

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 374 620**

51 Int. Cl.:
A61K 49/00 (2006.01)
A61K 39/00 (2006.01)
A61K 39/38 (2006.01)
A61K 39/02 (2006.01)
A61K 39/04 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **09170183 .9**
96 Fecha de presentación: **20.06.2001**
97 Número de publicación de la solicitud: **2133100**
97 Fecha de publicación de la solicitud: **16.12.2009**

54 Título: **ANTÍGENO MTB32A DE MYCOBACTERIUM TUBERCULOSIS CON EL SITIO ACTIVO INACTIVADO Y PROTEÍNAS DE FUSIÓN DE LOS MISMOS.**

30 Prioridad:
20.06.2000 US 597796
01.02.2001 US 265737 P

45 Fecha de publicación de la mención BOPI:
20.02.2012

45 Fecha de la publicación del folleto de la patente:
20.02.2012

73 Titular/es:
CORIXA CORPORATION
CSC THE UNITED STATES CORPORATION 2711
CENTERVILLE ROAD
WILMINGTON, DE 19808, US

72 Inventor/es:
Skeiky, Yasir;
Reed, Steven y
Alderson, Mark

74 Agente: **de Elzaburu Márquez, Alberto**

ES 2 374 620 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

Antígeno MTB32A de *Mycobacterium tuberculosis* con el sitio activo inactivado y proteínas de fusión de los mismos.

Campo de la invención

5 La presente invención se refiere a antígenos de *Mycobacterium sp.* En particular, se refiere a ácidos nucleicos que codifican antígenos individuales de *M. tuberculosis* que aumentan la sensibilidad serológica de los sueros de individuos infectados con tuberculosis, y a su uso en el diagnóstico, tratamiento y prevención de infección por tuberculosis.

Antecedentes de la invención

10 La tuberculosis es una enfermedad infecciosa crónica causada por infección con *M. tuberculosis* y otras especies de *Mycobacterium*. Es una enfermedad importante en los países en desarrollo, así como un problema creciente en las zonas desarrolladas del mundo, con aproximadamente 8 millones de casos nuevos y 3 millones de muertes cada año. Aunque la infección puede ser asintomática durante un periodo considerable de tiempo, la enfermedad se manifiesta lo más comúnmente como una inflamación aguda de los pulmones, dando como resultado fiebre y tos improductiva. Si no se trata, da como resultado típicamente complicaciones graves y muerte.

15 Aunque la tuberculosis puede controlarse generalmente usando una terapia de antibióticos extendida, dicho tratamiento no es suficiente para prevenir la difusión de la enfermedad. Los individuos afectados pueden ser asintomáticos, pero contagiosos, durante algún tiempo. Además, aunque el cumplimiento del régimen de tratamiento es crítico, el comportamiento del paciente es difícil de monitorizar. Algunos pacientes no completan el curso del tratamiento, lo que puede conducir a un tratamiento ineficaz y al desarrollo de resistencia a fármacos.

20 Para controlar la difusión de la tuberculosis, la vacunación eficaz y el diagnóstico temprano exacto de la enfermedad son de la máxima importancia. Actualmente, la vacunación con bacterias vivas es el método más eficaz para inducir inmunidad protectora. La micobacteria más común empleada con este fin es el bacilo de Calmette-Guerin (BCG), una cepa avirulenta de *M. bovis*. Sin embargo, la seguridad y eficacia del BCG es fuente de controversia y algunos países, tales como los Estados Unidos, no vacunan al público general con este agente.

25 El diagnóstico de la tuberculosis se consigue comúnmente usando un ensayo cutáneo, que implica la exposición intradérmica a tuberculina DPP (derivado de proteína purificado). Las respuestas de linfocitos T específicas de antígeno dan como resultado una induración mensurable en el sitio de inyección durante 48-72 horas después de la inyección, que indica exposición a antígenos micobacterianos. Sin embargo, la sensibilidad y especificidad han sido problemáticos con este ensayo, y los individuos vacunados con BCG no pueden distinguirse de los individuos infectados.

30 Aunque se ha mostrado que los macrófagos actúan como los efectores principales de la inmunidad por *Mycobacterium*, los linfocitos T son los inductores predominantes de dicha inmunidad. El papel esencial de los linfocitos T en la protección frente a infección por *Mycobacterium* se ilustra por la frecuente aparición de infección por *Mycobacterium* en pacientes de SIDA, debido al agotamiento de linfocitos T CD4⁺ asociado a la infección por virus de la inmunodeficiencia humana (VIH). Se ha mostrado que los linfocitos T CD4⁺ reactivos con *Mycobacterium* son potentes productores de interferón (IFN- γ) que, a su vez, se ha mostrado que desencadena los efectos antimicobacterianos de los macrófagos en ratones. Aunque el papel del IFN- γ en seres humanos está menos claro, los estudios han mostrado que la 1,25-dihidroxitamina D3, sola o en combinación con IFN- γ o factor alfa de necrosis tumoral, activa los macrófagos humanos para inhibir la infección por *M. tuberculosis*. Además, es conocido que el IFN- γ estimula a los macrófagos humanos a preparar 1,25 -dihidroxitamina D3. De forma similar, se ha mostrado que la interleucina-12 (IL-12) desempeña un papel en la estimulación de la resistencia a infección por *M. tuberculosis*. Para una revisión de la inmunología de infección por *M. tuberculosis*, véanse Chan y Kaufmann, "Tuberculosis: Pathogenesis, Protection and Control" (Bloom ed., 1994) y "Harrison's Principles of Internal Medicine", volumen 1, pág. 1004-1014 y 1019-1023 (14^a ed., Fauci *et al.*, eds., 1998).

45 Skeiky *et al.*, Infection and Immunity 1999, 67(8): 3998-4007 describe la clonación, expresión y evaluación inmunológica de dos supuestos antígenos de serina proteasa secretados de *Mycobacterium tuberculosis*, identificando la triada catalítica de histidina, ácido aspártico y serina (correspondiente a los residuos 95, 126 y 208 respectivamente de la SEQ ID N° 2 de la presente memoria).

50 En consecuencia, existe la necesidad de reactivos de diagnóstico mejorados y métodos mejorados para el diagnóstico, prevención y tratamiento de la tuberculosis.

Sumario de la invención

55 En un aspecto, la presente invención proporciona polipéptidos y proteínas de fusión que comprenden una versión mutada de MTB32A, en que uno, dos o tres de los tres aminoácidos histidina, aspartato o serina en el sitio activo se han mutado a un aminoácido diferente. En una realización, en Ra35FL, la serina en posición 183 se ha mutado a un residuo de alanina, creando Ra35FLMutSA. En una realización, el ADN que codifica Ra35FL se ha mutado

cambiando una T por G, dando como resultado una mutación de serina a alanina en el aminoácido 183 de la SEQ ID N° 4.

5 La presente invención está basada, en parte, en el descubrimiento por los inventores de que los polinucleótidos de fusión, polipéptidos de fusión o composiciones que contienen al menos dos secuencias de codificación heterólogas de *M. tuberculosis* o antígenos son altamente antigénicos y, tras administración a un paciente, aumentan la sensibilidad a sueros de tuberculosis. Además, las composiciones, polipéptidos y polinucleótidos de fusión son útiles como herramientas de diagnóstico en pacientes que pueden haberse infectado con *Mycobacterium*.

10 Los polipéptidos, polipéptidos de fusión y ácidos nucleicos de la invención se usan en ensayos *in vitro* e *in vivo* para detectar anticuerpos humorales o inmunidad mediada por célula contra *M. tuberculosis* para el diagnóstico de infección o la monitorización de la progresión de la enfermedad. Por ejemplo, los polipéptidos pueden usarse como agente de diagnóstico *in vivo* en forma de un ensayo cutáneo intradérmico. Los polipéptidos pueden usarse también en ensayos *in vitro* tales como ELISA con suero de paciente. Como alternativa, los ácidos nucleicos, composiciones y polipéptidos de fusión pueden usarse para crear anticuerpos anti-*M. tuberculosis* en un animal no humano. Los anticuerpos pueden usarse para detectar los antígenos diana *in vivo* e *in vitro*.

15 Los polipéptidos, polipéptidos de fusión y ácidos nucleicos pueden usarse como inmunógenos para generar o provocar una respuesta inmunitaria protectora en un paciente. Se usan polinucleótidos aislados o purificados para producir antígenos de polipéptido recombinante de fusión *in vitro*, que se administran entonces en forma de vacuna. Como alternativa, los polinucleótidos pueden administrarse directamente a un sujeto en forma de vacunas de ADN para causar la expresión de antígeno en el sujeto y la posterior inducción de una respuesta inmunitaria anti-*M. tuberculosis*. Por tanto, los polipéptidos y ácidos nucleicos de *M. tuberculosis* aislados o purificados de la invención pueden formularse en forma de composiciones farmacéuticas para administración a un sujeto en la prevención y/o el tratamiento de infección por *M. tuberculosis*. La inmunogenicidad de la proteína de fusión o antígenos puede potenciarse mediante la inclusión de un coadyuvante, así como polipéptidos de fusión adicionales, de *Mycobacterium* u otros organismos, tales como polipéptidos bacterianos, víricos o de mamífero. Pueden incluirse también polipéptidos adicionales en las composiciones, ligados o no ligados al polipéptido de fusión.

Breve descripción de los dibujos

La Figura 1 muestra el porcentaje de supervivencia de conejillos de Indias vacunados con poliproteína MTB72F.

La Figura 2 muestra las UFC de células de bazo (Fig. 2A) y células pulmonares después de inmunización con MTB72F, MTB59F, ADN de MTB72F o una composición que comprende los antígenos Ra12, TbH9, y Ra35.

30 La Figura 3 muestra un diagrama esquemático de MTB72F.

La Figura 4 muestra la secuencia nucleotídica y aminoacídica de Ra35 (195 aminoácidos de la porción N-terminal de MTB32A).

La Figura 5 muestra un alineamiento de las secuencias aminoacídicas de MTB72F y la versión mutada MTB72FMutSA.

35 La Figura 6 muestra un alineamiento de las secuencias aminoacídicas de Ra35/MTB32A madura (completa) y la versión mutada Ra35FLMutSA.

La Figura 7 muestra la supervivencia a largo plazo de conejillos de Indias vacunados con formulaciones de Mtb72F.

Descripción de realizaciones específicas

40 La presente invención se refiere a un polipéptido que comprende un polipéptido MTB32A de una especie de *Mycobacterium* del complejo de tuberculosis, en el que al menos un aminoácido en la triada de sitio activo del antígeno MTB32A (SEQ ID N° 2 o 4) se ha sustituido por un aminoácido diferente. Se proporciona también un ácido nucleico que codifica un polipéptido que comprende el antígeno MTB32A de una especie de *Mycobacterium* del complejo de tuberculosis, en el que al menos un aminoácido en la triada de sitio activo del antígeno MTB32A (SEQ ID N° 2 o 4) se ha sustituido por un aminoácido diferente.

45 Se proporciona adicionalmente un polipéptido de fusión que comprende un antígeno MTB39 (SEQ ID N° 12 o 14) de una especie de *Mycobacterium* del complejo de tuberculosis y un antígeno que comprende al menos 195 aminoácidos del extremo N de un antígeno MTB32A (SEQ ID N° 2 o 4) de una especie de *Mycobacterium* del complejo de tuberculosis, en el que un aminoácido de la triada del sitio activo del antígeno MTB32A (SEQ ID N° 2 o 4) se ha sustituido por un aminoácido diferente, y ácidos nucleicos que codifican dichos polipéptidos de fusión.

50 Estas composiciones, polipéptidos de fusión y los ácidos nucleicos que los codifican son por lo tanto útiles para provocar una respuesta protectora en pacientes, y para aplicaciones de diagnóstico.

Los antígenos de la presente invención pueden comprender adicionalmente otros componentes diseñados para potenciar la antigenicidad de los antígenos o para mejorar estos antígenos en otros aspectos, por ejemplo, el

aislamiento de estos antígenos mediante la adición de una extensión de residuos de histidina a un extremo del antígeno. Las composiciones, polipéptidos de fusión y ácidos nucleicos de la invención pueden comprender copias adicionales de antígenos o polipéptidos heterólogos adicionales de *Mycobacterium sp.*, tales como antígeno MTB8.4, antígeno MTB9.8, antígeno MTB9.9, antígeno MTB40, antígeno MTB41, antígenos 38-1, TbRa3, 38 kDa, DPEP, TbH4, DPPD, antígeno ESAT-6, antígeno complejo MTB85 (por ejemplo, MTB85b) o antígenos cristalinos y Erd14. Las composiciones, polipéptidos de fusión y ácidos nucleicos de la invención pueden comprender también polipéptidos heterólogos adicionales de otras fuentes no *Mycobacterium*. Por ejemplo, las composiciones y proteínas de fusión de la invención pueden incluir polipéptidos o ácidos nucleicos que codifican polipéptidos en los que el polipéptido potencia la expresión del antígeno, por ejemplo, NS 1, una proteína del virus de la gripe, o una porción inmunogénica de la misma (véanse, por ejemplo, los documentos WO99/4018 y WO93/04175). Los ácidos nucleicos de la invención pueden modificarse por ingeniería genética basándose en la preferencia de codón en la especie de elección, por ejemplo, seres humanos.

Las composiciones de la invención pueden ser ADN desnudo o las composiciones, por ejemplo polipéptidos, pueden comprender también coadyuvantes, por ejemplo, MPL, 3D-MPL, IFA, coadyuvantes AS tales como AS2, AS2', AS2'', AS4, AS6, ENHANZYN (Detox), QS21, CWS, TDM, AGP, CPG, Leif, saponina y miméticos de saponina y derivados de los mismos. Además, las composiciones de la invención pueden comprender BCG o Pvac como coadyuvante.

En la nomenclatura de la solicitud, Ra35 designa el extremo N de MTB32A (Ra35FL), que comprende al menos de aproximadamente 195 a 205 aminoácidos de MTB32A de *M. tuberculosis*, o la correspondiente región de otra especie de *Mycobacterium*. Ra12 designa el extremo C de MTB32A (Ra35FL), que comprende al menos aproximadamente los últimos 132 aminoácidos de MTB32A de *M. tuberculosis*, o la correspondiente región de otra especie de *Mycobacterium*.

A continuación, se proporcionan secuencias de algunos antígenos usados en las proteínas de fusión de la invención:

SEQ ID N° 1-4: MTB32A (Ra35FL o Ra35 madura), cuya secuencia se da a conocer también como SEQ ID N° 17 (ADNc) y SEQ ID N° 79 (proteína) en las solicitudes de patente de EE.UU. n° 08/523.436, 08/523.435, 08/658.800, 08/659.683, 08/818.112, 09/056.556 y 08/818.111 y en las solicitudes WO97/09428 y WO97/09429, véase también Skeiky *et al.*, *Infection and Immunity* 7: 3998-4007 (1999). El término MTB32A incluye también secuencias aminoacídicas de MTB32A en que uno cualquiera de los tres aminoácidos en la triada de sitio activo (His, Asp, Ser), por ejemplo, el residuo de serina en la posición aminoacídica 208 de la SEQ ID N° 2 o la posición aminoacídica 183 de la SEQ ID N° 4 se ha cambiado por otro aminoácido (por ejemplo, alanina, Ra35FLMutSA, véanse, por ejemplo, la Figura 6 y la SEQ ID N° 6).

SEQ ID N° 5 y 6: Ra35FLMut SA, la versión madura de RA35FL en la que el residuo de serina en la posición aminoacídica 183 de la SEQ ID N° 4 se ha cambiado por un residuo de alanina.

SEQ ID N° 7 y 8: Ra35, el extremo N de MTB32A (Ra35FL), que comprende al menos aproximadamente 195 aminoácidos del extremo N de MTB32A de *M. tuberculosis*, cuya secuencia nucleotídica y aminoacídica se dan a conocer en la Figura 4 (véanse también los aminoácidos 33-227 de la SEQ ID N° 2 y los aminoácidos 8 a 202 de la SEQ ID N° 4). El término Ra35 (N-terminal) incluye también secuencias aminoacídicas de Ra35 en que uno cualquiera de los tres aminoácidos en la triada de sitio activo (concretamente, His, Asp, Ser) se ha cambiado como se describe anteriormente.

SEQ ID N° 9 y 10: MTBRa12, el extremo C de MTB32A (Ra35FL), que comprende al menos aproximadamente 132 aminoácidos del extremo C de MTB32A de *M. tuberculosis* (véanse, por ejemplo, los aminoácidos 224 a 355 de la SEQ ID N° 2 y los aminoácidos 199 a 330 de la SEQ ID N° 4), cuya secuencia se da a conocer como SEQ ID N° 4 (ADN) y SEQ ID N° 66 (secuencia aminoacídica predicha) en la solicitud de patente de EE.UU. n° 09/072.967.

SEQ ID N° 11, 12, 13 y 14: MTB39 (TbH9), cuya secuencia se da a conocer como SEQ ID N° 106 (ADNc completo) y SEQ ID N° 107 (proteína completa) en las solicitudes de patente de EE.UU. n° 08/658.800, 08/659.683, 08/818.112 y 08/818.111 y en las solicitudes WO97/09428 y WO97/09429. La secuencia se da a conocer también como SEQ ID N° 33 (ADN) y SEQ ID N° 91 (aminoácido) en la solicitud de patente de EE.UU. n° 09/056.559.

A continuación, se proporcionan secuencias de algunas proteínas de fusión:

SEQ ID N° 15 y 16: MTB72F (Ra12-TbH9-Ra35), cuya secuencia se da a conocer como SEQ ID N° 1 (ADN) y SEQ ID N° 2 (proteína) en la solicitud de patente de EE.UU. n° 09/223.040 y en la solicitud PCT/US99/07717. El término MTB372F incluye también secuencias aminoacídicas de MTB72F en las que cualquiera de los tres aminoácidos en la triada de sitio activo en Ra35FL (concretamente, His, Asp o Ser) se ha cambiado como se describe anteriormente (véase, por ejemplo, MTB72FMutSA, Figura 5).

SEQ ID N° 17 y 18: MTB72FMutSA (Ra12-TbH9-Ra35MutSA), en la que, en el componente Ra35 de la proteína de fusión, la serina en la posición 710 se ha cambiado por alanina.

SEQ ID N° 19 y 20: TbH9-Ra35 (MTB59F), cuya secuencia se da a conocer como SEQ ID N° 23 (ADNc) y SEQ ID N° 24 (proteína) en la solicitud de patente de EE.UU. n° 09/287.849 y en la solicitud PCT/US99/07717.

A continuación, se proporcionan secuencias de algunos antígenos adicionales usados en las composiciones y proteínas de fusión de la invención:

5 SEQ ID N° 21 y 22: MTB8.4 (DPV), cuya secuencia se da a conocer como SEQ ID N° 101 (ADNc) y SEQ ID N° 102 (proteína) en las solicitudes de patente de EE.UU. n° 08/658.800, 08/659.683, 08/818.112 y 08/818.111 y en las solicitudes WO97/09428 y WO97/09429.

SEQ ID N° 23 y 24: MTB9.8 (MSL), cuya secuencia se da a conocer como SEQ ID N° 12 (ADN), SEQ ID N° 109 (secuencia aminoacídica predicha) y SEQ ID N° 110 a 124 (péptidos) en las solicitudes de patente de EE.UU. n° 08/859.381, 08/858.998, 09/073.009 y 09/073.010 y en las solicitudes PCT/US98/10407 y PCT/US98/10514.

10 SEQ ID N° 25, 26 y 27: MTB9.9A (MTI, también conocido como MTI-A), cuya secuencia se da a conocer como SEQ ID N° 3 y SEQ ID N° 4 (ADN) y SEQ ID N° 29 y SEQ ID N° 51 a 66 (péptido ORF para MTI) en las solicitudes de patente de EE.UU. n° 08/859.381, 08/858.998, 09/073.009 y 09/073.010 y en las solicitudes PCT/US98/10407 y PCT/US98/10514. Existen también otras dos variantes de MTI, llamadas MTI-B y MTI-C.

15 SEQ ID N° 28 y 29: MTB40 (HTCC n° 1), cuya secuencia se da a conocer como SEQ ID N° 137 (ADNc) y 138 (secuencia aminoacídica predicha) en las solicitudes de patente de EE.UU. n° 09/073.009 y 09/073.010 y en las solicitudes PCT/US98/10407 y PCT/US98/10514.

SEQ ID N° 30 y 31: MTB41 (MTCC n° 2), cuya secuencia se da a conocer como SEQ ID N° 140 (ADNc) y SEQ ID N° 142 (secuencia aminoacídica predicha) en las solicitudes de patente de EE.UU. n° 09/073.009 y 09/073.010 y en las solicitudes PCT/US98/10407 y PCT/US98/10514.

20 SEQ ID N° 32 y 33: ESAT-6, cuya secuencia se da a conocer como SEQ ID N° 103 (ADN) y SEQ ID N° 104 (secuencia aminoacídica predicha) en la solicitud de patente de EE.UU. n° 09/072.967. La secuencia de ESAT-6 se da a conocer también en la patente de EE.UU. n° 5.955.077.

SEQ ID N° 34 y 35: Tb38-1 o 38-1 (MTb11), cuya secuencia se da a conocer en la SEQ ID N° 46 (ADN) y SEQ ID N° 88 (aminoácido predicho) y en las solicitudes de patente de EE.UU. n° 09/072.96; 08/523.436; 08/523.435; 08/818,112 y 08/818.111 y en las solicitudes WO97/09428 y WO97/09429.

25 SEQ ID N° 36 y 37: TbRa3, cuya secuencia se da a conocer en la SEQ ID N° 15 (ADN) y la SEQ ID N° 77 (secuencia aminoacídica predicha) de las solicitudes WO 97/09428 y WO97/09429.

SEQ ID N° 38 y 39: 38 kDa, cuya secuencia se da a conocer en la SEQ ID N° 154 (ADN) y la SEQ ID N° 155 (secuencia aminoacídica predicha) en la solicitud de patente de EE.UU. n° 09/072.967. 38 kDa tiene dos formas alternativas con y sin el residuo de cisteína N-terminal.

30 SEQ ID N° 40 y 41: DPEP, cuya secuencia se da a conocer en la SEQ ID N° 52 (ADN) y la SEQ ID N° 53 (secuencia aminoacídica predicha) en las publicaciones WO97/09428 y WO97/09429.

SEQ ID N° 42 y 43: TbH4, cuya secuencia se da a conocer como la SEQ ID N° 43 (ADN) y la SEQ ID N° 81 (secuencia aminoacídica predicha) en las publicaciones WO97/09428 y WO97/09429.

35 SEQ ID N° 44 y 45: DPPD, cuya secuencia se da a conocer en la SEQ ID N° 240 (ADN) y la SEQ ID N° 241 (secuencia aminoacídica predicha) en el documento USSN 09/072.967 y en las solicitudes PCT/US99/03268 y PCT/US99/03265. Se muestra la forma secretada de DPPD de la presente memoria en la Figura 12 del documento PCT/US00/28095.

MTb82 (MTb867), cuya secuencia se da a conocer en las Figuras 8 (ADN) y 9 (aminoácido) del documento PCT/US00/2809.

40 Erd14 (MTb16), cuyas secuencias de ADNc y aminoacídica se dan a conocer en Verbon *et al.*, J. Bacteriology 174: 1352-1359 (1992).

Antígeno α cristalino, cuya secuencia se da a conocer en Verbon *et al.*, J. Bact. 174: 1352-1359 (1992).

Antígeno complejo 85, por ejemplo antígeno 85b, cuya secuencia se da a conocer en Content *et al.*, Infect. & Immunol. 59: 3205-3212 (1991).

45 A continuación, se proporcionan secuencias de algunas proteínas de fusión adicionales:

SEQ ID N° 46 y 47: DPV-MTI-MSL-MTCC n° 2 (MTb71F), cuya secuencia se da a conocer como SEQ ID N° 15 (ácido nucleico) y en la SEQ ID N° 16: (proteína) en la solicitud de patente de EE.UU. n° 09/287.849 y en la solicitud PCT/US99/07717.

