaagaag tgtgacgagt tggctgctaa gttgactgga    1200
ttgggattga acgctgttgc ttactacaga ggattggacg ttcccgkctt cccaacttcc    1260
    
```

10

ggtgatggtg ttgttqtgqc tactgacgct ttgatgactg gtttcactgq tgacttccgac 1320
 tccggtatcg actgtaacac ttgtgttact cagactggtg acttctcctt ggacccaact 1380
 ttactatcg agactactac tgttctcaa gacgctgttt ccagatccca aagaagaggt 1440
 gctactggta gaggaagatc cggatcttac agattcgtta ctccaggga aagaccatct 1500
 ggaatggtcg actcctccgt tttgtgtgaa tgttacgacg ctggttgtgc ttggtacgaa 1560
 ttgactccag ctgagactac tgttagattg agagcttact tgaaactcc aggattgcca 1620
 gtttgcacg accacttggg attctgggag tccgtttca ctggattgac tcacattgac 1680
 gctcactttt tgtcccaaac taagcaggtt ggtgacact ttccatactt ggttgcctac 1740
 caggctactg tttgtgctag agcacaagct ccaccaccat cttgggatca gatgtggaag 1800
 tgtttgatca gattgaagcc aactttgcac ggtccaactc cattgttcta cagattgggt 1860
 gctgttcaga acgagatcac tttgactcac ccaatcacta agttcgttat ggtttgcctg 1920
 tctgctgact tggaaagtgt tggttccgga tctggttcca tgtctacac ttggactggt 1980
 gctttgatca ctccatgtgc tgcigaagaa tccaagttgc caatcaacc attgtccaac 2040
 tcttgttga gacaccactc catggtttac tccactactt ccagatccgc ttcttgaga 2100
 cagaagaag ttacattcga cagattgcag qttttggacg accactacag agatgttttg 2160
 aaggagatga aggctaaggc ttccactggt aaggctagat tgttgtccat tgaggaggct 2220
 tghtaagtga ctccaccaca ctctgctaag tccaagtttg gttacggtgc taaggatgtt 2280
 agatccttgt cctccagagc tghtaaccac atcagatccg ttggggagga tttgttggag 2340
 gacactgaga ctccaatcga cactactatc atggttaaga acgaggtttt ctgtgttcaa 2400
 ccagagaag gtggaagaaa gccagctaga ttgatcgttt tcccagactt ggggtgttga 2460
 gtttgtgaga agatggcttt gtacgacgtt gtttccactt tcccacagc ttttatgga 2520
 ccactttacg gtttccaata ctccccaggc caaagagttg agttcttggg taacacttgg 2580
 aagttcaaga aatgtccaat gggattctcc tacaacacta gatgtttcga ctccactgtt 2640
 actgagaacg acatcagaac agaggagtc atctaccagt gttgtgactt ggctccagaa 2700
 gctagacaag ctatcaagtc cttgactgag agattgtaca tcggtgtcc tttgactaac 2760
 tccaaggac agaactgtgg ttacagaaga tgtagagctt ccggtgttt gactacttcc 2820
 tgtggttaaca ctttgacttg ttacttgaag gctactgctg cttgttagagc tgctaaattg 2880
 caggactgta ctatgttggg taacggtaac gacttggttg ttatctgtga gtccgctggg 2940
 actcaagag atgctgttcc tttgagagtt tccacagagg ctatgactag atactctgct 3000

```

ccacctggtg atccaccaca accagaatac gacttggagt tgatcacttc ctgttctctc 3060
aatgttgggtg ttgctcaccg tgcctccgga aagagaqttt actacttgac tagagaccca 3120
actactecat tggctagagc tgcctgggaa actgttagac acactccagt taactcctgg 3180
ttgggtaaca tcatcatgta cgtcccaact ttgtgggcta gaatgatctt gatgactcac 3240
ttcttctcca tcttgttggc tcaagagca ttggaaaagg ctttggactg tcagatttac 3300
ggtgcttgtt actccattga gccattggac ttgccacaga tcattgagag attgcacggg 3360
ttgtctgctt tctcttggc ctcttactcc cctgggaaa tcaacagagt tgcttctgt 3420
ttgagaaagt tgggtgttcc accattgaga gtttggagac acagagctag atccgttaga 3480
gctaagtgtt tgtcccaagg tggagagct gctacttgtg gtaagtactt gttcaactgg 3540
gctgttagaa caaagttaga gttgactcct attctgtctg ctcccaatt ggatttgtcc 3600
ggttggtttg ttgctggta caacggtggt gacatctacc actcttgtc cagagctaga 3660
ccaaga 3666

```

<210> 14
 <211> 3747
 5 <212> ADN
 <213> secuencia artificial

<220>
 10 <223> secuencia nucleotídica que codifica la fusión NS4A-3-4B-5B de HCV JA optimizada para la expresión en Pichia

<400> 14

```

atggggctccg ttgttatcgt tggtagaatc atcttgtctg gttctgggtc cgtcccaatt 60
actgcttact cccagcagac tagaggattg ttgggttqta tcatcacttc ctgtgactgg 120
agagacaaga accaagttag cggagaggtt caggttttgt cactgctac tcagtcttcc 180
ttggctactt gtgttaacgg tgtttgttgg actgtttacg ctggtgctgg tctataaact 240
ttggttggtc caaagggtcc aatcaactcag atgtacacaa acgttgacca ggatttgggt 300
ggttggccag ctccaccagg tgctagatcc atgactccat gtacttgtgg ttctccgac 360
ttgtacttgg ttaactagaca cgttgacgtt atcccagtta gaagaagagg agactccaga 420
ggatcttctg tgtccccaag accagtttct tacttgaagg gatcttccgg tgggtccattg 480
ttgtgtccat cgggtcacgt tgttggattt ttccagagctg ctgtttgtac tagaggtgtt 540
gctaaggctg ttgacttcat cccagttgag tccatggaga ctactatgag atccccagtt 600
ttcaactgac aactctcccc acctgctgtt ccacaaactt tccaagttgc tcaacttgcac 660

```

gctccaactg gttcttgtaa gtccactaag gttccagctg cttacgctgc tcaaggttac 720
 aaggcttttg ttttgaaccc atccggttget gctacttttg gtttcgggtgc ttacatgtct 780
 aaggctcacg gtattgagcc aaacatcaga actggtgcta gaactatcac tactggtggt 840
 cctattactt actccgctta cggaaagttt ttggtgacg gtggttgctc tgggtggtct 900
 tacgacatca tcactctgta cgagtgtcac tctactgact ggactactat cttgggtatc 960
 ggtactgttt tggaccaagc tgaactgct ggtgctagat tgggtgtttt ggtactgct 1020
 actccaccag gtccattac tgtccacac ccaaacatcg aggaagttgc tttgtctaac 1080
 actggagaga tcccattcta cggaaaggct atcccaattg aggtatcaa ggggtgtaga 1140
 cacttgattt tctgtcactc caagaagaag tgtgacgagt tggctgctaa gttgactgga 1200
 ttgggattga acgctgttgc ttactacaga ggattggacg tttccggtat cccaacttcc 1260
 ggtgatgttg ttgtgttgc tactgacgct ttgatgactg gtttcaactg tgaactcgac 1320
 tccgttatcg actgtaacac ttgtgttact cagactgttg acttctctt ggacccaact 1380
 ttcaactatg agactactac tgttctcaa gacgctgttt ccagatcca aagaagaggc 1440
 gctactggtg gaggaaagtc cggtatctac agattcgta ctccaggtga aagaccatct 1500
 ggaatgttcg actctccgt tttgtgtgaa tgttacgacg ctggttgtgc ttggtacgaa 1560
 ttgactccag ctgagactac tgttagattg agagottact tgaacactcc aggattgcca 1620
 gtttgtcaag accacttggg attctgggag tccgttttca ctggattgac tcacattgac 1680
 gctcactttt tgtcccaaac taagcagqct ggtgacaact ttccatactt ggttgcctac 1740
 caggctactg tttgtgctag agcacaagct ccaccaccat cttgggatca gatgtggaag 1800
 tgtttgatca gattgaagcc aactttgcac ggtccaactc cattgttga cagattgggt 1860
 gctgttcaga acgagatcac ttgactcac ccaatcacta agttcgttat ggcttgcctg 1920
 tctgtgact tggaaagttg ttcttqtatg gcttcaactg cttccattac ttccccattg 1980
 actactcaga acactttggt gttcaacatc ttgggaggat gggttgcagc tcaattgtcc 2040
 atgtctaca cctggactgg tgccttgatc actccatgct ctgctgaaga atccaagttg 2100
 ccaatcaacc cattgtccaa ctctttgttg agacaccact ccattggtta ctccactact 2160
 tccagatccg ctctcttgag acagaagaag gttacattcg acagattgca ggttttggac 2220
 gaccactaca gagatgtttt gaaggagatg aaggctaagg cttccactgt taaggctaga 2280
 ttgttgtcca ttgaggaggc ttgtaagttg actccaccac actctgctaa gtccaagttt 2340

```

ggttacqgtg ctaaggatgt tagatccttg tctccagag ctgtaacca catcaggtct 2400
gtttgggagg atttgttga ggacactgag actccaatcg acactactat catggctaag 2460
aacgaggttt tctgtgtca accagagaag ggtggaaga agccagctag attgatcgtt 2520
tcccagact tgggtgttag agtttgtgag aagatggctt tgtacgagct tgtttccact 2580
ttgccacagg ctgttatggg accctcttac ggtttccaat actccccagg acaaagagtt 2640
gagttccttg ttaacacttg gaagtccaag aatgtccaa tgggattctc ctacaacact 2700
agatgtttcg actccactgt tactgagaac gacatcaga cacaggaqtc catctaccag 2760
tqttgtgact tggctccaga agctagacaa gctatcaagt ccttgactga gagattgtac 2820
atcggcggtc ctttgactaa ctccaagga cagaactgtg gttacagaag atgtagagct 2880
tcgggtggtt tgactacttc ctgtgtaac acttgactt gttacttga ggctactgct 2940
gcttgtagag ctgctaaatt gcaggactgt actatgttg ttaacggtaa cgaactgggt 3000
gttatctgtg agtccgctgg tactcaagaa gatgctgctt ctttgagagt ttccacagag 3060
gctatgacta gatactctgc tccacctggt gatccaccac aaccagaata cgacttgag 3120
ttgatcactt cctgttctc caatgttgg gtgtctcag atgcttccgg aaagagagtt 3180
tactacttga ctagagacc aactactcca ttggctagag ctgcttggga aactgttaga 3240
cacactccag ttaactcctg gttgggtaac atcatcatgt acgctccaac tttgtgggt 3300
agaatgatct tgatgactca cttctctcc atcttgttg ctcaagagca attggaaaag 3360
gctttggact gtcagattta cgggtctgt tactccattg agccattgga cttgccacag 3420
atcattgaga gattgcaagg tttgtctgct ttctcttgc actcttactc ccttggtaga 3480
atcaacagag ttgcttctg tttgagaaag ttgggtgttc caccattgag agtttggaga 3540
cacagagcta gatccgttag agctaagtg ttgtcccaag gtggaagagc tgcacttct 3600
ggtaagtact tgttcaactg ggctgttaga acaaagtga agttgactcc tattctctct 3660
gcttcccaat tggatttgc cggttggtt gttgtgtgtt acaacgggtg tgacatctac 3720
cactcttgt ccagagctag accaaga 3747

```

<210> 15
 <211> 3666
 5 <212> ADN
 <213> secuencia artificial

<220>
 10 <223> secuencia nucleotídica que codifica la fusión NS4A-3-5B de HCV JA optimizada para la expresión en *E. coli*
 <400> 15

atggggagcg	ctgttattgt	tggccgcatt	attctgagcg	gtagcggtag	cgccccgatt	60
accgcctatt	ctcagcaaac	cegtggtctg	ctgggttga	ttattaccag	cctgaccggc	120
cgtgataaaa	atcaggtgga	tggcgaagt	caggttctga	gcaccgccac	ccagagcttt	180
ctggcgacct	gtgtgaatgg	tgtgtgctgg	accgtttatg	ccggtgcccg	tagcaaaacc	240
ctggcgggtc	cgaaagggtc	gattaccocg	atgtacacca	acgtggatca	ggatctggtt	300
ggttggccgg	cgccgcgggg	tgcccgtagc	atgaccccg	gtacctgtgg	tagcagcgat	360
ctgtatctgg	ttaccocgca	tgccgatggt	attccggttc	gtcgtcgtgg	tgatagccgt	420
ggtagcctgc	tgtctccggc	tcgggttagc	tatctgaaag	gtagcagcgg	tggtccgctg	480
ctgtgtccga	gcggtcatgt	ggtgggtatt	tttcgtgccg	ccgtttgtac	ccgtggtgtg	540
gcgaaagcgg	tggattttat	cccggttgaa	agcatggaaa	ccaccatgcg	tagccccgtg	600
ttaccgata	atagcagccc	gccggcggtt	ccgcagacct	ttcaggttgc	ccatctgcct	660
gcgccgaccg	gtagcggtaa	aagcaccaaa	gttccggcgg	cgtatgccgc	ccagggttat	720
aaagtgcctg	tgctgaatcc	gagcgtggcg	gcgaccctgg	gttttgggtc	ctatatgagc	780
aaagcccctg	gcattgaacc	gaacattcgt	accggcgctc	gtaccattac	caccgggtgg	840
ccgattacct	atagcgccta	cggcaaat	ctggcggtg	gtggctgtag	cggtgcccgc	900
tatgatatac	tcctctgtga	tgaatgccat	agcaccgatt	ggaccacct	tctgggtatt	960
ggcaccgctc	tggatcaggc	ggaaccgcc	ggtgcccgtc	tggttgttct	ggcgaccgca	1020
acgccgcggg	gtagcattac	cgtcccgcat	ccgaacattg	aagaagtggc	cctgagcaat	1080
accggcgaaa	ttccgtttta	tygcaagc	attccgatcg	aagcgattaa	aggcggccgt	1140
catctgattt	tttgccacag	caaaaaaaaa	tgtgatgaac	tggcggcgaa	actgaccggc	1200
ctgggtctga	atgccgtggc	gtattatcgt	ggtctggatg	tgagcgttat	tcggaccagc	1260
ggtgatgctg	ttgtgggtgg	gaccgatgoc	ctgatgaccg	gttttaccgg	cgattttgat	1320
agcgtgatcg	attgtaaac	ctgtgtgacc	cagaccgttg	attttagcct	ggaccogacc	1380
tttaccattg	aaaccaccac	cgtcccgocg	gatgccgtta	gccgtagcca	gcgtcgtggc	1440
gccaccggtc	gtggtcgtag	cggcatttat	cgtttttatg	cgccgggtga	acgtccgagc	1500
ggtatgcttg	atagcagcgt	gctgtgtgaa	tgttatgatg	ccggctgtgc	ctggtatgaa	1560
ctgaccocgg	cggaaacacc	cgttcgtctg	cgcgcgtatc	tgaatacgc	gggtctgccc	1620
glttgtcagg	atcatctgga	attctgggaa	agcgttttta	ccggcctgac	ccatattgat	1680

gcccattttc tgagccagac caaacagggc ggtgabaact ttccgtatct ggtggcgtat 1740
caggcgaccg tttgtgcccg tccccagggc ccgccgccga gctgggatca gatgtggaaa 1800
tgccgtattc gtcgaaacc gaccctgcat ggtccgaccc cgtgtgtgta tcgtctgggt 1860
gccgttcaga acgaaattac cctgaccctc ccgattacca aatttgtgat ggcgtgtatg 1920
agcgcggatc tggaaagtgt tggtagcggc agcggctcta tgagctatac ctggaccggc 1980
gccctgatta ccccggtgtc cgcggaagaa agcaaacctc cgattaaccc gctgtctaat 2040
agcctgtctc gtcctcatag catggtgtat agcaccacca gccgtagcgc cagcctgcgt 2100
cagaaaaaag tgaccttcga tcgtctgcaq gtgctggatg accattatcg tgatgtgctg 2160
aaaqaaatga aagcgaaagc gagcaccyct aaagcccgtc tgcgtctat tgaagaagcg 2220
tgtaaactga ccccgccgca tagcgccaaa agcaaatctg gctatggcgc caaagatgtt 2280
cgtagcctga gcagccgtc cgttaatcat attcgtagcg tctgggaaga tctgctgga 2340
gataccgaaa ccccgattga taccaccatc atggcgaaaa acgaagtgtt ttgtgttcag 2400
ccggaaaaag gtggtcgtaa accggcccgt ctgattgttt ttccggatct ggggtgtctg 2460
gtgtgtgaaa aaatggcgt gtaccgatgt gttagcacc ccgcccaggg ggttatgggt 2520
ccgagctatg gcttccagtc ttctccgggt cagcgtgttg aatttctggt gaacacctgg 2580
aaaagcaaaa aatgccgat gggcttcagc tataacacc cctgctttga tagcaccgtg 2640
accgaaaacg atatccgtac cgaagaaagc atttaccagt gctgtgatct ggcgccgga 2700
gcccgtcagg cgattaaaag cctgaccgaa cgcctgtata ttggcggtcc gctgaccaat 2760
aqcaaaagcc agaactgtgt ttatcgtcgt tgcctgcca gcggtgttct gaccaccagc 2820
tgtggtaata cctgacctg ctacctgaaa gcgaccgccg cctgtcgtgc cgccaaactg 2880
caggattgta ccatgctggt taacggcaat gatctgggtg tgatttgtga aagcgcggc 2940
accaggaag atgcccccag cctgcgcgtt ttaccgaaq cgatgaccgc ttatagcgc 3000
ccgcccgggt atccgccgca gccggaatat gatctggaac tgatcaccag ctgtagcagc 3060
aatgttgggt ttgcccatga tgcagcggc aaacgtgtgt attacctgac cgtgatccg 3120
accaccccgc tggcccgtgc cgcctgggaa accgttcgtc ataccccgt taatagctgg 3180
ctgggcaaca ttattatgta tgcccgaacc ctgtgggcc gtatgattct gatgaccac 3240
ttctttagca ctctgctggc ccaggaacag ctggaaaaag cgtcggattg ccagatttat 3300
ggcgcctgct atagcattga accgctggat ctgccgcaga ttattgaacg tctgcatgac 3360

ES 2 387 321 T3

```

ctgagcgccct ttagcctgca tagctactct ccgggtgaaa ttaatcgtgt ggcgagctgt 3420
ctgcgtaaac tgggtgttcc gccgctgcgt gtctggcgtc atcgtgcccg tagcgttcgt 3480
gccaaactgc tgtctcaggg tggccgtgcc gccacctgtg gtaatacct gtttaactgg 3540
gcggttcgta ccaaaactgaa actgaccccg attccggggc cgagccagct ggatctgagc 3600
ggttggtttg ttgccggtta taacggcggc gatctctatc atagcctgag ccgtgcccg 3660
ccgcgt 3666

```

<210> 16
 <211> 3747
 5 <212> ADN
 <213> secuencia artificial

<220>
 10 <223> secuencia nucleotídica que codifica la fusión NS4A-3-4B-5B de HCV JA optimizada para la expresión en *E. coli*

<400> 16