50 SEQ ID N° 48 y 49: DPV-MTI-MSL (MTb31F), cuya secuencia se da a conocer en la SEQ ID N° 18 (ADNc) y la SEQ ID N° 19 (proteína) en la solicitud de patente de EE.UU. n° 09/287.849 y en la solicitud PCT/US99/07717.

Cada una de las secuencias anteriores se da a conocer también en Cole *et al.* Nature 393: 537 (1998) y puede encontrarse, por ejemplo, en <http://www.sanger.ac.uk> y <http://www.pasteur.fr/mycdb/>. Las secuencias anteriores se dan a conocer en las solicitudes de patente de EE.UU. nº 08/523.435, 08/523.436, 08/658.800, 08/659.683, 08/818.111, 08/818.112, 08/942.341, 08/942.578, 08/858.998, 08/859.381, 09/056.556, 09/072.596, 09/072.967, 09/073.009, 09/073.010, 09/223.040, 09/1287.849 09/597.796 y en las solicitudes de patente PCT PCT/US00/28095; PCT/US98/10407, PCT/US98/10514, PCT/US99/03265, PCT/US99/03268, PCT/US99/07717 y los documentos WO97/09428 y WO97/09429, WO98/16645, WO98/16646.

Los antígenos descritos en la presente memoria incluyen variantes polimórficas y variaciones modificadas conservadoramente, así como homólogos de *Mycobacterium* entre cepas y entre especies. Además, los antígenos descritos en la presente memoria incluyen subsecuencias o secuencias truncadas. Las proteínas de fusión pueden contener también polipéptidos adicionales, opcionalmente péptidos heterólogos de *Mycobacterium* u otras fuentes. Estos antígenos pueden modificarse, por ejemplo, añadiendo secuencias peptídicas ligadoras como se describen a continuación. Estos péptidos ligadores pueden insertarse entre los uno o más polipéptidos que constituyen cada una de las proteínas de fusión.

Definiciones

"Polipéptido de fusión" o "proteína de fusión" designa una proteína que tiene al menos dos polipéptidos heterólogos de *Mycobacterium sp.* ligados covalentemente, directamente o mediante un ligador aminoacídico. Los polipéptidos que forman la proteína de fusión están típicamente ligados de extremo C a extremo N, aunque también pueden estar ligados de extremo C a extremo C, de extremo N a extremo N o de extremo N a extremo C. Los polipéptidos de la proteína de fusión pueden estar en cualquier orden. Este término designa también las variantes modificadas conservadoramente, variantes polimórficas, alelos, mutantes, subsecuencias, homólogos entre especies y fragmentos inmunogénicos de los antígenos que constituyen la proteína de fusión. Los antígenos de *Mycobacterium tuberculosis* se describen en Cole *et al.*, Nature 393: 537 (1998), que da a conocer el genoma completo de *Mycobacterium tuberculosis*. La secuencia completa de *Mycobacterium tuberculosis* puede encontrarse también en <http://www.sanger.ac.uk> y en <http://www.pasteur.fr/mycdb/> (MycDB). Pueden identificarse antígenos de otras especies de *Mycobacterium* que corresponden a los antígenos de *M. tuberculosis*, por ejemplo, usando algoritmos de comparación de secuencia como se describen en la presente memoria, u otros métodos conocidos por los expertos en la materia, por ejemplo, ensayos de hibridación y ensayos de unión de anticuerpo. Las proteínas de fusión de la invención pueden comprender también copias adicionales de un componente antigénico o fragmento inmunogénico del mismo.

En algunas realizaciones, los polipéptidos individuales de la proteína de fusión están en orden (extremo N a C) de grande a pequeño. Los antígenos grandes son de aproximadamente 30 a 150 kDa de tamaño, los antígenos medianos son de aproximadamente 10 a 30 kDa de tamaño y los antígenos pequeños son de aproximadamente menos de 10 kDa de tamaño. La secuencia que codifica el polipéptido individual puede ser tan pequeña como, por ejemplo, un fragmento inmunogénico tal como un epítipo de LTC individual que codifica aproximadamente 8 a 9 aminoácidos o, por ejemplo, un epítipo de LTA o linfocito B. El fragmento puede incluir también múltiples epítopos.

Los polipéptidos de fusión comprenden opcionalmente polipéptidos adicionales, por ejemplo, 3, 4, 5, 6 ó más polipéptidos, hasta aproximadamente 25 polipéptidos, opcionalmente polipéptidos heterólogos o polipéptidos homólogos repetidos, fusionados con al menos dos antígenos heterólogos. Los polipéptidos adicionales de la proteína de fusión derivan opcionalmente de *Mycobacterium* así como de otras fuentes, tales como otras fuentes bacterianas, víricas, invertebradas, vertebradas o de mamífero. Los polipéptidos individuales de la proteína de fusión pueden estar en cualquier orden. Como se describe en la presente memoria, la proteína de fusión puede estar también ligada a otras moléculas, incluyendo polipéptidos adicionales. Las composiciones de la invención pueden comprender también polipéptidos adicionales que no están ligados a las proteínas de fusión de la invención. Estos polipéptidos adicionales pueden ser polipéptidos heterólogos u homólogos.

El término "fusionado" designa el ligamiento covalente entre dos polipéptidos en una proteína de fusión. Los polipéptidos se unen típicamente mediante un enlace peptídico, directamente entre sí o mediante un ligador aminoacídico. Opcionalmente, los péptidos pueden unirse mediante ligamientos covalentes no peptídicos conocidos por los expertos en la materia.

"FL" designa completo, concretamente, un polipéptido que es de la misma longitud que el polipéptido de tipo silvestre.

El término "fragmento inmunogénico del mismo" designa un polipéptido que comprende un epítipo que es reconocido por linfocitos T citotóxicos, linfocitos T auxiliares o linfocitos B.

El término "especie de *Mycobacterium* del complejo de tuberculosis" incluye aquellas especies consideradas tradicionalmente como causantes de la enfermedad tuberculosis, así como especies de *Mycobacterium* ambientales y oportunistas que causan tuberculosis y enfermedad pulmonar en pacientes inmunocomprometidos, tales como pacientes con SIDA, por ejemplo, *M. tuberculosis*, *M. bovis*, o *M. africanum*, BCG, *M. avium*, *M. intracellulare*, *M. celatum*, *M. genavense*, *M. haemophilum*, *M. kansasii*, *M. simiae*, *M. vaccae*, *M. fortuitum* y *M. scrofulaceum* (véase,

por ejemplo, "Harrison's Principles of Internal Medicine", volumen 1, pág. 1004-1014 y 1019-1023 (14ª ed., Fauci *et al.*, eds., 1998).

Un coadyuvante designa los componentes de una vacuna o composición terapéutica que aumentan la respuesta inmunitaria específica al antígeno (véase, por ejemplo, Edelman, *AIDS Res. Hum Retroviruses* 8: 1409-1411 (1992)).

5 Los coadyuvantes inducen respuestas inmunitarias de respuesta de tipo Th1 y tipo Th2. Las citocinas de tipo Th1 (por ejemplo, IFN- γ , IL-2 e IL-12) tienden a favorecer la inducción de respuesta inmunitaria mediada por célula ante un antígeno administrado, mientras que las citocinas de tipo Th2 (por ejemplo, IL-4, IL-5, IL-6, IL-10 y TNF- β) tienden a favorecer la inducción de respuestas inmunitarias humorales.

"Ácido nucleico" designa desoxirribonucleótidos o ribonucleótidos y polímeros de los mismos en forma mono- o bicatenaria. El término comprende ácidos nucleicos que contienen análogos nucleotídicos o residuos o ligamientos de cadena principal modificados conocidos que son sintéticos, de origen natural y de origen no natural, que tienen propiedades de unión similares al ácido nucleico de referencia y que se metabolizan de manera similar a los nucleótidos de referencia. Los ejemplos de dichos análogos incluyen, sin limitación, fosforotioatos, fosforamidatos, fosfonatos de metilo, fosfonatos de metilo quirales, 2-O-metilribonucleótidos y ácidos nucleicos peptídicos (ANP).

10 A menos que se indique otra cosa, una secuencia de ácido nucleico particular comprende también implícitamente variantes modificadas conservadoramente del mismo (por ejemplo, sustituciones de codón degeneradas) y secuencias complementarias, así como la secuencia indicada explícitamente. Específicamente, las sustituciones de codón degeneradas pueden conseguirse generando secuencias en que la tercera posición de uno o más codones seleccionados (o todos) se sustituye por residuos de base mixta y/o desoxiinosina (Batzer *et al.*, *Nucleic Acid Res.* 19: 5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260: 2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8: 91-98 (1994)). El término ácido nucleico se usa intercambiamente con gen, ADNc, ARNm, oligonucleótido y polinucleótido.

Los términos "polipéptido", "péptido" y "proteína" se usan intercambiamente en la presente memoria para designar un polímero de residuos aminoacídicos. Los términos se aplican a polímeros aminoacídicos en que uno o más

25 residuos aminoacídicos son un mimético químico artificial de un correspondiente aminoácido de origen natural, así como a polímeros aminoacídicos de origen natural y polímeros aminoacídicos de origen no natural.

El término "aminoácido" designa aminoácidos de origen natural y aminoácidos sintéticos, así como análogos aminoacídicos y miméticos aminoacídicos que funcionan de manera similar a los aminoácidos de origen natural. Los aminoácidos de origen natural son aquellos codificados por el código genético, así como aquellos aminoácidos que se modifican después, por ejemplo, hidroxiprolina, -carboxiglutamato y O-fosfoserina. Análogos aminoacídicos designa compuestos que tienen la misma estructura química básica que un aminoácido de origen natural, concretamente, un carbono α que está unido a un hidrógeno, un grupo carboxilo, un grupo amino y un grupo R, por ejemplo, homoserina, norleucina, metionina-sulfóxido o metionina-metilsulfonio. Dichos análogos tienen grupos R modificados (por ejemplo, norleucina) o cadenas principales peptídicas modificadas, pero retienen la misma

30 estructura química básica que un aminoácido de origen natural. Miméticos aminoacídicos designa compuestos químicos que tienen una estructura que es diferente de la estructura química general de un aminoácido, pero que funcionan de manera similar a un aminoácido de origen natural.

Los aminoácidos pueden designarse en la presente memoria por cualquiera de los símbolos de tres letras o símbolos de una letra conocidos comúnmente recomendados por la IUPAC-IUB Biochemical Nomenclature Commission. Los nucleótidos, igualmente, pueden designarse por sus códigos de una letra comúnmente aceptados.

"Variantes modificadas conservadoramente" se aplica tanto a secuencias aminoacídicas como de ácido nucleico. Con respecto a secuencias de ácido nucleico particulares, las variantes modificadas conservadoramente designan aquellos ácidos nucleicos que codifican secuencias aminoacídicas idénticas o esencialmente idénticas, o cuando el ácido nucleico no codifica una secuencia aminoacídica, secuencias esencialmente idénticas. Debido a la degeneración del código genético, un gran número de ácidos nucleicos funcionalmente idénticos codifican cualquier proteína dada. Por ejemplo, los codones GCA, GCC, GCG y GCU codifican todos el aminoácido alanina. Por tanto, en cualquier posición en que la alanina esté especificada por un codón, el codón puede alterarse a cualquiera de los correspondientes codones descritos sin alterar el polipéptido codificado. Dichas variaciones de ácido nucleico son "variaciones silenciosas", que son una especie de variaciones modificadas conservadoramente. Cada secuencia de ácido nucleico en la presente memoria que codifica un polipéptido describe también cada variación silenciosa posible del ácido nucleico. Un experto reconocerá que cada codón de un ácido nucleico (excepto AUG, que normalmente es el único codón para metionina y TGG que normalmente es el único codón para triptófano) puede modificarse para proporcionar una molécula funcionalmente idéntica. En consecuencia, cada variación silenciosa de un ácido nucleico que codifica un polipéptido está implícita en cada secuencia descrita.

En cuanto a las secuencias aminoacídicas, un experto reconocerá que las sustituciones, deleciones o adiciones individuales en un ácido nucleico, péptido, polipéptido o secuencia proteica que alteran, añaden o eliminan un solo aminoácido o un pequeño porcentaje de aminoácidos en la secuencia codificada son una "variante modificada conservadoramente", en que la alteración da como resultado la sustitución de un aminoácido por un aminoácido químicamente similar. Las tablas de sustitución conservadora que proporcionan aminoácidos funcionalmente

similares son bien conocidas en la materia. Dichas variantes modificadas conservadoramente se añaden a y no excluyen variantes polimórficas, homólogos entre especies y alelos de la invención.

Los siguientes ocho grupos contienen cada uno aminoácidos que son sustituciones conservadoras entre sí:

- 1) alanina (A), glicina (G);
- 5 2) ácido aspártico (D), ácido glutámico (E);
- 3) asparagina (N), glutamina (Q);
- 4) arginina (R), lisina (K);
- 5) isoleucina (I), leucina (L), metionina (M), valina (V);
- 6) fenilalanina (F), tirosina (Y), triptófano (W);
- 10 7) serina (S), treonina (T); y
- 8) cisteína (C), metionina (M)

(véase, por ejemplo, Creighton, "Proteins" (1984)).

15 El término "heterólogo", cuando se usa con referencia a porciones de un ácido nucleico, indica que el ácido nucleico comprende dos o más subsecuencias que no se encuentran en la misma relación entre sí en la naturaleza. Por ejemplo, el ácido nucleico se produce típicamente por recombinación, y tiene dos o más secuencias de genes no relacionadas dispuestas para preparar un nuevo ácido nucleico funcional, por ejemplo, un promotor de una fuente y una región de codificación de otra fuente. De forma similar, proteína heteróloga indica que la proteína comprende dos o más subsecuencias que no se encuentran en la misma relación entre sí en la naturaleza (por ejemplo, una proteína de fusión).

20 La frase "hibrida selectiva (o específicamente) con" designa la unión, formación de dúplex o hibridación de una molécula sólo con una secuencia nucleotídica particular en condiciones de hibridación rigurosas cuando esa secuencia está presente en una mezcla compleja (por ejemplo, ADN o ARN celular total o de colección).

25 La frase "condiciones de hibridación rigurosas" designa condiciones en las que una sonda hibridará con su subsecuencia diana, típicamente en una mezcla compleja de ácidos nucleicos, pero no con otras secuencias. Las condiciones rigurosas son dependientes de la secuencia y serán diferentes en diferentes circunstancias. Las secuencias más largas hibridarán específicamente a temperaturas mayores. Se encuentra una guía extensa de la hibridación de ácidos nucleicos en Tijssen, "Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Probes, Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generalmente, las condiciones estrictas se seleccionan para ser de aproximadamente 5-10°C menores que el punto de fusión térmico (T_m) de la secuencia específica a una fuerza iónica y pH definidos. La T_m es la temperatura (a fuerza iónica, pH y concentración de ácido nucleico definidos) a la que un 50% de las sondas complementarias con la diana hibridan con la secuencia diana en equilibrio (ya que las secuencias diana están presentes en exceso, a T_m , el 50% de las sondas están ocupadas en equilibrio). Las condiciones rigurosas serán aquellas en que la concentración salina sea menor de aproximadamente 1,0 M de ión sodio, típicamente de aproximadamente 0,01 a 30 1,0 M de concentración de ión sodio (u otras sales) a pH 7,0 a 8,3 y la temperatura sea de al menos aproximadamente 30°C para sondas cortas (por ejemplo, 10 a 50 nucleótidos) y de al menos aproximadamente 60°C para sondas largas (por ejemplo, mayores de 50 nucleótidos). Las condiciones rigurosas pueden conseguirse también con la adición de agentes desestabilizantes tales como formamida. Para hibridación selectiva o específica, una señal positiva es al menos dos veces la hibridación de fondo, opcionalmente 10 veces. Las condiciones de 35 hibridación rigurosas ejemplares pueden ser las siguientes: 50% de formamida, 5x SSC y 1% de SDS, incubación a 42°C, o 5x SSC, 1% de SDS, incubación a 65°C, con lavado con 0,2x SSC y 0,1% de SDS a 65°C.

40 Los ácidos nucleicos que no hibridan entre sí en condiciones rigurosas son todavía sustancialmente idénticos si los polipéptidos que codifican son sustancialmente idénticos. Esto ocurre, por ejemplo, cuando se crea una copia de un ácido nucleico usando la degeneración de codón máxima permitida por el código genético. En dichos casos, los ácidos nucleicos hibridan típicamente en condiciones de hibridación moderadamente rigurosas. Las "condiciones de hibridación moderadamente rigurosas" ejemplares incluyen la hibridación en un tampón de 40% de formamida, NaCl 1 M, 1% de SDS a 37°C y un lavado con 1xSSC a 45°C. Una hibridación positiva es al menos dos veces la de fondo. Los expertos en la materia reconocerán fácilmente que pueden utilizarse condiciones de hibridación y lavado 45 alternativas para proporcionar condiciones de rigor similar.

50 "Anticuerpo" designa un polipéptido que comprende una región marco de un gen de inmunoglobulina o fragmentos de la misma que se unen a y reconocen específicamente un antígeno. Los genes de inmunoglobulina reconocidos incluyen los genes de la región kappa, lambda, alfa, gamma, delta, épsilon y mu, así como la miríada de genes de región variable de inmunoglobulina. Las cadenas ligeras se clasifican como kappa o lambda. Las cadenas pesadas

se clasifican como gamma, mu, alfa, delta o épsilon, que a su vez definen las clases de inmunoglobulina, IgG, IgM, IgA, IgD e IgE, respectivamente.

Una unidad estructural de inmunoglobulina ejemplar (anticuerpo) comprende un tetrámero. Cada tetrámero está compuesto por dos pares idénticos de cadenas polipeptídicas, teniendo cada par una cadena "ligera" (aproximadamente 25 kDa) y una cadena "pesada" (aproximadamente 50-70 kDa). El extremo N de cada cadena define una región variable de aproximadamente 100 a 110 o más aminoácidos responsable principalmente del reconocimiento de antígeno. Los términos cadena ligera variable (V_L) y cadena pesada variable (V_H) designan estas cadenas ligera y pesada, respectivamente.

Los anticuerpos existen, por ejemplo, como inmunoglobulinas intactas o como una serie de fragmentos bien caracterizados producidos mediante la digestión con diversas peptidasas. Por tanto, por ejemplo, la pepsina digiere un anticuerpo por debajo de los ligamientos disulfuro en la región de bisagra produciendo $F(ab)'_2$, un dímero de Fab que es por sí solo una cadena ligera unida a V_H-C_{H1} por un enlace disulfuro. El $F(ab)'_2$ puede reducirse en condiciones suaves para romper el ligamiento disulfuro en la región de bisagra, convirtiendo así el dímero $F(ab)'_2$ en un monómero Fab'. El monómero Fab' es esencialmente Fab con parte de la región de bisagra (véase "Fundamental Immunology" (Paul ed., 3ª ed. 1993). Aunque se definen diversos fragmentos de anticuerpo en términos de la digestión de un anticuerpo intacto, un experto apreciará que dichos fragmentos pueden sintetizarse *de novo* químicamente o usando metodología de ADN recombinante. Por tanto, el término anticuerpo, como se usa en la presente memoria, incluye también fragmentos de anticuerpo producidos mediante la modificación de anticuerpos enteros o aquellos sintetizados *de novo* usando metodologías de ADN recombinante (por ejemplo F_v monocatenario) o aquellos identificados usando colecciones de expresión en fago (véase, por ejemplo, McCafferty *et al.*, Nature 348: 552-554 (1990)).

Para la preparación de anticuerpos monoclonales o policlonales, puede usarse cualquier técnica conocida en la materia (véanse, por ejemplo, Kohler y Milstein, Nature 256: 495-497 (1975); Kozbor *et al.*, Immunology Today 4: 72 (1983); Cole *et al.*, pág. 77-96 en "Monoclonal Antibodies and Cancer Therapy" (1985)). Las técnicas para la producción de anticuerpos monocatenarios (patente de EE.UU. 4.946.778) pueden adaptarse para producir anticuerpos contra polipéptidos de esta invención. También pueden usarse ratones transgénicos, u otros organismos tales como otros mamíferos, para expresar anticuerpos humanizados. Como alternativa, puede usarse la tecnología de expresión en fago para identificar anticuerpos y fragmentos de Fab heteroméricos que se unen específicamente a antígenos seleccionados (véanse, por ejemplo, McCafferty *et al.*, Nature 348: 552-554 (1990); Marks *et al.*, Biotechnology 10: 779-783 (1992)).

La frase "se une específica (o selectivamente) a" un anticuerpo o "es específica (o selectivamente) inmunoreactivo con", cuando se designa una proteína o péptido, designa una reacción de unión que es indicativa de la presencia de la proteína en una población heterogénea de proteínas y otros productos biológicos. Por tanto, en las condiciones de inmunoensayo designadas, los anticuerpos especificados se unen a una proteína particular al menos a dos veces el fondo y no se unen sustancialmente en una cantidad significativa a otras proteínas presentes en la muestra. La unión específica a un anticuerpo en dichas condiciones puede requerir un anticuerpo que se selecciona por su especificidad por una proteína particular. Por ejemplo, los anticuerpos policlonales creados contra proteínas de fusión pueden seleccionarse para obtener sólo aquellos anticuerpos policlonales que sean específicamente inmunoreactivos con la proteína de fusión y no con componentes individuales de las proteínas de fusión. Esta selección puede conseguirse quitando los anticuerpos que reaccionan de forma cruzada con los antígenos individuales. Pueden usarse una variedad de formatos de inmunoensayo para seleccionar anticuerpos específicamente inmunoreactivos con una proteína particular. Por ejemplo, se usan rutinariamente inmunoensayos ELISA en fase sólida para seleccionar anticuerpos específicamente inmunoreactivos con una proteína particular (véase, por ejemplo, Harlow y Lane, "Antibodies, A Laboratory Manual" (1988), para la descripción de formatos de inmunoensayo y condiciones que pueden usarse para determinar la inmunoreactividad específica). Típicamente, una reacción específica o selectiva será al menos dos veces la señal o ruido de fondo y más típicamente más de 10 a 100 veces el fondo.

Los polinucleótidos pueden comprender una secuencia nativa (concretamente, una secuencia endógena que codifica un antígeno individual o una porción del mismo) o pueden comprender una variante de dicha secuencia. Las variantes polinucleotídicas pueden contener una o más sustituciones, adiciones, deleciones y/o inserciones de tal modo que la actividad biológica del polipéptido de fusión codificado no se reduzca respecto a un polipéptido de fusión que comprende antígenos nativos. Las variantes exhiben preferiblemente al menos aproximadamente un 70% de identidad, más preferiblemente al menos aproximadamente un 80% de identidad y lo más preferiblemente al menos aproximadamente un 90% de identidad con una secuencia polinucleotídica que codifica un polipéptido nativo o una porción del mismo.

Los términos "idéntico" o "identidad" porcentual, en el contexto de dos o más ácidos nucleicos o secuencias polipeptídicas, designan dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje especificado de residuos aminoacídicos o nucleótidos que son iguales (concretamente, 70% de identidad, opcionalmente 75%, 80%, 85%, 90% ó 95% de identidad en una región especificada) cuando se comparan y alinean para correspondencia máxima en una ventana de comparación, o región designada, medida usando uno de los algoritmos de comparación de secuencia siguientes o mediante alineamiento manual e inspección visual. Dichas

secuencias se dice que son “sustancialmente idénticas”. Esta definición designa también el complemento de una secuencia de ensayo. Opcionalmente, existe identidad en una región que es de al menos aproximadamente 25 a aproximadamente 50 aminoácidos o nucleótidos de longitud. Opcionalmente, existe identidad en una región que es al menos de aproximadamente 25 a aproximadamente 50 aminoácidos o nucleótidos de longitud, u opcionalmente en una región que es de 75-100 aminoácidos o nucleótidos de longitud.

Para comparación de secuencia, típicamente una secuencia actúa como secuencia de referencia, con la que se comparan las secuencias de ensayo. Cuando se usa un algoritmo de comparación de secuencia, se introducen las secuencias de ensayo y referencia en un ordenador, se diseñan las coordenadas de subsecuencia, si es necesario, y se diseñan los parámetros del programa de algoritmo de secuencia. Pueden usarse los parámetros de programa por defecto, o pueden diseñarse parámetros alternativos. El algoritmo de comparación de secuencia calcula entonces las identidades de secuencia porcentuales para las secuencias de ensayo respecto a la secuencia de referencia, basándose en los parámetros del programa.

Una “ventana de comparación”, como se usa en la presente memoria, incluye la referencia a un segmento de una cualquiera de la serie de posiciones contiguas seleccionadas del grupo constituido por 25 a 500, habitualmente de aproximadamente 50 a aproximadamente 200, más habitualmente de aproximadamente 100 a aproximadamente 150, en las que una secuencia puede compararse con una secuencia de referencia del mismo número de posiciones contiguas después de alinear óptimamente las dos secuencias. Los métodos de alineamiento de secuencias para comparación son bien conocidos en la materia. El alineamiento óptimo de secuencias para comparación puede realizarse, por ejemplo, mediante el algoritmo de homología local de Smith y Waterman, *Adv. Appl. Math.* 2: 482 (1981), por el algoritmo de alineamiento de homología de Needleman y Wunsch, *J. Mol. Biol.* 48: 443 (1970), mediante el método de búsqueda de similitud de Pearson y Lipman, *Proc. Nat'l. Acad. Sci. USA* 85: 2444 (1988), mediante implementaciones informatizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), o mediante alineamiento manual e inspección visual (véase, por ejemplo, “Current Protocols in Molecular Biology” (Ausubel *et al.*, ed. suplemento de 1995)).

Es un ejemplo de algoritmo útil PILEUP. El PILEUP crea un alineamiento de secuencia múltiple a partir de un grupo de secuencias relacionadas usando alineamientos pareados progresivos para mostrar la relación y la identidad de secuencia porcentual. Representa también un árbol o dendograma que muestra las relaciones de agrupamiento usadas para crear el alineamiento. El PILEUP usa una simplificación del método de alineamiento progresivo de Feng y Doolittle, *J. Mol. Evol.* 35: 351-360 (1987). El método usado es similar al método descrito por Higgins y Sharp, *CABIOS* 5: 151-153 (1989). El programa puede alinear hasta 300 secuencias, cada una de una longitud máxima de 5.000 nucleótidos o aminoácidos. El procedimiento de alineamiento múltiple empieza con el alineamiento pareado de las dos secuencias más similares, produciendo un agrupamiento de dos secuencias alineadas. Este agrupamiento se alinea entonces con la siguiente secuencia o agrupamiento de secuencias alineadas más relacionado. Se alinean dos agrupamientos de secuencias mediante una sencilla extensión del alineamiento pareado de dos secuencias individuales. Se consigue el alineamiento final mediante una serie de alineamientos pareados progresivos. El programa se ejecuta diseñando secuencias específicas y sus coordenadas aminoacídicas o nucleotídicas para regiones de comparación de secuencia y diseñando los parámetros del programa. Usando PILEUP, se compara una secuencia de referencia con otras secuencias de ensayo para determinar la relación de identidad de secuencia porcentual usando los siguientes parámetros: ponderación del hueco por defecto (3,00), ponderación de la longitud del hueco por defecto (0,10) y huecos finales ponderados. PILEUP puede obtenerse en el paquete de software de análisis de secuencia GCG, por ejemplo, la versión 7.0 (Devereaux *et al.*, *Nuc. Acids Res.* 12: 387-395 (1984)).

Son otro ejemplo de algoritmo que es adecuado para determinar la identidad de secuencia porcentual y la similitud de secuencia, los algoritmos BLAST y BLAST 2.0, que se describen en Altschul *et al.*, *Nuc. Acids Res.* 25: 3389-3402 (1977) y Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990), respectivamente. El software para efectuar análisis BLAST está públicamente disponible en el National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo implica en primer lugar identificar pares de secuencia de alta puntuación (HSP) identificando palabras cortas de longitud *W* en la secuencia de consulta, que puede coincidir con o satisfacer alguna puntuación umbral de valor positivo *T* cuando se alinea con una palabra de la misma longitud en una secuencia de base de datos. *T* se designa como el umbral de puntuación de palabra vecina (Altschul *et al.*, *supra*). Estos aciertos de palabra vecina iniciales actúan como semillas para iniciar búsquedas para encontrar HSP más largos que las contengan. Los aciertos de palabra se extienden en ambas direcciones a lo largo de cada secuencia en la medida en que la puntuación de alineamiento acumulativo pueda aumentarse. Las puntuaciones acumulativas se calculan usando, para secuencias nucleotídicas, los parámetros *M* (puntuación de recompensa para un par de residuos coincidentes; siempre > 0) y *N* (puntuación de penalización para residuos no coincidentes; siempre < 0). Para las secuencias aminoacídicas, se usa una matriz de puntuación para calcular la puntuación acumulativa. La extensión de los aciertos de palabra en cada dirección se detiene cuando: la puntuación de alineamiento acumulativo cae por debajo de la cantidad *X* desde su valor máximo conseguido; o se alcanza el final de cualquier secuencia. Los parámetros *W*, *T* y *X* del algoritmo BLAST determinan la sensibilidad y velocidad del alineamiento. El programa BLASTN (para secuencias nucleotídicas) usa por defecto una longitud de palabra (*W*) de 11, una expectativa (*E*) de 10, *M*=5, *N*= -4 y una comparación de ambas cadenas. Para secuencias aminoacídicas, el programa BLASTP usa por defecto una longitud de palabra de 3, una expectativa (*E*) de 10 y la matriz de

puntuación BLOSUM62 (véase Henikoff y Henikoff, Proc. Natl. Acad. Sci. USA 89: 10915 (1989)), alineamientos (B) de 50, expectativa (E) de 10, M= 5, N= -4 y una comparación de ambas cadenas.

El algoritmo BLAST efectúa también un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul, Proc. Nat'l. Acad. Sci. USA 90: 5873-5787 (1993)). Es una medida de similitud proporcionada por el algoritmo BLAST la suma menor de probabilidades (P(N)), que proporciona una indicación de la probabilidad de que una coincidencia entre dos secuencias nucleotídicas o aminoácídicas ocurra por casualidad. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la suma menor de probabilidades en una comparación del ácido nucleico de ensayo con el ácido nucleico de referencia es menor de aproximadamente 0,2, más preferiblemente menor de aproximadamente 0,01 y lo más preferiblemente menor de aproximadamente 0,001.

10 Composiciones polinucleotídicas

Como se usan en la presente memoria, los términos "segmento de ADN" y "polinucleótido" designan una molécula de ADN que se ha aislado exenta del ADN genómico total de una especie particular. Por lo tanto, un segmento de ADN que codifica un polipéptido designa un segmento de ADN que contiene una o más secuencias de codificación aunque está sustancialmente aislado, o purificado exento, del ADN genómico total de la especie de la que se obtiene el segmento de ADN. Se incluyen dentro de los términos "segmento de ADN" y "polinucleótido" los segmentos de ADN y fragmentos menores de dichos segmentos, y también vectores recombinantes, incluyendo, por ejemplo, plásmidos, cósmidos, fagémidos, fagos, virus y similares.

Como se entenderá por los expertos en la materia, los segmentos de ADN de esta invención pueden incluir secuencias genómicas, secuencias extragenómicas y codificadas por plásmido y segmentos génicos modificados por ingeniería genética menores que expresan, o pueden adaptarse para expresar, proteínas, polipéptidos, péptidos y similares. Dichos segmentos pueden ser aislados naturalmente o modificados sintéticamente por la mano del hombre.

Los términos "aislado", "purificado" o "biológicamente puro" designan por lo tanto material que está sustancial o esencialmente exento de componentes que normalmente lo acompañan cuando se encuentra en su estado nativo. Por supuesto, esto designa el fragmento de ADN como se aísla originalmente, y no excluye otras proteínas aisladas, genes o regiones de codificación añadidos después a la composición por la mano del hombre. La pureza y homogeneidad se determinan típicamente usando técnicas de química analítica tales como electroforesis en gel de poliacrilamida o cromatografía líquida de alta resolución. Se purifica sustancialmente una proteína que es la especie predominante presente en una preparación. Se separa un ácido nucleico aislado de los otros marcos abiertos de lectura que flanquean el gen y codifican proteínas distintas del gen.

Como se reconocerá por el experto, los polinucleótidos pueden ser monocatenarios (de codificación o no codificación) o bicatenarios, y pueden ser moléculas de ADN (genómico, ADNc o sintético) o ARN. Las moléculas de ARN incluyen moléculas de HnRNA, que contienen intrones y corresponden a una molécula de ADN de manera unívoca, y moléculas de ARNm que no contienen intrones. Las secuencias de codificación o no codificación adicionales pueden estar presentes, pero no es necesario, en un polinucleótido de la presente invención, y el polinucleótido puede estar ligado, pero no es necesario, a otras moléculas y/o materiales de soporte.

Los polinucleótidos pueden comprender una secuencia nativa (concretamente, una secuencia endógena que codifica un antígeno de *Mycobacterium* o una porción del mismo) o pueden comprender una variante o un equivalente funcional biológico o antigénico de dicha secuencia. Las variantes polinucleotídicas pueden contener una o más sustituciones, adiciones, deleciones y/o inserciones, como se describe adicionalmente a continuación, preferiblemente de tal modo que la inmunogenicidad del polipéptido codificado no se reduzca respecto a una proteína nativa. El efecto sobre la inmunogenicidad del polipéptido codificado puede evaluarse generalmente como se describe en la presente memoria. El término "variantes" comprende también genes homólogos de origen xenogénico.

Los polinucleótidos de la presente invención, o fragmentos de los mismos, independientemente de la longitud de la secuencia de codificación misma, pueden combinarse con otras secuencias de ADN tales como promotores, señales de poliadenilación, sitios de enzima de restricción adicionales, sitios de clonación múltiple, otros segmentos de codificación y similares, de tal modo que su longitud global puede variar considerablemente. Por lo tanto, se contempla que puede emplearse un fragmento de ácido nucleico de casi cualquier longitud, estando limitada preferiblemente la longitud total por la facilidad de preparación y uso en el protocolo de ADN recombinante pretendido. Por ejemplo, se contempla que son útiles en muchas implementaciones de esta invención segmentos de ADN ilustrativos con longitudes totales de aproximadamente 10.000, aproximadamente 5.000, aproximadamente 3.000, aproximadamente 2.000, aproximadamente 1.000, aproximadamente 500, aproximadamente 200, aproximadamente 100, aproximadamente 50 pares de bases de longitud y similares (incluyendo todas las longitudes intermedias).

Además, se apreciará por los expertos en la materia que, como resultado de la degeneración del código genético, hay muchas secuencias nucleotídicas que codifican un polipéptido como se describe en la presente memoria. Algunos de estos polinucleótidos portan una homología mínima con la secuencia nucleotídica de cualquier gen

nativo. No obstante, los polinucleótidos que varían debido a diferencias en el uso de codón se contemplan específicamente por la presente invención, por ejemplo polinucleótidos que están optimizados para selección de codón humana y/o de primate. Además, los alelos de los genes que comprenden las secuencias polinucleotídicas proporcionadas en la presente memoria están dentro del alcance de la presente invención. Los alelos son genes endógenos que se alteran como resultado de una o más mutaciones, tales como deleciones, adiciones y/o sustituciones de nucleótidos. El ARNm y proteína resultantes pueden tener, pero no es necesario, una estructura o función alterada. Los alelos pueden identificarse usando técnicas estándar (tales como hibridación, amplificación y/o comparación de secuencia de base de datos).

Identificación y caracterización de polinucleótidos

Los polinucleótidos pueden identificarse, prepararse y/o manipularse usando cualquiera de una variedad de técnicas bien establecidas. Como alternativa, los polinucleótidos pueden amplificarse a partir de ADNc preparado a partir de células que expresan las proteínas descritas en la presente memoria, tales como células de *M. tuberculosis*. Dichos polinucleótidos pueden amplificarse mediante reacción en cadena de la polimerasa (PCR). Para este enfoque, pueden diseñarse cebadores específicos de secuencia basándose en las secuencias proporcionadas en la presente memoria, y pueden adquirirse o sintetizarse.

Puede usarse una porción amplificada de un polinucleótido de la presente invención para aislar un gen completo de una colección adecuada (por ejemplo, una colección de ADNc de *M. tuberculosis*) usando técnicas bien conocidas. En dichas técnicas, se examina una colección (de ADNc o genómico) usando una o más sondas o cebadores polinucleotídicos adecuados para amplificación. Preferiblemente, se selecciona por tamaño una colección para incluir moléculas grandes. Las colecciones cebadas aleatoriamente pueden preferirse también para identificar regiones en 5' y en dirección 5' de genes. Se prefieren las colecciones genómicas para obtener intrones y extender las secuencias 5'.

Para técnicas de hibridación, puede marcarse una secuencia parcial (por ejemplo, traslado de muesca o marcaje terminal con ³²P) usando técnicas bien conocidas. Se examina entonces generalmente una colección bacteriana o de bacteriófago hibridando filtros que contienen colonias bacterianas desnaturalizadas (o céspedes que contienen placas de fago) con la sonda marcada (véase Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual" (1989)). Se seleccionan y expanden las colonias o placas de hibridación y se aísla el ADN para análisis posterior. Los clones de ADNc pueden analizarse para determinar la cantidad de secuencia adicional, por ejemplo, mediante PCR usando un cebador de la secuencia parcial y un cebador del vector. Pueden generarse mapas de restricción y secuencias parciales para identificar uno o más clones superpuestos. Puede determinarse entonces la secuencia completa usando técnicas estándar, que pueden implicar generar una serie de clones de deleción. Las secuencias superpuestas resultantes pueden ensamblarse entonces en una sola secuencia contigua. Puede generarse una molécula de ADNc completa ligando fragmentos adecuados usando técnicas bien conocidas.

Como alternativa, existen numerosas técnicas de amplificación para obtener una secuencia de codificación completa a partir de una secuencia de ADNc parcial. En dichas técnicas, la amplificación se efectúa generalmente mediante PCR. Puede usarse cualquiera de una variedad de kits comercialmente disponibles para efectuar la etapa de amplificación. Los cebadores pueden diseñarse usando, por ejemplo, software bien conocido en la materia. Los cebadores son preferiblemente de 22-30 nucleótidos de longitud, tienen un contenido de GC de al menos un 50% y se asocian con la secuencia diana a temperaturas de aproximadamente 68 a 72°C. La región amplificada puede secuenciarse como se describe anteriormente y ensamblarse las secuencias superpuestas en una secuencia contigua.

Una de dichas técnicas de amplificación es la PCR inversa (véase Triglia *et al.*, Nucl. Acids Res. 16: 8186 (1988)), que usa enzimas de restricción para generar un fragmento en la región conocida del gen. El fragmento se circulariza entonces mediante ligadura intramolecular y se usa como molde para PCR con cebadores divergentes derivados de la región conocida. En un enfoque alternativo, las secuencias adyacentes a una secuencia parcial pueden recuperarse mediante amplificación con un cebador de una secuencia ligadora y un cebador específico de una región conocida. Las secuencias amplificadas se someten típicamente a una segunda ronda de amplificación con el mismo cebador ligador y un segundo cebador específico de la región conocida. Se describe una variación de este procedimiento, que emplea dos cebadores que inician la extensión en direcciones opuestas a partir de la secuencia conocida, en el documento WO 96/38591. Otra de dichas técnicas es conocida como "amplificación rápida de los extremos de ADNc" o RACE. Esta técnica implica el uso de un cebador interno y un cebador externo, que hibridan con una región poliA o una secuencia de vector, para identificar secuencias que están 5' y 3' de una secuencia conocida. Las técnicas adicionales incluyen PCR de captura (Lagerstrom *et al.*, PCR Methods Applic. 1: 111-19 (1991)) y rastreo por PCR (Parker *et al.*, Nucl. Acids. Res. 19: 3055-60 (1991)). Pueden emplearse también otros métodos que emplean amplificación para obtener una secuencia de ADNc completa.

En ciertos casos, es posible obtener una secuencia de ADNc completa mediante el análisis de secuencias proporcionadas en una base de datos de marcadores de secuencia expresada (EST), tales como las disponibles en GenBank. Las búsquedas de EST superpuestos puede efectuarse generalmente usando programas bien conocidos (por ejemplo, búsquedas en NCBI BLAST), y dichos EST pueden usarse para generar una secuencia completa

contigua. Las secuencias de ADN completas pueden obtenerse también mediante el análisis de fragmentos genómicos.

Expresión polinucleotídica en células hospedadoras

5 En otras realizaciones de la invención, pueden usarse secuencias polinucleotídicas o fragmentos de las mismas que codifican polipéptidos de la invención, o proteínas de fusión, en moléculas de ADN recombinantes para dirigir la expresión de un polipéptido en células hospedadoras apropiadas. Debido a la degeneración inherente del código genético, pueden producirse otras secuencias de ADN que codifican sustancialmente la misma secuencia aminoacídica o una funcionalmente equivalente y estas secuencias pueden usarse para clonar y expresar un polipéptido dado.

10 Como se entenderá por los expertos en la materia, puede ser ventajoso en algunos casos producir secuencias nucleotídicas que codifican un polipéptido que posean codones de origen no natural. Por ejemplo, pueden seleccionarse codones preferidos por un hospedador procariótico o eucariótico particular para aumentar la velocidad de expresión de proteína o para producir un transcrito de ARN recombinante que tenga propiedades deseables, tales como una semivida que sea más larga que la de un transcrito generado a partir de la secuencia de origen natural.

15 Además, las secuencias polinucleotídicas de la presente invención pueden modificarse por ingeniería genética usando métodos conocidos generalmente en la materia para alterar las secuencias de codificación de polipéptido por una variedad de razones, incluyendo pero sin limitación, alteraciones que modifican la clonación, procesamiento y/o expresión del producto génico. Por ejemplo, la transposición de ADN mediante fragmentación aleatoria y reensamblado por PCR de fragmentos génicos y oligonucleótidos sintéticos puede usarse para modificar por ingeniería genética las secuencias nucleotídicas. Además, la mutagénesis dirigida a sitio puede usarse para insertar nuevos sitios de restricción, alterar los patrones de glucosilación, cambiar la preferencia de codón, producir variantes de corte y empalme o introducir mutaciones y demás.

20 En otra realización de la invención, las secuencias de ácido nucleico natural, modificado o recombinante pueden estar ligadas a una secuencia heteróloga para codificar una proteína de fusión. Por ejemplo, para cribar en colecciones peptídicas los inhibidores de la actividad polipeptídica, puede ser útil codificar una proteína quimérica que puede reconocerse por un anticuerpo comercialmente disponible. Una proteína de fusión puede estar modificada también por ingeniería genética para contener un sitio de escisión localizado entre la secuencia que codifica polipéptido y la secuencia de proteína heteróloga, de modo que el polipéptido pueda escindirse y purificarse del resto heterólogo.

25 Pueden sintetizarse secuencias que codifican un polipéptido deseado, total o parcialmente, usando métodos químicos bien conocidos en la materia (véase Caruthers, M. H. *et al.*, Nucl. Acids Res. Symp. Ser. pág. 215-223 (1980), Horn *et al.*, Nucl. Acids Res. Symp. Ser. pág. 225-232 (1980)). Como alternativa, la proteína misma puede producirse usando métodos químicos para sintetizar la secuencia aminoacídica de un polipéptido, o una porción del mismo. Por ejemplo, la síntesis peptídica puede efectuarse usando diversas técnicas en fase sólida (Roberge *et al.*, Science 269: 202-204 (1995)) y puede conseguirse la síntesis automatizada, por ejemplo, usando el sintetizador peptídico ABI 431A (Perkin Elmer, Palo Alto, CA).

30 Puede purificarse sustancialmente un péptido recién sintetizado mediante cromatografía líquida de alta resolución preparativa (por ejemplo, Creighton, "Proteins, Structures and Molecular Principles" (1983)) u otras técnicas comparables disponibles en la materia. La composición de los péptidos sintéticos puede confirmarse mediante análisis o secuenciación de aminoácidos (por ejemplo, el procedimiento de degradación de Edman). Adicionalmente, la secuencia aminoacídica de un polipéptido, o cualquier parte del mismo, puede alterarse durante la síntesis directa y/o combinarse usando métodos químicos con secuencias de otras proteínas, o cualquier parte de las mismas, para producir un polipéptido variante.

35 Para expresar un polipéptido deseado, las secuencias polipeptídicas que codifican el polipéptido, o equivalentes funcionales, pueden insertarse en un vector de expresión apropiado, concretamente, un vector que contiene los elementos necesarios para la transcripción y traducción de la secuencia de codificación insertada. Pueden usarse métodos que son bien conocidos por los expertos en la materia para construir vectores de expresión que contienen secuencias que codifican un polipéptido de interés y elementos de control transcripcional y traduccional apropiados. Estos métodos incluyen técnicas de ADN recombinante *in vitro*, técnicas sintéticas y recombinación genética *in vivo*. Dichas técnicas se describen en Sambrook *et al.*, "Molecular Cloning, A Laboratory Manual" (1989) y Ausubel *et al.*, "Current Protocols in Molecular Biology" (1989).

40 Puede utilizarse una variedad de sistemas de vector de expresión/hospedador que contienen y expresan secuencias polinucleotídicas. Estos incluyen, pero sin limitación, microorganismos tales como bacterias transformadas con vectores de expresión de ADN de bacteriófago, plásmido o cósmido recombinante; levadura transformada con vectores de expresión de levadura; sistemas de células de insecto infectados con vectores de expresión víricos (por ejemplo, baculovirus); sistemas de células de planta transformadas con vectores de expresión víricos (por ejemplo,

virus del mosaico de la coliflor, CaMV; virus del mosaico del tabaco, TMV) o con vectores de expresión bacterianos (por ejemplo, plásmidos Ti o pBR322); o sistemas de células animales.

Los “elementos de control” o “secuencias reguladoras” presentes en un vector de expresión son aquellas regiones no traducidas del vector: potenciadores, promotores, regiones 5' y 3' no traducidas, que interaccionan con proteínas celulares hospedadoras para llevar a cabo la transcripción y traducción. Dichos elementos pueden variar en su potencia y especificidad. Dependiendo del sistema de vector y hospedador utilizado, puede usarse cualquier número de elementos de transcripción y traducción adecuados, incluyendo promotores constitutivos e inducibles. Por ejemplo, cuando se clona en sistemas bacterianos, pueden usarse promotores inducibles tales como el promotor lacZ híbrido del fagémido PBLUESCRIPT (Stratagene, La Jolla, Calif.) o el plásmido PSPORT1 (Gibco BRL, Gaithersburg, MD) y similares. En sistemas de células de mamífero, se prefieren generalmente promotores de genes de mamífero o de virus de mamífero. Si es necesario generar una estirpe celular que contenga múltiples copias de la secuencia que codifica un polipéptido, pueden usarse ventajosamente vectores basados en SV40 o EBV con un marcador seleccionable apropiado.

En sistemas bacterianos, puede seleccionarse una serie de vectores de expresión dependiendo del uso pretendido del polipéptido expresado. Por ejemplo, cuando se necesitan grandes cantidades, por ejemplo para la inducción de anticuerpos, pueden usarse vectores que dirigen un alto nivel de expresión de proteínas de fusión que se purifican fácilmente. Dichos vectores incluyen, pero sin limitación, los vectores de clonación y expresión multifuncionales de *E. coli* tales como BLUESCRIPT (Stratagene), en los que la secuencia que codifica el polipéptido de interés puede estar ligada al vector en fase con secuencias de Met aminoterminal y los posteriores 7 residuos de β -galactosidasa, de modo que se produzca una proteína híbrida; vectores pIN (Van Heeke y Schuster, J. Biol. Chem. 264: 5503-5509 (1989)); y similares. Los vectores pGEX (Promega, Madison, Wis.) pueden usarse también para expresar polipéptidos ajenos en forma de proteínas de fusión con glutatión S-transferasa (GST). En general, dichas proteínas de fusión son solubles y pueden purificarse fácilmente a partir de células lisadas mediante adsorción sobre perlas de glutatión-agarosa seguida de elución en presencia de glutatión libre. Las proteínas preparadas en dichos sistemas pueden diseñarse para incluir sitios de escisión de heparina, trombina o factor XA de modo que el polipéptido clonado de interés pueda liberarse del resto GST a voluntad.