```

atggggagcg ttgttattgt tggccgcatt attctgagcg gtagcggtag cgcgccgatt 60
accgcctatt ctacgcaaac ccgtggtctg ctgggttgta ttattaccag cctgaccggc 120
cgtgataaaa atcagggtgga tggcgaagtt caggttctga gcaccgccac ccagagcttt 180
ctggcgacct gtgtgaatgg tgtgtgctgg accgtttatg ccggtgcccg tagcaaaacc 240
ctggcgggcc cgaaaggtec gattacccag atgtacacca acgtggatca ggatctggtt 300
ggttggcccg cgcgccgggg tgcctgtagc atgaccccg  gtacctgtgg tagcagcgat 360
ctgtatctgg ttaccctgca tgcctatggt attccgggtc gtcgtcgtgg tgatagccgt 420
ggtagcctgc tgtctccggc tccggtttagc tatctgaaag gtagcagcgg tggctccgtg 480
ctgtgtccga gcggtcatgt ggtgggtatt tttcgtgccg ccgtttgtac ccgtggtgtg 540
gcgaaagcgg tggattttat cccggttgaa agcatggaaa ccaccatgcg tagcccggtg 600
tttaccgata atagcagccc gccgggggtt ccgcagacct ttcaggttgc ccatctgcat 660
gcgcccaccg gtagcggtsa aagcaccaaa gttccggcgg cgtatgccgc ccagggttat 720
aaagtgtctg tgcctgaatcc gagcgtggcg gcgaccctgg gttttgggtg ctatatgagc 780
aaagcccatg gcattgacc gaacattcgt accggcggtc gtaccattac caccggtggc 840
ccgattacct atagcgccta cggcaaatct ctggcggatg gtagcgttag cgytggcgcc 900
tatgatatca tcatctgtga tgaatgccat agcaccgatt ggaccaccat tctgggtatt 960
ggcaccgkic tggatcaggc ggaaacggcc ggtgcccgtc tggttgttct ggcgaccgca 1020

```

acgccgccgg gtagcattac cgttccgcac ccgaacattg aagaagtggc cctgagcaat 1080
 accggcgaaa ttccgtttta tggcaasagcg attccgatcg aagcgattaa aggcggccgt 1140
 catctgattt ttggccacag caaaaaaaaaa tgtgatgaac tggcggcgaa actgaccggc 1200
 ctgggtctga atgcccgtggc gtattatcgt ggtctggatg tgaqcgttat tccgaccagc 1260
 ggtgatggtg ttgtggctggc gaccgatgcc ctgatgaccg gtcttaaccg cgattctgat 1320
 agcgtgatcg attgtaccac ctgtgtgacc cagaccggtg attttagcct ggaccgcacc 1380
 ttaccattg aaaccaccac cgttccgcag gatgccgtta gccgtagcca gcgtcgtggt 1440
 gccaccggtc gtggctctag cggcatttat cgttttctga cggcgggtga acgtccgagc 1500
 ggtatgtttg atagcagcgt gctgtgtgaa tgttatgatg ccggctgtgc ctggtatgaa 1560
 ctgaccccg cggaaaccac cgttcgtctg ccgcgcgata tgaatacgc ccggctctgcg 1620
 gtttctcagg atcactctga attctgggaa agcgttttta ccggcctgac ccatattgat 1680
 gcccaatttc tgaqccagac caaacaggcg ggtgataact ttccgtatct ggtggcgtat 1740
 caggcgaccg ttgtgcccg tgcaccggcg ccgcgcgcga gctgggatca gatgtggaaa 1800
 tgctgattc gtctgaaacc gaccctgcat ggtccgacc cgtctctgta tegtctgggt 1860
 gccgttcaga acgaaattac cctgaccat ccgatacca aatttctgat ggcgtgatg 1920
 agcgcgata tggaaattgt tagcctgatg gcgtttaccg ccagcattac cagcccgtg 1980
 accaccaga ataccctgct gtttaacatc ctgggcggtt gggctggcgc ccagctgct 2040
 atgagctata cctggaccgg tgcctgatt accccgtgtg ccgccgaaga aagcaaatc 2100
 ccgattacc cgtctctaa tagcctgctg cgtcatcata gcattggtga tagcaccacc 2160
 agccgtagcg ccagcctgcg tcagaaaaaa gtgaccttcg atcgtctgca ggtgctggat 2220
 gatcattatc gtgatgtgct gaaagaaatg aagcgaaaag ccagcaccgt taagcccgt 2280
 ctgctgtcta ttgaagaagc gtgtaaatc accccgccgc atagcgcaa aagcaaatc 2340
 ggctatggcg ccaagatgt tctagcctg agcagccgtg ccgttaatca tattcgtagc 2400
 gtgtgggag atctgctgga agataccgaa accccgattg ataccaccat catggcgaaa 2460
 aacgaagtgt ttgtgttca gccgaaaaa ggtggctgta aaccggccc tctgattgct 2520
 ttcccgatc tgggtgttg tgtgtgtgaa aaaatggcgc tgtacgatgt tgttagcacc 2580
 ctgccgcagg cggttatggg tccgagctat ggctttcagt attctccggg tcagcgtgtt 2640
 gaatttctg tgaaccctg gaaaagcaaa aatgcccga tgggttcag ctataacc 2700
 cgtgctttg atagcaccgt gaccgaaac gatatccgta ccgaagaaag ctttaccag 2760

ES 2 387 321 T3

tgctgtgata tggcgccgga agcccgtcag gcgattaaaa gcctgaccga acgcctgtat 2820
 attggcggtc cgtgaccbaa tagcaaaaggc cagaactgtg gttatcgtcg ttgtcgtgcc 2880
 agcgggtgtc tgaccaccag ctgtggtaat accctgacct gctacctgaa agcgaccggc 2940
 gcctgtcgtg ccgccaaact gcaggattgt accatgctgg ttaacggcaa tgatctgggtg 3000
 gtgatttgtg aaagcgcggg caccagga gaagccgcca gcctgcgctg tttaccgaa 3060
 gcqatgaccg gttatagcgc ccgcgggggt gatccggcgc agccgggata tgatctgaa 3120
 ctgatcaccg gctgtagcag caatgttggg gttgcccatg atgccagcgg taaacgtgtg 3180
 tattacctga ccgctgatcc gaccaccccg ctggcccgtg ccgcctggga aaccgttcgt 3240
 catacccggg ttaatagctg gctgggcaac attattatgt atgccccgac cctgtggggc 3300
 cgtatgattc tgatgaccba cttctttagc attctgtctg ccaggaaca gctggaaaa 3360
 gcgctggatt gccagattta tggcgcctgc tatagcattg aaccgctgga tctgccgcaq 3420
 attattgaac gctcgcctgg cctgagcgc tttagcctgc atagctactc tccgggtgaa 3480
 attaatcgtg tggcgagctg tctgcgtaa ctgggtgttc cgcgcctgctg tctctggcgt 3540
 catcgtgccg gtagcgttcg tgccaaactg ctgtctcagg gtggccgtgc cggcacctgt 3600
 ggtaaatacc tgtttaactg ggcggttcgt accaaactga aactgacccc gattccggcg 3660
 gcgagccagc tggatctgag cggttggttt gttgccggtt ataacggcgg cgatatctat 3720
 cataqccctg gccgtgcccg tccgcgt 3747

5 <210> 17
 <211> 46
 <212> ADN
 <213> secuencia artificial

10 <220>
 <223> oligonucleótido para mejorar la secuencia de consenso de kozak
 <400> 17

atgtacgtcg accaccatgg gcagcgtggt gattgtggc cggatc 46

15 <210> 18
 <211> 41
 <212> ADN
 <213> secuencia artificial

20 <220>
 <223> oligonucleótido para crear codones de parada
 <400> 18

25 catgtagcgg ccgctcatca tctaggcctg gccctggaca g 41

30 <210> 19
 <211> 36
 <212> ADN
 <213> secuencia artificial

<220>

ES 2 387 321 T3

<223> cebador

<400> 19

5 ggggggctag cgccacatg ggcagcgtgg tgattg

36

<210> 20

<211> 34

<212> ADN

10 <213> secuencia artificial

<220>

<223> cebador de PCR

15 <400> 20

gggggggtac cctcatcatc taggctggc cctg

34

REIVINDICACIONES

- 5 1. Proteína de fusión aislada, que comprende al menos tres polipéptidos NS, entre ellos los polipéptidos NS3, NS4A y NS5B, que se originan a partir de un virus de la hepatitis C (VHC), estando configurados dichos polipéptidos NS en dicha proteína de fusión en un orden que es distinto del orden en el que aparecen en la configuración nativa, estando situado el polipéptido NS4A en el término N y estando situado NS5B en el término C de dicha proteína de fusión.
- 10 2. Proteína de fusión aislada según la reivindicación 1, en la que la fusión entre cada uno de los polipéptidos NS es directa o a través de un ligador.
3. Proteína de fusión aislada según la reivindicación 1 o 2, en la que todos los polipéptidos NS se originan a partir de la misma cepa o aislado de VHC.
- 15 4. Proteína de fusión aislada según la reivindicación 1 o 2, en la que al menos dos de los polipéptidos NS se originan a partir de cepas o aislados de VHC diferentes.
- 20 5. Proteína de fusión aislada según cualquiera de las reivindicaciones 1 a 4, que comprende un polipéptido NS4A, un polipéptido NS3, un polipéptido NS4B y un polipéptido NS5B.
6. Proteína de fusión aislada según cualquiera de las reivindicaciones 1 a 5, en la que el polipéptido NS3 está modificado en comparación con un polipéptido NS3 nativo, para mostrar una actividad de proteasa significativamente reducida.
- 25 7. Proteína de fusión aislada según cualquiera de las reivindicaciones 1 a 6, en la que el polipéptido NS3 está modificado en comparación con un polipéptido NS3 nativo para mostrar una actividad de helicasa significativamente reducida.
- 30 8. Proteína de fusión aislada según cualquiera de las reivindicaciones 1 a 7, en la que el polipéptido NS5B está modificado en comparación con un polipéptido NS5B nativo para mostrar una actividad de ARN polimerasa dependiente de ARN significativamente reducida.
- 35 9. Proteína de fusión aislada según cualquiera de las reivindicaciones 1 a 8, en la que dicha proteína de fusión no comprende uno o más de los sitios de escisión reconocidos por NS3 normalmente presentes en un precursor poliproteico de VHC nativo en las uniones NS3/NS4A, NS4A/NS4B, NS4B/NS5A y NS5A/NS5B.
- 40 10. Proteína de fusión aislada según cualquiera de las reivindicaciones 1 a 9, en la que los polipéptidos NS4A, NS4B y/o NS5B se modifican para suprimir uno o más dominios hidrófobos que están implicados normalmente en el anclaje a la membrana de los polipéptidos nativos NS4A, NS4B y/o NS5B.
- 45 11. Proteína de fusión aislada según cualquiera de las reivindicaciones 1 a 10, en la que el polipéptido NS3 se origina a partir de la proteína NS3 nativa mostrada en la SEC ID nº: 1 que se modifica al menos de tal manera que:
- o retiene la porción que se extiende desde la posición 12 hasta la posición 56, desde la posición 70 hasta la posición 155, y desde la posición 218 hasta la posición 248;
 - o comprende la sustitución del resto His en la posición 57 por un resto de aminoácido diferente de una His, tal como un resto Ala;
 - 50 o comprende la sustitución del resto Thr en la posición 269 por un resto de aminoácido diferente de una Thr, tal como un resto Ala; y
 - o comprende la sustitución del resto Arg en la posición 464 por un resto de aminoácido diferente de una Arg, tal como un resto Ala; y
 - 55 o no comprende el resto Thr en la posición 631.
- 60 12. Proteína de fusión aislada según la reivindicación 11, en la que el polipéptido NS3 comprende la secuencia de aminoácidos mostrada en la SEC ID nº: 5.
- 65 13. Proteína de fusión aislada según cualquiera de las reivindicaciones 1 a 12, en la que el polipéptido NS4A se origina a partir de la proteína NS4A nativa mostrada en la SEC ID nº: 2 que se modifica al menos de tal manera que:
- o contiene la porción de la SEC ID nº: 2 desde la posición 21 hasta la posición 33;
 - o no contiene la porción de la SEC ID nº: 2 desde la posición 1 hasta la posición 20.

14. Proteína de fusión aislada según la reivindicación 13, en la que el polipéptido NS4A consiste esencialmente en la secuencia de aminoácidos mostrada en la SEC ID nº: 6 precedida de un resto Met iniciador.
- 5 15. Proteína de fusión aislada según cualquiera de las reivindicaciones 1 a 14, en la que el polipéptido NS4B se origina a partir de la proteína NS4B nativa mostrada en la SEC ID nº: 3 que se modifica al menos de tal manera que:
- o comprende la porción que se extiende desde el resto Ser en la posición 78 hasta el resto Leu en la posición 109;
 - 10 o no comprende el resto Cys en la posición 261.
16. Proteína de fusión aislada según la reivindicación 15, en la que el polipéptido NS4B opcional consiste esencialmente en la secuencia de aminoácidos tal como se muestra en la SEC ID nº: 7.
- 15 17. Proteína de fusión aislada según cualquiera de las reivindicaciones 1 a 16, en la que el polipéptido NS5B se origina a partir de la proteína nativa NS5B mostrada en la SEC ID nº: 4 que se modifica al menos de tal manera que:
- o comprende la porción que se extiende desde el resto Arg en la posición 155 hasta el resto Leu en la posición 182.
 - 20 o comprende la sustitución del resto Asp en la posición 220 por un resto de aminoácido diferente de una Asp, tal como un resto Asn;
 - o comprende la sustitución del resto Asp en la posición 318 por un resto de aminoácido diferente de una Asp, tal como un resto Asn;
 - 25 o no comprende la porción que se extiende desde el resto Trp en la posición 571 hasta el resto Arg en la posición 591.
- 30 18. Proteína de fusión aislada según la reivindicación 17, en la que el polipéptido NS5B comprende la secuencia de aminoácidos tal como se muestra en la SEC ID nº: 8.
19. Proteína de fusión aislada según cualquiera de las reivindicaciones 11 a 18, en la que dicha proteína de fusión comprende una secuencia de aminoácidos que es homóloga o idéntica a la secuencia de aminoácidos mostrada en la SEC ID nº: 9 o 10.
- 35 20. Molécula de ácido nucleico aislada que codifica la proteína de fusión según cualquiera de las reivindicaciones 1 a 19.
- 40 21. Molécula de ácido nucleico aislada según la reivindicación 20, que comprende una secuencia nucleotídica optimizada para la expresión en una célula hospedante de mamífero.
22. Molécula de ácido nucleico aislada según la reivindicación 21, que comprende una secuencia nucleotídica que es homóloga o idéntica a las secuencias nucleotídicas mostradas en la SEC ID nº: 11 o la SEC ID nº: 12.
- 45 23. Molécula de ácido nucleico aislada según la reivindicación 20, que tiene una secuencia nucleotídica optimizada para la expresión en una célula hospedante de levadura.
24. Molécula de ácido nucleico aislada según la reivindicación 23, en la que la célula hospedante de levadura se selecciona de entre el grupo constituido por *Saccharomyces cerevisiae*, *Saccharomyces pombe* y *Pichia pastoris*.
- 50 25. Molécula de ácido nucleico aislada según la reivindicación 24, que comprende una secuencia nucleotídica que es homóloga o idéntica a las secuencias nucleotídicas mostradas en la SEC ID nº: 13 o la SEC ID nº: 14.
- 55 26. Molécula de ácido nucleico aislada según la reivindicación 20, que tiene una secuencia nucleotídica optimizada para la expresión en una célula hospedante procarionta.
27. Molécula de ácido nucleico aislada según la reivindicación 26, que comprende una secuencia nucleotídica que es homóloga o idéntica a las secuencias nucleotídicas mostradas en la SEC ID nº: 15 o la SEC ID nº: 16.
- 60 28. Vector que comprende una o más copias de la molécula de ácido nucleico según cualquiera de las reivindicaciones 20 a 27.
29. Partícula vírica infecciosa que comprende la molécula de ácido nucleico según cualquiera de las reivindicaciones 20 a 27 o el vector de la reivindicación 28.
- 65

30. Célula hospedante que comprende la molécula de ácido nucleico según cualquiera de las reivindicaciones 20 a 27 o el vector de la reivindicación 28 o la partícula vírica infecciosa de la reivindicación 29.
- 5 31. Método para producir recombinantemente la proteína de fusión según cualquiera de las reivindicaciones 1 a 19, que comprende:
- (a) introducir el vector según la reivindicación 28 o la partícula vírica infecciosa según la reivindicación 29 en una célula hospedante para producir una célula hospedante transfectada o infectada según la reivindicación 30,
 - 10 (b) cultivar *in vitro* dicha célula hospedante transfectada o infectada, en condiciones adecuadas para el crecimiento de la célula hospedante,
 - (c) recuperar la proteína de fusión del cultivo celular, y
 - 15 (d) opcionalmente, purificar la proteína de fusión recuperada.
32. Composición que comprende la proteína de fusión según cualquiera de las reivindicaciones 1 a 19, la molécula de ácido nucleico según cualquiera de las reivindicaciones 20 a 27, el vector según la reivindicación 28, la partícula vírica infecciosa según la reivindicación 29, la célula hospedante según la reivindicación 30, o cualquier combinación de los mismos.
- 20 33. Proteína de fusión según cualquiera de las reivindicaciones 1 a 19, la molécula de ácido nucleico según cualquiera de las reivindicaciones 20 a 27, el vector según la reivindicación 28, la partícula vírica infecciosa según la reivindicación 29, la célula hospedante según la reivindicación 30 o la composición según la reivindicación 32, para su uso en el tratamiento o prevención de infecciones por VHC, enfermedades y estados patológicos asociados a VHC.
- 25 34. Proteína de fusión, molécula de ácido nucleico, vector, partícula vírica infecciosa, célula hospedante, o composición según la reivindicación 33, en los que dicho uso se lleva a cabo según una modalidad terapéutica de sensibilización y revacunación.
- 30 35. Composición según la reivindicación 33, en la que se usa para sensibilizar o estimular, o tanto sensibilizar como estimular, una respuesta inmunitaria anti-VHC.
- 35 36. Proteína de fusión, molécula de ácido nucleico, vector, partícula vírica infecciosa, célula hospedante, o composición según cualquiera de las reivindicaciones 33 a 35, en los que dicho uso se lleva a cabo conjuntamente con una quimioterapia con uno o más fármacos contra VHC.

ES 2 387 321 T3

```

.       790       800       810       820       830       840
.       |         |         |         |         |         |
DTETPIDTTIMAKNEVFCVQPEKGGGRKPARLIVFPDLGVRVCEKMALYDVVSTLPQAVMG
.                               (dominio antigénico D, NS5B)
.
.       850       860       870       880       890       900
.       |         |         |         |         |         |
PSYGFQYSPGQORVEFLVNTWKSKCPMGFSYNTRCFDSTVTENDIRTEESIYQCCLAPE
.                               D
.
.       910       920       930       940       950       960
.       |         |         |         |         |         |
ARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGNTLTCYLKATAACRAAKL
.
.       970       980       990       1000       1010       1020
.       |         |         |         |         |         |
QDCTMLVNGNDLVVICESAGTQEDAASLRVFTEAMTRYSAPPGDPPQPEYDLELITSCSS
.                               D
.
.       1030      1040      1050      1060      1070      1080
.       |         |         |         |         |         |
NVSVAHDASGKRVYYLTRDPTTPLARAAWETVRHTPVNSWLGNIIMYAPTLWARMILMTH
.
.       1090      1100      1110      1120      1130      1140
.       |         |         |         |         |         |
FFSILLAQEQLEKALDCQIYGACYSIEPLDLPQIIERLHGLSAFSLHSYSPGEINRVASC
.
.       1150      1160      1170      1180      1190      1200
.       |         |         |         |         |         |
LRKLGVPPLRVWRHRARSVRAKLLSQGGRAATCGKYLFNWAVRTKLKLTPIPAASQLDLS
.
.       1210      1220
.       |         |
GWFVAGYNGGDIYHSLSRARPR

```


ES 2 387 321 T3

```

.       730       740       750       760       770       780
.       |         |         |         |         |         |
SRSASLRQKKVTFDRLQVLDDHYRDVLEKEMKAKASTVKARLLSIEEACKLTPPHSAKSKF

.       790       800       810       820       830       840
.       |         |         |         |         |         |
GYGAKDVRSLSSRAVNHIRSVWEDLLEDTETPIDTTIMAKNEVFCVQPEKGGGRKPARLIV

.       850       860       870       880       890       900
.       |         |         |         |         |         |
FPDLGVRVCEKMALYDVVSTLPQAVMGPSYGFQYSPGQORVEFLVNTWKSNTKKCPMGFSYNT
(dominio antigénico D, NS5B)                                D

.       910       920       930       940       950       960
.       |         |         |         |         |         |
RCFDSTVTENDIRTEESIYQCCDLAPEARQAIKSLTERLYIGGPLTNSKGQNCGYRRCRA

.       970       980       990       1000      1010      1020
.       |         |         |         |         |         |
SGVLTTCGNTLTCYLKATAACRAAKLQDCTMLVNGDNDLVVICESAGTQEDAASLRVFTE
.                                     D

.       1030      1040      1050      1060      1070      1080
.       |         |         |         |         |         |
AMTRYSAPPGDPPQPEYDLELITSCSSNVSAHDASGKRVYYLTRDPTTPLARAAWETVR

.       1090      1100      1110      1120      1130      1140
.       |         |         |         |         |         |
HTPVNSWLGNIIIMYAPTLWARMILMTHFFSILLAQEQLKALDCQIYGACYSIEPLDLPO

.       1150      1160      1170      1180      1190      1200
.       |         |         |         |         |         |
I IERLHGLSAFSLHSYSPGEINRVASCLRKLGVPPLRVWRHRARSVRAKLLSQGGRATC

.       1210      1220      1230      1240
.       |         |         |         |
GKYLFNWAVRTKLKLTPIPAASQLDLSGWFEVAGYNGGDIYHSLSRARPR

```

Fig. 3
 Proteína de fusión 4A-3-5B_(Hs)
 optimizada para Homo sapiens

```

CGAATTGGGTACCGCCACCATGGGCAGCGTGGTGTATTGTGGGCCGGATCATCCTGAGCGG
1 -----+-----+-----+-----+-----+-----+-----+
GTTAACCCATGGCGGTGGTACCCGTGCGACCACTAACACCCGGCCTAGTAGGACTCGCC
      M G S V V I V G R I I L S G
61 CAGCGGCAGCGCCCCCATCACCGCCTACAGCCAGCAGACCAGAGGCCTGCTGGGCTGTAT
-----+-----+-----+-----+-----+-----+-----+
GTCGCCGTGCGCGGGGTAGTGGCGEATGTCGGTCTGCTGGTCTCCGGACGACCCGACATA
S G S A P I T A Y S Q Q T R G L L G C I
121 CATCACCGCCTGACCGGCAGAGACAAGAATCAGGTGGACGGCGAGGTGCAGGTGTGTGTC
-----+-----+-----+-----+-----+-----+-----+
GTAGTGGTGGGACTGGCCGTCTCTGTTCCTTAGTCCACCTGCCGCTCCACGTCCACGACAG
I T S L T G R D K N Q V D G E V Q V L S
181 CACCGCCACCCAGAGCTTCCTGGCCACCTGTGTGAATGCCGTGTGTGGACCGTGTACGC
-----+-----+-----+-----+-----+-----+-----+
GTGGCGGTGGTCTCGAAGGACCGGTGGACACACTTACCGCACACACCTGGCACATGGC
T A T Q S F L A T C V N G V C W T V Y A
241 CCGAGCCGGCAGCAAGACCCCTGGCCGGACCCAAGGGCCCCATCACCCAGATGTACACCAA
-----+-----+-----+-----+-----+-----+-----+
GCCTCGGCCGTGTTCTGCGGACCGGCCTGGGTTCGCCGGGTAGTGGTCTACATGTGGTT
G A G S K T L A G P K G P I T Q M Y T N
301 CGTGGACCAGGACCTGGTGGGCTGGCCCTGCCCTCTGCGCCAGAAGCATGACCCCTTG
-----+-----+-----+-----+-----+-----+-----+
GCACCTGGTCTCTGGACACCCGACCGGACGGGGAGGACCGCGGTCTTCGTACTGGGGAAC
V D Q D L V G W F A P P G A R S M T P C
361 TACCTGTGGCAGCAGCGACCTGTACCTGGTGACCAGACACGCCGATGTGATCCCTGTGAG
-----+-----+-----+-----+-----+-----+-----+
ATGGACACCGTCTGCTGCGTGGACATGGACCACTGGTCTGTGCGGCTACACTAGGGACACTC
T C G S S D L Y L V T R H A D V I P V R
421 GAGGAGAGCCGATAGCAGAGGCAGCCTGCTGTCTCCTAGACCCGTGTCTTACCTGAAGGG
-----+-----+-----+-----+-----+-----+-----+
CTCCTCTCCGCTATCGTCTCCGTGCGACGACAGAGGATCTGGGCACAGGATGGACTTCCC
R R G D S R G S L L S P R P V S Y L K G
481 CAGCAGCGGGCGGACCCCTGCTGTGCCCCAGCGGCCACGTGGTGGGCATCTTCAGAGCCGC
-----+-----+-----+-----+-----+-----+-----+
GTCGTGCGCCGCTGGGGACGACACGGGGTGGCCGGTGCACCCACCCGTAGAAGTCTCGGGC
S S G G P L L C P S G H V V G I F R A A
541 CGTGTGTACCAGAGGCGTGECCAAGGCCGTGGATTTCATCCCCGTGGAGAGCATGGAGAC
-----+-----+-----+-----+-----+-----+-----+
GCACACATGGTCTCCGCACCGGPTCCGGCACCTAAAGTAGGGGCACCTCTCGTACCTCTG
V C T R G V A K A V D F I P V E S M E T
601 CACCATGAGGAGCCCGTGTTCACCGACAATAGCAGCCCCCTGCCGTGCCTCAGACCTT
-----+-----+-----+-----+-----+-----+-----+
GTGGTACTCTCGGGGCACAAGTGGCTGTTATEGTCGGGGGACGGCACGGAGTCTGGAA
T M R S F V F T D N S S F P A V P Q T F
CCAGGTGGCCACCTGCACGCCCCACCGGCTCCGGCAGAGCACCAAGGTGCCAGCCGC
    