En la levadura *Saccharomyces cerevisiae*, pueden usarse una serie de vectores que contienen promotores constitutivos o inducibles tales como factor alfa, alcohol oxidasa y PGH. Para revisiones, véase Ausubel *et al.* (supra) y Grant *et al.*, Methods Enzymol. 153: 516-544 (1987).

En los casos en que se usan vectores de expresión de planta, la expresión de secuencias que codifican polipéptidos puede activarse por cualquiera de una serie de promotores. Por ejemplo, los promotores víricos tales como promotores 35S y 19S de CaMV pueden usarse solos o en combinación con la secuencia líder omega de TMV (Takamatsu, EMBO J. 6: 307-311 (1987)). Como alternativa, pueden usarse promotores de planta tales como la subunidad pequeña de RUBISCO o promotores de choque térmico (Coruzzi *et al.*, EMBO J. 3: 1671-1680 (1984); Broglie *et al.*, Science 224: 838-843 (1984) y Winter *et al.*, Results Probl. Cell Differ. 17: 85-105 (1991)). Estos constructos pueden introducirse en células de planta mediante transformación de ADN directa o transfección mediada por patógeno. Dichas técnicas se describen en una serie de revisiones generalmente disponibles (véase, por ejemplo, Hobbs en “McGraw Hill Yearbook of Science and Technology” pág. 191-196 (1992)).

Puede usarse también un sistema de insecto para expresar un polipéptido de interés. Por ejemplo, en uno de dichos sistemas, se usa el virus de la poliedrosis nuclear de *Autographa californica* (AcNPV) como vector para expresar genes ajenos en células de *Spodoptera frugiperda* o en *Trichoplusia larvae*. Las secuencias que codifican el polipéptido pueden clonarse en una región no esencial del virus, tal como el gen de poliedrina, y disponerse bajo el control del promotor de poliedrina. La inserción exitosa de la secuencia de codificación de polipéptido volverá al gen de poliedrina inactivo y producirá un virus recombinante que carece de la proteína de cubierta. Los virus recombinantes pueden usarse entonces para infectar, por ejemplo, células de *S. frugiperda* o *Trichoplusia larvae* en las que pueden expresarse el polipéptido de interés (Engelhard *et al.*, Proc. Natl. Acad. Sci. U.S.A. 91: 3224-3227 (1994)).

En células hospedadoras de mamífero, son generalmente viables una serie de sistemas de expresión basados en virus. Por ejemplo, en los casos en que se usa un adenovirus como vector de expresión, las secuencias que codifican un polipéptido de interés pueden ligarse con un complejo de transcripción/traducción adenovírico consistente en el promotor tardío y la secuencia líder tripartita. La inserción en una región E1 o E3 no esencial del genoma vírico puede usarse para obtener un virus viable que sea capaz de expresar el polipéptido en células hospedadoras infectadas (Logan y Shenk, Proc. Natl. Acad. Sci. U.S.A. 81: 3655-3659 (1984)). Además, pueden usarse potenciadores de la transcripción tales como potenciador del virus de sarcoma de Rous (RSV) para aumentar la expresión en células hospedadoras de mamífero.

Pueden usarse también señales de iniciación específicas para conseguir una traducción más eficaz de secuencias que codifican un polipéptido de interés. Dichas señales incluyen el codón de iniciación ATG y secuencias adyacentes. En los casos en que las secuencias que codifican el polipéptido, su codón de iniciación y secuencias en dirección 5' estén insertadas en el vector de expresión apropiado, pueden no ser necesarias señales de control transcripcional o traduccional adicionales. Sin embargo, en los casos en que se inserte sólo la secuencia de

codificación, o una porción de la misma, deben proporcionarse señales de control traduccional exógenas incluyendo el codón de iniciación ATG. Además, el codón de iniciación debe estar en el marco de lectura correcto para asegurar la traducción del inserto completo. Los elementos traduccionales exógenos y codones de iniciación pueden ser de diversos orígenes, tanto naturales como sintéticos. La eficacia de la expresión puede potenciarse mediante la inclusión de potenciadores que son apropiados para el sistema celular particular que se usa, tales como los descritos en la bibliografía (Scharf *et al.*, Results Probl. Cell Differ. 20: 125-162 (1994)).

Además, puede elegirse una cepa de célula hospedadora por su capacidad de modular la expresión de las secuencias insertadas o de procesar la proteína expresada de la forma deseada. Dichas modificaciones del polipéptido incluyen, pero sin limitación, acetilación, carboxilación, glucosilación, fosforilación, lipidación y acilación. El procesamiento postraduccional que escinde una forma "prepro" de la proteína puede usarse también para facilitar una correcta inserción, plegamiento y/o función. Pueden elegirse diferentes células hospedadoras tales como CHO, HeLa, MDCK, HEK293 y WI38, que tienen una maquinaria celular específica y mecanismos característicos para dichas actividades postraduccionales, para asegurar la correcta modificación y procesamiento de la proteína ajena.

Para la producción a largo plazo con alto rendimiento de proteínas recombinantes, se prefiere generalmente la expresión estable. Por ejemplo, las estirpes celulares que expresan establemente un polinucleótido de interés pueden transformarse usando vectores de expresión que pueden contener orígenes de replicación víricos y/o elementos de expresión endógenos y un gen marcador seleccionable en el mismo o en un vector separado. Después de la introducción del vector, pueden dejarse crecer las células durante 1-2 días en un medio enriquecido antes de cambiarse a medio selectivo. El fin del marcador seleccionable es conferir resistencia a la selección, y su presencia permite el crecimiento y recuperación de células que expresan exitosamente las secuencias introducidas. Los clones resistentes de células transformadas establemente pueden proliferar usando técnicas de cultivo de tejido apropiadas para el tipo celular.

Puede usarse cualquier número de sistemas de selección para recuperar estirpes celulares transformadas. Estos incluyen, pero sin limitación, genes de timidina cinasa de herpesvirus simple (Wigler *et al.*, Cell 11: 223-32 (1977)) y adenina fosforribosiltransferasa (Lowy *et al.*, Cell 22: 817-23 (1990)) que pueden emplearse en células tk.sup. o aprt.sup., respectivamente. También pueden usarse la resistencia a antimetabolito, antibiótico o herbicida como base para la selección, por ejemplo, dhfr que confiere resistencia a metotrexato (Wigler *et al.*, Proc. Natl. Acad. Sci. U.S.A. 77: 3567-70 (1980)); npt, que confiere resistencia a aminoglucósidos, neomicina y G-418 (Colbere-Garapin *et al.*, J. Mol. Biol. 150: 1-14 (1981)); y als o pat, que confieren resistencia a clorsulfurón y fosfotricina acetiltransferasa, respectivamente (Murry, *supra*). Se han descrito genes seleccionables adicionales, por ejemplo, trpB, que permite a las células utilizar indol en lugar de triptófano, o hisD, que permite a las células utilizar histinol en lugar de histidina (Hartman y Mulligan, Proc. Natl. Acad. Sci. U.S.A. 85: 8047-51 (1988)). Recientemente, el uso de marcadores visibles ha ganado popularidad con marcadores tales como antocianinas, β -glucuronidasa y su sustrato GUS y luciferasa y su sustrato luciferina, usados ampliamente no sólo para identificar transformantes, sino también para cuantificar la cantidad de expresión de proteína transitoria o estable atribuible a un sistema de vector específico (Rhodes *et al.*, Methods Mol. Biol. 55: 121-131 (1995)).

Aunque la presencia/ausencia de expresión del gen marcador sugiere que el gen de interés está también presente, su presencia y expresión pueden tener que confirmarse. Por ejemplo, si la secuencia que codifica un polipéptido se inserta en una secuencia de gen marcador, las células recombinantes que contienen secuencias pueden identificarse por la ausencia de función de gen marcador. Como alternativa, puede disponerse un gen marcador en serie con una secuencia de codificación de polipéptido bajo el control de un solo promotor. La expresión del gen marcador en respuesta a la inducción o selección indica habitualmente la expresión del gen en serie también.

Como alternativa, las células hospedadoras que contienen y expresan una secuencia polinucleotídica deseada pueden identificarse mediante una variedad de procedimientos conocidos por los expertos en la materia. Estos procedimientos incluyen, pero sin limitación, hibridaciones ADN-ADN o ADN-ARN y técnicas de bioensayo de proteína o inmunoensayo que incluyen tecnología basada en membrana, disolución o chip para la detección y/o cuantificación de ácido nucleico o proteína.

Son conocidas en la materia una variedad de protocolos para detectar y medir la expresión de productos codificados por polinucleótidos, usando anticuerpos policlonales o monoclonales específicos del producto. Los ejemplos incluyen ensayo de inmunosorción ligado a enzima (ELISA), radioinmunoensayo (RIA) y clasificación celular activada por fluorescencia (FACS). Puede preferirse un inmunoensayo basado en anticuerpo monoclonal de dos sitios que utiliza anticuerpos monoclonales reactivos con dos epítopos no interferentes en un polipéptido dado para algunas aplicaciones, pero puede emplearse también un ensayo de unión competitiva. Se describen estos y otros ensayos, entre otros lugares, en Hampton *et al.*, "Serological Methods, a Laboratory Manual" (1990) y Maddox *et al.*, J. Exp. Med. 158: 1211-1216 (1983).

Son conocidas una amplia variedad de marcadores y técnicas de conjugación por los expertos en la materia y pueden usarse en diversos ensayos de ácido nucleico y aminoácido. Los medios para producir sondas de hibridación o PCR marcadas para detectar secuencias relacionadas con polinucleótidos incluyen oligomarcaje, traslado de muesca, marcaje terminal o amplificación por PCR usando un nucleótido marcado. Como alternativa, las secuencias, o cualquier porción de las mismas, pueden clonarse en un vector para la producción de una sonda de

ARNm. Dichos vectores son conocidos en la materia, están comercialmente disponibles y pueden usarse para sintetizar muestras de ARN *in vitro* mediante la adición de una ARN polimerasa apropiada tal como T7, T3 o SP6 y nucleótidos marcados. Estos procedimientos pueden realizarse usando una variedad de kits comercialmente disponibles. Las moléculas marcadoras o indicadores adecuados que pueden usarse incluyen radionucleidos, enzimas, agentes fluorescentes, quimioluminiscentes o cromogénicos, así como sustratos, cofactores, inhibidores, partículas magnéticas y similares.

Las células hospedadoras transformadas con una secuencia polinucleotídica de interés pueden cultivarse en condiciones adecuadas para la expresión y recuperación de la proteína del cultivo celular. La proteína producida por una célula recombinante puede secretarse o estar contenida intracelularmente dependiendo de la secuencia y/o el vector usados. Como se entenderá por los expertos en la materia, los vectores de expresión que contienen polinucleótidos de la invención pueden diseñarse para contener secuencias señal que dirijan la secreción del polipéptido codificado a través de una membrana celular procariótica o eucariótica. Pueden usarse otras construcciones recombinantes para unir secuencias que codifican un polipéptido de interés con una secuencia nucleotídica que codifica un dominio polipeptídico que facilitará la purificación de proteínas solubles. Dichos dominios facilitadores de la purificación incluyen, pero sin limitación, péptidos quelantes de metal tales como módulos de histidina-triptófano que permiten la purificación sobre metales inmovilizados, dominios de proteína A que permiten la purificación sobre inmunoglobulina inmovilizada, y el dominio utilizado en el sistema de purificación por extensión/afinidad FLAGS (Immunex Corp., Seattle, Wash.). La inclusión de secuencias ligadoras escindibles tales como aquellas específicas de factor XA o enterocinasa (Invitrogen, San Diego, Calif.) entre el dominio de purificación y el polipéptido codificado puede usarse para facilitar la purificación. Uno de dichos vectores de expresión proporciona la expresión de una proteína de fusión que contiene un polipéptido de interés y un ácido nucleico que codifica 6 residuos de histidina que preceden a una tiorredoxina o sitio de escisión de enterocinasa. Los residuos de histidina facilitan la purificación en IMIAC (cromatografía de afinidad por ión metálico inmovilizado) como se describe en Porath *et al.*, Prot. Exp. Purif. 3: 263-281 (1992), mientras que el sitio de escisión de enterocinasa proporciona un medio para purificar el polipéptido deseado a partir de la proteína de fusión. Se proporciona una discusión de vectores que contienen proteínas de fusión en Kroll *et al.*, DNA Cell Biol. 12: 441-453 (1993).

Además de los métodos de producción recombinante de la invención, los polipéptidos de la invención y fragmentos de los mismos pueden producirse mediante síntesis peptídica directa usando técnicas en fase sólida (Merrifield, J. Am. Chem. Soc. 85: 2149-2154 (1963)). La síntesis de proteína puede efectuarse usando técnicas manuales o mediante automatización. La síntesis automatizada puede conseguirse, por ejemplo, usando el sintetizador peptídico Applied Biosystems 431A (Perkin Elmer). Como alternativa, pueden sintetizarse químicamente diversos fragmentos por separado y combinarse usando métodos químicos para producir la molécula completa.

Técnicas de suministro de polinucleótido *in vivo*

En realizaciones adicionales, se introducen constructos genéticos que comprenden uno o más de los polinucleótidos de la invención en células *in vivo*. Esto puede conseguirse usando cualquiera de una variedad de enfoques bien conocidos, varios de los cuales se exponen a continuación con fines de ilustración.

1. ADENOVIRUS

Uno de los métodos preferidos para el suministro *in vivo* de una o más secuencias de ácido nucleico implica el uso de un vector de expresión de adenovirus. "Vector de expresión de adenovirus" pretende incluir aquellos constructos que contienen secuencias de adenovirus suficientes para (a) apoyar el empaquetamiento del constructo y (b) expresar un polinucleótido que se ha clonado en el mismo en orientación de codificación o no codificación. Por supuesto, en el contexto de un constructo de no codificación, la expresión no requiere que el producto génico se sintetice.

El vector de expresión comprende una forma modificada por ingeniería genética de un adenovirus. El conocimiento de la organización genética del adenovirus, un virus de ADN bicatenario lineal de 36 kb, permite la sustitución de grandes trozos de ADN adenovírico por secuencias ajenas de hasta 7 kb (Grunhaus y Horwitz, 1992). En contraposición con los retrovirus, la infección adenovírica de células hospedadoras no da como resultado la integración cromosómica, porque el ADN adenovírico puede replicarse de manera episómica sin genotoxicidad potencial. También, los adenovirus son estructuralmente estables y no se ha detectado redistribución genómica después de una extensa amplificación. Los adenovirus pueden infectar virtualmente todas las células epiteliales independientemente de su etapa de ciclo celular. Hasta ahora, la infección adenovírica parece estar ligada sólo a enfermedad leve tal como enfermedad respiratoria aguda en seres humanos.

El adenovirus es particularmente adecuado para uso como vector de transferencia génica debido a su genoma de tamaño medio, facilidad de manipulación, alta titulación, amplio intervalo de células diana y alta infectividad. Ambos extremos del genoma vírico contienen 100-200 repeticiones invertidas de pares de bases (ITR) que son elementos en cis necesarios para la replicación y empaquetamiento de ADN vírico. Las regiones temprana (E) y tardía (L) del genoma contienen diferentes unidades de transcripción que están divididas por el inicio de replicación del ADN vírico. La región E1 (E1A y E1B) codifica proteínas responsables de la regulación de la transcripción del genoma vírico y unos pocos genes celulares. La expresión de la región E2 ((E2A y E2B) da como resultado la síntesis de

proteínas para la replicación del ADN vírico. Estas proteínas están implicadas en la replicación de ADN, la expresión de gen tardío y el bloqueo de células hospedadoras (Renan, 1990). Los productos de los genes tardíos, incluyendo la mayoría de las proteínas de cápsida vírica, se expresan sólo después de un procesamiento significativo de un solo transcrito primario procedente del promotor tardío principal (MLP). El MLP (localizado a 16,8 u.m.) es particularmente eficaz durante la fase tardía de la infección, y todos los ARNm procedentes de este promotor poseen una secuencia líder tripartita 5' (TPL) que los hace ARNm preferidos para traducción.

En un sistema actual, se generan adenovirus recombinantes a partir de recombinación homóloga entre un vector lanzadera y un vector provírico. Debido a la posible recombinación entre dos vectores províricos, pueden generarse adenovirus de tipo silvestre en este proceso. Por lo tanto, es crítico aislar un solo clon de virus de una placa individual y examinar su estructura genómica.

La generación y propagación de los vectores adenovíricos actuales, que son deficientes de replicación, depende de una estirpe celular auxiliar única, designada 293, que se transformó a partir de células de riñón embrionarias humanas mediante fragmentos de ADN de Ad5 y que expresa constitutivamente proteínas E1 (Graham *et al.*, 1977). Puesto que la región E3 es prescindible en el genoma de adenovirus (Jones y Shenk, 1978), los vectores adenovíricos actuales, con la ayuda de células 293, portan ADN ajeno en las regiones E1, D3 o ambas (Graham y Prevec, 1991). En la naturaleza, los adenovirus pueden empaquetar aproximadamente un 105% del genoma de tipo silvestre (Ghosh-Choudhury *et al.*, 1987), proporcionando capacidad para aproximadamente 2 kb extra de ADN. Combinado con las aproximadamente 5,5 kb de ADN que son reemplazables en las regiones E1 y E3, la capacidad máxima del vector adenovírico actual es menor de 7,5 kb, o aproximadamente un 15% de la longitud total del vector. Más de un 80% del genoma vírico del adenovirus permanece en la cadena principal del vector y es la fuente de citotoxicidad portada por vector. También la deficiencia de replicación del virus con E1 eliminado es incompleta. Por ejemplo, se ha observado pérdida de expresión génica vírica con los vectores actualmente disponibles a altas multiplicidades de infección (MOI) (Mulligan, 1993).

Las estirpes celulares auxiliares pueden derivar de células humanas tales como células de riñón embrionarias humanas, células musculares, células hematopoyéticas u otras células mesenquimáticas o epiteliales embrionarias humanas. Como alternativa, las células auxiliares pueden derivar de células de otras especies de mamífero que toleran adenovirus humanos. Dichas células incluyen, por ejemplo, células Vero u otras células mesenquimáticas o epiteliales embrionarias de mono. Como se indica anteriormente, la estirpe celular auxiliar actualmente preferida es 293.

Recientemente, Racher *et al.* (1995) dieron a conocer métodos mejorados para cultivar células 293 y propagar adenovirus. En un formato, se cultivan agregados de células naturales inoculando células individuales en matraces de centrifugación siliconizados de 1 litro (Techne, Cambridge, RU) que contienen 100-200 ml de medio. Después de agitar a 40 rpm, se estima la viabilidad celular con azul de tripano. En otro formato, se emplean microportadores Fibra-Cel (Bibby Sterlin, Stone, RU) (5 g/l) del modo siguiente. Se añade un inóculo celular, resuspendido en 5 ml de medio, al portador (50 ml) en un matraz Erlenmeyer de 250 ml y se deja estacionario, con agitación ocasional, durante 1 a 4 h. Se reemplaza entonces el medio por 50 ml de medio reciente y se inicia la agitación. Para la producción de virus, se dejan crecer las células hasta aproximadamente un 80% de confluencia, después de dicho tiempo se reemplaza el medio (hasta 25% del volumen final) y se añade adenovirus a una MOI de 0,05. Se dejan los cultivos estacionarios durante una noche, después de lo cual se aumenta el volumen al 100% y se comienza la agitación durante otras 72 h.

Aparte del requisito de que el vector adenovírico sea defectivo de replicación, o al menos condicionalmente defectivo, no se cree la naturaleza del vector adenovírico sea crucial para la práctica exitosa de la invención. El adenovirus puede ser cualquiera de los 42 serotipos o subgrupos A-F diferentes conocidos. El adenovirus de tipo 5 del subgrupo C es el material de partida preferido para obtener un vector adenovírico defectivo de replicación condicional para uso en la presente invención, puesto que el adenovirus de tipo 5 es un adenovirus humano sobre el que se conoce una gran cantidad de información bioquímica y genética, y se ha usado históricamente para la mayoría de construcciones que emplean adenovirus como vector.

Como se afirma anteriormente, el vector típico según la presente invención es defectivo de replicación y no tendrá una región E1 adenovírica. Por tanto, lo más conveniente será introducir el polinucleótido que codifica el gen de interés en la posición de la que se han eliminado las secuencias de codificación de E1. Sin embargo, la posición de la inserción del constructo en las secuencias adenovíricas no es crítica para la invención. El polinucleótido que codifica el gen de interés puede insertarse también en lugar de la región E3 eliminada en vectores de reemplazo de E3 como se describe por Karlsson *et al.* (1986) o en la región E4 donde una estirpe celular auxiliar o virus auxiliar complementa el defecto de E4.

El adenovirus es fácil de cultivar y manipular y exhibe un amplio intervalo hospedador *in vitro* e *in vivo*. Este grupo de virus puede obtenerse con altas titulaciones, por ejemplo, 10^9 - 10^{11} unidades de formación de placa por ml, y son altamente infecciosos. El ciclo vital del adenovirus no requiere la integración en el genoma de la célula hospedadora. Los genes ajenos suministrados por los vectores adenovíricos son episómicos, y por lo tanto, tienen una baja genotoxicidad para células hospedadoras. No se han reseñado efectos secundarios en estudios de vacunación con

adenovirus de tipo silvestre (Couch *et al.*, 1963; Top *et al.*, 1971), demostrando su seguridad y potencial terapéutico como vector de transferencia génica *in vivo*.

Los vectores adenovíricos se han usado en la expresión génica eucariótica (Levrero *et al.*, 1991; Gómez-Foix *et al.*, 1992) y el desarrollo de vacunas (Grunhaus y Horwitz, 1992; Graham y Prevec, 1992). Recientemente, estudios animales han sugerido que podrían usarse adenovirus recombinantes para terapia génica (Stratford-Perricaudet y Perricaudet, 1991; Stratford-Perricaudet *et al.*, 1990; Rich *et al.*, 1993). Los estudios para administrar adenovirus recombinantes a diferentes tejidos incluyen instilación traqueal (Rosenfeld *et al.*, 1991; Rosenfeld *et al.*, 1992), inyección muscular (Ragot *et al.*, 1993), inyecciones intravenosas periféricas (Herz y Gerard, 1993) e inoculación estereotáctica en el cerebro (Le Gal La Salle *et al.*, 1993).

2. RETROVIRUS

Los retrovirus son un grupo de virus de ARN monocatenarios caracterizados por la capacidad de convertir su ARN en ADN bicatenario en células infectadas mediante un proceso de transcripción inversa (Coffin, 1990). El ADN resultante se integra establemente entonces en cromosomas celulares en forma de provirus y dirige la síntesis de proteínas víricas. La integración da como resultado la retención de las secuencias génicas víricas en la célula receptora y sus descendientes. El genoma retrovítico contiene tres genes: gag, pol y env, que codifican proteínas de cápsida, enzima polimerasa y componentes de cubierta, respectivamente. Una secuencia encontrada en dirección 5' del gen gag contiene una señal de empaquetamiento del genoma en viriones. Están presentes dos secuencias repetidas terminales largas (LTR) en los extremos 5' y 3' del genoma vírico. Estas contienen secuencias promotoras fuertes y potenciadoras y son también necesarias para la integración en el genoma de la célula hospedadora (Coffin, 1990).

Para construir un vector retrovítico, se inserta un ácido nucleico que codifica una o más secuencias oligonucleotídicas o polinucleotídicas de interés en el genoma vírico en lugar de ciertas secuencias víricas para producir un virus que es defectivo de replicación. Para producir viriones, se construye una estirpe celular de empaquetamiento que contiene los genes gag, pol y env pero sin LTR ni componentes de empaquetamiento (Mann *et al.*, 1983). Cuando se introduce un plásmido recombinante que contiene un ADNc, junto con las LTR retrovíticas y las secuencias de empaquetamiento, en esta estirpe celular (mediante precipitación con fosfato de calcio, por ejemplo), las secuencias de empaquetamiento permiten al transcrito de ARN del plásmido recombinante empaquetarse en partículas víricas, que se secretan entonces al medio de cultivo (Nicolas y Rubenstein, 1988; Temin, 1986; Mann *et al.*, 1983). Se recoge entonces el medio que contiene los retrovirus recombinantes, opcionalmente concentrado, y se usa para transferencia génica. Los vectores retrovíticos son capaces de infectar una amplia variedad de tipos celulares. Sin embargo, la integración y expresión estables requieren la división de células hospedadoras (Paskind *et al.*, 1975).