```

661 -----+-----+-----+-----+-----+-----+-----+
 GGTCCACCGGGTGGACGTGCGGGGGTGGCCGAGGCCGTTCTCGTGGTTCACGGTCCGGG
Q V A H L H A F T G S G K S T K V P A A
 CTACGCCGCCAGGGCTACAAGGTGCTGGTGTGAATCCCAGCGTGGCCGCCACCCTGGG
 721 -----+-----+-----+-----+-----+-----+-----+
 GATGCGGGGGTCCCGATGTTCCACGACCACGACTTAGGGTCGCACCGGGCGGTGGGACCC
Y A A Q G Y K V L V L N P S V A A T L G
 CTTCCGGCGCTACATGAGCAAGGCCACGGCATCGAGCCCAATATCCGGACCGGAGTGG
 781 -----+-----+-----+-----+-----+-----+-----+
 GAAGCCCGGGATGTACTCGTTCGGGTGCCGTAGCTCGGGTTATAGGCCTGGCCCTCACTC
F G A Y M S K A H G I E P N I R T G V R
 GACCATCACACAGGGCGGCCCTATCACCTACAGCGCCTACGGCAAGTTCCTGGCCGACGG
 841 -----+-----+-----+-----+-----+-----+-----+
 CTGGTAGTGGTGTCCGCCGGGATAGTGGATGTCCGGATGCCGTTCAAGGACCGGCTGCC
T I T T G G P I T Y S A Y G K F L A D G
 CCGCTGTAGCGGGAGCCCTACGACATCATCTGTGACGAGTGCCACAGCACCGATTG
 901 -----+-----+-----+-----+-----+-----+-----+
 GCCGACATCGCCGCCCTCGATGCTGTAGTAGTAGACTGCTCAACGGTGTCTGGCTAAC
G C S G G A Y D I I I C D E C H S T D W
 GACCACCATCCTGGGCATCGGCACCGTGC TGGACCAGGCCGAGACCGCCGGAGCCAGACT
 961 -----+-----+-----+-----+-----+-----+-----+
 CTGGTGGTAGGACCCGTAGCCGTGGCACGACCTGGTCCGGCTCTGGCGGCCCTCGGTCTGA
T T I L G I G T V L D Q A E T A G A R L
 GGTGGTGTGGCCACAGCCACACCCCTGGCAGCATCACCGTGC CCCCACCCCAACATCGA
 1021 -----+-----+-----+-----+-----+-----+-----+
 CCACCACGACCGGTGTGGTGGGGGACCGTGTAGTGGCACGGGGTGGGGTTGTAGCT
V V L A T A T P P G S I E V P H P N I E
 GGAGGTGGCCCTGAGCAACACCGGCCGAGATCCCTTCTACGGCAAGGCCATCCCTATCGA
 1081 -----+-----+-----+-----+-----+-----+-----+
 CCTCCACCGGGACTCGTTGTGGCCGCTCTAGGGGAAGATGCCGTTCCGGTAGGGATAGCT
E V A L S N T G E I P F Y G K A I P I E
 GGCCATCAAGGGCGGCAGACACCTGATCTTCTGCCACACCAAGAAGAAGTGTGACGAGCT
 1141 -----+-----+-----+-----+-----+-----+-----+
 CCGGTAGTTCCCGCCGTCTGTGGACTAGAAGACGGTGTGCTTCTTCTCACACTGCTCGA
A I K G G R H L I F C H S K K K C D E L
 GGCCGCCAAGCTGACCGGCCTGGGCCTGACGCCCTGGCCTACTACAGAGGCCTGGACGT
 1201 -----+-----+-----+-----+-----+-----+-----+
 CCGGGCGTTCGACTGGCCGGACCCGGACTTGGCGCACCGGATGATGTCTCCGGACCTGCA
A A K L T G L G L N A V A Y Y R G L D V
 GTCCGTGATCCCTACCAGCGCGATGTGGTGGTGGTGGCCACCGACGCCCTGATGACCGG
 1261 -----+-----+-----+-----+-----+-----+-----+
 CAGGCACTAGGGATGGTCCCGCTACACCACCACCACCGGTGGCTGCCGGGACTACTGGCC
S V I P T S G D V V V V A T D A L M T G
 CTTACCGGGCGATTCGACAGCGTGATCGACTGTAATACCTGTGTGACCCAGACCGTGG
 1321 -----+-----+-----+-----+-----+-----+-----+
 GAAGTGGCCGCTAAAGCTGTCCGACTAGCTGACATTATGGACACACTGGGTCTGGCACCT
F T G D F D S V I D C N T C V T Q T V D
 CTTACGCTGGACCCACCTTCACCATCGAGACCACCACCGTGGCACAGGATGCCGTGTC
 1381 -----+-----+-----+-----+-----+-----+-----+
 GAAGTCCGACCTGGGGTGAAGTGGTAGCTCTGGTGGTGGCACGGTGTCTACGGCACAG
F S L D P T F T I E T T V F Q D A V S

CAGAAGCCAGAGAAGAGGGCGCCACCGGCAGAGGCAGAAGCGGCATCTACAGATTCGTGAC
 1441-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GTCTTCGGTCTCTTCTCCGCGGTGGCCGTCTCCGTCTTCGCCGTAGATGTCTAAGCACTG
R S Q R R G A T G R G R S G I Y R F V T
 CCCTGGCGAGAGACCCAGCGGCATGTTTCGATAGCAGCGTGTGTGTGAGTGCTACGACGC
 1501-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GGGACCGCTCTCTGGGTCCGCTACAAGCTATCGTCCGACGACACACTCACGATGCTGGC
P G E R P S G M F D S S V L C E C Y D A
 CGGCTGTGCCTGGTACGAGCTGACCCAGCEGAGACCACAGTGAGGCTGAGGGCCTACCT
 1561-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GCCGACACGGACCATGCTCGACTGGGGTCCGGCTCTGGTGTCACTCCGACTCCCGGATGGA
G C A W Y E L T P A E T T V R L R A Y L
 GAACACCCCTGGCCTGCCTGTGTGCCAGGATCACCTGGAGTTCGTTGGGAGAGCGTGTTTAC
 1621-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CTTGTGGGGACCGGACGGACACCGGTCTTAGTGGACCTCAAGACCCCTCTCGCACAATG
N T P G L P V C Q D H L E F W E S V F T
 CGGCCTGACCCACATCGATGCCCACTTCTGAGCCAGACCAAGCAGGCGGGGACAACCT
 1681-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GCCGACTGGGTGTAGCTACGGGTGAAAGACTCGGTCTGGTTCGTCGCGCCGCTGTGAA
G L T H I D A H F L S Q T K Q A G D N F
 CCCCTACCTGGTGGCCTACCAGGCCACCGTGTGTGCCAGAGCCAGGCCCCCTCCCCCAG
 1741-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GGGGATGGACCACCGGATGGTCCGGTGGCACACACGGTCTCGGGTCCGGGGAGGGGGGTC
P Y L V A Y Q A T V C A R A Q A P P P S
 CTGGGACCAGATGTGGAAGTGCCTGATCAGGCTGAAGCCCACCCTCGACGGCCCTACCCC
 1801-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GACCCGTGGTCTACACCTTCACGGACTAGTCCGACTTCGGGTGGGACGTGCCGGGATGGGG
W D Q M W K C L I R L K P T L H G P T P
 CCTGCTGTACAGACTGGGCGCCGTGCAGAATGAGATCACCCCTGACCCACCCCTATCACCAA
 1861-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GGACGACATGTCTGACCCGCGGCACGTCTTACTCTAGTGGGACTGGGTGGGATAGTGGTT
L L Y R L G A V Q N E I T L T H P I T K
 GTTCGTGATGGCCTGTATGAGCGCCGACCTGGAGGTGGTGGGAGCGGCTCCGCTCCAT
 1921-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CAAGCACTACCGGACATACTCGCGGCTGGACCTCCACCACCCGTCGCCGAGGCGGAGGTA
F V M A C M S A D L E V V G S G S Q S M
 GAGCTACACCTGGACAGGGCGCCCTGATCACACCCTGTGCCGCGGAGGAGCAAGCTGCC
 1981-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CTCGATGTGGACCTGTCCGCGGACTAGTGTGGGACACGGCGGCTCCCTCTGTTCCGACGG
S Y T W T G A L I T P C A A E E S K L P
 CATCAACCCCTGAGCAATAGCCTGCTGAGGCACCACAGCATGGTGTACAGCACCACCTC
 2041-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GTAGTTGGGGGACTCGTTATCCGACCACTCCGTGGTGTGCTACCACATGTGCTGGTGGAG
I N P L S N S L L R H H S M V Y S T T S
 CAGAAGCGCCAGCCTGAGGCAGAAGAAGGTGACCTTCGACAGGCTGCAGGTGCTGGACGA
 2101-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GTCTTCGGGTCCGACTCCGTCTTCTTCCACTGGAAGCTGTCCGACGTCCACGACCTGCT
R S A S L R Q K K V T F D R L Q V L D D
 CCACTACAGGGACGTGCTGAAGGAGATGAAGGCCAAGGCCAGCACCCGTGAGGGCCAGACT
 2161-----+-----+-----+-----+-----+-----+-----+-----+-----+

GGTGATGTCCCTGCACGACTTCCTCTACTTCCGGTTCCGGTTCGTGGCACTTCCGGTCTGA
H Y R D V L K E M K A K A S T V K A R L
 GCTGTCTATCGAGGAGGCCTGTAAGCTGACCCCCCTCACAGCGCCAAGAGCAAGTTCCG
 2221-----+-----+-----+-----+-----+-----+-----+-----+
 CGACAGATAGCTCCTCCGGACATTCGACTGGGGGGGAGTGTCCGGTTCCTCGTTCAAGCC
L S I E E A C K L T P P H S A K S K F G
 CTACGGCGCCAAGGATGTGAGAAGCCTGAGCAGCAGAGCCGTGAACCACATCCGGTCTGT
 2281-----+-----+-----+-----+-----+-----+-----+-----+
 GATGCCGCGGTTCTTACACTCTTCGGACTCGTCGTCTCGGCACTTGGTGTAGGCCAGACA
Y G A K D V R S L S R A V N H I R S V
 GTGGGAGGATCTGCTGGAGGATACCGAGACCCCCATCGACACCACAATCATGGCCAAGAA
 2341-----+-----+-----+-----+-----+-----+-----+-----+
 CACCCCTCTAGACGACCTCCTATGGCTCTGGGGGTAGCTGTGGTGTAGTACCGGTTCTT
W E D L L E D T E T P I D T T I M A K N
 CGAGGTGTCTGCGTGCAGCCCGAGAAGGGCGGAAGAAAGCCCGCCAGGCTGATCGTGTT
 2401-----+-----+-----+-----+-----+-----+-----+-----+
 GCTCCACAAGACGCACGTCGGGCTCTTCCCGCCTCTTTCGGGCGGTCCGACTAGCACAA
E V F C V Q P E K G G R K P A R L I V F
 CCCTGACCTGGGAGTGAGACTGTGTGACAAGATGGCCCTGTACGACGTCGTCTCCACCCT
 2461-----+-----+-----+-----+-----+-----+-----+-----+
 GGGACTGGACCCTCACTCTCACACACTCTTCTACCGGACATGCTGCACCACAGGTGGGA
F D L G V R V C E K M A L Y D V V S T L
 GCCCCAGGCCGTGATGGGCCCCAGCTACGGCTTCAGTACAGCCCTGGCCAGAGAGTGGAA
 2521-----+-----+-----+-----+-----+-----+-----+-----+
 CGGGGTCCGGCACTACCCGGGTCGATGCCGAGGTCAATGTGGGACCGGTCCTCACT
P Q A V M G P S Y G F Q V S P G Q R V E
 GTTCTGGTGAACACCTGGAAGAGCAAGAAATGCCCCATGGGCTTCAGCTACAACACCCG
 2581-----+-----+-----+-----+-----+-----+-----+-----+
 CAAGGACCACTTGTGGACCTTCTCGTTCFTTACGGGTACCCGAAGTCGATGTTGTGGGC
F L V N T W K S K K C P M G F S Y N T R
 GTGCTTCGACAGCACAGTGAECGAGAACGACATCAGGACCGAGGAGTCCATCTACCAGTG
 2641-----+-----+-----+-----+-----+-----+-----+-----+
 CACGAAGCTGTGCTGCTCACTGGCTCTTGTGTAGTCTCGGCTCCTCAGGTAGATGGTCAC
C F D S T V T E N D I R T E E S I Y Q C
 CTGTGACCTGGCCCCGAGGCCAGACAGGCCATCAAAAGCCTGACCCAGCGGCTGTACAT
 2701-----+-----+-----+-----+-----+-----+-----+-----+
 GACACTGGACCAGGGGCTCCGGTCTGTCCGGTAGTTTTCCGACTGGCTCGCCGACATGTA
C D L A P E A R Q A I K S L T E R L Y I
 CGGCGGACCTCTGACCAACAGCAAGGGCCAGAACTGTGGCTACAGAAGATGTAGGSCCAG
 2761-----+-----+-----+-----+-----+-----+-----+-----+
 GCCGCCTGGAGACTGCTTGTGCTTCCCGGCTTGACACCGATGTCTTCTACATCCGGGTC
G G F L T N S K G Q N C G Y R R C R A S
 CGGCGTGTGACCACCTCTGTGGCAACACCCCTGACCTGTTACCTGAAGGCCACCGCCGC
 2821-----+-----+-----+-----+-----+-----+-----+-----+
 GCCGCACGACTGTTGGAGAACACCGTGTGGGACTGGACAATGGACTTCCGGTGGCGGCG
G V L T T S C G N T L T C Y L K A T A A
 CTGTAGAGCCGCCAAACTGCAGGACTGTACCATGCTGGTGAACGGCAACGACCTGGTGGT
 2881-----+-----+-----+-----+-----+-----+-----+-----+
 GACATCTCGGCGGTTTGACGTCCTGACATGGTACGACCACTTGCCTGTTGCTGGACCACCA
C R A A K L Q D C T M L V N G H D L V V

GATCTGTGAGAGCGCCGGCACCAGGAGGATGCCGCCTCCCTGAGAGTGTTTACCGAGGC
 2941-----+-----+-----+-----+-----+-----+-----+
 CTAGACACTCTCCCGCCCGTGGGTCCCTACGGCCGGAGGGACTCTCACAAATGGCTCCG
I C E S A G T Q E D A A S L R V F T E A
 CATCACCAGATACACCGCCCTCCCGGGGACCCTCCCCAGCCCGAGTACGATCTGGAGCT
 3001-----+-----+-----+-----+-----+-----+-----+
 GTACTGGTCTATGTCCCGGGGAGGGCCGCTGGGAGGGGTCCGGCTCATGCTAGACCTCGA
M T R Y S A P P G D P P Q P E Y D L E L
 GATCACCAGCTGCTCCAGCAATGTGGGCGTGGCTCACGACGCCAGCGCAAGAGAGTGTA
 3061-----+-----+-----+-----+-----+-----+-----+
 CTAGTGGTCCGACGAGGTGCTTACACCCCGCACCGAGTGCCTGGGGTGGCCGTTCTCTCACAT
I T S C S S N V G V A H D A E G K R V Y
 CTACCTGACCAGGGACCCACCACCCCTCTGGCCAGAGCCGCTGGGAGACAGTGAGACA
 3121-----+-----+-----+-----+-----+-----+-----+
 GATGCACTGGTCCCTGGGGTGGTGGGGAGACCGGTCCTGGCGGACCCTCTGTCACTCTGT
Y L T R D P T T P L A R A A W E T V R H
 CACCCCGTGAACAGCTGGCTGGGCAACATCATCATGTACGCCCTACCTGTGGGGCCAG
 3181-----+-----+-----+-----+-----+-----+-----+
 GTGGGGCACTTGTGACCGACCCGTTGTAGTAGTACATGCGGGGATGGGACACCCGGTC
T P V N S W L G N I I M Y A P T L W A R
 AATGATCCTGATGACCEACTTCTTCAGCATCCTGCTGGCCAGGAGCAGCTGGAGAAGGC
 3241-----+-----+-----+-----+-----+-----+-----+
 TFACTAGGACTACTGGGTGAAGAAGTCGTAGGACGACCGGGTCCCTCGTCGACCTCTCCG
M I L M T H F F S I L L A Q E Q L E K A
 COTGGACTGCCAGATCTACGGCCCTGCTACAGCATCGAGCCTCTGGACCTGCCTCAGAT
 3301-----+-----+-----+-----+-----+-----+-----+
 GGACCTGACGGTCTAGATGCCCGGGACGATGTCTGTAGCTCGGAGACCTGGACGGAGTCTA
L D C Q I Y G A C Y S I E P L D L F Q I
 CATCCAGAGACTGCCACGCCCTGAGCGCCTTCAGCCTGCACAGCTACAGCCCAGGGCAGAT
 3361-----+-----+-----+-----+-----+-----+-----+
 GTAGCTCTCTGACGTGCCGGACTCGCGGAAGTCGGACGTGTGATGTGCGGGTCCGCTCTA
I E R L H G L S A F S L H S Y S P G E I
 CAATAGAGTGGCCAGCTGCCTGAGAAAGCTGGGCGTCCACCTCTGAGAGTGTGGCGGCA
 3421-----+-----+-----+-----+-----+-----+-----+
 GTTATCTCACCGGTCGACGGACTCTTTGACCCCGCACGGTGGGAGACTCTCACACCGCCGT
N R V A S C L R K L G V P P L R V W R H
 CAGAGCCAGATCTGTGAGGGCCAAGCTGCTGTCCCAGGGCGGCAGGGCCGCCACCTGTGG
 3481-----+-----+-----+-----+-----+-----+-----+
 GTCTCGGTCTAGACACTCCCGGTTGACCGACAGGGTCCCGCGTCCCGGGTGCACACC
R A R S V R A K L L S Q G G R A A T C G
 CAAGTACCTGTTCAACTGGGCGGTGAGGACAAAGCTGAAGCTGACACCCATCCCTGCCCG
 3541-----+-----+-----+-----+-----+-----+-----+
 GTTCATGGACAAGTTGACCCGGCACTCCTGTTTCGACTTCGACTGTGGGTAGGGACGGCG
K Y L F N W A V R T K L K L T F I P A A
 CAGCCAGCTGGACCTGAGCGGCTGGTTCGTGGCCGGCTACAATGGCCGGGACATCTACCA
 3601-----+-----+-----+-----+-----+-----+-----+
 GTCGGTCCGACCTGGACTCGCCGACCAAGCACCGGCCGATGTTACCGCCGCTGTAGATGGT
S Q L D L S G W F V A G Y N G G D I Y H

CAGCCTGTCCAGGGCCAGGCCTAGATGATGAGGAGCTCCAGCTT
3661-----+-----+-----+-----+-----
GTCGGACAGGTCCCGGTCCGGATCTACTACTCCTCGAGGTCGAAA
S L S R A R P R * *

Fig. 4
 Proteína de fusión 4A-3-4B-5B (Hs) optimizada para Homo sapiens

```

        KpnI      NcoI
GGGCGAATTGGGTACCGCCACCATGGGCAGCGTGGTGATTGTGGGCCGGATCATCCTGAG
-----+-----+-----+-----+-----+-----+-----+
CCCGCTTAACCCATGGCGGGTGGTACCCGTEGCACCACTAACACCCGGCCCTAGTAGGACTC
        M G S V V I V G R I I L S
CGGCAGCGGCAGCGCCCCCATCACCGECTACAGCCAGCAGACCAGAGGCCCTGCTGGGCTG
-----+-----+-----+-----+-----+-----+-----+
GCGTGGCGGTGCGGGGGTGGTGGCGGATGTCGCTCGTCTGGTCTCCGGACGACCCGAC
        G S G S A F I T A Y S Q Q T R G L L G C
TATCATCACCAGCCTGACCGGCAGAGACAAGAATCAGGTGGACGGCGAGGTGCAGGTGCT
-----+-----+-----+-----+-----+-----+-----+
ATAGTAGTGGTTCGACTGGCCGTCTCTGTCTTAGTCCACCTGCCGCTCCACGTCCACGA
        I I T S L T G R D K N Q V D G E V Q V L
GTCCACCGCCACCCAGAGCTTCCTGGCCACCTGTGTGAATGGCGTGTGTTGGACCGTGA
-----+-----+-----+-----+-----+-----+-----+
CAGGTGGCGGTGGGTCTCGAAGGACCGGTGGACACACTTACCGCACACAACCTGGCACAT
        S T A T Q S F L A T C V N G V C W T V Y
CGCCGGAGCCCGCAGCAAGACCTGCCCGGACCCAAGGGCCCCATCACCCAGATGTACAC
-----+-----+-----+-----+-----+-----+-----+
GCGGCCCTCGGCCCTCGTTCCTGGGACCGCCCTGGGTTCCCGGGGTAGTGGGTCTACATGTG
        A G A G S K T L A G P K G P I T Q M Y T
CAACGTGGACCAGGACCTGGTGGGCTGGCCTGCCCTCCTGGCGCCAGAAGCATGACCCC
-----+-----+-----+-----+-----+-----+-----+
GTTGCACCTCGTCTGGACCACCCGACCGGACGGGGAGGACCGCGGTCTTCGTACTGGGG
        N V D Q D L V G W P A P F G A R S M T P
TTGCACCTGTGGCAGCAGCGACCTGTACCTGGTGACCAGACACGCGGATGTGATCCCTGT
-----+-----+-----+-----+-----+-----+-----+
AACATGGACACCCCTCGTCTGGACATGGACCACTGGTCTGTGGCGCTACACTAGGGACA
        C T C G S S D L Y L V T R H A D V I P V
GAGGAGGAGAGGCBATAGCAGAGGCAGCCTGCTGCTCCTAGACCCGTGTCTTACCTGAA
-----+-----+-----+-----+-----+-----+-----+
CTCCTCCTCTCCGCTATCGTCTCCGTCCGACGACAGAGGATCTGGGCACAGGATGGACTT
        R R R G D S R G S L L S P R P V S Y L K
GGGCAGCAGCGCGGACCCCTGCTGTGCCCCAGCGGCCACGTGGTGGGCATCTTCAGAGC
-----+-----+-----+-----+-----+-----+-----+
CCCGTCTGTGCGCGCTGGGACGACACGGGTCGCGGGTGCACCACCCGTAGAAGTCTCG
        G S S G G P L L C P S G H V V G I F R A
CGCCGTGTGTACCAGAGGCGTGGCCAAAGCCGTGGATTCATCCCCGTGGAGAGCATGGA
-----+-----+-----+-----+-----+-----+-----+
GCGGCACACATGGTCTCCGCACCGGTTCGCGCACCTAAAGTAGGGGCACCTCTCGTACCT
        A V C T R G V A K A V D F I F V E S M E
GACCACCATGAGGAGCCCCGTGTTACCGACAATAGCAGCCCCCTGCCGTGCCCTCAGAC
-----+-----+-----+-----+-----+-----+-----+
CTGCTGGTACTCCTCGGGGCACAAGTGGCTGTATCGTCCGGGGGACGGCACGAGTCTG
        T T M R S P V F T D N S S P P A V P Q T
CTTCCAGGTGGCCACCTGCACGCCCCACCGGCTCCGGCAAGAGCACCAAGGTGCCAGC
-----+-----+-----+-----+-----+-----+-----+
GAAGGTCCACCGGGTGGACGTGCGGGGGTGGCCGAGGCCGTTCTCGTGGTCCACGGTCC
    
```

F Q V A H L H A P T G S G K S T K V P A
 CGCCTACGCCGCCAGGGCTACAAGGTGCTGGTGTCTGAATGCCAGCGTGGCCGCCACCCCT
 721 -----+-----+-----+-----+-----+-----+-----+
 GCGGATGCGGGCGGGTCCCGATGTTCCACGACCACGACTTAGGGTCCGACCGGGCGGTGGGA
A Y A A Q G Y K V L V L N P S V A A T L
 GGGCTTCGGCGCCTACATGAGCAAGGCCACGGCATCGAGCCCAATATCCGGACCGGAGT
 781 -----+-----+-----+-----+-----+-----+-----+
 CCCGAAGCCCGCGGATGTACTCGTTCGGGGTCCGTAGCTCGGGTTATAGGCCTGGCCTCA
G F G A Y M S K A N G I E P N I R T G V
 GAGGACCATCACACAGGGCGGCCCTATCACCTACAGCGCCTACGGCAAGTTCTTGGCCGA
 841 -----+-----+-----+-----+-----+-----+-----+
 CTCCTGGTAGTGGTGTCCCGCGGGATAGTGGATGTCGGCGGATGCCGTTCAAGGACCGGCT
R T I T T G G P I T Y S A Y G K E L A D
 CGGCGGCTGTAGCGGCGGAGCCTACGACATCATCTGTGACGAGTGCCACAGCACCGA
 901 -----+-----+-----+-----+-----+-----+-----+
 GCGCGCGACATCGCCGCCTCGGATGCTGTAGTAGTAGACACTGCTCACGGTGTCTGGCT
G G C S G G A Y D I I I C D E C H S T D
 TTGGACCACCATCCTGGGCATCGGCACCGTGTGGACCAGGCCGAGACCGCCGGAGCCAG
 961 -----+-----+-----+-----+-----+-----+-----+
 AACCTGGTGGTAGGACCCGTAGCCGTGGCAGACCTGGTCCGGCTCTGGCGGCCCTCGGTC
W T T I L G I E T V L D Q A E T A G A R
 ACTGGTGGTGTGGCCACAGCCACACCCCTGGCAGCATCACCGTGCCCCACCCCAACAT
 1021 -----+-----+-----+-----+-----+-----+-----+
 TGACCACCACGACCGGTGTCCGGTGTGGGGACCGTCTAGTGGCACGGGTGGGGTTGTA
L V V L A T A T P P G S I T V P H P N I
 CGAGGAGGTGGCCCTGAGCAACACCGGCGAGATCCCCTTCTACGGCAAGGCCATCCCTAT
 1081 -----+-----+-----+-----+-----+-----+-----+
 GCTCCTCCACCGGACTCGTGTGGCCGCTCTAGGGGAAGATGCCCTTCCCGTAGGGATA
E E V A L S N T G E I P F Y G K A I P I
 CGAGGCCATCAAGGGCGGCAGACACCTGATCTTCTGCCACAGCAAGAAGAAGTGTGACGA
 1141 -----+-----+-----+-----+-----+-----+-----+
 GCTCCGGTAGTTCGCCCGTCTGTGGACTAGAAGACGGTGTCTCTCTCACACTGCT
E A I K G G R H L I F C H S K K K C D E
 GCTGGCCGCCAAGCTGACCGGCCGCGGCTGAAACGCCGTGGCCCTACTACAGAGGCCTGGA
 1201 -----+-----+-----+-----+-----+-----+-----+
 CGACCGCCGCTTCGACTGGCCGGACCCGGACTTCCGGCACCCGATGATGTCTCCGGACCT
L A A K L T G L G L N A V A Y Y R G L D
 CGTGTCCGTGATCCCTACCAGCGCGGATGTGGTGGTGGTGGCCACCGACGCCCTGATGAC
 1261 -----+-----+-----+-----+-----+-----+-----+
 GCACAGGCACTAGGGATGGTCCGCGCTACACCACCACCACCGGTGGCTGCGGGACTACTG
V S V I P T S G D V V V V A T D A L M T
 CGGCTTACCCGGCGATTTCGACAGCGTGATCGACTGTAATACCTGTGTGACCCAGACCGT
 1321 -----+-----+-----+-----+-----+-----+-----+
 GCCGAAGTGGCCGCTAAAGCTGTCCGACTAGCTGACATTATGGACACACTGGGTCTGGCA
G F T G D F D S V I D C N T C V T Q T V
 GGACTTCAGCCTGGACCCACCTTCACCATCGAGACCACCACCGTGCCACAGGATGCCGT
 1381 -----+-----+-----+-----+-----+-----+-----+
 CCTGAAGTCCGACCTGGGGTGAAGTGGTAGCTCTGGTGGTGGCACGGTGTCTACGGCA
D F S L D P T F T I E T T V P Q D A V
 GTCCAGAAGCCAGAGAAGAGGCCGCCACCGCCAGAGGCAGAAGCGGCATCTACAGATTCGT
 1441 -----+-----+-----+-----+-----+-----+-----+

CAGGTCTTCGGTCTCTTCTCCGCGGTGGCCGCTCCCGTCTTCGCCGTAGATGTCTAAGCA
S R S Q R R G A T G R G R S G I Y R F V
 GACCCCTGGCGAGAGACCCAGCGGCATGTTTCGATAGCAGCGTGTGTGAGTGTCTACGA
 1501-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CTGGGSACCGCTCTCTGGGTCCGCGTACAAGCTATCGTCCGACGACACACTCAGGATGCT
T P G E R F S G M F D S S V L C E C Y D
 CGCCGGCTGTGCCTGGTACGAGCTGACCCAGCCGAGACCACAGTGAGGCTGAGGGCCTA
 1561-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GCGCCCGACACGGACCATGCTCGACTGGGGTCCGGTCTGGTGTCACTCCGACTCCCGGAT
A G C A W Y E L T P A E T T V R L R A Y
 CCTGAACACCCCTGGCCTGCCTGTGTGCCAGGATCACCTGGAGTTCTGGGAGAGCGTGT
 1621-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GGACTTGTGGGGACCCGACGGACACACGGTCCCTAGTGGACCTCAAGACCCCTCTCGCACAA
L N T P G L P V C Q D H L E F W E S V F
 TACCGGCCTGACCCACATCGATGCCCACTTTEGAGCCAGACCAAGCAGGCCCGGCGACAA
 1681-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 ATGGCCGGACTGGGTGTAGCTACGGGTGAAGACTCGGTCTGGTTCGTCCGGCCCGCTGTT
T G L T H I D A H F L S Q T K Q A G D N
 CTTCCTTACCTGGTGGCCTACCAGGCCACCGTGTGTGCCAGAGCCAGGCCCTCCCCC
 1741-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GAAGGGGATGGACCACCGGATGGTCCGGTGGCACACACGGTCTCGGGTCCGGGGAGGGGG
F P Y L V A Y Q A T V C A R A Q A P P P
 CAGCTGGGACCAGATGTGGAAGTGCCCTGATCAGGCTGAAGCCACCCCTGCACGGCCCTAC
 1801-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GTGCACCCCTGGTCTACACCTTCACGGACTAGTCCGACTTCGGGTGGGACGTGCCGGGATG
S W D Q M W K C L I R L K P T L H G P T
 CCCCCTGCTGTACAGACTGGGCGCCGTGCAGAATGAGATCACCCCTGACCCACCCCTATCAC
 1861-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GGGGGACGACATGTCTGACCCGCGGCACGCTTACTCTAGTGGGACTGGGTGGGATAGTG
P L L Y R L G A V Q N E I T L T H P T T
 CAAGTTCGTGATGGCCTGTATGACCGCCGACCTEGAGGTGGTGTCCCTGATGGCCTTCAC
 1921-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GTTCAAGCACTACCGACATACTCGCGGCTGGACCCTCCACCACAGGGACTACCGGAAGTG
K F V M A C M S A D L E V V S L M A F T
 CGCCAGCATCACAAGCCCCCTGACCACCCAGAATACCCTGCTGTTCAACATCCTGGGCGG
 1981-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GCGGTCTAGTGTTCGGGGGACTGGTGGGTCTTATGGGACGACAAGTTGTAGGACCCCGCC
A S I T S P L T T Q N T L L F N I L G G
 CTGGGTGGCCGCCAGCTGTCCATGAGCTACACCTGGACAGGGCGCCCTGATCACACCCTG
 2041-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GACCCACCGGCGGGTCCGACAGGTAAGTCTGATGTGGACCTGTCCGCGGGACTAGTGTGGGAC
W V A A Q L S M S Y T W T G A L I T P C
 TGCCGCCGAGGAGAGCAAGCTGCCCATCAACCCCTGAGCAATAGCCTGCTGAGGCACCA
 2101-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 ACGGCGGCTCCTCICGTTCCGACGGGTAGTTGGGGGACTCGTTATCCGACGACTCCGCGGT
A A E E S K L F I N P L S N S L L R H H
 CAGCATGGTGTACAGCACCCACCTCCAGAAGCGCCAGCCTGAGGCAGAAGAAGGTGACCTT
 2161-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GTCGTACCACATGTCTGTGGTGGAGGTCTTCBCGGTCCGACTCCGTCTTCTTCCACTGGAA
S M V Y S T T S R S A S L R Q K K V T F

2221 CGACAGGCTGCAGGTGCTGGACGACCCTACAGGGACGTGCTGAAGGAGATGAAGGCCAA
 GCTGTCCGACGTCACGACCTGCTGGTGTATGTCCTGCACGACTTCCTCTACTTCGGTT
D R L Q V L D D H Y R D V L K E M K A K

2281 GGCCAGCACCGTGAAGGCCAGACTGCTGTCTATCGAGGAGGCCTGTAAGCTGACCCCCC
 CCGGTCGTGGCACTTCCGGTCTGACGACAGATAGCTCCTCCGGACATTCGACTGGGGGG
A S T V K A R L L S I E E A C K L T P P

2341 TCACAGCGCCAAGAGCAAGTTCGGCTACGGCCCAAGGATGTGAGAAGCCTGAGCAGCAG
 AGTGTCCGGTTCCTCGTTCAGCCGATGCCGGTTCCTACACTCTTCGGACTCGTCTC
H S A K S K F G Y G A K D V R S L S S R

2401 AGCCGTGAACCACATCCGGTCTGTGTGGGAGGATCTGCTGGAGGATACCGAGACCCCAT
 TCGGCACTTGGTGTAGGCCAGACACACCCCTCCTAGACGACCTCCTATGGCTCTGGGGTA
A V N H I R S V W E D L L E D T E T P I

2461 CGACACCACAATCATGGCCAAGAACGAGGTGTTCTGCGTGCAGCCCGAGAAGGGCGGAG
 GCTGTGGTGTAGTACCGGTCTTGTCCACCAAGACGCACGTCCGGCTCTTCCCGCCTC
D T T I M A K N E V F C V Q P E K G G R

2521 AAAGCCCGCCAGGCTGATCGTGTCCCTGACCTGGGAGTGAGAGTGTGTGAGAAGATGGC
 TTCCGGGCGGTCCGACTAGCACAAAGGGACTGGACCCCTCACTCTCACACACTCTTCTACCG
K P A R L I V F P D L G V R V C E K H A

2581 CCTGTACGACGTGGTGTCCACCCTGCCCAAGCCCGTGTGSGGCCAGCTACGGCTCCA
 GGACATGCTGCACCACAGGTGGGACGGGTCCGGCACTACCCGGGGTCGATGCCGAGGT
L Y D V V S T L P Q A V M G P S Y G P Q

2641 GTACAGCCCTGGCCAGAGAGTGGAGTTCCTGTTGAACACCTGGAAGAGCAAGAAATGCC
 CATGTCCGGACCGGTCTCTCACCTCAAGGACCACTTGTGGACCTTCTCGTCTTTACGGG
Y S P G Q R V E F L V N T W K S K K C P
 CATGGGCTTCAGCTACAACACCCGGTGTCTCGACAGCACAGTGACCGAGAACGACATCAG

2701 GTACCCGAAGTCGATGTTGTGGGCCAGGAGCTGTGCTGTCACTGGCTCTGTCTGTAGTC
M G F S Y N T R C F D S T V T E N D I R

2761 GACCGAGGAGTCCATCTACCAGTGTGTGACCTGGCCCCGAGGCCAGACAGGCCATCAA
 CTGGCTCCTCAGGTAGATGGTTCACGACACTGGACCGGGGCTCCGGTCTGTCCGGTAGTT
T E E S I Y Q C C D L A P E A R Q A I K

2821 AAGCCTGACCGAGCGGCTGTACATCGGCGGACCTCTGACCAACAGCAAGGGCCAGAACTG
 TTCGGACTGGCTCGCCGACATGTAGCCGCTGGAGACTGGTTGTGTTCCCGGTCTTGAC
S L T E R L Y I G G P L T N S K G Q N C

2881 TGGCTACAGAAGATGTAGGGCCAGCGGCTGTGACCACTCTTGTGGCAACACCCCTGAC
 ACCGATGTCTTCTACATCCCGTCCGCCACGACTGGTGGAGAACACCGTGTGGGACTG
G Y R R C R A S G V L T T S C G N T L T

2941 CTGTTACCTGAAGGCCACCGCCGCTGTAGAGCCGCCAAACTGCAGGACTGTACCATGCT
 GACAATGGACTTCCGGTGGCGGCGGACATCTCGGGGTTTGACGTCTGACATGGTACGA
C Y L K A T A A C R A A K L Q D C T M L

3001 GGTGAACGGCAACGACCTGGTGGTGATCTGTGAGAGCGCCGGCACCAGGAGGATGCCGGC
 -----+-----+-----+-----+-----+-----+-----+-----+
 CCACTTGCCGTTGCTGGACCACCCTAGACACTCTCGCGGCCGTGGGTCTCTACGGCG
V N G N D L V V I C E S A G T Q E D A A
 CTCCCTGAGAGTGTTTACCGAGGCCATGACCAGATACAGCGCCCTCCCGGGCACCCTCC
 3061 -----+-----+-----+-----+-----+-----+-----+-----+
 GAGGGACTCTCACAAATGGCTCCGGTACTGGTCTATGTGCGGGGAGGGCCGCTGGGAGG
S L R V F T E A M T R Y S A P P G D P P
 CCAGCCCGAGTACCATCTGGAGCTGATCACCAGCTGCTCCAGCAATGTGGGCGTGGCTCA
 3121 -----+-----+-----+-----+-----+-----+-----+-----+
 GGTCCGGCTCATGCTAGACCTCGACTAGTGGTCCGACGAGGTCGTTACACCCCGCACCAGT
Q P E Y D L E L I T S C B S N V G V A H
 CGACGCCAGCGGCAAGAGAGTGTACTACCTGACCAGGGACCCCACCACCCTCTGGCCAG
 3181 -----+-----+-----+-----+-----+-----+-----+-----+
 GCTGCGGTGCGCGTTCCTCTCACATGATGGACTGGTCCCTGGGGTGGTGGGAGACCGGTC
D A S G K R V Y Y L T R D P F T T P L A R
 AGCCGCTGGGAGACAGTGAGACACACCCCGTGAACAGCTGGCTGGGCAACATCATCAT
 3241 -----+-----+-----+-----+-----+-----+-----+-----+
 TCGCGGACCCTCTGTCACTCTGTGTGGGGGCACTTGTGACCGACCCGTTGTAGTAGTA
A A W E T V R H T P V N S W L G N I I M
 GTACGCCCTACCTGTGGGCCAGAATGATCCTGATGACCCACTTCTTCAGCATCCTGCT
 3301 -----+-----+-----+-----+-----+-----+-----+-----+
 CATGCGGGGATGGGACACCCGGTCTTACTAGGACTACTGGGTGAAGAAGTCGTAGGACGA
Y A P T L W A R M I L M T H F F S I L L
 GGCCAGGAGCAGCTGGAGAAGGCCCTGGACTGCCAGATCTACGGCGCCTGCTACAGCAT
 3361 -----+-----+-----+-----+-----+-----+-----+-----+
 CCGGGTCCCTCGTCGACCTCTTCCGGGACCTGACGGTCTAGATGCCCGGACGATGTCGTA
A Q E Q L E K A L D C Q I Y G A C Y S I
 CGAGCCTCTGGACCTGCCTCAGATCATCGAGAGACTGCACGGCCTGAGCGCCTTCAGCCT
 3421 -----+-----+-----+-----+-----+-----+-----+-----+
 GCTCGGAGACCTGGACGGAGTCTAGTAGCTCTCTGACGTGCCGGACTCGCGGAAGTCGGA
E P L D L P Q I I E R L H G L S A F S L
 GCACAGCTACAGCCCAGGCGAGATCAATAGAGTGGCCAGCTGCCTGAGAAAGCTGGGCGT
 3481 -----+-----+-----+-----+-----+-----+-----+-----+
 CGTGTGATGTCGGGTCCGCTCTAGTTATCTCACCGGTCCGACGGACTCTTCGACCCGCA
H S Y S P G E I N R V A S C L R K L G V
 GCCACCTCTGAGAGTGTGGCGGCACAGAGCCAGATCTGTGAGGGCCAGCTGCTGTCCCA
 3541 -----+-----+-----+-----+-----+-----+-----+-----+
 CCGTGGAGACTCTCACACCCCGGTGCTCGGTCTAGACACTCCCGGTTCCGACACAGGCT
F P L R V W R H R A R S V R A K L L S Q
 GGGCGGCAGGGCCGCCACCTGTGGCAAGTACCTGTTCAACTGGGCCCTGAGGACAAAGCT
 3601 -----+-----+-----+-----+-----+-----+-----+-----+
 CCCCGGTCGCCGGGTGGACACCGTTTATGGACAAGTTGACCGGGCACTCCTGTTTCGA
G G R A A T C G K Y L F N W A V R T K L
 GAAGCTGACACCCATCCCTGCCGCCAGCCAGCTGGACCTGAGCGGCTGGTTCGTTGCCGG
 3661 -----+-----+-----+-----+-----+-----+-----+-----+
 CTTCGACTGTGGCTAGCGACGGCGGTCCGGTCCGACCTGGACTCCCGACCAAGCACCGGCC
K L T P I F A A S Q L D L S G W F V A G
 CTACAATGGCGGCGACATCTACCACAGCCTGTCCAGGGCCAGGCCTAGATGATGAGGAGC
 3721 -----+-----+-----+-----+-----+-----+-----+-----+

GATGTTACCGCCGCTGTAGATGGTGTCCGGACAGGTCCCAGGATCTACTACTCCTCG
Y N G G D I Y H S L S R A R P R * _

TCCAGCTTTTGGTCCC
3781-----+-----
AGGTCGAAAACAAGGG

Fig. 5
 Proteína de fusión 4A-3-5B optimizada para la expresión en: *Pichia pastoris*

```

ATGGGGTCCGTTGTTATCGTTGGTAGAATCATCTTGTCTGGTTCTGGTCCGGCTCCAATT
1  -----+-----+-----+-----+-----+-----+-----+
TACCCAGGCAACAATAGCAACCATCTTAGTAGAACAGACCAAGACCAAGCGAGGTTAA
M G S V V I V G R I I L S G S G S A P I
ACTGCTTACTCCCAGCAGACTAGAGGATTGTTGGGTTGTATCATCACTTCCTTGACTGGT
61  -----+-----+-----+-----+-----+-----+-----+
TGACGAATGAGGGTCGTCTGATCTCCFAACAACCAACATAGTAGTGAAGGAACTGACCA
T A Y S Q Q T R G L L G C I I T S L F G
AGAGACAAGAACCAAGTTGACGGGAGAGGTTTCAGGTTTGTCCACTGCTACTCAGTCTTTC
121 -----+-----+-----+-----+-----+-----+-----+
TCTCTGTTCTGGTTCAACTGCCTCTCCAAGTCCAAAACAGGTGACGATGAGTCAGAAAG
R D K N Q V D G E V Q V L S T A T Q S F
TTGGCTACTTGTGTTAACGGTGTGTTGTTGGACTGTTTACGCTGGTGCTGGTTCTAAAAC
181 -----+-----+-----+-----+-----+-----+-----+
AACCGATGAACACAATTGCCACAAACAACCTGACAAATGCGACCACGACCAAGATTTGA
L A T C V N G V C W T V Y A G A G S K T
TTGGCTGGTCCAAAGGTTCCAATCACTCAGATGTACACAAAGTTGACCAGGATTGGTT
241 -----+-----+-----+-----+-----+-----+-----+
AACCGACCAGGTTCCAGGTTAGTGAGTCTACATGTGTTTGAACCTGGTCCCTAAACCAA
L A G P K G P I T Q M Y T N V D Q D L V
GGTTGGCCAGTCCACCAGGTGCTAGATCCATGACTCCATGTACTTGTGGTTCCCTCCGAC
301 -----+-----+-----+-----+-----+-----+-----+
CCAACCGGTCGAGGTGGTCCACGATCTAGGTACTGAGGTACATGAACACCAAGGAGGCTG
G W P A P P G A R S M T P C T C G S S D
TTGTACTTGGTTACTAGACACCGTCACTTATCCAGTTAGAAGAAGAGGAGACTCCAGA
361 -----+-----+-----+-----+-----+-----+-----+
AACATGAACCAATGATCTGTGGGACTGCAATAGGGTCAATCTTCTTCTCCTCTGAGGTCT
L Y L V T R H A D V I P V R R R G D S R
GGATCTTTGTTGTTCCCAAGACCAGTTTCTTACTTGAAGGGATCTTCCGGTGGTCCATTG
421 -----+-----+-----+-----+-----+-----+-----+
CCTAGAACAACAGGGGTTCTGGTCAAAGAATGAACCTCCCTAGAAGGCCACCAGGTAAC
G S L L S P R P V S Y L K G S S G G P L
TTGTGTCCATCCGGTCAAGTTGTTGGTATTTTTCAGAGCTGCTGTTTGTACTAGAGGTGT
481 -----+-----+-----+-----+-----+-----+-----+
AACACAGGTAGGCCAGTGCACAACCATATAAAGTCTCGACGACAAACATGATCTCCACAA
L C P S G H V V G I F R A A V C T R C V
GCTAAGGETGTGACTTCATCCAGTTGAGTCCATGGAGACTACTATGAGATCCCCAGTT
541 -----+-----+-----+-----+-----+-----+-----+
CGATTCCGACAACCTGAAGTAGGGTCAACTCAGGTACCTCTGATGATACTCTAGGGGTCAA
A K A V D F I P V E S M E T T M R S P V
TTCACTGACAACCTTCCCCACCTGCTGTTCACAAACTTTCCAAGTTGCTCACTTGCAT
601 -----+-----+-----+-----+-----+-----+-----+
AAGTGACTGTTGAGAAGGGTGGACGACAAGGTGTTTGAAGGTTCAACGAGTGAACGTA
F T D N S S P P A V P Q T F Q V A H L H
GCTCCAACCTGGTTCTGGTAAGTCCACTAAGGTTCCACCTGCTTACGCTGCTCAAGGTTAC
    