Se ha desarrollado recientemente un enfoque novedoso diseñado para permitir una orientación específica de vectores retrovíticos basado en la modificación química de un retrovirus mediante la adición química de residuos de lactosa a la cubierta vírica. Esta modificación podría permitir la infección específica de hepatocitos mediante receptores de sialoglicoproteína.

Se ha diseñado un enfoque diferente para la orientación de retrovirus recombinantes en el que se usaron anticuerpos biotinilados contra una proteína de cubierta retrovítica y contra un receptor celular específico. Se acoplaron los anticuerpos mediante los componentes de biotina usando estreptavidina (Roux *et al.*, 1989). Usando anticuerpos contra antígenos del complejo principal de histocompatibilidad de clase I y clase II, se demostró la infección de una variedad de células humanas que portan esos antígenos de superficie con un virus ecotrópico *in vitro* (Roux *et al.*, 1989).

3. VIRUS ADENOASOCIADOS

El AAV (Ridgeway, 1988; Hermonat y Muzycska, 1984) es un parovirus descubierto como contaminación de disoluciones madre adenovíricas. Es un virus ubicuo (los anticuerpos están presentes en el 85% de la población humana de los EE.UU.) que no se ha ligado a ninguna enfermedad. Se clasifica también como dependovirus, porque sus replications dependen de la presencia de un virus auxiliar tal como adenovirus. Se han aislado cinco serotipos, de los cuales el AAV-2 es el mejor caracterizado. El AAV tiene un ADN lineal monocatenario que está encapsulado en las proteínas de cápsida VP1, VP2 y VP3 formando un virión icosaédrico de 20 a 24 nm de diámetro (Muzyczka y McLaughlin, 1988).

El ADN de AAV es de aproximadamente 4,7 kb de longitud. Contiene dos marcos abiertos de lectura y está flanqueado por dos ITR. Hay dos genes principales en el genoma de AAV: *rep* y *cap*. El gen *rep* codifica proteínas responsables de replications víricas, mientras que *cap* codifica la proteína de cápsida VP1-3. Cada ITR forma una estructura de horquilla en forma de T. Estas repeticiones terminales son los únicos componentes en cis esenciales de AAV para la integración cromosómica. Por lo tanto, el AAV puede usarse como vector con todas las secuencias de codificación víricas eliminadas y reemplazadas por el módulo de genes para suministro. Se han identificado tres promotores víricos y nombrado p5, p19 y p40, según su posición en el mapa. La transcripción a partir de p5 y p19 da

como resultado la producción de proteínas rep, y la transcripción de p40 produce las proteínas de cápsida (Hermonat y Muzyczka, 1984).

5 Hay varios factores que alentaron a los investigadores a estudiar la posibilidad de usar rAAV como vector de expresión. Uno es que los requisitos para suministrar un gen para integrar en el cromosoma hospedador son sorprendentemente pocos. Es necesario tener las ITR de 145 pb, que son sólo un 6% del genoma de AAV. Esto deja espacio en el vector para ensamblar una inserción de ADN de 4,5 kb. Aunque esta capacidad de carga puede evitar que AAV suministre genes grandes, es ampliamente adecuada para suministrar los constructos de no codificación de la presente invención.

10 El AAV es también una buena elección de vehículo de suministro debido a su seguridad. Hay un mecanismo de rescate relativamente complicado: no sólo los adenovirus de tipo silvestre, sino también los genes de AAV son necesarios para movilizar el rAAV. Igualmente, el AAV no es patógeno y no está asociado a ninguna enfermedad. La eliminación de secuencias de codificación víricas minimiza las reacciones inmunitarias ante la expresión génica vírica y, por lo tanto, el rAAV no provoca una respuesta inflamatoria.

4. OTROS VECTORES VÍRICOS COMO CONSTRUCTOS DE EXPRESIÓN

15 Pueden emplearse otros vectores víricos como constructos de expresión en la presente invención para el suministro de secuencias oligonucleotídicas o polinucleotídicas a una célula hospedadora. Pueden emplearse vectores derivados de virus tales como virus Vaccinia (Ridgeway, 1988; Coupar *et al.*, 1988), lentivirus, poliovirus y herpesvirus. Ofrecen varios rasgos atractivos para diversas células de mamífero (Friedmann, 1989; Ridgeway, 1988; Coupar *et al.*, 1988; Horwich *et al.*, 1990).

20 Con el reciente reconocimiento de los virus de hepatitis B defectivos, se ha conseguido un nuevo conocimiento de la relación entre estructura y función de diferentes secuencias víricas. Los estudios *in vitro* mostraron que los virus podían retener la capacidad de empaquetamiento dependiente del auxiliar y la transcripción inversa a pesar de la eliminación de hasta el 80% de su genoma (Horwich *et al.*, 1990). Esto sugería que podían reemplazarse grandes porciones del genoma por material genético ajeno. El hepatotropismo y la persistencia (integración) eran propiedades particularmente atractivas para la transferencia génica dirigida al hígado. Chang *et al.* (1991) introdujeron el gen de cloranfenicol acetiltransferasa (CAT) en genoma del virus de hepatitis B de pato en lugar de las secuencias de codificación de polimerasa, superficie y presuperficie. Se cotransfectó con virus de tipo silvestre en una estirpe celular de hepatoma aviar. Se usaron medios de cultivo que contenían altas titulaciones del virus recombinante para infectar hepatocitos de pato joven primarios. Se detectó la expresión génica de CAT estable durante al menos 24 días después de la transfección (Chang *et al.*, 1991).

5. VECTORES NO VÍRICOS

35 Para efectuar la expresión de las secuencias oligonucleotídica o polinucleotídica de la presente invención, el constructo de expresión debe suministrarse a una célula. Este suministro puede conseguirse *in vitro*, como en procedimientos de laboratorio para transformar estirpes celulares, o *in vivo* o *ex vivo*, como en el tratamiento de ciertos estados patológicos. Como se describe anteriormente, es un mecanismo preferido para el suministro mediante infección vírica en el que el constructo de expresión se encapsula en una partícula vírica infecciosa.

40 Una vez se ha suministrado el constructo de expresión a la célula, puede colocarse y expresarse en diferentes sitios el ácido nucleico que codifica las secuencias oligonucleotídicas o polinucleotídicas. En ciertas realizaciones, el ácido nucleico que codifica el constructo puede integrarse establemente en el genoma de la célula. Esta integración puede estar en la localización y orientación específica mediante recombinación homóloga (reemplazo génico) o puede integrarse en una localización aleatoria no específica (aumento génico). En aún otras realizaciones adicionales, el ácido nucleico puede mantenerse establemente en la célula en forma de un fragmento episómico separado de ADN. Dichos segmentos de ácido nucleico o "episomas" codifican secuencias suficientes para permitir el mantenimiento y replicación independientemente o sincronizados con el ciclo de la célula hospedadora. Cómo se suministra el constructo de expresión a una célula y dónde permanece en la célula el ácido nucleico depende del tipo de constructo de expresión empleado.

45 En ciertas realizaciones de la invención, el constructo de expresión que comprende una o más secuencias oligonucleotídicas o polinucleotídicas puede consistir simplemente en ADN recombinante desnudo o plásmidos. La transferencia del constructo puede efectuarse mediante cualquiera de los métodos mencionados anteriormente que permeabilizan física o químicamente la membrana celular. Esto es particularmente aplicable para la transferencia *in vitro*, pero puede aplicarse también al uso *in vivo*. Dubensky *et al.* (1984) inyectaron exitosamente ADN de poliomavirus en forma de precipitados de fosfato de calcio en el hígado y bazo de ratones adultos y recién nacidos, demostrando una replicación vírica activa e infección aguda. Benvenisty y Reshef (1986) demostraron también que la inyección intraperitoneal directa de plásmidos precipitados con fosfato de calcio da como resultado la expresión de los genes transfectados. Se prevé que el ADN que codifica un gen de interés pueda transferirse también de manera similar *in vivo* y expresar el producto génico.

Otra realización de la invención para transferir un constructo de expresión de ADN desnudo a células puede implicar el bombardeo de partículas. Este método depende de la capacidad de acelerar microproyectiles recubiertos con

ADN a alta velocidad, permitiéndoles perforar las membranas celulares y entrar en las células sin matarlas (Klein *et al.*, 1987). Se han desarrollado varios dispositivos para acelerar partículas pequeñas. Uno de dichos dispositivos se basa en una descarga de alto voltaje para generar una corriente eléctrica, que a su vez proporciona la fuerza motriz (Yang *et al.*, 1990). Los microproyectiles usados han consistido en sustancias biológicamente inertes tales como perlas de wolframio u oro.

Se han bombardeado *in vivo* los órganos seleccionados, incluyendo hígado, piel y tejido muscular de ratas y ratones (Yang *et al.*, 1990; Zelenin *et al.*, 1991). Esto puede requerir la exposición quirúrgica del tejido o células para eliminar cualquier tejido intermedio entre la pistola y el órgano diana, concretamente, tratamiento *ex vivo*. De nuevo, el ADN que codifica un gen particular puede suministrarse mediante este método y seguir incorporado a la presente invención.

Composiciones polipeptídicas

La presente invención proporciona, en otros aspectos, composiciones polipeptídicas.

Pueden identificarse generalmente las porciones inmunogénicas usando técnicas bien conocidas, tales como las resumidas en Paul, "Fundamental Immunology", 3ª ed., 243-247 (1993) y referencias citadas en el mismo. Dichas técnicas incluyen examen en polipéptidos de la capacidad de reaccionar con anticuerpos específicos de antígeno, antisueros y/o estirpes de linfocitos T o clones. Como se usa en la presente memoria, los antisueros y anticuerpos son "específicos de antígeno" si se unen específicamente a un antígeno (concretamente, reaccionan con la proteína en un ELISA u otro inmunoensayo, y no reaccionan detectablemente con proteínas no relacionadas). Dichos antisueros y anticuerpos pueden prepararse como se describe en la presente memoria y usando técnicas bien conocidas. Una porción inmunogénica de una proteína de *Mycobacterium sp.* es una porción que reacciona con dichos antisueros y/o linfocitos T a un nivel que no es sustancialmente menor que la reactividad del polipéptido completo (por ejemplo, en un ensayo ELISA y/o de reactividad de linfocitos T). Dichas porciones inmunogénicas pueden reaccionar en dichos ensayos a un nivel que es similar a o mayor que la reactividad del polipéptido completo. Dichos exámenes pueden efectuarse generalmente usando métodos conocidos por los expertos en la materia, tales como los descritos en Harlow y Lane, "Antibodies: A Laboratory Manual" (1988). Por ejemplo, puede inmovilizarse un polipéptido sobre un soporte sólido y ponerse en contacto con sueros del paciente para permitir la unión de los anticuerpos en los sueros al polipéptido inmovilizado. Los sueros no unidos pueden retirarse entonces y detectarse los anticuerpos unidos usando, por ejemplo, proteína A marcada con ¹²⁵I.

Los polipéptidos pueden prepararse usando cualquiera de una variedad de técnicas bien conocidas. Los polipéptidos recombinantes codificados por secuencias de ADN como se describen anteriormente pueden prepararse fácilmente a partir de secuencias de ADN usando cualquiera de una variedad de vectores de expresión conocidos por los expertos en la materia. Puede conseguirse la expresión en cualquier célula hospedadora apropiada que se ha transformado o transfectado con un vector de expresión que contiene una molécula de ADN que codifica un polipéptido recombinante. Las células hospedadoras adecuadas incluyen procariotas, levaduras y células eucarióticas superiores tales como células de mamífero y células de planta. Preferiblemente, las células hospedadoras empleadas son *E. coli*, levadura o una estirpe celular de mamífero tal como COS o CHO. Pueden concentrarse en primer lugar los sobrenadantes de sistemas de hospedador/vector adecuados que secretan proteína o polipéptido recombinante en el medio de cultivo usando un filtro comercialmente disponible. Después de la concentración, puede aplicarse el concentrado a una matriz de purificación adecuada tal como una matriz de afinidad o una resina de intercambio iónico. Finalmente, pueden emplearse una o más etapas de HPLC inversa para purificar adicionalmente un polipéptido recombinante.

Los polipéptidos de la invención, fragmentos inmunogénicos de los mismos y otras variantes que tienen menos de aproximadamente 100 aminoácidos, y generalmente menos de aproximadamente 50 aminoácidos, pueden generarse también mediante medios sintéticos, usando técnicas bien conocidas por los expertos en la materia. Por ejemplo, dichos polipéptidos pueden sintetizarse usando cualquiera de las técnicas en fase sólida disponibles comercialmente, tales como el método de síntesis en fase sólida de Merrifield, en el que se añaden secuencialmente aminoácidos a una cadena aminoácida creciente. Véase Merrifield, J. Am. Chem. Soc. 85: 2149-2146 (1963). El equipo para la síntesis automatizada de polipéptidos está comercialmente disponible de suministradores tales como Perkin Elmer/Applied BioSystems Division (Foster City, CA), y puede hacerse funcionar según las instrucciones del fabricante.

En ciertas realizaciones específicas, un polipéptido puede ser una proteína de fusión como se describe en la presente memoria, o comprende al menos una proteína de fusión como se describe en la presente memoria y una secuencia no relacionada. Un asociado de fusión puede, por ejemplo, ayudar a proporcionar epítopos auxiliares de T (un asociado de fusión inmunitario), preferiblemente epítopos auxiliares de T reconocidos por seres humanos, o puede ayudar a expresar la proteína (un potenciador de la expresión) a rendimientos mayores que la proteína recombinante nativa. Ciertos asociados de fusión preferidos son asociados de fusión tanto inmunitarios como potenciadores de la expresión. Pueden seleccionarse otros asociados de fusión para aumentar la solubilidad de la proteína o para posibilitar orientar la proteína a los compartimentos intracelulares deseados. Aún otros asociados de fusión incluyen marcadores de afinidad, que facilitan la purificación de la proteína.

Las proteínas de fusión pueden prepararse generalmente usando técnicas estándar, incluyendo la conjugación química. Preferiblemente, se expresa una proteína de fusión como proteína recombinante, que permite la producción de niveles aumentados respecto a una proteína no fusionada en un sistema de expresión. Brevemente, las secuencias de ADN que codifican los componentes polipeptídicos pueden ensamblarse separadamente, y ligarse en un vector de expresión apropiado. El extremo 3' de la secuencia de ADN que codifica un componente polipeptídico está ligado, con o sin un ligador peptídico, al extremo 5' de una secuencia de ADN que codifica un segundo componente polipeptídico de modo que los marcos de lectura de las secuencias estén en fase. Esto permite la traducción en una sola proteína de fusión que retiene la actividad biológica de ambos componentes polipeptídicos.

Puede emplearse una secuencia ligadora peptídica para separar el primer y segundo componentes polipeptídicos por una distancia suficiente para asegurar que cada polipéptido se pliega en sus estructuras secundaria y terciaria. Dicha secuencia ligadora peptídica se incorpora a la proteína de fusión usando técnicas estándar bien conocidas en la materia. Las secuencias ligadoras peptídicas adecuadas pueden elegirse basándose en los siguientes factores: (1) su capacidad de adoptar una conformación extendida flexible; (2) su incapacidad de adoptar una estructura secundaria que pudiera interaccionar con epítopos funcionales en el primer y segundo polipéptidos; y (3) la falta de residuos hidrófobos o cargados que podrían reaccionar con los epítopos funcionales polipeptídicos. Las secuencias ligadoras peptídicas preferidas contienen residuos de Gly, Asn y Ser. Pueden usarse también otros residuos aminoácidos neutros tales como Thr y Ala en la secuencia ligadora. Las secuencias aminoácidas que pueden emplearse con provecho como ligadores incluyen las dadas a conocer en Maratea *et al.*, Gene 40: 39-46 (1985); Murphy *et al.*, Proc. Natl. Acad. Sci. USA 83: 8258-8262 (1986); patentes de EE.UU. n° 4.935.233 y patente de EE.UU. n° 4.751.180. La secuencia ligadora puede ser generalmente de 1 a aproximadamente 50 aminoácidos de longitud. Las secuencias ligadoras no son necesarias cuando el primer y segundo polipéptidos tienen regiones aminoácidas N-terminales no esenciales que pueden usarse para separar los dominios funcionales y prevenir la interferencia estérica.

Las secuencias de ADN ligadas están ligadas operativamente a elementos reguladores transcripcionales o traduccionales adecuados. Los elementos reguladores responsables de la expresión de ADN están localizados sólo 5' de la secuencia de ADN que codifica los primeros polipéptidos. De forma similar, los codones de terminación necesarios para terminar la traducción y las señales de terminación de la transcripción están presentes sólo en 3' de la secuencia de ADN que codifica el segundo polipéptido.

Se proporcionan también proteínas de fusión. Dichas proteínas comprenden un polipéptido como se describe en la presente memoria junto con una proteína inmunogénica no relacionada. Preferiblemente, la proteína inmunogénica es capaz de provocar una respuesta de evocación. Los ejemplos de dichas proteínas incluyen proteínas de tétanos, tuberculosis y hepatitis (véase, por ejemplo, Stoute *et al.*, New Engl. J. Med. 336: 86-91 (1997)).

En las realizaciones preferidas, un asociado de fusión inmunitario deriva de proteína D, una proteína de superficie de la bacteria gramnegativa *Haemophilus influenzae* B (documento WO 91/18926). Preferiblemente, un derivado de proteína D comprende aproximadamente el primer tercio de la proteína (por ejemplo, los primeros 100-110 aminoácidos N-terminales) y un derivado de proteína D puede estar lipidado. En ciertas realizaciones preferidas, los primeros 109 residuos de un asociado de fusión de lipoproteína D están incluidos en el extremo N para proporcionar al polipéptido epítopos de linfocitos T exógenos adicionales y para aumentar el nivel de expresión en *E. coli* (funcionando por tanto con un potenciador de la expresión). La cola lipídica asegura una presentación óptima del antígeno a células presentadoras de antígeno. Otros asociados de fusión incluyen la proteína no estructural del virus de la gripe NS1 (hemaglutinina). Típicamente, se usan los 81 aminoácidos N-terminales, aunque pueden usarse diferentes fragmentos que incluyan epítopos auxiliares de T.

En otra realización, el asociado de fusión inmunitario es la proteína conocida como LYTA, o una porción de la misma (preferiblemente una porción C-terminal). La LYTA deriva de *Streptococcus pneumoniae*, que sintetiza una *N*-acetil-L-alanina amidasa conocida como amidasa LYTA (codificada por el gen *LytA*; Gene 43: 265-292 (1986)). La LYTA es una autolisina que degrada específicamente ciertos enlaces en la cadena principal de peptidoglicano. El dominio C-terminal de la proteína LYTA es responsable de la afinidad por la colina o algunos análogos de colina tales como DEAE. Esta propiedad se ha explotado para el desarrollo de plásmidos que expresan C-LYTA de *E. coli* útiles para la expresión de proteínas de fusión. Se ha descrito la purificación de proteínas híbridas que contienen el fragmento C-LYTA en el extremo amino (véase *Biotechnology* 10: 795-798 (1992)). En una realización preferida, puede incorporarse una porción repetida de LYTA a una proteína de fusión. Se encuentra una porción repetida en la región C-terminal a partir del residuo 178. Una porción repetida particularmente preferida incorpora los residuos 188-305.

En general, se aíslan los polipéptidos (incluyendo proteínas de fusión) y polinucleótidos como se describen en la presente memoria. Un polipéptido o polinucleótido "aislado" es aquel que se retira de su entorno original. Por ejemplo, una proteína de origen natural se aísla si se separa de algunos o todos los materiales coexistentes en el sistema natural. Preferiblemente, dichos polipéptidos son al menos aproximadamente un 90% puros, más preferiblemente al menos aproximadamente un 95% puros y lo más preferiblemente al menos aproximadamente un 99% puros. Un polinucleótido se considera aislado, por ejemplo, si se clona en un vector que no es parte del entorno natural.

Composiciones farmacéuticas

En realizaciones adicionales, la presente invención se refiere a la formulación de una o más de las composiciones de polinucleótidos y/o polipéptidos dadas a conocer en la presente memoria en disoluciones farmacéuticamente aceptables o fisiológicamente aceptables para administración a una célula o un animal, solos o en combinación con una o más de otras modalidades de terapia. Dichas composiciones son también útiles para usos de diagnóstico.

Se entenderá también que, si se desea, el segmento de ácido nucleico o composiciones de ARN, ADN o ANP que expresan un polipéptido como se da a conocer en la presente memoria pueden administrarse en combinación con otros agentes también tales como, por ejemplo, otras proteínas o polipéptidos o diversos agentes farmacéuticamente activos. De hecho, no hay virtualmente límite para los otros componentes que pueden incluirse también, dado que los agentes adicionales no causan un efecto adverso significativo tras el contacto con las células diana o tejidos hospedadores. Por tanto, las composiciones pueden suministrarse junto con diversos otros agentes según sea necesario en un caso particular. Dichas composiciones pueden purificarse a partir de células hospedadoras u otras fuentes biológicas, o como alternativa, pueden sintetizarse químicamente como se describe en la presente memoria. Igualmente, dichas composiciones pueden comprender adicionalmente composiciones de ARN o ADN sustituido o derivatizado.

La formulación de excipientes y disoluciones portadoras farmacéuticamente aceptables es bien conocida por los expertos en la materia, como el desarrollo de regímenes de dosificación y tratamiento adecuados para usar las composiciones particulares descritas en la presente memoria en una variedad de regímenes de tratamiento incluyendo, por ejemplo, administración y formulación oral, parenteral, intravenosa, intranasal e intramuscular.

1. SUMINISTRO ORAL

En ciertas aplicaciones, las composiciones farmacéuticas dadas a conocer en la presente memoria pueden suministrarse mediante administración oral a un animal. Como tales, estas composiciones pueden formularse con un diluyente inerte o con un portador comestible asimilable, o pueden incluirse en cápsulas de gelatina de cubierta dura o blanda, o pueden comprimirse en comprimidos, o pueden incorporarse directamente al alimento de la dieta.

Los compuestos activos pueden incluso incorporarse a excipientes y usarse en forma de comprimidos ingeribles, comprimidos bucales, trociscos, cápsulas, elixires, suspensiones, jarabes, obleas y similares (Mathiowitz *et al.*, 1997; Hwang *et al.*, 1998; patente de EE.UU. 5.641.515; patente de EE.UU. 5.580.579 y patente de EE.UU. 5.792.451. Los comprimidos, trociscos, píldoras, cápsulas y similares pueden contener también lo siguiente: un aglutinante como goma de tragacanto, goma arábiga, almidón de maíz o gelatina; excipientes tales como fosfato dicálcico; un agente disgregante tal como almidón de maíz, almidón de patata, ácido alginico y similares; un lubricante tal como estearato de magnesio y puede añadirse un agente edulcorante tal como sacarosa, lactosa o sacarina o un agente aromatizante tal como menta piperita, aceite de gaulteria o aroma de cereza. Cuando la forma unitaria de dosificación es una cápsula, puede contener, además de los materiales del tipo anterior, un portador líquido. Pueden estar presentes diversos otros materiales como recubrimientos o para modificar de otro modo la forma física de la unidad de dosificación. Por ejemplo, los comprimidos, píldoras o cápsulas pueden estar recubiertos con goma laca, azúcar o ambos. Un jarabe o elixir pueden contener el compuesto activo sacarosa como agente edulcorante, metil- y propilparabenos como conservantes, un tinte y aromatizantes tales como aroma de cereza o naranja. Por supuesto, cualquier material usado en la preparación de cualquier forma unitaria de dosificación debe ser farmacéuticamente puro y sustancialmente no tóxico en las cantidades empleadas. Además, los compuestos activos pueden incorporarse a preparaciones y formulaciones de liberación sostenida.

Típicamente, estas formulaciones pueden contener al menos aproximadamente un 0,1% del compuesto activo o más, aunque el porcentaje de ingrediente(s) activo(s) puede variar, por supuesto, y puede estar convenientemente entre aproximadamente 1 ó 2% y aproximadamente 60 ó 70% o más del peso o volumen de la formulación total. Naturalmente, la cantidad de principio(s) activo(s) en cada composición terapéuticamente útil puede prepararse de tal modo que se obtenga una dosificación adecuada a cualquier dosis unitaria dada del compuesto. Se contemplarán por un experto en la materia de preparar dichas formulaciones farmacéuticas factores tales como solubilidad, biodisponibilidad, semivida biológica, vía de administración, vida útil del producto, así como otras consideraciones farmacológicas y, como tales, pueden ser deseables una variedad de dosificaciones y regímenes de tratamiento.

Para administración oral, las composiciones de la presente invención pueden incorporar como alternativa uno o más excipientes en forma de un colutorio, dentífrico, comprimido bucal, pulverizador oral o formulación administrada por vía oral sublingual. Por ejemplo, puede prepararse un colutorio incorporando el ingrediente activo en la cantidad necesaria a un disolvente apropiado, tal como una disolución de borato de sodio (disolución de Dobell). Como alternativa, el ingrediente activo puede incorporarse a una disolución oral tal como aquella que contiene borato de sodio, glicerina y bicarbonato de potasio, o dispersarse en un dentífrico, o añadirse en una cantidad terapéuticamente eficaz a una composición que puede incluir agua, aglutinantes, abrasivos, agentes aromatizantes, agentes espumantes y humectantes. Como alternativa, las composiciones pueden conformarse en forma de comprimido o disolución que puede disponerse bajo la lengua o disolverse de otro modo en la boca.

2. SUMINISTRO INYECTABLE

5 En ciertas circunstancias, será deseable suministrar las composiciones farmacéuticas dadas a conocer en la presente memoria por vía parenteral, intravenosa, intramuscular o incluso intraperitoneal como se describe en la patente de EE.UU. n° 5.543.158; la patente de EE.UU. n° 5.641.515 y la patente de EE.UU. n° 5.399.363. Pueden prepararse disoluciones de los compuestos activos en forma de base libre o sales farmacológicamente aceptables en agua adecuadamente mezcladas con un tensioactivo tal como hidroxipropilcelulosa. Pueden prepararse también dispersiones en glicerol, polietilenglicoles líquidos y mezclas de los mismos y en aceites. En condiciones ordinarias de almacenamiento y uso, estas preparaciones contienen un conservante para evitar el crecimiento de microorganismos.

10 Las formas farmacéuticas adecuadas para uso inyectable incluyen disoluciones o dispersiones acuosas estériles y polvos estériles para la preparación extemporánea de disoluciones o dispersiones inyectables estériles (patente de EE.UU. 5.466.468). En todos los casos, la forma debe ser estéril y debe ser fluida en la medida en que exista una fácil inyectabilidad. Debe ser estable en las condiciones de fabricación y almacenamiento y debe conservarse frente a la acción contaminante de microorganismos, tales como bacterias y hongos. El portador puede ser un disolvente o medio de dispersión que contiene, por ejemplo, agua, etanol, poliol (por ejemplo, glicerol, propilenglicol y polietilenglicol líquido y similares), mezclas adecuadas de los mismos y/o aceites vegetales. La fluidez apropiada puede mantenerse, por ejemplo, mediante el uso de un recubrimiento tal como lecitina, mediante el mantenimiento del tamaño de partícula necesario en el caso de dispersión y mediante el uso de tensioactivos. La prevención de la acción de los microorganismos puede facilitarse mediante diversos agentes antibacterianos y antifúngicos, por ejemplo, parabenos, clorobutanol, fenol, ácido sórbico, timerosal y similares. En muchos casos, será preferible incluir agentes isotónicos, por ejemplo, azúcares o cloruro de sodio. La absorción prolongada de las composiciones inyectables puede ocasionarse mediante el uso en las composiciones de agentes que retardan la absorción, por ejemplo, monoestearato de aluminio y gelatina.

25 Para administración parenteral en una disolución acuosa, por ejemplo, la disolución debe estar adecuadamente tamponada en caso necesario y volverse en primer lugar el diluyente líquido isotónico con suficiente disolución salina o glucosa. Estas disoluciones acuosas particulares son especialmente adecuadas para administración intravenosa, intramuscular, subcutánea e intraperitoneal. A este respecto, será conocido por los expertos en la materia un medio acuoso estéril que pueda emplearse a la vista de la presente divulgación. Por ejemplo, puede disolverse una dosificación en 1 ml de disolución isotónica de NaCl y añadirse a 1.000 ml de fluido de hipodermocisis o inyectarse en el sitio de infusión propuesto (véase, por ejemplo, "Remington's Pharmaceutical Sciences", 15ª edición, pág. 1035-1038 y 1570-1580). Aparecerá necesariamente cierta variación en la dosificación dependiendo de la afección del sujeto que se esté tratando. La persona responsable de la administración determinará, en cualquier caso, la dosis apropiada para el sujeto individual. Además, para administración humana, las preparaciones deberían satisfacer los estándares de esterilidad, pirogenicidad y seguridad y pureza generales requeridos por la Oficina de estándares biológicos de la FDA

40 Las disoluciones inyectables estériles se preparan mediante la incorporación de los compuestos activos en la cantidad necesaria al disolvente apropiado con algunos de los otros ingredientes enumerados anteriormente, según sea necesario, seguido de esterilización por filtración. Generalmente, las dispersiones se preparan incorporando diversos ingredientes activos esterilizados a un vehículo estéril que contiene el medio de dispersión básico y los otros ingredientes necesarios de los enumerados anteriormente. En el caso de polvos estériles para la preparación de disoluciones inyectables estériles, los métodos preferidos de preparación son técnicas de secado a vacío y liofilización que proporcionan un polvo del ingrediente activo más cualquier ingrediente deseado adicional a partir de una disolución del mismo anteriormente esterilizada por filtración.

45 Las composiciones dadas a conocer en la presente memoria pueden formularse en forma neutra o de sal. Las sales farmacéuticamente aceptables incluyen las sales de adición de ácido (formadas con los grupos amino libres de la proteína) y las que se forman con ácidos inorgánicos tales como, por ejemplo, ácidos clorhídrico o fosfórico, o ácidos orgánicos tales como acético, oxálico, tartárico, mandélico y similares. Las sales formadas con los grupos carboxilo libres pueden derivar también de bases inorgánicas tales como, por ejemplo, hidróxidos de sodio, potasio, amonio, calcio o férrico, y bases orgánicas tales como isopropilamina, trimetilamina, histidina, procaína y similares. Tras la formulación, se administrarán las disoluciones de manera compatible con la formulación de dosificación y en tal cantidad que sea terapéuticamente eficaz. Las formulaciones se administran fácilmente en una variedad de formas de dosificación tales como disoluciones inyectables, cápsulas de liberación de fármaco y similares.

55 Como se usa en la presente memoria "portador" incluye todos y cada uno de los disolventes, medios de dispersión, vehículos, recubrimientos, diluyentes, agentes antibacterianos y antifúngicos, agentes isotónicos y retardadores de la absorción, tampones, disoluciones portadoras, suspensiones, coloides y similares. El uso de dichos medios y agentes para sustancias farmacéuticamente activas es bien conocido en la materia. Excepto en la medida en que cualquier medio o agente convencional sea incompatible con el ingrediente activo, se contempla su uso en las composiciones terapéuticas. Pueden incorporarse también ingredientes activos suplementarios a las composiciones.

60 La frase "farmacéuticamente estable" designa entidades moleculares y composiciones que no producen una reacción alérgica o similar adversa cuando se administran a un ser humano. La preparación de una composición

acuosa que contiene una proteína como ingrediente activo es bien conocida en la materia. Típicamente, se preparan dichas composiciones en forma de inyectables, en disoluciones o suspensiones líquidas; pueden prepararse también formas sólidas adecuadas para disolución o suspensión en líquido antes de la inyección. La preparación puede también emulsionarse.

5 3. SUMINISTRO NASAL

En ciertas realizaciones, las composiciones farmacéuticas pueden suministrarse mediante pulverizadores intranasales, de inhalación y/u otros vehículos de suministro por aerosol. Se han descrito métodos para suministrar genes, ácidos nucleicos y composiciones peptídicas directamente a los pulmones mediante pulverizadores por aerosol, por ejemplo, en la patente de EE.UU. n° 5.756.353 y la patente de EE.UU. n° 5.804.212. Igualmente, son también bien conocidos en las técnicas farmacéuticas el suministro de fármacos usando resinas microparticuladas intranasales (Takenaga *et al.*, 1998) y compuestos de lisofosfatidilglicerol (patente de EE.UU. n° 5.725.871). Igualmente, se describe el suministro transmucoso de fármaco en forma de una matriz de soporte de politetrafluoroetileno en la patente de EE.UU. n° 5.780.045

4. SUMINISTRO MEDIADO POR LIPOSOMAS, NANOCÁPSULAS Y MICROPARTÍCULAS

15 En ciertas realizaciones, los inventores contemplan el uso de liposomas, nanocápsulas, micropartículas, microesferas, partículas lipídicas, vesículas y similares para la introducción de las composiciones de la presente invención en células hospedadoras adecuadas. En particular, las composiciones de la presente invención pueden formularse para suministro encapsuladas en una partícula lipídica, un liposoma, una vesícula, una nanoesfera o una nanopartícula o similar.

20 Dichas formulaciones pueden preferirse para la introducción de formulaciones farmacéuticamente aceptables de los ácidos nucleicos o constructos dados a conocer en la presente memoria. La formación y uso de liposomas son generalmente conocidos por los expertos en la materia (véase, por ejemplo, Couvreur *et al.*, 1977; Couvreur, 1988; Lasic, 1998; que describen el uso de liposomas y nanocápsulas en la terapia de antibióticos orientada para infecciones y enfermedades bacterianas intracelulares). Recientemente, se han desarrollado liposomas con estabilidad sérica y semividas de circulación mejoradas (Gabizon y Papahadjopoulos, 1988; Allen y Choun, 1987; patente de EE.UU. n° 5.741.516). Adicionalmente, se han revisado diversos métodos de liposomas y preparaciones similares a liposomas como portadores de fármaco potenciales (Takakura, 1998; Chandran *et al.*, 1997; Margalit, 1995; patente de EE.UU. n° 5.567.434; patente de EE.UU. n° 5.552.157; patente de EE.UU. n° 5.565.213; patente de EE.UU. n° 5.738.868 y patente de EE.UU. n° 5.795.587.

30 Los liposomas se han usado exitosamente con una serie de tipos celulares que son normalmente resistentes a la transfección mediante otros procedimientos, incluyendo suspensiones de linfocitos T, cultivos de hepatocitos primarios y células PC 12 (Renneisen *et al.*, 1990; Muller *et al.*, 1990). Además, los liposomas están exentos de las limitaciones de longitud de ADN que son típicas en sistemas de suministro basados en virus. Los liposomas se han usado eficazmente para introducir genes, fármacos (Heath y Martin, 1986; Heath *et al.*, 1986; Balazsovits *et al.*, 1989; Fresta y Puglisi, 1996), agentes radioterapéuticos (Pikul *et al.*, 1987), enzimas (Imaizumi *et al.*, 1990a; Imaizumi *et al.*, 1990b), virus (Faller y Baltimore, 1984), factores de transcripción y efectores alostéricos (Nicolau y Gersonde, 1979) en una variedad de estirpes celulares cultivadas y animales. Además, se han completado varios ensayos clínicos exitosos que examinaban la eficacia del suministro de fármacos mediado por liposomas (López-Berestein *et al.*, 1985a; 1985b; Coune, 1988; Sculier *et al.*, 1988). Además, varios estudios sugieren que el uso de liposomas no está asociado a respuestas autoinmunitarias, toxicidad ni localización gonádica después del suministro sistémico (Mori y Fukatsu, 1992).

Los liposomas se forman a partir de fosfolípidos que se dispersan en un medio acuoso y forman espontáneamente vesículas bicapa concéntricas multilamelares (también denominadas vesículas multilamelares (VML). Las VML tienen generalmente diámetros de 25 nm a 4 µm. La sonicación de VML da como resultado la formación de vesículas unilamelares pequeñas (VUP) con diámetros en el intervalo de 200 a 500 Å, que contienen una disolución acuosa en el núcleo.

Los liposomas tienen semejanza con membranas celulares y se contemplan para uso relacionado con la presente invención como portadores de las composiciones peptídicas. Son ampliamente adecuados ya que pueden atrapar tanto sustancias solubles en agua como en lípidos, concretamente, en los espacios acuosos y en la bicapa misma, respectivamente. Es posible que los liposomas portadores de fármaco puedan emplearse incluso para el suministro específico de sitio de agentes activos modificando selectivamente la formulación liposómica.

Además de las enseñanzas de Couvreur *et al.* (1977; 1988), puede utilizarse la siguiente información en la generación de formulaciones liposómicas. Los fosfolípidos pueden formar una variedad de estructuras distintas de liposomas cuando se dispersan en agua, dependiendo de la relación molar de lípido a agua. A relaciones bajas, el liposoma es la estructura preferida. Las características físicas de los liposomas dependen del pH, fuerza iónica y la presencia de cationes divalentes. Los liposomas pueden mostrar baja permeabilidad a sustancias iónicas y polares, pero a temperaturas elevadas experimentan una transición de fase que altera notablemente su permeabilidad. La transición de fase implica un cambio desde una estructura ordenada estrechamente empaquetada conocida como

estado de gel a una estructura menos ordenada empaquetada suelta conocida como estado de fluido. Esto ocurre a una temperatura de transición de fase característica y da como resultado un aumento de la permeabilidad a los iones, azúcares y fármacos.

5 Además de la temperatura, la exposición a proteínas puede alterar la permeabilidad de los liposomas. Ciertas proteínas solubles tales como citocromo c se unen a, deforman y penetran la bicapa, causando así cambios en la permeabilidad. El colesterol inhibe esta penetración de proteínas, aparentemente al empaquetar los fosfolípidos más estrechamente. Se contempla que las formaciones de liposoma más útiles para suministro de antibióticos e inhibidores contendrán colesterol.

10 La capacidad de atrapar solutos varía entre los diferentes tipos de liposomas, por ejemplo, las VML son moderadamente eficaces para atrapar solutos, pero las VUP son extremadamente ineficaces. Las VUP ofrecen la ventaja de homogeneidad y reproducibilidad en la distribución por tamaño, sin embargo, y se ofrece un compromiso entre tamaño y eficacia de atrapamiento por las vesículas unilamelares grandes (VUG). Estas se preparan mediante evaporación de éter y son 3 a 4 veces más eficaces en el atrapamiento de soluto que las VML.

15 Además de las características del liposoma, es un determinante importante en el atrapamiento de compuestos las propiedades fisicoquímicas del compuesto mismo. Los compuestos polares se atrapan en los espacios acuosos y los compuestos no polares se unen a la bicapa lipídica de la vesícula. Los compuestos polares se liberan mediante permeación o cuando se rompe la bicapa, pero los compuestos no polares permanecen unidos a la bicapa a menos que se desestabilice por temperatura o exposición a lipoproteínas. Ambos tipos muestran velocidades de salida máximas a la temperatura de transición de fase.

20 Los liposomas interactúan con células mediante cuatro mecanismos diferentes: endocitosis por células fagocíticas del sistema reticuloendotelial tales como macrófagos y neutrófilos; adsorción a la superficie celular, por fuerzas hidrófobas débiles no específicas o electrostáticas o por interacciones específicas con componentes de la superficie celular; fusión con la membrana celular plasmática mediante inserción de la bicapa lipídica del liposoma en la membrana plasmática, con liberación simultánea de los contenidos liposómicos en el citoplasma, y por transferencia de lípidos liposómicos a membranas celulares o subcelulares o viceversa, sin asociación con los contenidos liposómicos. A menudo es difícil determinar cuál mecanismo es operativo y pueden funcionar más de uno a la vez.

30 El destino y disposición de los liposomas inyectados por vía intravenosa depende de sus propiedades físicas, tales como tamaño, fluidez y carga superficial. Pueden persistir en tejidos durante horas o días, dependiendo de su composición, y las semividas en la sangre están en el intervalo de minutos a varias horas. Los liposomas grandes, tales como VML y VUG, se incorporan rápidamente por las células fagocíticas del sistema reticuloendotelial, pero la fisiología del sistema circulatorio limita la salida de dichas especies grandes en la mayoría de los sitios. Sólo pueden salir en lugares en que existan grandes aberturas o poros en el endotelio capilar, tales como en los sinusoides del hígado o bazo. Por tanto, estos órganos son el sitio de captación predominante. Por otro lado, las VUP muestran una distribución de tejido más amplia, pero siguen secuestrándose en gran medida en el hígado y el bazo. En general, este comportamiento *in vivo* limita la orientación potencial de los liposomas a sólo aquellos órganos y tejidos accesibles a su gran tamaño. Estos incluyen sangre, hígado, bazo, médula ósea y órganos linfoides.

40 La orientación no es generalmente una limitación en términos de la presente invención. Sin embargo, en el caso de que se desee una orientación específica, están disponibles métodos para conseguir esto. Pueden usarse anticuerpos para unirse a la superficie liposómica y dirigir el anticuerpo y sus contenidos de fármaco a receptores antigénicos específicos localizados en una superficie de tipo celular particular. Los determinantes de carbohidrato (componentes de superficie celular glucoproteicos o glucolipídicos que desempeñan un papel en el reconocimiento, interacción y adhesión célula-célula) pueden usarse también como sitios de reconocimiento, ya que tienen potencial para dirigir los liposomas a tipos celulares particulares. La mayoría de veces, se contempla que se usaría la inyección intravenosa de preparaciones liposómicas, pero también son concebibles otras vías de administración.

45 Como alternativa, la invención proporciona formulaciones en nanocápsula farmacéuticamente aceptables de las composiciones de la presente invención. Las nanocápsulas pueden atrapar generalmente compuestos de modo estable y reproducible (Henry-Michelland *et al.*, 1987; Quintanar-Guerrero *et al.*, 1998; Douglas *et al.*, 1987). Para evitar los efectos secundarios debidos a una sobrecarga polimérica intracelular, deben diseñarse dichas partículas ultrafinas (tamaño de aproximadamente 0,1 μm) usando polímeros capaces de degradarse *in vivo*. Se contemplan para uso en la presente invención nanopartículas de poli(cianoacrilato de alquilo) biodegradables que satisfacen estos requisitos. Dichas partículas pueden prepararse fácilmente como se describe en (Couvreur *et al.*, 1980; 1988; zur Muhlen *et al.*, 1998; Zambaux *et al.* 1998; Pinto-Alphandry *et al.*, 1995 y la patente de EE.UU. n° 5.145.684).

Vacunas

55 En ciertas realizaciones preferidas de la presente invención, se proporcionan vacunas. Las vacunas comprenderán generalmente una o más composiciones farmacéuticas, tales como las discutidas anteriormente, en combinación con un inmunoestimulante. Un inmunoestimulante puede ser cualquier sustancia que potencie una respuesta inmunitaria (de anticuerpo y/o mediada por célula) ante un antígeno exógeno. Los ejemplos de inmunoestimulantes incluyen coadyuvantes, microesferas biodegradables (por ejemplo, poli(galactida láctica)) y liposomas (en los que se

incorpora el compuesto, véase, por ejemplo, Fullerton, patente de EE.UU. nº 4.235.877). La preparación de vacuna se describe generalmente, por ejemplo, en Powell y Newman, eds., "Vaccine Design (the subunit and adjuvant approach)" (1995). Las composiciones farmacéuticas y vacunas dentro del alcance de la presente invención pueden contener también otros compuestos, que pueden ser biológicamente activos o inactivos.

5 Las vacunas ilustrativas pueden contener ADN que codifica uno o más de los polipéptidos descritos anteriormente, de tal modo que el polipéptido se genere *in situ*. Como se observa anteriormente, el ADN puede estar presente en cualquiera de una variedad de sistemas conocidos por los expertos en la materia, incluyendo sistemas de expresión de ácido nucleico, bacterias y sistemas de expresión vírica. Son bien conocidas en la materia numerosas técnicas de suministro génico, tales como las descritas por Rolland, Crit. Rev. Therap. Drug Carrier Systems 15: 143-198 (1998) y las referencias citadas en el mismo. Los sistemas de expresión de ácido nucleico apropiados contienen las secuencias de ADN necesarias para expresión en el paciente (tales como una señal promotora y terminadora adecuadas). Los sistemas de suministro bacterianos implican la administración de una bacteria (tal como bacilo de Calmette-Guerin) que expresa una porción inmunogénica del polipéptido sobre su superficie celular o secreta dicho epítipo. En una realización preferida, el ADN puede introducirse usando un sistema de expresión vírico (por ejemplo, virus Vaccinia u otro poxvirus, retrovirus o adenovirus), que puede implicar el uso de un virus no patogénico (defectivo) competente de replicación. Se dan a conocer sistemas adecuados, por ejemplo, en Fisher-Hoch *et al.*, Proc. Natl. Acad. Sci. USA 86: 317-321 (1989); Flexner *et al.*, Ann. N.Y. Acad. Sci. 569: 86-103 (1989); Flexner *et al.*, Vaccine 8: 17-21 (1990); patentes de EE.UU. nº 4.603.112, 4.769.330 y 5.017.487; WO 89/01973; patente de EE.UU. nº 4.777.127; GB 2.200.651; EP 0.345.242; WO 91/02805; Berkner, Biotechniques 6: 616-627 (1988); Rosenfeld *et al.*, Science 252: 431-434 (1991); Kolls *et al.*, Proc. Natl. Acad. Sci. USA 91: 215-219 (1994); Kass-Eisler *et al.*, Proc. Natl. Acad. Sci. USA 90: 11498-11502 (1993); Guzmán *et al.*, Circulation 88: 2838-2848 (1993) y Guzmán *et al.*, Cir. Res. 73: 1202-1207 (1993). Las técnicas para incorporar ADN a dichos sistemas de expresión son bien conocidas por los expertos en la materia. El ADN puede ser también "desnudo" como se describe, por ejemplo, en Ulmer *et al.*, Science 259: 1745-1749 (1993) y se revisa por Cohen, Science 259: 1691-1692 (1993). La captación de ADN desnudo puede aumentarse recubriendo el ADN sobre perlas biodegradables, que se transportan eficazmente en las células. Resultará evidente que una vacuna puede comprender tanto un componente polinucleotídico como polipeptídico. Dichas vacunas pueden proporcionar una respuesta inmunitaria potenciada.

Resultará evidente que una vacuna puede contener sales farmacéuticamente aceptables de los polinucleótidos y polipéptidos proporcionados en la presente memoria. Dichas sales pueden prepararse a partir de bases no tóxicas farmacéuticamente aceptables incluyendo bases orgánicas (por ejemplo, sales de aminas primarias, secundarias y terciarias y aminoácidos básicos) y bases inorgánicas (por ejemplo, sales de sodio, potasio, litio, amonio, calcio y magnesio).

Aunque puede emplearse cualquier portador adecuado conocido por los expertos en la materia en las composiciones de vacuna de esta invención, el tipo de portador variará dependiendo del modo de administración. Las composiciones de la presente invención pueden formularse para cualquier manera apropiada de administración incluyendo, por ejemplo, administración tópica, oral, nasal, intravenosa, intracraneal, intraperitoneal, subcutánea o intramuscular. Para administración parenteral, tal como inyección subcutánea, el portador comprende preferiblemente agua, disolución salina, alcohol, una grasa, una cera o un tampón. Para administración oral, puede emplearse cualquiera de los portadores anteriores o un portador sólido tal como manitol, lactosa, almidón, estearato de magnesio, sacarina de sodio, talco, celulosa, glucosa, sacarosa y carbonato de magnesio. Pueden emplearse también microesferas biodegradables (por ejemplo, polilactato-poliglicolato) como portadores para las composiciones farmacéuticas de esta invención. Se dan a conocer microesferas biodegradables adecuadas, por ejemplo, en las patentes de EE.UU. nº 4.897.268; 5.075.109; 5.928.647; 5.811.128; 5.820.883; 5.853.763; 5.814.344 y 5.942.252. Puede emplearse también un portador que comprende los complejos de partícula-proteína descritos en la patente de EE.UU. nº 5.928.647, que son capaces de inducir una respuesta de linfocitos T citotóxicos limitada de clase I en un hospedador.