```

661 -----+-----+-----+-----+-----+
CGAGGTTGACCAAGACCATTCCAGGTGATTCCAAGGTCGACGAATGCGACGAGTTCCAATG
A P T G S G K S T K V P A A Y A A Q G Y

721 -----+-----+-----+-----+-----+
AAGGTTTTGGTTTTGAACCCATCCGTTGCTGCTACTTTGGGTTTCGGTGCTTACATGTCT
TTCCAAAACCAAACCTTGGGTAGGCAACGACGATGAAACCCAAAGCCACGAATGTACAGA
K V L V L N P S V A A T L G F G A Y M S

781 -----+-----+-----+-----+-----+
AAGGCTCACGGTATTGAGCCAAACATCAGAACTGGTGTAGAACTATCACTACTGGTGGT
TTCCGAGTGCCATAACTCGGTTTGTAGTCTTGACCACAATCTTGATAGTGATGACCACCA
K A H G I E P N I R T G V R T I T T G G

841 -----+-----+-----+-----+-----+
CCTATTACTTACTCCGCTTACGGAAAGTTTTTGGCTGACGGTGGTTGTTCTGGTGGTGGT
GGATAATGAATGAGGCGAATGCCTTTCAAAACCGACTGCCACCAACAAGACCACCACGA
P I T Y S A Y G K F L A D G G C S G G A

901 -----+-----+-----+-----+-----+
TACGACATCATCTGTGACGAGTGTCACTCTACTGACTGGACTACTATCTTGGGTATC
ATGCGTAGTAGTAGACTGCTCACAGTGAGATGACTGACCTGATGATAGAACCCATAG
Y D I I I C D E C H S T D W T T I L G I

961 -----+-----+-----+-----+-----+
GGTACTGTTTTGGACCAAGCTGAAACTGCTGGTGTAGATTGGTTGTTTTGGCTACTGCT
CCATGACAAAACCTGGTTCGACTTTGACGACCACGATCTAACCAACAAAACCGATGACGA
G T V L D Q A E T A G A R L V V L A T A

1021 -----+-----+-----+-----+-----+
ACTCCACCAGGTTCCATTACTGTTCCACACCCAAACATCGAGGAAGTTGCTTTGTCTAAC
TGAGGTGGTCCAAGGTAATGACAAGGTGTGGGTTTGTAGCTCCTTCAACGAAACAGATTG
T P P G S I T V P H P N I E E V A L S N

1081 -----+-----+-----+-----+-----+
ACTGGAGAGATCCCATTTCTACGGAAAGGCTATCCCAATTGAGGCTATCAAGGGTGGTAGA
TGACCTCTCTAGGGTAAGATGCCTTTCCGATAGGGTAACTCCGATAGTTCACCACATCT
T G E I P F Y G K A I P I E A I K G G R

1141 -----+-----+-----+-----+-----+
CACTTGATTTTTCTGTCACTCCAAGAAGAAGTGTGACGAGTTGGCTGCTAAGTTGACTGGA
GTGAACATAAAGACAGTGAGGTTCTTCTTCACACTGCTCAACCGACGATTCAACTGACCT
TTGGGATIGAACGCTGTTGCTTACTACAGAGGATTGGACGTTCCGTTATCCCAACTTCC
1201 -----+-----+-----+-----+-----+
AACCTAACTTGCGACAACGAATGATGTCTCCTAACCTGCAAAGCCAATAGGGTTGAAGG
L G L N A V A Y Y R G L D V S V I F T S

1261 -----+-----+-----+-----+-----+
GGTGATGTTGTTGTTGTTGCTACTGACGCTTTGATGACTGCTTTCACTCGTGACTTCGAC
CCTACTACAACAACAACACGATGACTGCGAAACTACTGACCAAAGTGACCACTGAAGCTG
G D V V V V A T D A L M T G F T G D F D

1321 -----+-----+-----+-----+-----+
TCCGTTATCGACTGTAACACTTGTGTTACTCAGACTGTTGACTTCTCCTTGGACCCAACT
AGGCAATAGCTGACATTGTGAACACAATGAGTCTGACAACCTGAAAGGAACCTGGGTTGA
S V I D C N T C V T Q T V D F S L D P T

1381 -----+-----+-----+-----+-----+
TTCATATCGAGACTACTACTGTTCTCAAGACGCTGTTTCCAGATCCCAAAGAAGAGGT
AAGFGATAGCTCTGATGATGACAAGGAGTTCTGCGACAAGGCTAGGGTTTCTTCTCCA
F T I E T T T V P Q D A V S R S Q R R G

GCTACTGGTAGAGGAAGATCCGGTATCTACAGATTTCGTTACTCCAGGTGAAAGACCATCT
1441-----+-----+-----+-----+-----+-----+-----+-----+
CGATGACCATCTCCTTCTAGGCCATAGATGTCTAAGCAATGAGGTCCACTTTCTGGTAGA
A T G R G R S G I Y R F V T P G E R P S

GGAAATGTTCCGACTCCTCCGTTTTGTGTGAATGTTACGACGCTGGTGTGCTTGGTACGAA
1501-----+-----+-----+-----+-----+-----+-----+-----+
CCTTACAAGCTGAGGAGGCCAAAACACACTTACAATGCTGCGACCAACACGAACCATGCTT
G M F D S S V L C E C Y D A G C A W Y E

TTGACTCCAGCTGAGACTACTGTTAGATTGAGAGCTTACTTGAACACTCCAGGATTGCCA
1561-----+-----+-----+-----+-----+-----+-----+-----+
AACTGAGGTCCGACTCTGATGACAATCTAACTCTCGAATGAACTTGTGAGGTCCCTAACGGT
L T P A E T T V R L R A Y L N T P G L P

GTTTGTCAAGACCACTTGGAAATCTGGGAGTCCGTTTTCACTGGATTGACTCACATTGAC
1621-----+-----+-----+-----+-----+-----+-----+-----+
CAAACAGTTCTGGTGAACCTTAAGACCCTCAGGCCAAAAGTGACCTAACTGAGTGTAACTG
V C Q D H L E F W E S V F T G L T H I D

GCTCACTTTTTGTCCCAAATAAGCAGGCTGGTGACAACFTTCCATACTTGGTGTGCTTAC
1681-----+-----+-----+-----+-----+-----+-----+-----+
CGAGTGAAAACAGGGTTTTGATTTCGTCCGACCACCTGTTGAAAGGTATGAACCAACGAATG
A H F L S Q T K Q A G D N F P Y L V A Y

CAGGCTACTGTTGTGCTAGAGCACAAGCTCCACCACCCTCTGGGATCAGATGTGGAAG
1741-----+-----+-----+-----+-----+-----+-----+-----+
GTCCGATGACAAAACAGGATCTCGTGTTCGAGGTGGTGGTAGAACCTTAGTCTACACCTTC
Q A T V C A R A Q A P F P S W D Q M W K

TGTTTGATCAGATTGAAGCCAACTTTGCACGGTCCAACTCCATTGTTGTACAGATTGGGT
1801-----+-----+-----+-----+-----+-----+-----+-----+
ACAACTAGTCTAACTTCGGTTGAAACGTGCCAGGTGAGGTAACAACATGTCTAACCCA
C L I R L K P T L H G P T P L L Y R L G

GCTGTTCCAGAACGAGATCACTTTGACTCACCCAATCACTAAGTTCGTTATGGCTTGCAIG
1861-----+-----+-----+-----+-----+-----+-----+-----+
CGACAAGTCTTGCTCTAGTGAACCTGAGTGGTGTAGTGATTCAAGCAATACCGAACGTAC
A V Q N E I T L T H P I T K F V M A C H

TCTGCTGACTTGAAGTTGTTGGTTCGGATCTGGTCCATGTCTTACACTTGGACTGGT
1921-----+-----+-----+-----+-----+-----+-----+-----+
AGACGACTGAACCTTCAACAACCAAGGCCTAGACCAAGGTACAGGATGTGAACCTGACCA
S A D L E V V G S G S G S M S Y T W T G

GCTTTGATCACTCCATGTGCTGCTGAAGAATCCAAGTTGCCAATCAACCCATTGTCCAAC
1981-----+-----+-----+-----+-----+-----+-----+-----+
CGAACTAGTGAGGTACACGACGACTTCTTAGGTTCAACGGTTAGTTGGGTAACAGGTTG
A L I E P C A A E E S K L P I N P L S N

TCTTTGTTGAGACACCACCTCCATGGTTTACTCCACTACTTCCAGATCCGCTTCCCTTGAGA
2041-----+-----+-----+-----+-----+-----+-----+-----+
AGAAACAACTCTGTGGTGGGTACCAAATGAGGTGATGAAGGTCTAGGCCAAGGAACCTCT
S L L R H K S M V Y S T T S R S A S L R

CAGAAGAAGGTTACATTCGACAGATTGCAGGTTTTGGACGACCACTACAGAGATGTTTTG
2101-----+-----+-----+-----+-----+-----+-----+-----+
GTCTTCTCCAATGTAAGCTGTCTAACGTCCAAAACCTCGTGGTGTGCTCTACAAAAC
Q K K V T F D R L Q V L D D H Y R D V L

AAGGAGATGAAGGCTAAGGCTTCCACTGTTAAGGCTAGATTGTTGTCCATTGAGGAGGCT
2161-----+-----+-----+-----+-----+-----+-----+-----+
TTCTCTACTTCCGATTCCGAAGGTGACAATTCGGATCTAACAAACAGGTAACCTCCTCCGA

K E M K A K A S T V K A R L L S I E E A
 TGTAAGTTGACTCCACCACACTCTGCTAAGTCCAAGTTTGGTTACGGTGCTAAGGATGTT
 2221-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ACATTCAACTGAGGTGGTGTGAGACGATTCAGGTTCAAACCAATGCCACGATTCCTACAA
C K L T P P H S A K S K F G Y G A K D V
 AGATCCTTGTCTCCAGAGCTGTTAACCACATCAGATCCGTTTGGGAGGATTTGTTGGAG
 2281-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
TCTABGAACAGGAGGTCTCGACAATTGGTGTAGTCTAGGCCAAACCCTCCTAAACAACCTC
R S L S S R A V N H I R S V W E D L L E
 GACACTGAGACTCCAATCGACACTACTATCATGGCTAAGAACGAGGTTTTCTGTGTTCAA
 2341-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CTGTGACTCTGAGGTTAGCTGTGATGATAGTACCGATTCTTGCTCCAAAAGACACAAGTT
D T E T P I D T T I M A K N E V F C V Q
 CCAGAGAAGGGTGAAGAAAGCCAGCTAGATTGATCGTTTTCCAGACTTGGGTGTTAGA
 2401-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GGTCTCTTCCACCTTCTTTCCGGTGCATCTAACTAGCAAAAGGGTCTGAACCCACATCT
P E K G G R K P A R L I V F P D L G V R
 GTTTGTGAGAAGATGGCTTTGTACGACGTTGTTTCCACTTTGCCACAGGCTGTTATGGGA
 2461-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CAACACTCTTCTACCGAAACATGCTGCAACAAAGGTGAAACGGTGTCCGACAATACCCT
V C E K M A L Y D V V S T L P Q A V M G
 CCATCTTACGGTTTCCAATACTCCCCAGGACAAAGAGTTGAGTTCTTGGTTAACACTTGG
 2521-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GGTAGAATGCCAAAGGTTATGAGGGGCTCTGTTTCTCAACTCAAGAACCAATTGTGAACC
P S Y G F Q Y S P G Q R V E F L V N T W
 AAGTCCAAGAAATGTCCAATGGGATTCCTTACAACACTAGATGTTTCGACTCCACTGTT
 2581-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
TTCAGGTTCTTACAGGTTACCCTAAGAGGATGTTGTGATCTACAAGCTGAGGTGACAA
K S K K C P M G F S Y N T R C F D S T V
 ACTGAGAACGACATCAGAACAGAGGAGTCCATCTACCAGTGTGTTGACTTGGCTCCAGAA
 2641-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
TGACTCTTGTGTAGTCTTGTCTCCTCAGGTAGATGGTCACAACACTGAACCGAGGTTCTT
T E N D I R T E E S I Y Q C C D L A P E
 GCTAGACAAGCTATCAAGTCCCTTACTGAGAGATTGTACATCGGTGGTCCCTTGGACTAAC
 2701-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
CGATCTGTTGATAGTTCAGGAACTGACTCTCTAACATGTAGCCACCAGGAAACTGATTG
A R Q A I K S L T E R L Y I G G P L T N
 TCCAAGGGACAGAACTGTGGTTACAGAAGATGTAGAGCTTCCGGTGTGTTTACTACTTCC
 2761-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
AGGTTCCCTGTCTTGACACCAATGTCTTCTACATCTCGAAGGCCACAAAAGTGTGAAGG
S K G Q N C G Y R R C R A S G V L T T S
 TGTGGFAACACTTTGACTTGTACTTGAAGGCTACTGCTGCTTGTAGAGCTGCTAAATG
 2821-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
ACACCATTGTGAAACTGAACAATGAACCTCCGATGACGACGAACATCTCGACGATTTAAC
C G N T L T C Y L K A T A A C R A A K L
 CAGGACTGTACTATGTTGGTTAACGGTAACGACTTGGTTGTTATCTGTGAGTCCGCTGGT
 2881-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
GTCCTGACATCATAACAACCAATTGCCATTGCTGAACCAACAATAGACACTCAGCCGACCA
Q D C T M L V N G N D L V V I C E S A G
 ACTCAAGAAGATGCTGCTTCTTTGAGAGTTTTCCAGAGGCTATGACTAGATACTCTGCT

2941-----+-----+-----+-----+-----+
 TGAGTTCTTCTACGACGAAGAACTCTCAAAAGTGTCTCCGATACTGATCTATGAGACGA
T_Q_E_D_A_A_S_L_R_V_F_T_E_A_M_T_R_Y_S_A

CCACCTGGTGATCCACCACAACCAGAATACGACTTGGAGTTGATCACTTCCTGTTCCTCC
 3001-----+-----+-----+-----+-----+
 GGTGGACCACTAGGTGGTGTGGTCTTATGCTGAACCTCAACTAGTGAAGGACAAGGAGG
P_P_G_D_P_P_Q_P_E_Y_D_L_E_L_I_T_S_C_S_S

AATGTTGGTGTGCTCAGGATGCTTCCGGAAAGAGAGTTTACTACTTGACTAGAGACCCA
 3061-----+-----+-----+-----+-----+
 TTACAACCACAACGAGTGTCTAGGAAGGCTTTCTCTCAAATGATGAAGTCTCTGGGT
N_V_G_V_A_H_D_A_S_G_K_R_V_Y_Y_L_T_R_D_P

ACTACTCCATTGGCTAGAGCTGCTTGGGAAACTGTTAGACACACTCCAGTAACTCCTGG
 3121-----+-----+-----+-----+-----+
 TGATGAGGTAACCGATCTCGACGAACCCCTTTGACAATCTGTGTGAGGTCAATFGAGGACC
T_T_P_L_A_R_A_A_W_E_T_V_R_H_T_P_V_N_S_W

TTGGGTAACATCATCATGTACGCTCCAACCTTTGTGGGCTAGAATGATCTTGATGACTCAC
 3181-----+-----+-----+-----+-----+
 AACCCATTGTAGTAGTACATGCGAGGTTGAAACACCCGATCTTACTAGAAGTACTGAGTG
L_G_N_I_I_M_Y_A_P_T_L_W_A_R_M_I_L_M_T_H

TTCTTCTCCATCTTGTGGCTCAAGAGCAATTGGAAAAGGCTTTEGACTGTCAGATTTAC
 3241-----+-----+-----+-----+-----+
 AAGAAGAGGTAGAACAACCGAGTTCCTCGTTAACCTTTTCCGAAACTGACAGTCTAAATG
F_F_S_I_L_L_A_Q_E_Q_L_E_K_A_L_D_C_Q_I_Y

GGTGCTTGTACTCCATTGAGCCATTGGACTTGGCACAGATCATTGAGAGATTGCACGGT
 3301-----+-----+-----+-----+-----+
 CCACGAACAATGAGGTAACCTCGTAACCTGAACGGTGTCTAGTAACCTCTTAACGTGCCA
G_A_C_Y_S_I_E_P_L_D_L_P_Q_I_I_E_R_L_H_G

TTGTCTGCTTTCTCTTTGCACTCTTACTCCCCTGGTGAATCAACAGAGTTGCTTCCTGT
 3361-----+-----+-----+-----+-----+
 AACAGACGAAAGAGAAACGTGAGAATGAGGGGACCACTTTAGTTGTCTCAACGAAGGACA
L_S_A_F_S_L_H_B_Y_S_P_G_E_I_N_R_V_A_S_C

TTGAGAAAAGTTGGGTGTTCCACCATTGAGAGTTTGGACACACAGAGCTAGATCCCGTTAGA
 3421-----+-----+-----+-----+-----+
 AACTCTTTCAACCCACAAGGTGGTAACCTCAAACTCTGTGTCTCGATCTAGGCAATCT
L_R_K_L_G_V_P_P_L_R_V_W_R_H_R_A_R_S_V_R

GCTAAGTTGTTGTCCCAAGGTGGAAGAGCTGCTACTTGTGGTAAGTACTTGTCAACTGG
 3481-----+-----+-----+-----+-----+
 CGATTCAACAACAGGGTTCCACCTTCTCGACGATGAACACCATTCAAGCAAGTTGACC
A_K_L_L_S_Q_G_G_R_A_A_T_C_G_K_Y_L_F_N_W

GCTGTTAGAACAAGTTGAAGTTGACTCCATTCCTGCTGCTTCCCAATTGGATTTGTCC
 3541-----+-----+-----+-----+-----+
 CGACAATCTTGTTCACCTCAACTGAGGATAAGGACGACGAAGGGTTAACCTAACAGG
A_V_R_T_K_L_K_L_T_P_I_P_A_A_S_Q_L_D_L_S

GGTTGGTTTGTGCTGGTTACAACGGTGGTGACATCTACCACCTTTTGTCCAGAGCTAGA
 3601-----+-----+-----+-----+-----+
 CCAACCAACAACGACCAATGTTGCCACCACTGTAGATGGTGAGAAACAGGTCTCGATCT
G_W_F_V_A_G_Y_N_G_G_D_I_Y_H_S_L_S_R_A_R

CCAAGATAATAGACTAGTGGATCC
 3661-----+-----+-----+-----+-----+
 GGTTCTATTATCTGATCACCTAGG
P_R_*_*

Fig. 6

Proteína de fusión 4A-3-4B-5B optimizada para la expresión en *Pichia pastoris*

```

ATGGGGTCCGGTTGTTATCGTTGGTAGAATCATCTTGTCTGGTTCIGGTTCCGCTCCAATT
1  -----+-----+-----+-----+-----+-----+-----+
TACCCAGGCAACAATAGCAACCATCTTAGTAGAACAGACCAAGACCAAGGCGAGGTTAA
M C S V V I V G R I I L S G S G S A F I
ACTGCTTACTCCCAGCAGACTAGAGGATTGTTGGGTTGTATCATCACTTCCTTGACTGGT
61  -----+-----+-----+-----+-----+-----+-----+
TGACGAATGAGGGTGGTCTGATCTCCTAACCAACCAACATAGTASTGAAGGAACTGACCA
T A Y S Q Q T R G L L G C I I T S L T G
AGAGACAAGAACCAAGTTGACGGAGAGGTTTCAGGTTTTGTCCACTGCTACTCAGTCTTC
121  -----+-----+-----+-----+-----+-----+-----+
TCTCTGTTCTTGGTTCAACTGCCTCTCCAAGTCCAAAACAGGTGACGATGAGTCAGAAG
R D K N Q V D G E V Q V L S T A T Q S F
TTGGCTACTTGTGTTAACGGTGTTTGTTGGACTGTTTACGCTGGTGTGGTTCIAAACT
181  -----+-----+-----+-----+-----+-----+-----+
AACCGATGAACACAATTGCCACAAACAACCTGACAAATGCGACCACGACCAAGATTTTGA
L A T C V N G V C W T V Y A G A G S K T
TTGGCTGGTCCAAAGGGTCCAATCACTCAGATGTACACAAACGTTGACCAGGATTTGGTT
241  -----+-----+-----+-----+-----+-----+-----+
AACCGACCAGGTTTCCCAGGTTAGTGAGTCTACATGTGTTTGCAACTGGTCTAAACCAA
L A G P K G P I T Q M Y T N V D Q D L V
GGTTGGCCAGCTCCACCAGGTGCTAGATCCATGACTCCATGTACTTGTGGTTCCTCCGAC
301  -----+-----+-----+-----+-----+-----+-----+
CCAACCGGTTCGAGGTGGTCCACGATCTAGGTACTGAGGTACATGAACACCAAGGAGGCTG
G W F A P P G A R S M T P C T C G S S D
TTGTACTTGGTTACTAGACACGCTCAGCTTATCCCAGTTAGAAGAAGAGGAGACTCCAGA
361  -----+-----+-----+-----+-----+-----+-----+
AACATGAACCAATGATCTGTGCGACTGCAATAGGGTCAATCTTCTCTCCTCTGAGGTCT
L Y L V T R H A D V I P V R R R G D S R
GGATCTTTGTTGTCCCAAGACCAGTTTCTTACTTGAAGGGATCTTCCGGTGGTCCATTG
421  -----+-----+-----+-----+-----+-----+-----+
CCTAGAACAACAGGGGTTCCTGGTCAAAGAATGAACTTCCCTAGAAGGCCACCAGTTAAC
G S L L S P R P V S Y L K G S S G G P L
TTGTGTCCATCCGGTCACGTTGTTGGTATTTTCAGAGCTGCTGTTTGTACTAGAGGTGTT
481  -----+-----+-----+-----+-----+-----+-----+
AACACAGGTAGGCCAGTGCAACAACCATAAAAGTCTCGACGACAAACATGATCTCCACAA
L C P S G H V V G I F R A A V C T R G V
GCTAAGGCTGTTGACTTCATCCCAGTTGAGTCCATGGAGACTACTATGAGATCCCCAGTT
541  -----+-----+-----+-----+-----+-----+-----+
CGATTCCGACAACCTGAGTAGGGTCAACTCAGGTACCTCTGATGATACTCTAGGGGTCAA
A K A V D F I P V E S M E T T M R S P V
TTCCTGACAACCTCTTCCCCACCTGCTGTTCCACAACTTTCCAAGTTGCTCACTTGCAT
601  -----+-----+-----+-----+-----+-----+-----+
AAGTGACTGTTGAGAAGGGGTGGACGACAGGTTGTTGAAAGGTTCAACGAGTGAACGTA
F T D N S S P P A V F Q T F Q V A H L H
GCTCCAACCTGGTCTGGTAAGTCCACTAAGGTTCCAGCTGCTTACGCTGCTCAAGGTTAC
661  -----+-----+-----+-----+-----+-----+-----+
CGAGGTTGACCAAGACCATTTCAGGTGATTCCAAGGTCGACGAATGCGACGAGTTCCAATG

```

A P T G S G K S T K V P A A Y A A Q G Y
 AAGGTTTTGGTTTTGAACCCATCCGTTGCTGCTACTTTGGGTTTCGGTGCCTTACATGTCT
 721 -----+-----+-----+-----+-----+-----+-----+-----+
TTCCAAAACCAAACTTGGGTAGGCAACGACGATGAAACCCAAAGCCACGAATGTACAGA
K V L V L N P S V A A T L G F G A Y M S
 AAGGCTCACGGTATTGAGCCAAACATCAGAACTGGTGTAGAACTATCACTACTGGTGGT
 781 -----+-----+-----+-----+-----+-----+-----+
TCCGAGTGCATAACTCGGTTTGTAGTCTTGACCACAATCTTGATAGTGATGACCACCA
K A H G I E P N I R T G V R T I T T G G
 CCTATTACTTACTCCGCTTACGGAAAGTTTTTGGCTGACGGTGGTTGTTCTGGTGGTGC
 841 -----+-----+-----+-----+-----+-----+-----+
GGATAATGAATGAGGCGAATGCCTTTCAAAAACCGACTGCCACCAACAAGACCACCACGA
P I T Y S A Y G K F L A D G C C S G G A
 TACGACATCATCATCTGTGACGAGTGTCACTCTACTGACTGGACTACTATCTTGGGTATC
 901 -----+-----+-----+-----+-----+-----+-----+
ATGCTGTAGTAGTAGACACTGCTCACAGTGAGATGACTGACCTGATGATAGAACCCATAG
Y D I I I C D E C H S T D W T T I L G I
 GGTACTGTTTTGGACCAAGCTGAAACTGCTGGTGTAGATTGGTTGTTTTGGCTACTGCT
 961 -----+-----+-----+-----+-----+-----+-----+
CCATGACAAAACCTGGTTCGACTTTGACGACCACGATCTAACCAACAAAACCGATGACGA
G T V L D Q A E T A G A R L V V L A T A
 ACTCCACCAGGTTCCATTACTGTTCCACACCCAAACATCGAGGAAGTIGCTTTGTCCTAAC
 1021 -----+-----+-----+-----+-----+-----+-----+
TGAGTGGTCCAAGGTAATGACAAGGTGTGGGTTTTSTAGCTCCTTCAACGAAACAGATTG
T P P G S I T V P H P N I E E V A L S N
 ACTGGAGAGATCCCATTCTACGGAAAGGCTATCCCAATTGAGGCATCAAGGGTGSTAGA
 1081 -----+-----+-----+-----+-----+-----+-----+
TGACCTCTCTAGGGTAAGATGCCTTTCCGATAGGGTTAACTCCGATAGTTCCCACCATCT
T G E I P F Y G K A I P I E A I K G G R
 CACTTGATTTTCTGTCACTCCAAGAAGAAGTGTGACGAGTTGGCTGCTAAGTTGACTGGA
 1141 -----+-----+-----+-----+-----+-----+-----+
GTGAACTAAAAGACAGTGAGGTTCTTCTTACACTGCTCAACCGACGATTCAACTGACCT
H L I F C H S K K K C D E L A A K L T G
 TTGGGATTGAACGCTGTTGCTTACTACAGAGGATTGGACGTTTCCGTTATCCCAACTTCC
 1201 -----+-----+-----+-----+-----+-----+-----+
AACCTAACTTGGGACAACGAATGATGTCTCCTAACCTGCAAAGGCAATAGGTTGAAGG
L G L N A V A Y Y R G L D V S V I P T S
 GGTGATGTTGTTGTTGTTGCTACTGACGCTTTGATGACTGGTTTCACTGGTACTTCGAC
 1261 -----+-----+-----+-----+-----+-----+-----+
CCACTACAACAACAACAACGATGACTGCGAAACTACTGACCAAAGTGACCACCTGAAGCTG
G D V V V V A T D A L M T G F T G D F D
 TCGTTATCGACTGTAACACTTGTGTTACTCAGACTGTTGACTTCTCCTTGGACCCAACT
 1321 -----+-----+-----+-----+-----+-----+-----+
AGGCAATAGCTGACATTGTGAACACAATGAGTCTGACAACCTGAAGAGGAACCTGGGTTGA
S V I D C N T C V T Q T V D F S L D P T
 TTCCTATCGAGACTACTACTGTTCTCCTCAAGACGCTGTTTCCAGATCCCAAAGAAGAGGT
 1381 -----+-----+-----+-----+-----+-----+-----+
AAGTGATAGCTCTGATGATGACAAGGAGTCTGCGACAAGGTCCTAGGTTTCTTCTCCA
F T I E T T T V P Q D A V S R S Q R R G
 GCTACTGGTAGAGGAAGATCCGGTATCTACAGATTCGTTACTCCAGGTGAAAGACCATCT

1441-----+-----+-----+-----+-----+-----+-----+
 CGATGACCATCTCCTTCTAGGCCATAGATGTCTAAGCAATGAGGTCCACTTTCTGGTACA
A T G R G R S G I Y R F V T P G E R F S
 GGAATGTTCCGACTCCTCCGTTTTGTGTGAATGTTACGACGCTGGTTGTGCTTGCTACGAA
 1501-----+-----+-----+-----+-----+-----+-----+
 CCTTACAAGCTGAGGAGGC AAAACACACTTACAATGCTGCGACCAACACGAACCATGCTT
G M F D S S V L C E C Y D A G C A W Y E
 TTGACTCCAGCTGAGACTACTGTTAGATTGAGAGCTTACTTGAACACITCCAGGATTGCCA
 1561-----+-----+-----+-----+-----+-----+-----+
 AACTGAGGTGACTCTGATGACAATCTAACTCTCGAATGAACTTGTGAGGTCCCTAACGGT
L T P A E T T V R L R A Y L N T P G L P
 GTTTGTCAAGACCACTTGGAAATTCBGGAGTCCGTTTTTCACTGGATTGACTCACATTGAC
 1621-----+-----+-----+-----+-----+-----+-----+
 CAAACAGTTCGGTGAACCTAAGACCCCTCAGGCCAAAAGTGACCTAAGTGTAACTG
V C Q D H L E F W E S V F T G L T H I D
 GCTCACTTTTTGTCCCAAAC TAAGCAGGCTGGTGACAACCTTCCATACTTGGTTGCTTAC
 1681-----+-----+-----+-----+-----+-----+-----+
 CGAGTGAAAACAGGGTTTGATTCGTCGACCACTGTTGAAAGGTATGAACCAACGAATG
A H F L S Q T K Q A G D N F P Y L V A Y
 CAGGCTACTGTTTGTGCTAGAGCACAAGCTCCACCACCATCTTGGGATCAGATGTGGAAG
 1741-----+-----+-----+-----+-----+-----+-----+
 GTCCGATGACAAACACGATCTCGTSTTCGAGGTGGTGGTAGAACCCCTAGTCTACACCTTC
Q A T V C A R A Q A F P P S W D Q M W K
 TGTTTGATCAGATTGAAGCCAACCTTTCACGGTCCAACCTCCATTGTTGTACAGATTGGGT
 1801-----+-----+-----+-----+-----+-----+-----+
 ACAAACTAGTCTAACTTCGGTTGAAACGTTGCCAGGTTGAGGTAACAACATGTCTAACCCA
C L I R L K P T L H G P T F L L Y R L G
 GCTGTTCAGAACGAGATCACTTTGACTCACCCAATCACTAAGTTCGTTATGGCTTGCATG
 1861-----+-----+-----+-----+-----+-----+-----+
 CGACAAGTCTTGCTCTAGTGAAACTGAGTGGGTTAGTGATTCAAGCAATACCGAACGTAC
A V Q N E I T L T H P I T K F V M A C M
 TCTGCTGACTTGAAGTTGTTTCCTTGATGGCTTTCCTGCTTCCATTACTTCCCATTG
 1921-----+-----+-----+-----+-----+-----+-----+
 AGCGACTGAACCTTCAACAAGGAACCTACCGAAAGTGACGAAGTAATGAAGGGGTAAC
S A D L E V V S L M A F T A S I T S P L
 ACTACTCAGAACACTTTGTTGTTCAACATCTTGGGAGGATGSGTTCAGCTCAATTGTCC
 1981-----+-----+-----+-----+-----+-----+-----+
 TGATGAGTCTTGTAACAACAAGTTGTAGAACCCTCCTACCCACCGTCCAGTTAACAGG
T T Q N T L L E N I L G G W V A A Q L S
 ATGTCTACACTTGGACTGGTGCTTTGATCACTCCATGTGCTGCTGAAGAATCCAAGTTG
 2041-----+-----+-----+-----+-----+-----+-----+
 TACAGGATGTGAACCTGACCACGAAACTAGTGAGGTACACGACGACTTCTTAGGTTCAAC
M S Y T W T G A L I T P C A A E E S K L
 CCAATCAACCCATTGTCCAACCTTTGTTGAGACACCCTCCATGGTTACTCCACTACT
 2101-----+-----+-----+-----+-----+-----+-----+
 GGTAGTGGGTAACAGGTTGAGAAACAACCTCTGTGGTGAGGTACCAAATGAAGTGATGA
P I N P L S N S L L R H H S M V Y S T T
 TCCAGATCCGCTTCCTTGAGACAGAAGAAGGTTACATTCCGACAGATTGCAGGTTTGGAC
 2161-----+-----+-----+-----+-----+-----+-----+
 AGGTCTAGGCCGAAGGAACCTCTGTCTTCTTCCAATGTAAGCTGTCTAACGTCCAAAACCTG
S R S A S L R Q K K V T F D R L Q V L D

2221 GACCACTACAGAGATGTTTTGAAGGAGATGAAGGCTAAGGCTTCCACTGTAAAGGCTAGA
 CTGGTGTATGCTCTACAAAACCTTCCCTCTACTTCCGATTCGGAAGGTGACAATTCGGATCT
 D H Y R D V L K E M K A K A S I V K A R
 TTGTTGTCCATTGAGGAGGCTTGTAAAGTGTACTCCACCACACTCTGCTAAGTCCAAGTTT
 2281 AACAAACAGGTAACCTCCTCCGAACATTCAACTGAGGTGGTGTGAGACGATTCAAGTTCAAA
 L L S I E E A C K L T P P H S A K S K F
 GGTACGGTGTCTAAGGATGTTAGATCCTTGTCTCCAGAGCTGTTAACCACATCAGGTCT
 2341 CCAATGCCACGATTCTTACAATCTAGEAACAGGAGGTCTCGACAATTGGTGTAGTCCAGA
 G Y G A K D V R S L S S R A V N H I R S
 GTTTGGGAGGATTTGTTGGAGGACACTGAGACTCCAATCGACACTACTATCATGGCTAAG
 2401 CAAACCCCTCCTAAACAACCTCCTGTGACTCTGAGGTAGCTGTGATGATAGTACCGATT
 V W E D L L E D T E T P I D T T I M A K
 AACGAGGTTTTCTGTGTTCAACCAGAGAAGGGTGGAAAGAAAGCCAGCTAGATTGATCGTT
 2461 TTGCTCCAAAAGACACAAGTTGGTCTCTTCCACCTTCTTTCGGTCTGATCTAACTAGCAA
 N E V F C V Q P E K G G R K P A R L I V
 TTCCCAGACTTGGGTGTTAGAGTTTGTGAGAAGATGGCTTGTACGACGTTGTTCCACT
 2521 AAGGGTCTGAACCCACAATCTCAACACTCTTCTACCGAACATGCTGCAACAAAGGTGA
 F P D L G V R V C E K M A L Y D V V S T
 TTGCCACAGGCTGTTATGGGACCATCTTACGGTTTCCAATACITCCCCAGGACAAAGAGTT
 2581 AACGGTGTCCGACAATACCTGGTAGAATGCCAAAGGTTATGAGGGTCTCTGTTCTCAA
 L P Q A V H G P S Y G F Q Y S P G Q R V
 GAGTCTTGGTTAACACTTGGAACTCCAAGAAATGTCCAATGGGATTCTCCTACAACACT
 2641 CTCAAGAACCAATTGIGAACCTTCAGGTTCTTACAGGTTACCCTAAGAGGATGTTGTGA
 E F L V N T W K S K K C P M G F S Y N T
 AGATGTTTCGACTCCACTGTTACTGAGAACGACATCAGAACAGAGGAGTCCATCTACCAG
 2701 TCTACAAAGCTGAGGTGACAAATGACTCTTGTCTGATGCTTGTCTCCTCAGGTAGATGGTC
 R C F D S T V T E N D I R T E E S I Y Q
 TGTTGTGACTTGGCTCCAGAAGCTAGACAAGCTATCAAGTCTTGTACTGAGAGATTGTAC
 2761 ACAACACTGAACCGAGGTCTTCGATCTGTTTCGATAGTTCAGGAAGTACTCTCTAACATG
 C C D L A P E A R Q A I K S L T E R L Y
 ATCGGTGGTCTTGTACTACTCCAAGGGACAGAAGTGTGGTTACAGAAGATGTAGAGCT
 2821 TAGCCACCAGGAAACTGATTGAGGTTCCCTGTCTTGACACCAATGTCTTCTACATCTCGA
 I G G P L T N S K G Q N C G Y R R C R A
 TCCGGTGTGTTTGTACTACTTCCCTGTGGTAACACTTGTACTTGTACTTGAAGGCTACTGCT
 2881 AGGCCACAAAACCTGATGAAGGACACCATTGTGAACTGAACAATGAACCTCCGATGACGA
 S G V L F T S C G N T L T C Y L K A T A
 GCTTGTAGAGCTGCTAAATTGCAGGACTGTACTATGTTGGTTAACGGTAACGACTTGGTT
 2941

CGAACATCTCGACGATTTAACGTCTTGACATGATACAACCAATTGCCATTGCTGAACCAA
A C R A A K L Q D C T M L V N G N D L V

3001 GTTATCTGTGACTCCGCTGGTACTCAAGAAGATGCTGCTTCTTTGAGAGTTTTACAGAG
 +-----+-----+-----+-----+-----+-----+-----+-----+
 CAATAGACACTCAGGCGACCATGAGTTCTTCTACGACGAAGAAACTCTCAAAGTGTCTC
V I C E S A G T Q E D A A S L R V F T E

3061 GCTATGACTAGATACTCTGCTCCACCTGGTGATCCACCACAACCAGAATACGACTTGGAG
 +-----+-----+-----+-----+-----+-----+-----+-----+
 CCATACTGATCTATGAGACGAGGTGGACCACTAGGTGGTGTGGTCTTTATGCTGAACCTC
A M T R Y S A P P G D P P Q P E Y D L E

3121 TTGATCACTTCCTGTTCCCAATGTTGGTGTGGTCTCACGATGCTTCCGAAAGAGAGTT
 +-----+-----+-----+-----+-----+-----+-----+-----+
 AACTAGTGARGGACAAGGAGGTTACAACCACAACGAGTGTACGAAGGCCCTTCTCTCAR
L I T S C S S N V G V A H D A S G K R V

3181 TACTACTTGACTAGAGACCCAACTACTCCATTGGCTAGAGCTGCTTGGGAAACTGTTAGA
 +-----+-----+-----+-----+-----+-----+-----+-----+
 ATGATGAACTGATCTCTGGGTTGATGAGGTAACCGATCTCGACGAACCTTTGACAATCT
Y Y L T R D F T T P L A R A A W E T V R

3241 CACACTCCAGTTAACTCCTGGTGGGTAACATCATCATGTACGCTCCAACCTTTGTGGGCT
 +-----+-----+-----+-----+-----+-----+-----+-----+
 GTGTGAGGTCAATTGAGGACCAACCCATTGTAGTAGTACATGCGAGGTTGAACACCCCGA
H T P V N S W L G N I I M Y A P T L W A

3301 AGAATGATCTTGATGACTCACTTCTTCTCCATCTTGTGGCTCAAGAGCAATTGGAAAAG
 +-----+-----+-----+-----+-----+-----+-----+-----+
 TCTTACTAGAACTACTGAGTGAAGAAGAGGTAGAACAACCGAGTTCGTTAACCTTTTC
R M I L M T H F F S I L L A Q E Q L E K

3361 GCTTTGGACTGTGAGATTTACGGTGGTGTACTCCATTGAGCCATTGGACTTGCCACAG
 +-----+-----+-----+-----+-----+-----+-----+-----+
 CGAAACCTGACAGTCTAATGCCACGAACAATGAGGTAACCTCGGTAACTGAAACGGTGT
A L D C Q I Y G A C Y S I E P L D L P Q

3421 ATCATTGAGAGATTCACGGTTGTCTGCTTTCTCTTTGCACCTTFACTCCCCTGGTGAA
 +-----+-----+-----+-----+-----+-----+-----+-----+
 TAGTAACTCTCTAACGTGCCAAACAGACGAAGAGAGAAACGTGAGARTGAGGGGACCACTT
I I E R L H G L S A F B L H S Y S P G E

3481 ATCAACAGAGTTGCTTCCCTGTTGAGAAAAGTTGGGTGTTCCACCATTGAGAGTTTGGAGA
 +-----+-----+-----+-----+-----+-----+-----+-----+
 TAGTTGTCTCAACGAAGGACAAACTCTTTCAACCCACAAGGTGGTAACTCTCAAACCTCT
I N R V A S C L R K L G V P P L R V W R

3541 CACAGAGCTAGATCCGTTAGAGCTAAGTTGTTGTUCCAAGGTGGAAGAGCTGCTACTTGT
 +-----+-----+-----+-----+-----+-----+-----+-----+
 GTGTCTCGATCTAGGCAATCTCGATTCAACAACAGGGTCCACCTTCTCGACGATGAACA
H R A R S V R A K L L S Q G G R A A T C

3601 CGTAAGTACTTGTCAACIGGGCTGTTAGAACAAGTTCAAGTTGACTCCTATTCTGCT
 +-----+-----+-----+-----+-----+-----+-----+-----+
 CCATTTCATGAACAAGTTGACCCGACAATCTTGTTCAACTTCAACTGAGGATAAGGACGA
G K Y L F N W A V R T K L K L T P I F A

3661 GCTTCCCAATTGGATTTGTCCGGTTGGTTTGTGCTGGTTACAACGGTGGTGACATCTAC
 +-----+-----+-----+-----+-----+-----+-----+-----+
 CGAAGGGTTAACCTAACAGGCCAACCAACAACGACCAATGTTGCCACCACTGTAGATG
A S Q L D L S G W F V A G Y N G G D I Y

CACTCTTTGTCCAGAGCTAGACCAAGATAATAGACTAGTGGATCC
3721-----+-----+-----+-----+-----+
GTGAGAAACAGGTCTCGATCTGGTTCATTACTGATCACCTAGG
H S L S R A R F R * * *

Fig. 7
 Proteína de fusión 4A-3-5B optimizada para Escherichia coli

```

AGGGCGAATTGGGTACCATGGGGAGCGTTGTATFGTTGGCCGCAATTATTCTGAGCGGTA
1  -----+-----+-----+-----+-----+-----+
TCCCGCTTAACCCATGGTACCCCTCGCAACAATAACAACCGGCGTAATAAGACTCGCCAT
      M G S V V I V G R I I L S G S

GCGGTAGCGCCCGATTACCGCCTATTCTCAGCAAACCCGGTGGTCTGCTGGGTTGTATTA
61  -----+-----+-----+-----+-----+-----+
CGCCATCGCGGGGCTAATGGCGGATAAGAGTCGTTTGGGCACCAGACGACCCAAACATAAT
      G S A P I T A Y S Q Q T R G L L G C I I

TTACCAGCCTGACCGGCCGTGATAAAAAATCAGGTGGATGGCGAAGTTCAGGTTCTGAGCA
121 -----+-----+-----+-----+-----+-----+
AATGGTCCGACTGGCCGGCACTATTTTTAGTCCACCTACCGCTTCAAGTCCAAGACTCGT
      T S L T G R D K N Q V D G E V Q V L S T

CCGCCACCCAGAGCTTTCGGCGACCTGTGTGAATGGTGTGTGCTGGACCGTTTATGCCG
181 -----+-----+-----+-----+-----+-----+
GGCGGTGGGTCTCGAAAGACCGCTGGACACACTTACCACACACGACCTGGCAAATACGGC
      A T Q S F L A T C V N G V C W T V Y A G

GTGCCGGTAGCAAAACCCGTGGCGGGTCCGAAAGGTCCGATTACCCAGATGTACACCAACG
241 -----+-----+-----+-----+-----+-----+
CACGGCCATCGTTTTGGGACCGCCAGGCTTCCAGGCTAATGGGTCTACATGTGGTTGC
      A G S K T L A G P K G P I T Q M Y T N V

TGGATCAGGATCTGGTTGGTTGGCCGGCCCGCCGGGTGCCCGTAGCATGACCCCGTGTA
301 -----+-----+-----+-----+-----+-----+
ACCTAGTCCTAGACCAACCAACCGGCCCGCGCCGCGCCACGGGCATCGTACTGGGGCACAT
      D Q D L V G W P A P P G A R S M T P C T

CCTGTGGTAGCAGCGATCTGTATCTGGTTACCCGTCATGCCGATGTTATTCCGGTTCGTC
361 -----+-----+-----+-----+-----+-----+
GGACACCATCGTCGCTAGACATAGACCAATGGGCAGTACGGCTACAATAAGGCCAAGCAG
      C G S S D L Y L V T R H A D V I P V R R

TCGTGGTGATAGCCGTGGTAGCCTGCTGTCTCCGCGTCCGGTTAGCTATCTGAAAGGTA
421 -----+-----+-----+-----+-----+-----+
CAGCACCATATCGGCACCATCGGACGACAGAGGGCGCAGGCCAATCGATAGACTTTCAT
      R G D S R G S L L S P R F V S Y L K G S

GCAGCGGTGGTCCGCTGCTGTGTCCGAGCGGTCATGTGGTGGGTATTTTTCGTGCCGCCG
481 -----+-----+-----+-----+-----+-----+
CGTCGCCACCAGGCGACGACACAGGCTCGCCAGTACACCACCCATAAAAAGCACGGCGGC
      S G G P L L C P S G H V V G I F R A A V

TTTGTACCCGTGGTGTGGCGAAAGCGGTGGATTTTATCCCGGTTGAAAGCATGGAACCA
541 -----+-----+-----+-----+-----+-----+
AAACATGGGCACCAACCGCTTTCGCCACCTAAAATAGGGCCAACCTTTCGTACCTTTGGT
      C T R G V A K A V D F I P V E S M E T T

CCATGCGTAGCCCGGTGTTTACCGATAATAGCAGCDCGCGGCGGTTCCGCAGACCTTTC
601 -----+-----+-----+-----+-----+-----+
EGTACGCATCGGGCCACAAATGGCTATTATCGTCGGGCGGCGCCAAAGGCGTCTGGAAG
      M R S P V F T D N S S P P A V P Q T F Q

AGGTTGCCCATCTGCATGCGCCGACCGGTAGCGGTAAGCAACCAAGTTCGCGCGGCGT
661 -----+-----+-----+-----+-----+-----+
TCCAACGGGTAGACGTACGCGGCTGGCCATCGCCATTTTCGTGGTTTCAAGGCCGCGCA
    