Dichas composiciones pueden comprender también tampones (por ejemplo, disolución salina tamponada neutra o disolución salina tamponada con fosfato), carbohidratos (por ejemplo, glucosa, manosa, sacarosa o dextranos), manitol, proteínas, polipéptidos o aminoácidos tales como glicina, antioxidantes, bacteriostáticos, agentes quelantes tales como EDTA o glutatión, coadyuvantes (por ejemplo, hidróxido de aluminio), solutos que vuelven la formulación isotónica, hipotónica o débilmente hipertónica con la sangre de un receptor, agentes de suspensión, agentes espesantes y/o conservantes. Como alternativa, las composiciones de la presente invención pueden formularse en forma de un liofilizado. Los compuestos pueden encapsularse también en liposomas usando tecnología bien conocida.

Puede emplearse cualquiera de una variedad de inmunoestimulantes en las vacunas de esta invención. Por ejemplo, puede incluirse un coadyuvante. La mayoría de coadyuvantes contienen una sustancia diseñada para proteger al antígeno de un catabolismo rápido, tal como hidróxido de aluminio o aceite mineral, y un estimulante de respuestas inmunitarias tal como lípido A, especies de *Bordetella pertussis* o *Mycobacterium* o proteínas derivadas de *Mycobacterium*. Por ejemplo, puede usarse *M. vaccae* deslipidada desglucolipidada ("pVac"). En otra realización, se usa BCG como coadyuvante. Además, la vacuna puede administrarse a un sujeto expuesto anteriormente a BCG. Los coadyuvantes adecuados están comercialmente disponibles, por ejemplo, como coadyuvante completo y coadyuvante completo de Freund (Difco Laboratories, Detroit, MI); coadyuvante Merck 65 (Merck and Company,

5 Inc., Rahway, NJ); AS-2 y derivados de los mismos (SmithKline Beecham, Filadelfia, PA); CWS, TDM, Leif, sales de aluminio tales como gel de hidróxido de aluminio (alúmina) o fosfato de aluminio; sales de calcio, hierro o cinc; una suspensión insoluble de tirosina acilada; azúcares acilados; polisacáridos derivatizados catiónica o aniónicamente; polifosfazenos; microesferas biodegradables; monofosforil-lípido A y quil A. Pueden usarse también como

10 En las vacunas proporcionadas en la presente memoria, la composición de coadyuvante se diseña preferiblemente para inducir una respuesta inmunitaria predominantemente de tipo Th1. Los altos niveles de citocinas de tipo Th1 (por ejemplo, IFN- γ , TNF- α , IL-2 e IL-12) tienden a favorecer la inducción de respuestas inmunitarias mediadas por células ante un antígeno administrado. En contraposición, los altos niveles de citocinas de tipo Th2 (por ejemplo, IL-4, IL-5, IL-6 e IL-10) tienden a favorecer la inducción de respuestas inmunitarias humorales. Después de la aplicación de una vacuna como se proporciona en la presente memoria, un paciente soportará una respuesta

15 inmunitaria que incluye respuestas de tipo Th1 y Th2. En una realización preferida, en la que una respuesta es predominantemente de tipo Th1, el nivel de citocinas de tipo Th1 aumentará en una mayor medida que el nivel de citocinas de tipo Th2. Los niveles de estas citocinas pueden evaluarse fácilmente usando ensayos estándar. Para una revisión de las familias de citocinas, véase Mosmann y Coffman, Ann. Rev. Immunol. 7: 145-173 (1989).

20 Los coadyuvantes preferidos para uso para provocar una respuesta de tipo predominantemente Th1 incluyen, por ejemplo, una combinación de monofosforil-lípido A, preferiblemente monofosforil-lípido A 3-des-O-acilado (3D-MPL), junto con una sal de aluminio. Los coadyuvantes de MPL están disponibles en Corixa Corporation (Seattle, WA; véanse las patentes de EE.UU. n° 4.436.727; 4.877.611; 4.866.034 y 4.912.094). Los oligonucleótidos que contienen CpG (en los que el dinucleótido CpG no está metilado) inducen también predominantemente una respuesta Th1. Dichos oligonucleótidos son bien conocidos y se describen, por ejemplo, en los documentos WO 96/02555, WO 99/33488 y patentes de EE.UU. n° 6.008.200 y 5.856.462. Se describen también secuencias de ADN

25 inmunoestimulantes, por ejemplo, por Sato *et al.*, Science 273: 352 (1996). Otro coadyuvante preferido comprende una saponina tal como Quil A, o derivados de la misma, incluyendo QS21 y QS7 (Aquila Biopharmaceuticals Inc., Framingham, MA); escina; digitonina; o saponinas de *Gypsophila* o *Chenopodium quinoa*. Otras formulaciones preferidas incluyen más de una saponina en las combinaciones de coadyuvante de la presente invención, por ejemplo, combinaciones de al menos dos del siguiente grupo que comprende QS21, QS7, quil A, β -escina o digitonina.

30 Como alternativa, las formulaciones de saponina pueden combinarse con vehículos de vacuna compuestos por quitosano u otros polímeros policatiónicos, partículas de polilactida y polilactida-co-glicolida, matriz polimérica basada en poli-N-acetilglucosamina, partículas compuestas por polisacáridos o polisacáridos modificados químicamente, liposomas y partículas basadas en lípidos, partículas compuestas por monoésteres de glicerol, etc. Las saponinas pueden formularse también en presencia de colesterol para formar estructuras particuladas tales como liposomas o ISCOM. Además, las saponinas pueden formularse junto con un éter o éster de polioxietileno en

35 una disolución o suspensión no particulada o en una estructura particulada tal como un liposoma paucilamellar o ISCOM. Las saponinas pueden formularse también con excipientes tales como Carbopol® para aumentar la viscosidad, o pueden formularse en un polvo seco con un excipiente en polvo tal como lactosa.

40 En una realización preferida, el sistema coadyuvante incluye la combinación de un monofosforil-lípido A y un derivado de saponina, tal como la combinación de QS21 y coadyuvante 3D-MPL®, como se describe en el documento WO 94/00153, o una composición menos reactogénica en que el QS21 se inactiva con colesterol, como se describe en el documento WO 96/33739. Otras formulaciones preferidas comprenden una emulsión de aceite en agua y tocoferol. Se describe en el documento WO 95/17210 otra formulación coadyuvante particularmente preferida que emplea QS21, coadyuvante 3D-MPL® y tocoferol en una emulsión de aceite en agua.

45 Otro sistema coadyuvante potenciado implica la combinación de un oligonucleótido que contiene CpG y un derivado de saponina, particularmente la combinación de CpG y QS21 como se da a conocer en el documento WO 00/09159. Preferiblemente, la formulación comprende adicionalmente una emulsión de aceite en agua y tocoferol.

50 Otros coadyuvantes preferidos incluyen Montanide ISA 720 (Seppic, Francia), SAF (Chiron, California, Estados Unidos), ISCOMS (CSL), MF-59 (Chiron), la serie de coadyuvantes SBAS (por ejemplo, SBAS-2, AS2', AS2'', SBAS-4 o SBAS6, disponibles en SmithKline Beecham, Rixensart, Bélgica), Detox (Corixa, Hamilton, MT), RC-529 (Corixa, Hamilton, MT) y otros 4-fosfatos de aminoalquilglucosaminida (AGP), tales como aquellos descritos en las solicitudes de patente de EE.UU. en tramitación de n° de serie 08/853.826 y 09/074.720 y coadyuvantes de polioxietiléneter tales como los descritos en el documento WO 99/52549A1.

Otros coadyuvantes preferidos incluyen moléculas coadyuvantes de fórmula general (I): HO(CH₂CH₂O)_n-A-R, en la que n es 1-50, A es un enlace o -C(O)-, R es alquilo C₁₋₅₀ o fenilalquilo C₁₋₅₀.

55 Una realización de la presente invención consiste en una formulación de vacuna que comprende un polioxietiléneter de fórmula general (I), en la que n es entre 1 y 50, preferiblemente 4-24, lo más preferiblemente 9; el componente R es alquilo C₁₋₅₀, preferiblemente alquilo C₄₋₂₀ y lo más preferiblemente alquilo C₁₂ y A es un enlace. La concentración de polioxietiléneteres debería estar en el intervalo de 0,1-20%, preferiblemente de 0,1-10%, y lo más preferiblemente en el intervalo de 0,1-1%. Los polioxietiléneteres preferidos se seleccionan del siguiente grupo: polioxietilén-9-

lauriléter, polioxietilen-9-esteariléter, polioxietilen-8-esteariléter, polioxietilen-4-lauriléter, polioxietilen-35-lauriléter y polioxietilen-23-lauriléter. Los polioxietilenéteres tales como polioxietilenlauriléter se describen en el Índice Merck (12ª edición, entrada 7717). Estas moléculas coadyuvantes se describen en el documento WO 99/52549.

5 El polioxietilenéter según la fórmula general (I) anterior puede combinarse, si se desea, con otro coadyuvante. Por ejemplo, es una combinación coadyuvante preferida con CpG como se describe en la solicitud de patente de RU en tramitación GB 9820956.2.

10 Puede prepararse cualquier vacuna proporcionada en la presente memoria usando métodos bien conocidos que dan como resultado una combinación de un antígeno, un potenciador de la respuesta inmunitaria y un portador o excipiente adecuado. Las composiciones descritas en la presente memoria pueden administrarse como parte de una formulación de liberación sostenida (concretamente, una formulación tal como una cápsula, esponja o gel (compuesta por polisacáridos, por ejemplo) que efectúa una liberación lenta de compuesto después de la administración. Dichas formulaciones pueden prepararse generalmente usando tecnología bien conocida (véase, por ejemplo, Coombes *et al.*, Vaccine 14: 1429-1438 (1996)) y administrarse, por ejemplo, por vía oral, rectal o implante subcutáneo, o mediante implante en el sitio diana deseado. Las formulaciones de liberación sostenida pueden 15 contener un polipéptido, polinucleótido o anticuerpo dispersado en una matriz portadora y/o contenido en un depósito rodeado por una membrana controladora de la velocidad.

20 Los portadores para uso en dichas formulaciones son biocompatibles y pueden ser también biodegradables; preferiblemente, la formulación proporciona un nivel relativamente constante de liberación de compuesto activo. Dichos portadores incluyen micropartículas de poli(lactida-co-glicolida), poliacrilato, látex, almidón, celulosa, dextrano y similares. Otros portadores de liberación retardada incluyen biovectores supramoleculares que comprenden un núcleo hidrófilo no líquido (por ejemplo, un polisacárido u oligosacárido reticulado) y, opcionalmente, una capa externa que comprende un compuesto anfífilo, tal como un fosfolípido (véanse, por ejemplo, la patente de EE.UU. nº 5.151.254 y las solicitudes PCT WO 94/20078, WO/94/23701 y WO 96/06638). La cantidad de compuesto activo contenida en una formulación de liberación sostenida depende del sitio de implantación, la velocidad y 25 duración esperada de la liberación y la naturaleza de la afección a tratar o prevenir.

30 Las vacunas y composiciones farmacéuticas pueden presentarse en envases de dosis unitaria o dosis múltiples, tales como ampollas o viales sellados. Dichos envases están preferiblemente sellados herméticamente para conservar la esterilidad de la formulación hasta su uso. En general, las formulaciones pueden almacenarse en forma de suspensiones, disoluciones o emulsiones en vehículos oleosos o acuosos. Como alternativa, puede almacenarse una vacuna o composición farmacéutica en estado liofilizado que requiere sólo la adición de un portador líquido estéril inmediatamente antes del uso.

Ejemplos

35 Los siguientes ejemplos se proporcionan sólo a modo de ilustración y no a modo de limitación. Los expertos en la materia reconocerán fácilmente una variedad de parámetros no críticos que podrían cambiarse o modificarse para proporcionar resultados esencialmente similares.

Ejemplo de antecedentes 1:

Vacunación de conejillos de Indias con la proteína de fusión MTB72F y composiciones con antígenos individuales

Se inmunizaron conejillos de Indias con coadyuvante solo (SBAS1, SBAS2, o ASAS7 más $Al(OH)_3$), proteína de fusión MTB72 en coadyuvante o TbH9 más composición antigénica Ra35.

40 **Métodos:**

Grupos: 1) SBAS1

2) SBAS2

3) SBAS7 + $Al(OH)_3$

4) TbH9+Ra35 + SBAS 1

45 5) TbH9 + Ra35 + SBAS2

6) TbH9 + Ra35 + SBAS7($Al(OH)_3$)

7) MTB72F en SBAS 1

8) MTB72F en SBAS2

9) MTB72F en SBAS7+ $Al(OH)_3$

50 10) PBS

11) BCG

Dosificación: 4 µg de cada uno de TbH9 y Ra35

8 µg de MTB72F.

5 Protocolo: 1ª inmunización, 2ª inmunización aproximadamente 3 semanas después, 3ª inmunización aproximadamente dos semanas y media después.

Antes de la exposición: DTH (hipersensibilidad de tipo retardado, usada para determinar la antigenicidad; 10 µg de antígeno)

Exposición: Aerosol con ~30 ufc de cepa Erdman

Monitorización después de la exposición: pérdida de peso y muerte (~6 meses después de la exposición).

10 *Resultados:*

1. DTH

Reacción positiva a los antígenos inmunizantes. Las reacciones a antígenos individuales o la proteína de fusión fueron comparables. La reactividad del ensayo cutáneo ante PPD se observó sólo con los grupos inmunizados con BCG.

15 2. Protección:

Los conejillos de Indias vacunados con la proteína de fusión MTB72F proporcionaban protección en comparación con los inmunizados con una mezcla de antígenos (véase la Figura 1).

Ejemplo de antecedentes 2:

Vacunación de ratón con la proteína de fusión MTB72F y composiciones con antígenos individuales

20 Como se describe anteriormente, se inmunizaron ratones con coadyuvante solo (SBAS2, SBAS2', SBAS2" o SBAS6), proteína de fusión MTB72F en coadyuvante, ADN de MTB72F, proteína de fusión MTB59F en coadyuvante o composición antigénica TbH9, Ra35 y Ra12.

Métodos:

Grupos: 1) MTB72F+ SBAS2

25 2) MTB72F + SBAS2'

3) MTB72F + SBAS2"

4) MTB72F + SBAS6

5) Ra12+ TbH9 + Ra35 en SEAS2

6) MTB59F en SBAS2

30 7) SBAS2

8) MTB72F + *M. vaccae* deslipidada desglucolipidada

9) ADN de MTB72F

10) MTB72F + IFA

11) MTB72F + BCG

35 12) *M. vaccae* deslipidada desglucolipidada

13) BCG

14) disolución salina

15) MTB72F +SBAS2 (formulación propia).

8 animales por grupo

Esquema de inmunización: primera inmunización, segunda inmunización aproximadamente 3 semanas después, tercera inmunización aproximadamente 3 semanas después.

Exposición a aerosol aproximadamente 3 meses después de la primera dosis

5 Se aislaron células de bazo o pulmón y se cultivaron; recuento de UFC de los cultivos aproximadamente 3 semanas después de la siembra.

Dosis: 8 µg de MTB72F, 6,56 µg de MTB59F, o 1,52, 4,3 y 2,24 µg, respectivamente, de Ra12, TbH9 y Ra35 mezclados.

Resultados:

10 De los coadyuvantes AS, AS2" + MTB72F dieron la mejor protección tanto en bazo como en pulmón en esta serie de experimentos (véanse las Figuras 2A y 2B). MTB72F dio ~1 log mejor protección que MTB59F tanto en bazo como en pulmón en este conjunto de experimentos, indicando que Ra12 proporciona un beneficio adicional. La mezcla de 12/H9/35 + AS2 dio una mejor protección que MTB72F en este experimento. El ADN de MTB72F dio la mejor protección en este experimento, particularmente en el bazo (>2 log). La protección era comparable en el pulmón con la de proteína MTB72F + AS2" en este experimento.

15 Ejemplo de antecedentes 3:

Vacunación de conejillos de Indias con proteína de fusión MTB72F y composiciones con antígenos individuales

Como se describe anteriormente, se inmunizaron conejillos de Indias con coadyuvante solo (SBAS2, SBAS2', SBAS2" o SBAS6), proteína de fusión MTB72F en coadyuvante, ADN de MTB72F, proteína de fusión MTB59F en coadyuvante o composición antigénica TbH9, Ra35 y Ra12.

20 *Métodos:*

Grupos: 1) MTB72F + SBAS2

2) MTB72F + SBAS2'

3) MTB72F + SBAS2"

4) MTB72F + SBAS6

25 5) Ra12+ TbH9 + Ra35 en SBAS2

6) MTB59F en SBAS2

7) SBAS2

8) MTB72F + pvac

9) ADN de MTB72F

30 10) MTB72F +IFA

11) MTB72F + BCG

12) BCG

13) disolución salina

14) *M. vaccae* deslipidada desglucolipidada

35 *Antígenos:*

Se formularon los antígenos en equivalentes molares

5 animales por grupo

El volumen de inyección por dosis es de 250 µl (IM) que contiene

MTB72F 20 µg

40 Ra12, TbH9, Ra35 3,8, 10,8 y 5,6 µg

MTB59F 16,4 µg

Esquema:

1ª inmunización, 2ª inmunización aproximadamente tres semanas después, 3ª inmunización aproximadamente tres semanas después.

Exposición: ~ Un mes y medio después de la primera inmunización.

5 *Resultados:*

~ 38 semanas después de la exposición.

Grupos	Vivos	Estado
G1. MTB72F + AS2	1/5	[pérdida de peso]
G2. MTH72F + AS2'	2/5	[sin ganancia de peso]
G3. MTB72F + AS2"	3/5	[parecen estar bien, pero sin ganancia de peso]
G4. MTB72F + AS6	2/5	[ambos con ganancia de peso]
G5. MTBRa12+H9+Ra35 +AS2	4/5	[uno quizá un poco enfermo, pero dos con ganancia]
G6. MTB59F + AS2	2/5	[ambos pierden un poco]
G7. AS2 2/5	2/5	[ambos pierden]
G8. MTB72F + pVac	1/5	[no parece demasiado bien]
G9. ADN de MTB72F	3/5	[todos se mantienen bien]
G10. MTB72F + IFA	2/5	[van bien]
G11. MTB72F + BCG	5/5	[comen muy bien]
G12 BCG	4/5	[van bien]
G13. Disolución salina	Todos muertos	
G14 pVac	2/5	[sin ganancia de peso]

A las 50 semanas después de la exposición, aunque un 80% (4/5) de los conejillos de Indias inmunizados con BCG + Mtb72F seguían vivos, sólo un 20% (1/5) de los inmunizados con BCG solo estaban vivos. A las 85 semanas, 4/5 de los conejillos de Indias inmunizados con BCG + Mtb72F seguían vivos y sanos (véase la Figura 7).

10 Ejemplo de antecedentes 4:

Protección a largo plazo

Como se describe anteriormente, se inmunizaron conejillos de Indias con coadyuvante solo (AS2 o AS2"), proteína de fusión MTB72F en coadyuvante, composición antigénica TbH9, Ra35 y Ra12, o una variedad de antígenos individuales en coadyuvante.

15 *Métodos*

GRUPOS	DOSIS DE ANTÍGENO
1. AS2" + MTB39 (TbH9)	20 µg/250 µl (IM)
2. AS2" + MTB8.4 (DPV)	20 µg
3. AS2" + MTB9.9 (MTI)	20 µg
4. AS2" + MTB41 (MTCC nº2)	20 µg
5. AS2" + MTB40 (HTCC nº1)	20 µg
6. AS2" + MTB9.8 (MSL)	20 µg
7. AS2" + MTB72F	20 µg

ES 2 374 620 T3

- 8. AS2" + Ra12+TbH9 + Ra35 (equivalente molar) 3,8 µg + 10,8 µg + 5,6 µg
- 9. AS2" + MTB71F + MTB72F+HTCC n°1 20 µg + 20 µg + 10 µg
- 10. AS2" + Ra12 20 µg
- 11. BCG
- 12. AS2"
- 13. AS2 + MTB72F
- 14. AS2+ Ra12+TbH9+Ra35
- 15. AS2

Ejemplo de antecedentes 5:

Vacunación de mono con proteína de fusión MTB72F y composiciones con antígenos individuales

5 Como se describe anteriormente, se inmunizaron monos con la proteína de fusión MTB72F en coadyuvante SBAS2, o composición antigénica MTB8.4 en coadyuvante, o una mezcla de MTB72F y MTB8.4.

Métodos:

Grupos:

- 1. Disolución salina
- 2. BCG
- 10 3. MTB8.4/AS2
- 4. MTB72F/AS2
- 5. MTB72F/AS2 (un brazo) + MTB8.4/AS2 (el otro brazo)
- 40 µg de cada antígeno

Resultados:

15 A las 8 semanas después de la exposición, los monos inmunizados con BCG muestran signos de infección.

Los datos actuales durante 16 semanas después de la exposición revelan la siguiente tendencia:

Los grupos inmunizados con MTB72F (4 y 5) mantienen sus pesos y tienen bajos valores de ESR comparados con el grupo 3 (inmunización con MTB8.4) (Tablas 1 y 2).

Tabla 1

20 Estudio de vacuna profiláctica en monos cinomolgos con MTB8.4 y MTB72F formuladas en AS2 20 semanas después de la exposición

Grupos	ID	Cambio de peso neto (kg)	Rayos X de tórax (inicio)	Estado
AS2	1398K	-24%	Neu, bil, prog (sem 8)	Vivo
	4437B	-33%	Neu, bil, prog (sem 4)	Muerto
	2959G	-8,30%	Neu, bil, prog (sem 4)	Vivo
	605AE	-14,00%	Neu, der, estable (sem 8)	Vivo
BCG	3436A	-15,00%	Neg.	Vivo
	3642G	Más 4,5%	Neu, der, prog (sem 8)	Vivo
	1190H	0%	Neg.	Vivo

ES 2 374 620 T3

Grupos	ID	Cambio de peso neto (kg)	Rayos X de tórax (inicio)	Estado
	10511	-30%	Neu, der, prog (sem 8)	Muerto
MTB8.4	3665C	-25%	Neu, der, prog (sem 8)	Muerto
	2200F	-18,00	Neu, der, estable (sem 8)	Vivo
	1654J	-33,00%	Neu, der, prog (sem 4)	Muerto
	4141C	-33%	Neu, der, prog (sem 4)	Muerto
MTB72F	3061C	Muerto después de la exposición IT		
	1228G	Más 3,6%	Bron, bil, estable durante 3 meses (sem 8)	Vivo
	3462E	-2,20%	Neg.	Vivo
	4254C	Más 1,21	Neu, der, estable durante 3 meses (sem 4)	Vivo
	4496A	Más 7%	Neu, der, estable durante 1 mes (sem 8)	Vivo
MTB8.4	4422C	-39,00%	Neu, bil, prog (sem 4)	Muerto
MTB72F	4416A	Más 11%	Neu, der, estable durante 2 meses (sem 12)	Vivo
	2734E	Más 12,5%	Sosp infil der, estable durante 3 meses (sem 8)	Vivo

Tabla 2

Estudio de vacuna profiláctica en monos cinomolgos con MTB8.4 y MTB72F formuladas en AS2

Semanas después de la exposición						
Grupos	ID	ESR				Rayos X de tórax de 16 semanas
		4	8	12	16	
AS2	1398K	3	3	10	19	Neu, bil progr.
	4437B	10	20	3		Muerto
	2959G	6	3	3	0	Neu, der, progr.
	605AE	1	4	7	3	Neu, der, estable
	3436A	0	8	7	15	Neg.
BCG	3642G	0	0	0	0	Neu, der, progr.
	1190H	1	0	2	0	Neg.
	10511	0	8	22	7	Neu, bil, con progr. ad. Muerto
	3665C	12	30	19		Muerto
MTB8.4	2200F	1	7	2	0	Neu, der, prog
	1654J	20	8	21	7	Neu, bi. c/prog. ad.
	4141C	13	8	2	15	Neu, bi, c/prog. ad.
	3061C*	Murió después de la exposición IT				
MTB72F	1228G	0	1	20	0	Ahora estable
	3462E	0	0	0	0	Neg.
	4254C	13	0	0	0	Neu, ahora estable

ES 2 374 620 T3

	4496A	5	1	0	5	Neu, der, c/prog. ad.
MTB8.4I	4422C	10	3	0		Muerto
MTB72F	4416A	6	0	1	0	Neu, ahora estable
	2734E	0	0	0	0	Sosp inf, ahora estable

Ejemplo de antecedentes 6:

Experimento de cebado con BCG en monos

5 Se inmunizaron con BCG 4 grupos de 5 animales por grupo y después reposo, se inmunizaron entonces como se describe anteriormente y se expusieron. Se usará el siguiente protocolo:

Grupos	Nº de animales	de Antígeno inmunizante	Dosis de antígeno
1. Nada	5	AS2	
2. BCG	5	AS2	
3. BCG	5	MTB72F	40 µg
4. BCG	4	Ra12+TbH9+Ra3	Equiv. molar de antígenos en la dosis de MTB72F
5. BCG	4	MTB72F + MTB71F + MTB40	40 µg de MTB72F 40 µg de MTB71F 20 µg de MTB40

Todos los antígenos se formulan en AS2.