```

V A H L H A P T G S G K S T K V P A A Y
 ATGCCGCCAGGGTTATAAAGTGCTGGTGCTGAATCCGAGCGTGGCGGGCACCCTGGSTT
 721 -----+-----+-----+-----+-----+-----+-----+-----+-----+
TACGGCGGGTCCCAATATTTACGACCACGACTTAGGCTCGCACCGCCGCTGGGACCCAA
A A Q G Y K V L V L N P S V A A T L G F
 TTGGTGCTATATGAGCAAAGCCCATGGCATTGAACCGAACATTTCGTACCGGCGTTCGTA
 781 -----+-----+-----+-----+-----+-----+-----+-----+-----+
AACCACGGATATACTCGTTTCGGGTACCGTAAC TTGGCTTGTAAAGCATGSCCGCAAGCAT
G A Y M S K A H G I E P N I R T G V R T
 CCATTACCACCGGTGGCCCGATTACCTATAGCGCCTACGGCAAATTTCTGGCGGATGGTG
 841 -----+-----+-----+-----+-----+-----+-----+-----+-----+
GGTAATGGTGGCCACCGGGCTAATGGATATCGCGGATGCCGTTTAAAGACCGCCTACCAC
I T T G G P I T Y S A Y G K F L A D G G
 GCTGTAGCGGTGGCGCCTATGATATCATCATCTGTGATGAATGCCATAGCACCGATTGGA
 901 -----+-----+-----+-----+-----+-----+-----+-----+-----+
CGACATCGCCACCGGGATACTATAGTAGTAGACACTACTTACGGTATCGTGGCTAACCT
C S G G A Y D I I I C D E C H S T D W T
 CCACCAITCTGGGTATTGGCACCGTTCTGGATCAGGCGGAAACCGCCGGTGGCCGTTCTGG
 961 -----+-----+-----+-----+-----+-----+-----+-----+-----+
GGTGGTAAGACCCATAACCGTGGCAAGACCTAGTCCGCCTTTGGCGGCCACGGGCAGACC
T I L G I G T V L D Q A E T A G A R L V
 TTGTTCTGCGGACCGCAAACGCCCGCGGGTAGCATTACCGTTCGGCATCCGAACATTGAAG
 1021 -----+-----+-----+-----+-----+-----+-----+-----+-----+
AACAAGACCGCTGGCGTTGCGGGCGGCCATCGTAATGGCAAGGCGTAGGCTTGTAACTTC
V L A T A T P P G S I T V P H P N I E E
 AAGTGGCCCTGAGCAATACCGGCGAAATTCGGTTTTATGGCAAAGCGATTCCGATCGAAG
 1081 -----+-----+-----+-----+-----+-----+-----+-----+-----+
TTACCGGGACTCGTTATGGCCGCTTTAAGGCAAAATACCGTTTCGCTAAGGCTAGCTTC
V A L S N T G E I P F Y G K A I P I E A
 CGATTAAACCGCGCGTCACTGATTTTTTGGCACAGCAAAAAAAAAATGTEATGAACTGG
 1141 -----+-----+-----+-----+-----+-----+-----+-----+-----+
GCPAATTTCCGCCCGGAGTAGACTAAAAAACGGTGTGCTTTTTTTTTTACACTACTTGACC
I K G G R H L I F C H S K K K C D E L A
 CGSCGAAACTGACCGGTCTGGGTCTGAATGCCGTGGCGTATTATCGTGGTCTGGATGTGA
 1201 -----+-----+-----+-----+-----+-----+-----+-----+-----+
GCCGCTTTGACTGGCCAGACCCAGACTTACGGCACCGCATATAGCACCCAGACTTACACT
A K L T G L G L N A V A Y Y R G L D V S
 GCGTTATTCGACCAGCGGTGATGTTGTTGTGGTGGCGACCGATGCCCTGATGACCGGTT
 1261 -----+-----+-----+-----+-----+-----+-----+-----+-----+
CGCAATAAGGCTGGTCCCACTACAACAACACCACCGCTGGCTACGGGACTACTGGCCAA
V I P T S G D V V V V A T D A L M T G F
 TTACCGCGGATTTGATAGCGTGATCGATTGTAACACCTGTGTGACCCAGACCGTTGATF
 1321 -----+-----+-----+-----+-----+-----+-----+-----+-----+
AATGGCCGCTAAACTATCCCACTAGCTAACATTGTGGACACACTGGGTCTGGCAACTAA
T G D F D S V I D C N T C V T Q T V D F
 TTAGCCTGGACCCGACCTTTACCATTGAAACCACCACCGTTCGGCAGGATGCCGTTAGCC
 1381 -----+-----+-----+-----+-----+-----+-----+-----+-----+
AATCGACCTCGGCTGGAATGGTAAC TTGGTGGTGGCAAGGCGTCTACGGCAATCGG
S L D P T F T I E T T T V P Q D A V S R
 GTAGCCAGCGTCTGGTGCCACCGGTCTGGTCTAGCGGCATTTATCGTTTTGTGACCG

1441-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CATCGGTCCGAGCACCACGGTGGCCAGCACCAGCATCGCCGTAAATAGCAAAACACTGGC
S Q R R G A T G R G R S G I Y R F V T P
 CGGGTGAACGFCGAGCGGTATGTTTGATAGCAGCGTGTGTGAATGTTATGATGCCG
 1501-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GCCCACTTGCAGGCTCGCCATACAACTATCGTCCGACGACACACTTACAATACTACGGC
G E R P S G M F D S S V L C E C Y D A G
 GCTGTGCCTGGTATGAACTGACCCGGCGGAAACCACCGTTCGTCTGCCCGCGTATCTGA
 1561-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CGACACGGACCATACTTGGACTGGGGCCGCCTTTGGTGGCAAGCAGACCGCCGATAGACT
C A W Y E L T P A E T T V R L R A Y L N
 ATACGCCCGGGTCTGCCGGTTTGTCCAGGATCATCTGGAATTTCTGGGAAGCGTTTTTACCG
 1621-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 TATCGGGCCAGACGGCCAAACAGTCCCTAGTAGACCTTAAGACCCCTTTCCGCAAAAATGGC
T F G L P V C Q D H L E F W E S V F T G
 GCCTGACCCATATTGATGCCCATTTTCTGAGCCAGACCAACAGCGGGGTGATAACTTTC
 1681-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CGGACTGGGTATAACTACGGGTAAAAGACTCGGTCGGTTTGTCCGCCACTATTGAAAG
L T H I D A H F L S Q T K Q A G D N F P
 CGTATCTGGTGGCGTATCAGGCGACCGTTTGTGCCCGTGCCAGGGCCGGCCGGCGAGCT
 1741-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GCATAGACCACCGCATAGTCCGCTGGCAAACACGGGCACGGGTCCGCGGGCCGGCTCGA
Y L V A Y Q A T V C A R A Q A P P P S W
 GGGATCAGATGTGGAATGCCTGATTCGTCTGAAACCGACCCCTGCATGGTCCGACCCCGC
 1801-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CCCTAGTCTACACCTTTACGGACTAAGCAGACTTTGGCTGGGACGTACCAGGCTGGGGCG
D Q M W K C L I R L K P T L H G P T P L
 TGCTGTATCGTCTGGGTGCCGTTTCAGAACGAAATTACCCTGACCCATCCGATCACCAAAT
 1861-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 ACGACATAGCAGACCCACGGCAAGTCTTGCCTTAATGGGACTGGGTACGCTAGTGGTTTA
L Y R L C A V Q N E I T L T H P I T K F
 TTGTGATGGCGTGTATGAGCGCCGATCTGGAAGTTGTTGGTAGCGGTAGCGGCTCTATGA
 1921-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 AACACTACCGCACATACTCGCGGCTAGACCTTCAACAACCATCGCCATCGCCGAGATACT
V H A C M S A D L E V V G S G S G S B M S
 GCTATAACCTGGACCGGTGCCCTGATTACCCCGTGTGCCCGCCGAAGAAAGCAAACTGCCGA
 1981-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CGATATGGACCTGGCCACGGGACTAATGGGSCACACGGCGGCTTCTTTCCGTTTGACGGCT
Y T W T G A L I T P C A A E E S K L P I
 TTAACCCGCTGTCTAATAGCCTGCTGGCTCATCATAGCATGGTGTATAGCACCACCAGCC
 2041-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 AATTGGGCGACGATTATCGGACGACGCGAGTAGTATCGTACCACATATCGTGGTGGTCCG
N P L S H S L L R H H S M V Y S T T S R
 GTAGCGCCAGCCTGCGTCAGAAAAAGTGACCTTCGATCGTCTGCAGGTGCTGGATGATC
 2101-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CATCGCGGTCCGACGCGAGTCTTTTTCAC TGGAACTAGCAGACGTCACACGCTACTAG
S A S L R Q K K V T F D R L Q V L D D H
 ATTATCGTGATGTGCTGAAAGAAATGAAAGCGAAAGCGAGCACCGTTAAAGCCCGTCTGC
 2161-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 TAATAGCACTACACGACTTCTTTACTTTTCGCTTTTCGCTCGTGGCAATTTCCGGGCAGACG

2941-----+-----+-----+-----+-----+-----+-----+
AAACACTTTTCGGGCGCGTGGGTCTCTACGGCGGTCCGGACCGGCAAAAATGGCTTCGCT
C E S A G T Q E D A A S L R V F T E A M
TGACCCETTATAGCGCCCGCGGGTGATCCGCCGAGCCGGAATATGATCTGGAAGTGA
3001-----+-----+-----+-----+-----+-----+-----+
ACTGGGCAATATCGGGGGCGGCCACTAGCGGGGTCCGGCTTATACTAGACCTTGACT
T R Y S A P P G D P P Q P E Y D L E L I
TCACCAGCTGTAGCAGCAATGTGGTGTGGCCATGATGCCAGCGGTAAACGTTGTATT
3061-----+-----+-----+-----+-----+-----+-----+
AGTGGTGCACATCGTTCGTTACAACCACAACGGGTACTACGGTCCCATTTGCACACATAA
T S C S S N V G V A H D A S G K R V Y Y
ACCTGACCCGTTGATCCGACCACCCCGCTGGECCBTGCCCGCTGGGAAAACCGTTTCGTATA
3121-----+-----+-----+-----+-----+-----+-----+
TGGACTGGGCACTAGGCTGGTGGGGCGACCGGGCACGGCGGACCCCTTTGGCAAGCAGTAT
L T R D P T T P L A R A A W E T V R H T
CCCCGGTTAATAGCTGGCTGGGCAACATTATTATGTATGCCCCGACCCCTGTGGGCCCGTA
3181-----+-----+-----+-----+-----+-----+-----+
GGGGCCAATTATCGACCGACCCGTTGTAATAATACATACGGGGCTGGGACACCCGGGCAT
P V N S W L G N I I M Y A P T L W A R M
TGATTCTGATGACCCACTTCTTTAGCATTCTGTGGCCAGGAACAGCTGGAAAAGCGC
3241-----+-----+-----+-----+-----+-----+-----+
ACTAAGACTACTGGGTGAAGAAATCGTAAGACGACCGGGTCTTGTGACCTTTTTCGCG
I L M T H F F S I L L A Q E Q L E K A L
TGGATTGCCAGATTTATGGCGCCTGCTATAGCATTGAACCGCTGGATCTGCCGCAGATTA
3301-----+-----+-----+-----+-----+-----+-----+
ACCTAACGGTCTAATACCGCGGACGATATCGTAACTTGGCGACCTAGACCGCGTCTAAT
D C Q I Y G A C Y S I E P L D L P Q I I
TTGAACGTCTGCATGGCCTGAGCGCCTTTAGCCTGCATAGCTACTCTCCGGGTGAAAITA
3361-----+-----+-----+-----+-----+-----+-----+
AACTTGCAGACGTACCGGACTCGCGGAAATCGGACGTATCGATGAGAGGCCCACTTTAAT
E R L H G L S A F S L H S Y S P G E I N
ATCGTGTGGCGAGCTGTCTGGTAAACTGGGTGTTCCGCCGCTGCGTGTCTGGCGTCTC
3421-----+-----+-----+-----+-----+-----+-----+
TAGCACACCGCTCGACAGACGCATTTGACCCACAGGGCGCGACGCACAGACCGCAGTAG
R V A S C L R K L G V P P L R V W R H R
GTGCCCGTAGCGTTTCGTGCCAAACTGCTGTCTCAGGGTGGCCGTGCCGCCACCTGTGGTA
3481-----+-----+-----+-----+-----+-----+-----+
CACGGGCATCGCAAGCACGGTTTGACCACAGAGTCCCACCGGCACGGCGGTGGACACCAT
A R S V R A K L L S Q G G R A A T C G K
AATACCTGTTTAACTGGGCGGTTTCGTACCAAACCTGAAACTGACCCCGATTCGGCGGGCGA
3541-----+-----+-----+-----+-----+-----+-----+
TTATGGACAAATGACCCGCCAAGCATGTTTACTTTGACTGGGGCTAAGGCCCGCCGCT
Y L F N W A V R T K L K L T P I P A A S
GCCAGCTGGATCTGAGCGGTTGGTTTGTGGUGGTATATAACGGCGGGGATATCTATCATA
3601-----+-----+-----+-----+-----+-----+-----+
CGGTGCACCTAGACTCGCCAACCAAACACGGCCAATATTGCCCGCGCTATAGATAGTAT
Q L D L S G W F V A G Y N G G D I Y H S
GCCTGAGCCGTGCCCGTCCGCGTTAATAAACTAGTGGATCCCGAGCTCCAGCTTTGTTC
3661-----+-----+-----+-----+-----+-----+-----+
CGGACTCGGCACGGGCAGGCGCAATTTATGATCACCTAGGCCTCGAGGTCGAAACAGG
L S R A R F R + +

ES 2 387 321 T3

CTAG
3721----
GATC

Fig. 8
 Proteína de fusión 4A-3-4B-5B optimizada para Escherichia coli

```

1  GGGCGAATTGGGTACCATGGGGAGCGFTGTTATGTTGGCCGCATTATTCTGAGCGGTAG
-----+-----+-----+-----+-----+-----+-----+-----+
CCCCTTAACCCATGGTACCCCTCGCAACAATAACAACCGGCGTAATAAGACTCGCCATC
    M G S V V I V G R I I L S G S

61  CGGTAGCGCCCCGATTACCGCCTATTCTCAGCAACCCGTTGGTCTGCTGGGTGTATTAT
-----+-----+-----+-----+-----+-----+-----+-----+
GCCATCGCGGGCTAATGGCGGATAGAGTCTGTTTGGGCACCAGACGACCCAACATAATA
    G S A P I T A Y S Q Q T R G L L G C I I

121  TACCAGCCTGACCGGCCGTGATAAAAATCAGGTGGATGGCGAAGTTCAGGTTCTGAGCAC
-----+-----+-----+-----+-----+-----+-----+-----+
ATGGTTCGGACTGGCCGGCACTATTTTGTCCACCTACCGCTTCAAGTCCAAGACTCGTG
    T S L T G R D K N Q V D G E V Q V L S T

181  CGCCACCAGAGCTTCTGGCGACCTGTGTGAATGGTGTGTCTGGACCGTTTATGCCGG
-----+-----+-----+-----+-----+-----+-----+-----+
GGGTGGGTCTCGAAAGACCGCTGGACACACTTACCACACAGACCTGGCAAATACGGCC
    A T Q S F L A T C V N G V C W T V Y A G

241  TGCCGGTAGCAAACCCCTGGCGGGTCCGAAAGGTCCGATTAECCAGATGTACACCAACGT
-----+-----+-----+-----+-----+-----+-----+-----+
ACGGCCATCGTTTTGGGACCGCCAGGCTTCCAGGCTAATGGGTCTACATGTGGTTGCA
    A G S K T L A G P K G P I T Q M Y T N V

301  GGATCAGGATCTGGTTGGTTGGCCGGCGCCGGGGTGCCCGTAGCATGACCCCGTGTAC
-----+-----+-----+-----+-----+-----+-----+-----+
CCTAGTCTAGACCAACCAACCGCCCGCGGCCACGGGCATCGTACTGGGGCACATG
    D Q D L V G W P A P P G A R S M T P C T

361  CTGTGGTAGCAGCGATCTGTATCTGGTTACCCGTCATGCCGATGTTATCCGGTTCGTCC
-----+-----+-----+-----+-----+-----+-----+-----+
GACACCATCGTCTAGACATAGACCAATGGGCAGTACGGCTACAATAAGGCCAAGCAGC
    C G S S D L Y L V T R H A D V I P V R R

421  TCGTGGTGATAGCCGTGGTAGCCTGCTGTCTCCGCGTCCGGTATGCTATCTGAAAGGTAG
-----+-----+-----+-----+-----+-----+-----+-----+
AGCACCCTATCGGCACCATCGGACGACAGAGGCGCAGGCCAATCGATAGACTTCCATC
    R G D S R G S L L S P R P V S Y L K G S

481  CAGCGGTGGTCCGCTGCTGTGTCCGAGCSGTCTGTGGTGGGTATTTTTCGTCCCGCGT
-----+-----+-----+-----+-----+-----+-----+-----+
GTCGCCACCAGGCGACGACAGGCTCGCCAGTACACCACCCATAAAAAGCACCGCGGCA
    S G G P L L C P S G H V V G I F R A A V

541  TTGTACCCGTGGTGTGGCGAAAGCGGTGGATTTTATCCCGGTTCAAAGCATGGAAACCAC
-----+-----+-----+-----+-----+-----+-----+-----+
AACATGGGCACCACACCGCTTTCGCCACCTAAAATAGGGCCAACTTTCGTACCTTTGGTG
    C T R G V A K A V D F I P V E S M E T T

601  CATGCGTAGCCCGGTGTTTACCGATAATAGCAGCCCGCGCGGTTCCGCAGACCTTTCA
-----+-----+-----+-----+-----+-----+-----+-----+
GTACGCATCGGGCCACAAATGGCTATTATCGTCCGCGCGCCGCAAGGCGTCTGGAAAGT
    M R S P V F T D N S S P P A V P Q T F Q

661  G5TTGCCCATCTGCATGCGCGACCGGTAGCGGTAAAAGCACCAAAGTTCCGGCGGGCGTA
-----+-----+-----+-----+-----+-----+-----+-----+
    
```

CCAACGGGTAGACGTACGCGCTGCCATCGCCATTTTCGTGGTTTCAAGGCCGCCGCAT
V A H L H A P T G S G K S T K V P A A Y
 TGCCGCCAGGGTTATAAAGTGCTGGTGCTGAATCCGAGCGTGGCGGGACCCTGGGTTT
 721 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 ACGCGGGTCCCAATATTTACGACCACGACTTAGGCTCGCACCGCCGCTGGGACCCAAA
A A Q G Y K V L V L N P S V A A T L G F
 TGGTGCCTATATGAGCAAAGCCCATGGCATTGAACCGAACATTTCGTACCGGGCTTCGTAC
 781 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 ACCACGGATATACTCGTTTCGGGTACCGTAACTTGGCTTGTAAAGCATGGCCGCAAGCATG
G A Y M S K A H G I E P N I R T G V R T
 CATTACCACCGGTGGCCCGATTACCTATAGCGCCTACGGCAAATTTCTGGCGGATGGTGG
 841 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 GTAATGGTGGCCACCGGGCTAATGGATATCGCGGATGCCGTTTAAAGACCGCCTACCACC
I T T G G P I T Y S A Y G K F L A D G G
 CTGTAGCGGTGGCCCTATGATATCATCATCTGTGATGAATGCCATAGCACCGATTGGAC
 901 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 GACATCGCCACCGGATAGTATAGTACTACTTACGGTATCGTGGCTAACCTG
C S G G A Y D I I I C D E C H S T D W T
 CACCATTCTGGGTATTGGCACCGTTCGTGGATCAGGGGAAACCGCCGGTGCCTGGTGGT
 961 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 GTGGTAAGACCCATAACCGTGGCAAGACCTAGTCCGCTTTGGCGGCCACGGGCAGACCA
T I L G I G T V L D Q A E T A C A R L V
 TGTCTGGCGACCGCAACCGCCCGGGTAGCATTACCGTTCGCGCATCCGAACATTGAAGA
 1021 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 ACAAGACCCCTGGGTTGGCGCGGCCCATCGTAATGGCAAGCGTAGGCTTGTAACTTCT
V L A T A T P P G S I T V P H P N I E E
 AGTGGCCCTGAGCAATACCGCGGAAATCCGTTTTATGGCAAAGCGATTCCGATCGAAGC
 1081 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 TCACCGGGACTCGTTATGGCCGCTTTAAGGCAAAATACCGTTTCGCTAAGGCTAGCTTCG
V A L S N T G E I P F Y G K A I P I E A
 GATTAAGGGCGCCGTCATCTGATTTTTTGGCACAGCAAAAAAATGTGATGAAGTGGC
 1141 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 CTAATTTCCGCGCGCAGTAGACTAAAAACGGTGTGCTTTTTTTTTTACACTACTTGACCG
I K G G R H L I F C H S K K K C D E L A
 GCGAAACTGACCGGTCTGGGTCTGAATGCCGTGGCGTATTATCGTGGTCTGGATGTGAG
 1201 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 CCGCTTTGACTGGCCAGACCCAGACTTACGGCACCGCATAATAGCACCCAGACCTACACTC
A K L T G L G L N A V A Y Y R G L D V S
 CGTTATTCGACCAGCGGTGATGTTGTTGTGGTGGCGACCGATGCCCTGATGACCGGTT
 1261 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 GCAATAAGGCTGGTCCGCACTACAACAACACCACCGCTGGCTACGGGACTACTGGCCAAA
V I P T S G D V V V V A T D A L M T C F
 TACCGGGATTTTGATAGCGTGATCGATTGTAACACCTGTGTGACCCAGACCGTTGATT
 1321 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 ATGGCCGCTAAAACCTATCGCACTAGCTAACATTGTGGACACACTGGGTCTGGCAACTAAA
T G D F D S V I D C N T C V T Q T V D F
 TAGCCTGGACCCGACCTTTACCATTGAAACCACCACCGTTCGCGAGGATGCCGTTAGCCG
 1381 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 ATCGGACCTGGGCTGGAAATGGTAACTTTGGTGGTGGCAAGGCGTCTACGGCAATCGGC
S L D P T P T I E T T T V P Q D A V S R

TAGCCAGCGTCGTGGTGCCACCGGTCGTGGTCGTAGCGGCATTTATCGTTTTGTGACGCC
 1441-----+-----+-----+-----+-----+-----+-----+-----+
 ATCGGTCCGAGCACCACGGTGGCCAGCACCAGCATCGCCGTAATAGCAAAACACTGCGG
S Q R R G A T G R G R S G I Y R F V T P
 GGGTGAACGTCCGAGCGGTATGTTTGATAGCAGCGTCTGTGTGAATGTTATGATGCCGG
 1501-----+-----+-----+-----+-----+-----+-----+-----+
 CCCACTTGCAGGCTCGCCATACAAACTATCGTCCGACGACACACTTACAATACTACGGCC
G E R P S G M F D S S V L C E C Y D A G
 CTGTGCCCTGGTATGAACTGACCCCGGGGAAACCACCGTTCGTCTGCCGCGGTATCTGAA
 1561-----+-----+-----+-----+-----+-----+-----+-----+
 GACACGGACCATACTTACTGAGGGCCGCTTTGGTGGCAAGCAGACGCGCGCATAGACTT
C A W Y E L T P A E T T V R L R A Y L N
 TACGCCGGGTCTGCCGTTTGTGTCAGGATCATCTGGAATTCGGGAAAGCGTTTTTACCGG
 1621-----+-----+-----+-----+-----+-----+-----+-----+
 ATGCGGCCAGACGGCCAAACAGTCCCTAGTAGACCTAAGACCTTTTCGCAAAAATGCC
T F G L P V C Q D H L E F W E S V F T G
 CCTGACCCATATTGATGCCCATTTTCGAGCCAGACAAACAGGGCGGGTGATAACTTTCC
 1681-----+-----+-----+-----+-----+-----+-----+-----+
 GGACTGGGTATAACTACGGGTAAAAGACTCGGTCTGGTTTTGTCCGCCACTATTGAAAGG
L T H I D A H F L S Q T K Q A G D N F P
 GTATCTGGTGGCGTATCAGGCGACCGTTTGTGCCCGTGCCAGGCGCCGCGCGGAGCTG
 1741-----+-----+-----+-----+-----+-----+-----+-----+
 CATAGACCACCGCATAGTCCGCTGGCAACACGGGCACGGGTCCGCGGGCGGGCGGCTCGAC
Y L V A Y Q A T V C A R A Q A P P P S W
 GGATCAGATGTGGAAATGCCTGATTCGTCTGAAACCGACCTGCATGGTCCGACECCCGCT
 1801-----+-----+-----+-----+-----+-----+-----+-----+
 CCTAGTCTACACCTTTACGGACTAAGCAGACTTGGCTGGGACGTACCAGGCTGGGGCGA
D Q M W K C L I R L K P T L H G P T P L
 GCTGTATCGTCTGGGTGCCGTTTCAGAACGAAATTACCCGACCCATCCGATCACCAATT
 1861-----+-----+-----+-----+-----+-----+-----+-----+
 CGACATAGCAGACCCACGGCAAGTCTTGCTTTAATGGGACTGGGTAGGCTAGTGSTTTAA
L Y R L G A V Q N E I T L T H P I T K F
 TGTGATGGCGTGTATGAGCGCCGATCTGGAGTTGTTAGCCTGATGGCGTTTACCGCCAG
 1921-----+-----+-----+-----+-----+-----+-----+-----+
 ACACTACCGCACATACTCGCGGCTAGACCTTCAACAATCGGACTACCGCAAATGGCGGTC
V M A C M S A D L E V V S L M A F T A S
 CATTACEAGCCCGCTGACCACCCAGAATACCCCTGCTGTTTAAACATCCTGGGCGGTTGGGT
 1981-----+-----+-----+-----+-----+-----+-----+-----+
 GTAATGGTCCGGCGACTGGTGGTCTTATGGGACGACAAATTGTAGGACCCGCCAACCCA
I T S P L T T Q N T L L F N I L G G W V
 GGCGGCCAGCTGTCTATGAGCTATACCTGGACCGGTGCCCTGATTACCCCGTGTGCCGC
 2041-----+-----+-----+-----+-----+-----+-----+-----+
 CCGCCGGGTCCGACAGATACTCGATATGGACCTGGCCACGGGACTAATGGGGCACACGGCG
A A Q L S M S Y T W T G A L I T P C A A
 CGAAGAAAGCAAACCTGCCGATTAACCCGCTGTCTAATAGCCTGCTGCGTCATCATAGCAT
 2101-----+-----+-----+-----+-----+-----+-----+-----+
 GCTTCTTTGTTTTGACGGCTAATTGGGGCAGAGATTATCGGACGACGCAGTAGTATCGTA
E E S K L P I N P L S N S L L R H H S M
 GGTGTATAGCACCACCAAGCCGTAGCGCCAGCCTGCGTCAGAAAAAGTGACCTTCGATCG
 2161-----+-----+-----+-----+-----+-----+-----+-----+
 CCACATATCGTGGTGGTCCGATCGCGTCCGACGAGTCTTTTTTCACTGGAAGCTAGC

V Y S T T S R S A S L R Q K K V T F D R
 TCTGCAGGTGCTGGATGATCATTATCGTGATGTGCTGAAAGAAATGAAAGCGAAAGCGAG
 2221-----+-----+-----+-----+-----+-----+-----+-----+
 AGACGTCCACGACCTACTAGTAATAGCACTACAGGACTTCTTTACTTTTCGCTTTTCGCTC
L Q V L D D H Y R D V L K E M K A K A S
 CACCGTTAAAGCCCGTCTGCTGTCTATTGAAGAAGCGTGTAAGTACCCCGCCGCATAG
 2281-----+-----+-----+-----+-----+-----+-----+-----+
 GTGGCAATTTCCGGCAGACGACAGATAACTTCTTCGCACATTTGACTGGGGCGGGCGTATC
T V K A R L L S I E E A C K L T P P H S
 CGCCAAAAGCAAATTTGGCTATGGCGCCAAAGATGTTTCGTAGCCTGAGCAGCCGTGCCGT
 2341-----+-----+-----+-----+-----+-----+-----+-----+
 GGGGTTTTCGTTTTAAACCGATACCGCGTTTCTACAAGCATCGGACTCGTCCGGCAGGCA
A K S K F G Y G A K D V R S L S S R A V
 TAATCATATTCGTAGCGTGTGGGAAGATCTGCTGGAAGATACCGAAACCCCGATTGATAC
 2401-----+-----+-----+-----+-----+-----+-----+-----+
 ATTAGTATAAGCATCGCACACCCTTCTAGACGACCTTCTATGGCTTTGGGGCTAACTATG
N N I R S V W E D L L E D T E T P I D T
 CACCATCATGGCGAAAACGAAGTGTTTTGTGTTTCAGCCGAAAAGGTGGTTCGTAAACC
 2461-----+-----+-----+-----+-----+-----+-----+-----+
 GTGGTAGTACCGCTTTTGTCTCACAAAACACAAGTCGGCCTTTTCCACCAGCATTGG
T I M A K N E V F C V Q P E K G R K P
 GGCCCGTCTGATTGTTTTTCCGGATCTGGGTGTTTCGTGTGTGTGAAAAAATGGCGCTGTA
 2521-----+-----+-----+-----+-----+-----+-----+-----+
 CCGGGCAGACTAACAAAAGGCCTAGACCCACAAGCACACACTTTTTTACCGCGACAT
A R L I V F P D L G V R V C E K M A L Y
 CGATGTTGTTAGCACCCCTGCCGCRGGCGGTTATGGGTCCGAGCTATGGCTTTTCAGTATTC
 2581-----+-----+-----+-----+-----+-----+-----+-----+
 GCTACAACAATCGTGGGACGGCGTCCGCCATACCCAGGCTCGATACCGAAAGTCATAAG
D V V S T L P Q A V M G P S Y G F Q Y S
 TCCGGTCCAGCGTGTGAATTTCTGGTGAACACCTGGAAAAGCAAAAATGCCCGATGGG
 2641-----+-----+-----+-----+-----+-----+-----+-----+
 AGGCCAGTCCGACAACTTAAAGACCACTTGTGGACCTTTTCGTTTTTACGGGCTACCC
P G Q R V E F L V N T W K S K K C P M G
 CTTACGCTATAACACCCGCTGCTTTGATAGCACCGTGACCGAAAACGATATCCGTACCGA
 2701-----+-----+-----+-----+-----+-----+-----+-----+
 GAAGTCCGATATGTGGGCGACGAAACTATCGTGGCACTGGCTTTTGCTATAGGCATGGCT
F S Y N T R C F D S T V T E N D I R T E
 AGAAAGCATTACAGTGCTGTGATCTGGCGCCGGAAGCCCGTCAGGCGATTAAAAGCCT
 2761-----+-----+-----+-----+-----+-----+-----+-----+
 TCTTTCGTAATGGTCACGACACTAGACCGCGGCTTCGGGCACTCCGCTAATTTTCGGA
E S I Y Q C C D L A F E A R Q A I K S L
 GACCGAACGCCTGTATATTGGCGGTCGCTGACCAATAGCAAAGGCCAGAAGTGTGGTTA
 2821-----+-----+-----+-----+-----+-----+-----+-----+
 CTGGCTTGGGACATATAACCGCCAGGCGACTGGTTATCGTTTCCGGTCTTGACACCAAT
T E R L Y I G G F L T N S K C Q N C G Y
 TCGTCTGTGTGCTGCCAGCGGTGTTCTGACCACCAGCTGTGGTAATACCCTGACCTGCTA
 2891-----+-----+-----+-----+-----+-----+-----+-----+
 AGCAGCAACAGCAGGTCGCCACAAAGACTGGTGGTCCGACACCATTATGGGACTGGACGAT
R R C R A S G V L T T S C G N T L T C Y
 CCTGAAAGCGACCGCCGCTGTCGTGCCGCCAAACTGCAGGATTGTACCATGCTGGTTAA

2941-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GGACTTTCGCTGGCGGGACAGCACGGCGGTTTGACGTCCCTAACATGGTACGACCAATT
L K A T A A C R A A K L Q D C T M L V N
 CGGCAATGATCTGGTGGTGATTGTGAAAGCGCCGGCACCCAGGAAGATGCCGCCAGCCT
 3001-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GCCGTTACTAGACCACCCTAAACAUTTCGCGGGCGTGGGTCTTCTACGGCGGTCCGA
G N D L V V I C E S A G T Q E D A A S L
 GCGCGTTTTTACCGAAGCGATGACCCGTTATAGCGCCCGCGCGGTGATCCGCGCCAGCC
 3061-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CGCGCAAAAATGGCTTCGCTACTGGGCAATATCCCGGGCGGCCACTAGCGCGCGTCCG
R V F T E A M T R Y S A F P G D P P Q P
 CGAATATGATCTGGAACGATCACCAGCTGTAGCAGCAATGTTGGTGTGCCCATGATGC
 3121-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CCTTATACTAGACCTTGACTAGTGGTCGACATCGTGGTTACAACCACAACGGGTACTACG
E Y D L E L I T S C S S N V G V A H D A
 CAGCGGTAACGTTGTATTACCTGACCCGTTGATCCGACCACCCCGCTGGCCCGTCCCGC
 3181-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GTCGCCATTTGCACACATAATGGACTGGGCACTAGGCTGGTGGGGCGACCGGGCACGGG
S G K R V Y Y L T R D P T T P L A R A A
 CTGGAAACCGTTCTGCATACCCCGTTAATAGCTGGCTGGGCAACATTATTATGTATGC
 3241-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GACCCCTTGGCAAGCAGTATGGGGCCCAATTATCGACCGACCCGTTGTAATAATACATACG
W E T V R H T P V N S W L G N I I M Y A
 CCGGACCCGTGGGGCCCGTATGATTCTGATGACCCACTTCTTTAGCATCTGCTGGCCCA
 3301-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GGGCTGGGACACCCGGGCATACTAAGACTACTGGGTGAAGAAATCGTAAGACGACCGGGT
P T L W A R M I L M T H F F S I L L A Q
 GGAACAGCTGGAAAAGCGCTGGATTGCCAGATTTATGGCGCCTGCTATAGCATTGAACC
 3361-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CCTTGTGACCTTTTTCGCGACCTAACGGTCTAAATACCGGGACGATATCGTAACCTGG
E Q L E K A L D C Q I Y G A C Y S I E P
 GCTGGATCTGCCGAGATTATTGAACGCTGCGATGGCCTGAGCGCCTTTAGCCTGCATAG
 3421-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CGACCTAGACGGCGTCTAATAACTTGCAGACGTACCGGACTCGCGGAAATCGGACGATC
L D L P Q I I E R L H C L S A F S L H S
 CTACTCTCCGGGTGAAATTAATCGTGTGGCGAGCTGCTGCGTAAACTGGGTGTTCGGCC
 3481-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GATGAGAGGGCCACTTTAATTAGCACACCGCTCGACAGACGATTTGACCCACAAGGCGG
Y S P G E I N R V A S C L R K L G V P P
 GCTGCGTGTCTGGCGTATCGTGGCCGTTAGCGTTTCGTGCCAAACTGCTGTETCAGGGTGG
 3541-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CGACGACAGACCGCAGTAGCACGGGCATCGCAAGCACGGTTTGACGACAGAGTCCACCC
L R V W R H R A R S V R A K L L S Q G G
 CCGTCCCGCCACCTGTGGTAAATACCTGTTTAACTGGGCGGTTTCGTACCAAACCTGAAACT
 3601-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 GGCACGGCGGTGGACACCATTTATGGACAAATGACCCGCAAGCATGGTTTGACTTTGA
R A A T C G K Y L F N W A V R T K L K L
 GACCCCGATTCEGGCGGGCAGCCAGCTGGATCTGAGCGGTTGGTTTGTTCGGGTTATAA
 3561-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
 CTGGGCTAAGGCGCGGCTCGGTGACCTAGACTCGCCAACCAACAACGGCCATATT
T P I P A A S Q L D L S G W F V A G Y N

```
CGGCGGCGATATCTATCATAGCCTGAGCCGTGCCCGTCCGCGTTAATAAACTAGTGGATC
3721-----+-----+-----+-----+-----+-----+
GCCGCCGCTATAGATAGTATCGGACTCGGCACGGCCAGGCGCAATTATTTGATCACCTAG
G G D I Y H S L S R A R P R * * .

CGGAGCTCCAGCTTTTGTCC
3781-----+-----+-----
GCCTCGAGGTCGAAAACAAGGG
```

Fig. 9

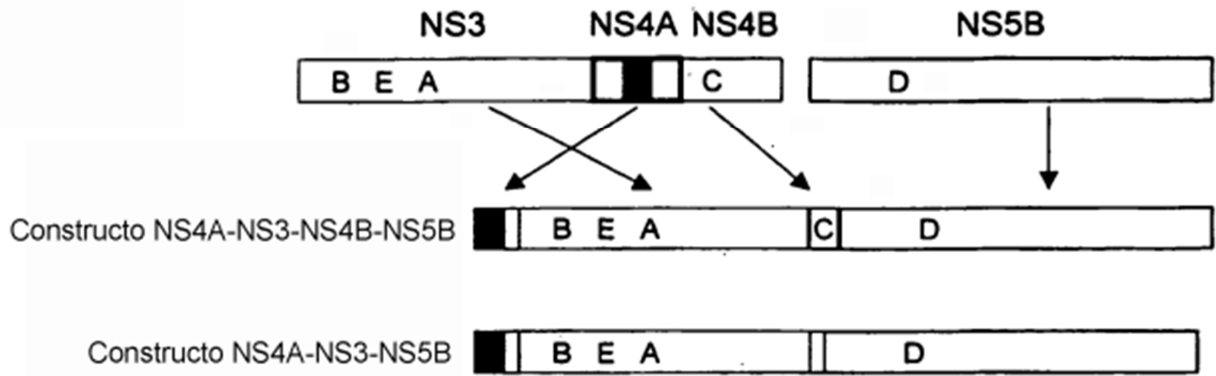


Fig. 10

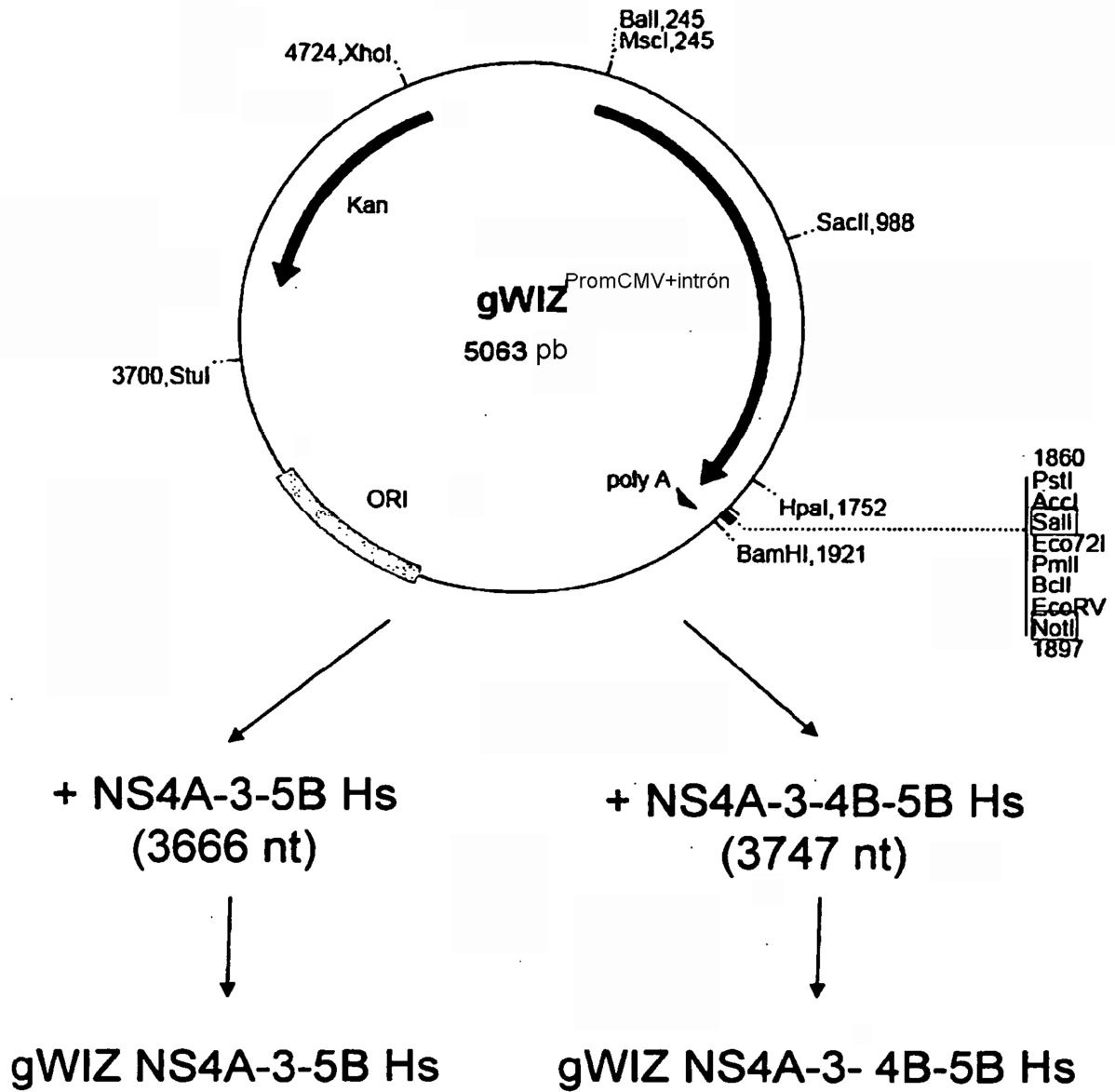
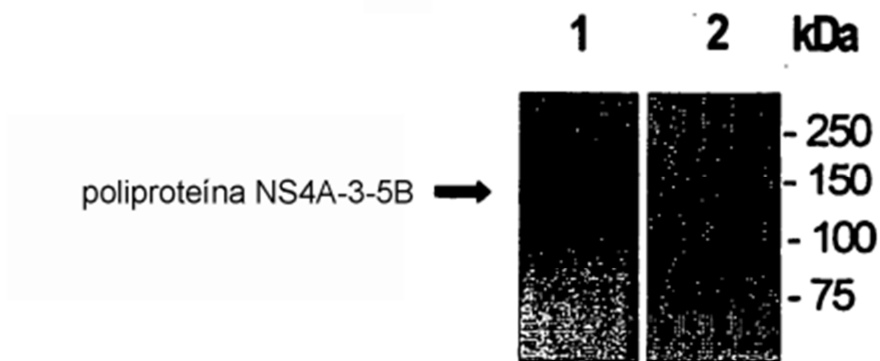
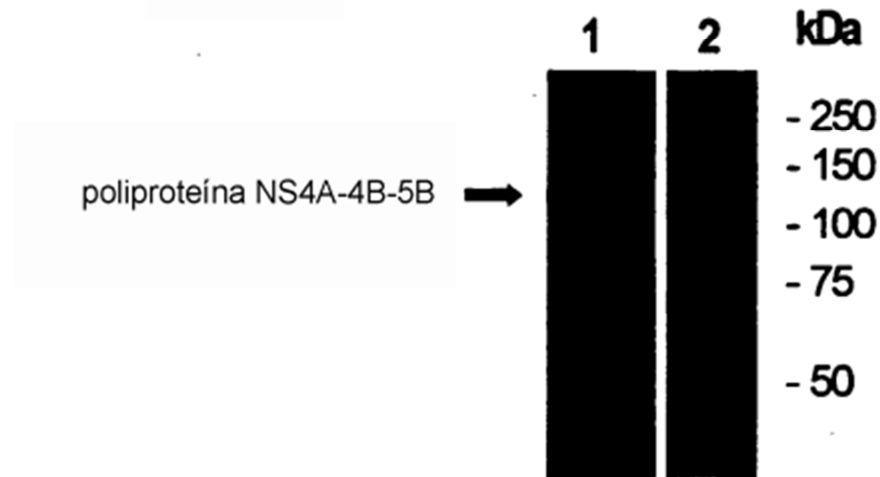


Fig. 11

(A)



(B)



(C)

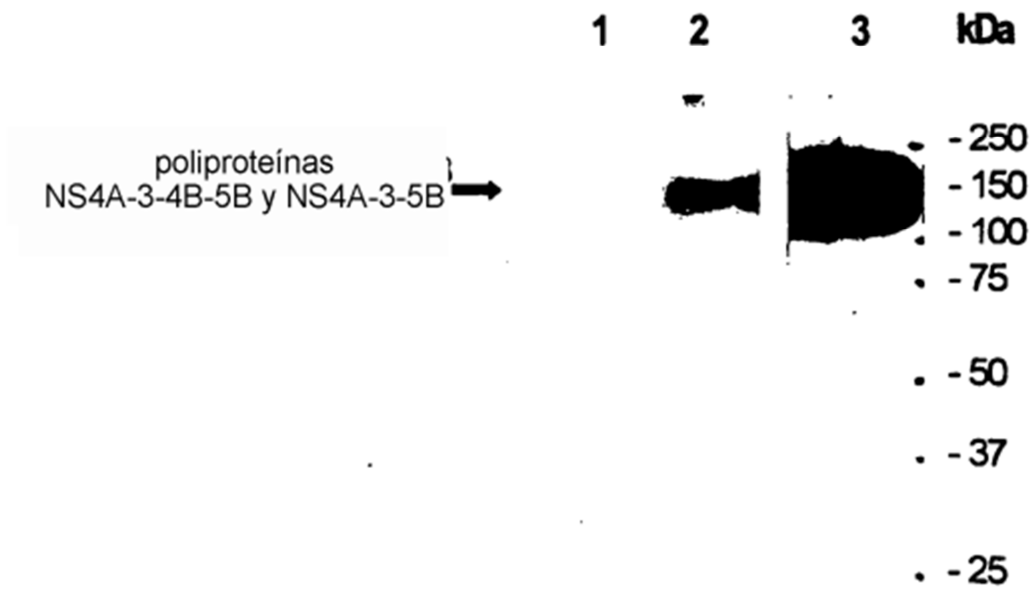


Fig. 12

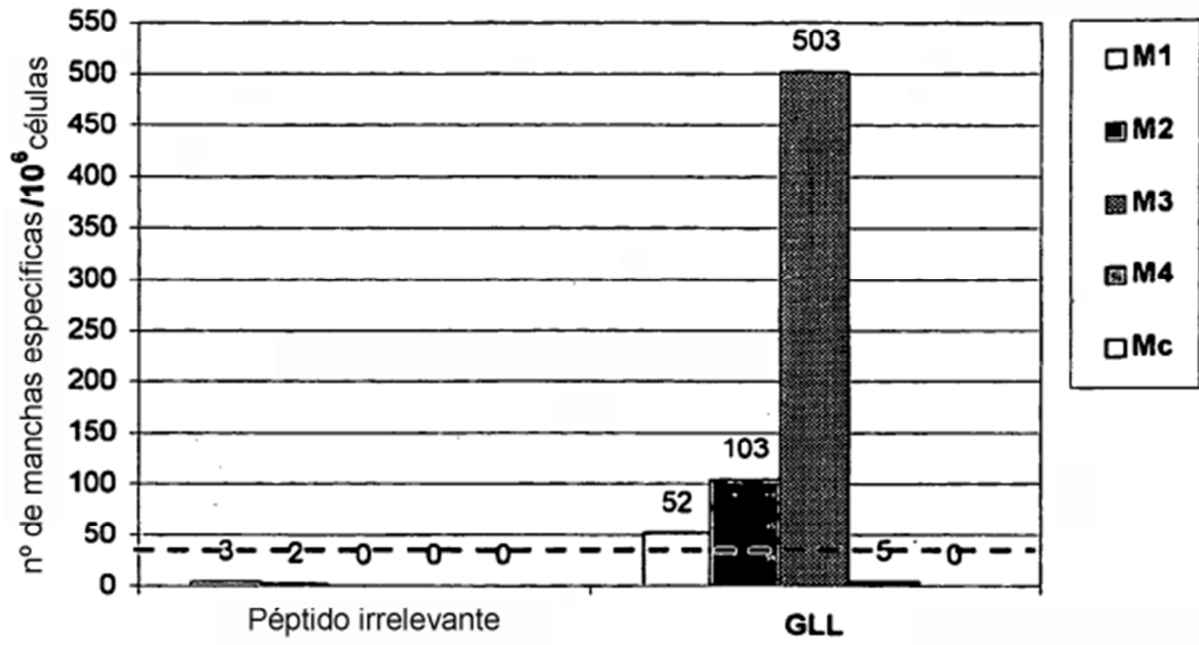


Figura 13

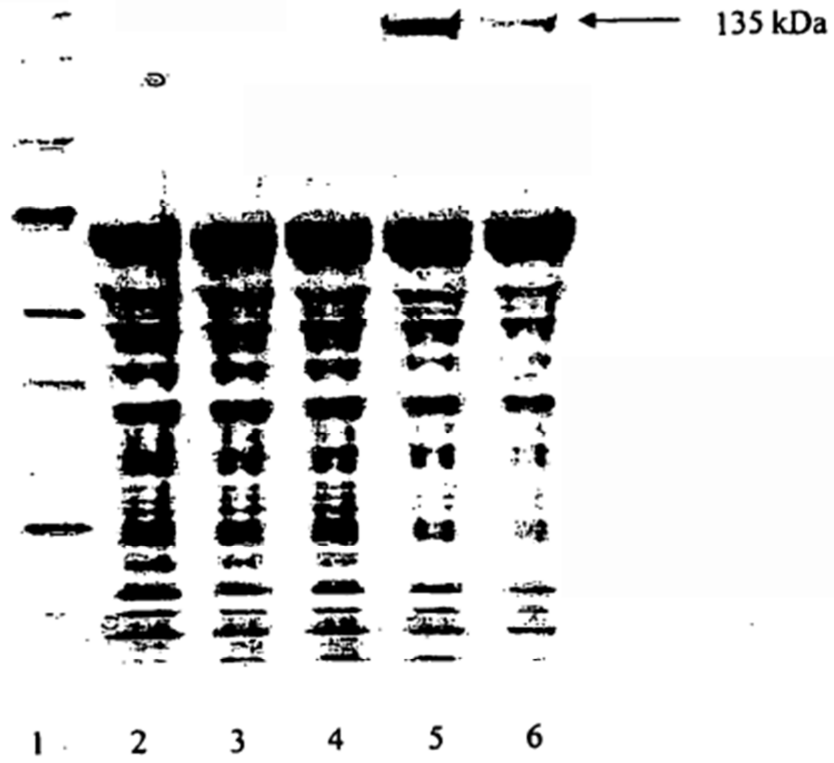


Figura 14