Los grupos 4 y 5 tienen 4 animales cada uno. Dos de los monos inmunizados con BCG murieron.

Grupos	Nº de animales	Antígeno inmunizante	Antígenos para ensayos de proliferación de linfocitos T y producción de citocina
1. Nada	5	AS2	PHA, PPD, MTB72F, MTB71F, HTCC n°1, DPV, MTCC n°2, Ra12, TbH9, Ra35, MSL, MTI
2. BCG	5	AS2	PHA, PPD, MTB72F, MTB71F, HTCC n°1, DPV, MTCC n°2, Ra12, TbH9, Era35, MSL, MTI
3. BCG	5	MTB72F	PHA, PPD, MTB72F, Ra12, TbH9, Ra35
4. BCG	4	Ra12+TbH9+Ra35	PHA, PPD, MTB72F, Ra12, TbH9, Ra35
5. BCG	4	MTB72F + MTB71F + MTB40	PHA, PPD, MTB72F, MTB71F, HTCC n°1, DPV, MTCC-2, Ra12, TbH9, Ra35, MSL, MTI

Ejemplo 7: Construcción de Ra35MutSA y MTB72FMutSA

10 La expresión de Mtb72f da típicamente como resultado ciertos productos de degradación. Además, la expresión de las secuencias completas de la forma madura o completa de Ra35 (Mtb32A) en *E. coli* ha sido difícil. El producto expresado era sólo visible después de inmunotransferencia con un Ab policlonal de conejo anti-Ra35, indicativo de bajos niveles de expresión de proteína. Incluso entonces, se detectaron múltiples especies específicas (bandas)

indicativas de degradación autocatalítica del antígeno recombinante. Esto se supuso debido a la expresión de Ra35FL en *E. coli* en forma biológicamente activa.

5 Se ha mostrado anteriormente que era posible expresar Ra35FL en forma de dos mitades superpuestas que comprenden la mitad N-terminal (extremo N de Ra35, denominado Ra35) y la mitad C-terminal (extremo C de Ra35, denominado Ra12). Para potenciar y estabilizar la expresión de toda la molécula de Ra35, se introdujo una sola mutación puntual en uno de los residuos en la triada de sitio activo (sustitución de Ser por Ala; véanse la Figuras 6). Esta forma mutagenizada de Mtb32A puede expresarse ahora fácilmente a altos niveles en forma estable. Además, para estabilizar la expresión de Mtb72F, se incorporó una sola sustitución nucleotídica (T a G, dando como resultado un cambio de Ser a Ala en la posición 710 del polipéptido de fusión) en la secuencia de Mtb72F en la posición nucleotídica 2128 (véase la Figura 5).

Esta estabilización se consigue también fácilmente mutagenizando uno cualquiera, dos o los tres residuos que comprenden la triada de sitio activo en Ra35FL, Ra35 o Mtb72F u otras proteínas de fusión que comprenden Ra35 (His, Asp o Ser). La mutagénesis puede efectuarse usando cualquier técnica conocida por el experto en la materia.

Ejemplo 8: Inmunización de ratones con Ra35FLMutSA-TbH9 y MTB72FMutSA

15 Se inmunizaron 8 ratones por grupo con las composiciones enumeradas a continuación, que incluyen el coadyuvante AS2A. Se expusieron entonces los ratones a *Mycobacterium tuberculosis* y se midió la supervivencia de los ratones.

Grupo	Concentración de proteína o ADN
1. Proteína Mtb72f	1,5 mg/ml
2. ADN de Mtb72f	1,2 mg/ml
3. Proteína Mtb72f-85b	0,6 mg/ml
4. ADN de Mtb72f-85b	1,1 mg/ml
5. Proteína Mtb72f-MTI	1,3 mg/ml
6. ADN de Mtb72f-MTI	1,1 mg/ml
7. Proteína Mtb72fMutSA	1,7 mg/ml
8. Proteína MTB3AMutSA-TbH9	2,4 mg/ml
9. BCG	
10. AS2	
11. Vector solo	1,5 mg/ml
12. Disolución salina	

Listado de secuencias

- 20
- <110> Skeiky, Yasir
 Reed, Steven
 Alderson, Mark
 Corixa Corporation
- 25 <120> Proteínas de fusión de *Mycobacterium tuberculosis*
- <130> CRX-P457EP3D1
- <150> US 09/597.796
- <151> 20-06-2000
- <150> US 60/265.737

ES 2 374 620 T3

<151> 01-02-2001
 <160> 49
 <170> PatentIn versión 2.1
 5 <210> 1
 <211> 1872
 <212> ADN
 <213> *Mycobacterium tuberculosis*

10 <220>
 <223> MTB32A (Ra35FL)
 <220>
 <221> base_modificada
 <222> (1)...(1872)

15 <223> n= g, a, c o t

<400> 1
 gactacggtg gtgtagaaaa atcctgcccg cgggaccctt aaggctggga caatttctga 60
 tagctacccc gacacaggag gttacgggat gagcaattcg cgccgcccgt cactcaggtg 120
 gtcattgggtg ctgagcgtgc tggctgccgt cgggctgggc ctggccacgg cgccggccca 180
 ggcggcccccg ccggccttgt cgcaggaccg gttcgcggac ttccccgcgc tgcccctcga 240
 cccgtccgcg atggctcgccc aagtggcgcc acagggtggtc aacatcaaca ccaaactggg 300
 ctacaacaac gccgtggggc cggggaccgg catcgtcatc gatcccaacg gtgtcgtgct 360
 gaccaacaac cacgtgatcg cgggcgccac cgacatcaat gcgttcagcg tcggctccgg 420
 ccaaacctac ggcgtcgatg tggctgggta tgaccgcacc caggatgtcg cggtgctgca 480
 gctgctcggt gccggtggcc tggctcggc ggcgatcggg ggccggcgtc cggtttgtga 540
 gccgctcgtc gcgatgggca acagcgggtg gcagggcggg acgcccctg cggtgccctg 600
 caggggtggtc gcgctcggcc aaaccgtgca ggcgtcggat tcgctgaccg gtgccgaaga 660
 gacattgaac gggttgatcc agttcgatcg cgcaatccag cccggtgatt cgggcggggc 720
 cgctcgtcaac ggcctaggac aggtggctcg tatgaacacg gccgctccg ataacttcca 780
 gctgtcccag ggtgggcagg gattcgccat tccgatcggg caggcgtggt cgatcgcggg 840
 ccaaatccga tcgggtgggg ggtcaccac cgttcataat gggcctaccg ccttccctcg 900
 cttgggtggt gtcgacaaca acggcaacgg cgcacgagtc caacgcgtg tcggaagcgc 960
 tccggcggca agtctcggca tctccaccgg cgacgtgat accgcggtc acggcgcctc 1020
 gatcaactcg gccaccgca tggcggacgg gcttaacggg catcatccc gtgacgtcat 1080
 ctcggtgaac tggcaaacca agtcgggagg cacgcgtaca gggaaactga cattggccga 1140
 gggacccccg gcctgatttg tcgcggtatc caccgcggg ccggccaatt ggattggcgc 1200
 cagccgtgat tgccgctgga gcccccgagt tccgtctccc gtgcgcgtgg cattgtggaa 1260
 gcaatgaacg aggcagaaca cagcgttgag caccctccc tgaggggcag ttacgtcga 1320
 ggcgggtggt tcgagcatcc ggatgccaag gacttcggca gcgcccggc cctgcccggc 1380
 gatccgacct ggtttaagca cggcgtctt tacgaggtg tggctccggg gttcttcgac 1440
 gccagcgcgg acggttccgn cgatctgctg ggactcatc atcgccctga ctacctgac 1500
 tggcttggca tcgactgcat ctggtgccgc cgttcctacg actcaccgct gcgagcggc 1560
 ggttacgaca ttcgagact ctacaagggt ctgcccgaat tcggaccgt cgacgatcc 1620
 gtcgcccctg tcgacaccgc tcaccggcga ggtatccgca tcatcaccga cctgggtgat 1680
 aatcacacct cggagtcgca cccctggtt caggagtccc gccgcgacc agacggacc 1740
 tacgggtgact attacgtgtg gagcgacacc agcagcgtc acaccgacg ccggatcatc 1800
 ttcgtcgaca ccgaagagtc gaactggtca ttcgatcctg tccgccgaca gtttactg 1860
 gcaccgattc tt 1872

<210> 2
 20 <211> 355
 <212> PRT

<213> *Mycobacterium tuberculosis*

<220>

<223> MTB32A (Ra35FL)

5

<400> 2

```

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

Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His Val
85      90      95
Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln
100     105     110
Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val Ala
115     120     125
Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile Gly
130     135     140
Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser Gly
145     150     155     160
Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala Leu
165     170     175
Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu Thr
180     185     190
Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp Ser
195     200     205
Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn Thr
210     215     220
Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe Ala
225     230     235     240
Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser Gly
245     250     255
Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly Leu
260     265     270
Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val Val
275     280     285
Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val Ile
290     295     300
Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala Asp
305     310     315     320
Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp Gln
325     330     335
Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu Gly
340     345     350
Pro Pro Ala
355

```

<210> 3

ES 2 374 620 T3

<211> 1002
 <212> ADN
 <213> *Mycobacterium tuberculosis*

5 <220>
 <223> MTB32A (Ra35 madura)

<400> 3

```

catatgcatc accatcacca tcacgccccg ccggccttgt cgcaggaccg gttcgccgac 60
ttccccgcgc tgcccctcga cccgtccgcg atggtcgccc aagtggggcc acaggtggtc 120
aacatcaaca ccaaactggg ctacaacaac gccgtgggcg cggggaccgg catcgtcatc 180
gateccaacg gtgtcgtgct gaccaacaac cacgtgatcg cgggcgccac cgacatcaat 240
gcgttcagcg tcggctccgg ccaaactac ggcgtcgatg tggtcgggta tgaccgcacc 300
caggatgtcg cggtgctgca gctgcgcggt gccggtggcc tgccgtcggc ggcgatcggc 360
ggcggcgtcg cggttggtga gcccgctcgc gcgatgggca acagcgggtg gcagggcgga 420
acgccccgtg cggtcctcgg cagggtggtc gcgctcggcc aaaccgtgca ggcgtcggat 480
tcgctgaccg gtgccgaaga gacattgaac gggttgatcc agttcgatgc cgcgatccag 540
cccggtgagg cgggcggggc cgtcgtcaac ggcctaggac aggtggtcgg tatgaacacg 600
gccgctccg ataactcca gctgtcccag ggtgggcagg gattcgccat tccgatcggg 660
caggcgatgg cgatcgcggg ccagatccga tcgggtgggg ggtcaccac cgttcatatc 720
gggcctaccg ccttcctcgg cttgggtggt gtcgacaaca acggcaacgg cgacagatc 780
caacgcgtgg tcgggagcgc tccggcggca agtctcggca tctccaccgg cgacgtgatc 840
accgcggtcg acggcgctcc gatcaactcg gccaccgcga tggcggacgc gcttaacggg 900
catcatcccg gtgacgtcat ctcggtgacc tggcaaacca agtcgggcgg cacgcgtaca 960
gggaacgtga cattggccga gggacccccc gcctgagaat tc 1002
    
```

10 <210> 4
 <211> 330
 <212> PRT
 <213> *Mycobacterium tuberculosis*

15 <220>
 <223> MTB32A (Ra35 madura)

<400> 4

ES 2 374 620 T3

Met His His His His His His Ala Pro Pro Ala Leu Ser Gln Asp Arg
1 5 10 15
Phe Ala Asp Phe Pro Ala Leu Pro Leu Asp Pro Ser Ala Met Val Ala
20 25 30
Gln Val Gly Pro Gln Val Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn
35 40 45
Asn Ala Val Gly Ala Gly Thr Gly Ile Val Ile Asp Pro Asn Gly Val
50 55 60
Val Leu Thr Asn Asn His Val Ile Ala Gly Ala Thr Asp Ile Asn Ala
65 70 75 80
Phe Ser Val Gly Ser Gly Gln Thr Tyr Gly Val Asp Val Val Gly Tyr
85 90 95
Asp Arg Thr Gln Asp Val Ala Val Leu Gln Leu Arg Gly Ala Gly Gly
100 105 110
Leu Pro Ser Ala Ala Ile Gly Gly Gly Val Ala Val Gly Glu Pro Val
115 120 125
Val Ala Met Gly Asn Ser Gly Gly Gln Gly Gly Thr Pro Arg Ala Val
130 135 140
Pro Gly Arg Val Val Ala Leu Gly Gln Thr Val Gln Ala Ser Asp Ser
145 150 155 160
Leu Thr Gly Ala Glu Glu Thr Leu Asn Gly Leu Ile Gln Phe Asp Ala
165 170 175
Ala Ile Gln Pro Gly Asp Ser Gly Gly Pro Val Val Asn Gly Leu Gly
180 185 190
Gln Val Val Gly Met Asn Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser
195 200 205
Gln Gly Gly Gln Gly Phe Ala Ile Pro Ile Gly Gln Ala Met Ala Ile
210 215 220
Ala Gly Gln Ile Arg Ser Gly Gly Gly Ser Pro Thr Val His Ile Gly
225 230 235 240
Pro Thr Ala Phe Leu Gly Leu Gly Val Val Asp Asn Asn Gly Asn Gly
245 250 255
Ala Arg Val Gln Arg Val Val Gly Ser Ala Pro Ala Ala Ser Leu Gly
260 265 270
Ile Ser Thr Gly Asp Val Ile Thr Ala Val Asp Gly Ala Pro Ile Asn
275 280 285
Ser Ala Thr Ala Met Ala Asp Ala Leu Asn Gly His His Pro Gly Asp
290 295 300
Val Ile Ser Val Thr Trp Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly
305 310 315 320
Asn Val Thr Leu Ala Glu Gly Pro Pro Ala
325 330

<210>

5

ES 2 374 620 T3

<211> 1002

<212> ADN

<213> Secuencia artificial

5 <220>

<223> Descripción de la secuencia artificial: Ra35FLMutSA

<400> 5

catatgcatc accatcacca tcacgccccg cgggccttgt cgcaggaccg gttcgcggac 60
ttccccgcgc tgcccctcga cccgtccgcg atggtcgccc aagtggggcc acaggtggtc 120
aacatcaaca ccaaactggg ctacaacaac gccgtgggcg ccgggaccgg catcgtcatc 180

10

gatcccaacg gtgtcgtgct gaccaacaac cacgtgatcg cgggcgccac cgacatcaat 240
gcggttcagcg tcggctccgg ccaaacctac ggcgtcgatg tggtcgggta tgaccgcacc 300
caggatgtcg cggtgctgca gctgcgcggt gccggtggcc tgccgtcggc ggcgatcggg 360
ggcggcgtcg cggttggtga gcccgctcgtc gcgatgggca acagcgggtg gcagggcggg 420
acgccccgtg cggtgccctgg caggggtggtc gcgctcggcc aaaccgtgca ggcgtcggat 480
tcgctgaccg gtgccgaaga gacattgaac gggttgatcc agttcgatgc cgcgatccag 540
cccggtgatg cgggcggggc cgtcgtcaac ggcctaggac aggtggtcgg tatgaacacg 600
gccgctccg ataacttcca gctgtcccag ggtgggcagg gattcgccat tccgatcggg 660
caggcgatgg cgatcgcggg ccagatccga tcgggtgggg ggtcaccac cgttcatatc 720
gggcctaccg ccttcctcgg ctgggtggt gtcgacaaca acggcaacgg cgcacgagtc 780
caacgcgtgg tcgggagcgc tccggcggca agtctcggca tctccaccgg cgacgtgatc 840
accgcggtcg acggcgtcc gatcaactcg gccaccgcga tggcggacgc gcttaacggg 900
catcatcccg gtgacgtcat ctcggtgacc tggcaaacca agtcgggcgg cacgcgtaca 960
gggaacgtga cattggccga gggacccccg gcctgagaat tc 1002

<210> 6

<211> 330

<212> PRT

15 <213>

Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Ra35FLMutSA

20 <400>

6

Met His His His His His His Ala Pro Pro Ala Leu Ser Gln Asp Arg
 1 5 10 15
 Phe Ala Asp Phe Pro Ala Leu Pro Leu Asp Pro Ser Ala Met Val Ala
 20 25 30
 Gln Val Gly Pro Gln Val Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn
 35 40 45
 Asn Ala Val Gly Ala Gly Thr Gly Ile Val Ile Asp Pro Asn Gly Val
 50 55 60
 Val Leu Thr Asn Asn His Val Ile Ala Gly Ala Thr Asp Ile Asn Ala
 65 70 75 80
 Phe Ser Val Gly Ser Gly Gln Thr Tyr Gly Val Asp Val Val Gly Tyr
 85 90 95
 Asp Arg Thr Gln Asp Val Ala Val Leu Gln Leu Arg Gly Ala Gly Gly
 100 105 110
 Leu Pro Ser Ala Ala Ile Gly Gly Gly Val Ala Val Gly Glu Pro Val
 115 120 125
 Val Ala Met Gly Asn Ser Gly Gly Gln Gly Gly Thr Pro Arg Ala Val
 130 135 140
 Pro Gly Arg Val Val Ala Leu Gly Gln Thr Val Gln Ala Ser Asp Ser
 145 150 155 160
 Leu Thr Gly Ala Glu Glu Thr Leu Asn Gly Leu Ile Gln Phe Asp Ala
 165 170 175
 Ala Ile Gln Pro Gly Asp Ala Gly Gly Pro Val Val Asn Gly Leu Gly
 180 185 190
 Gln Val Val Gly Met Asn Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser
 195 200 205
 Gln Gly Gly Gln Gly Phe Ala Ile Pro Ile Gly Gln Ala Met Ala Ile
 210 215 220
 Ala Gly Gln Ile Arg Ser Gly Gly Gly Ser Pro Thr Val His Ile Gly
 225 230 235 240
 Pro Thr Ala Phe Leu Gly Leu Gly Val Val Asp Asn Asn Gly Asn Gly
 245 250 255
 Ala Arg Val Gln Arg Val Val Gly Ser Ala Pro Ala Ala Ser Leu Gly
 260 265 270
 Ile Ser Thr Gly Asp Val Ile Thr Ala Val Asp Gly Ala Pro Ile Asn
 275 280 285
 Ser Ala Thr Ala Met Ala Asp Ala Leu Asn Gly His His Pro Gly Asp
 290 295 300
 Val Ile Ser Val Thr Trp Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly
 305 310 315 320
 Asn Val Thr Leu Ala Glu Gly Pro Pro Ala
 325 330

<210> 7

<211> 585

ES 2 374 620 T3

<212> ADN

<213> *Mycobacterium tuberculosis*

<220>

5 <223> Extremo N Ra35 de MTB32A (Ra35FL)

<400> 7

```

gccccgccgg ccttgtcgca ggaccggttc gccgacttcc ccgcgctgcc cctcgacccg 60
tccgcgatgg tcgccaagt ggggccacag gtggtcaaca tcaacaccaa actgggctac 120
aacaacgccg tgggcgccgg gaccggcatc gtcatcgatc ccaacggtgt cgtgctgacc 180
aacaaccacg tgatcgccgg cgccaccgac atcaatgctg tcagcgtcgg ctccggccaa 240
acctacggcg tcgatgtggt cgggtatgac cgcaccagg atgtcgcggt gctgcagctg 300
cgcggtgccg gtggcctgcc gtcggcggcg atcgggtggc gcgtcgcggt tggtgagccc 360
gtcgtcgcga tgggcaacag cgggtggcag ggcggaacgc cccgtgcggt gcctggcagg 420
gtggtcgcgc tcggccaaac cgtgcaggcg tcggattcgc tgaccggtgc cgaagagaca 480
ttgaacgggt tgatccagtt cgatgcccg atccagccc gtgaggcggg cgggcccgtc 540
gtcaacggcc taggacaggt ggtcggtatg aacacggccg cgfcc 585
    
```

<210> 8

10 <211> 195

<212> PRT

<213> *Mycobacterium tuberculosis*

<220>

15 <223> Extremo N Ra35 de MTB32A (Ra35FL)

<400> 8

ES 2 374 620 T3

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

<210> 9

<211> 447

5 <212> ADN

<213> *Mycobacterium tuberculosis*

<220>

<223> Extremo C Ra12 o MTBRa12 de MTB32A (Ra35FL)

10

<400> 9

```

cggtatgaac  acggccgcgt  ccgataaactt  ccagctgtcc  caggggtgggc  agggattcgc  60
cattccgatac  gggcaggcga  tggcgatcgc  gggccagatc  cgatcgggtg  gggggtcacc  120
caccgttcat  atcgggccta  ccgccttctc  cggcttgggt  gttgtcgaca  acaacggcaa  180
cggcgcacga  gtccaacgcg  tggtcgggag  cgctccggcg  gcaagtctcg  gcatctccac  240
cggcgacgtg  atcaccgcg  tcgacggcgc  tccgatcaac  tcggccaccg  cgatggcgga  300
cgcgcttaac  gggcatcatc  ccggtgacgt  catctcggtg  aactggcaaa  ccaagtcggg  360
cggcacgcgt  acagggaaac  tgacattggc  cgagggacc  ccggcctgat  ttcgtcgygg  420
ataccaccgg  ccggccggcc  aattgga  447
  
```

<210> 10

ES 2 374 620 T3

<211> 132
 <212> PRT
 <213> *Mycobacterium tuberculosis*

5 <220>
 <223> Extremo C Ra12 o MTBRa12 de MTB32A (Ra35FL)

<400> 10
 Thr Ala Ala Ser Asp Asn Phe Gln Leu Ser Gln Gly Gly Gln Gly Phe
 1 5 10 15
 Ala Ile Pro Ile Gly Gln Ala Met Ala Ile Ala Gly Gln Ile Arg Ser
 20 25 30
 Gly Gly Gly Ser Pro Thr Val His Ile Gly Pro Thr Ala Phe Leu Gly
 35 40 45
 Leu Gly Val Val Asp Asn Asn Gly Asn Gly Ala Arg Val Gln Arg Val
 50 55 60
 Val Gly Ser Ala Pro Ala Ala Ser Leu Gly Ile Ser Thr Gly Asp Val
 65 70 75 80
 Ile Thr Ala Val Asp Gly Ala Pro Ile Asn Ser Ala Thr Ala Met Ala
 85 90 95
 Asp Ala Leu Asn Gly His His Pro Gly Asp Val Ile Ser Val Asn Trp
 100 105 110
 Gln Thr Lys Ser Gly Gly Thr Arg Thr Gly Asn Val Thr Leu Ala Glu
 115 120 125
 Gly Pro Pro Ala
 130

10 <210> 11
 <211> 851
 <212> ADN
 <213> *Mycobacterium tuberculosis*

15 <220>
 <223> MTB39 (TbH9)
 <220>
 <221> base_modificada
 <222> (767)

20 <223> n= g, a, c o t
 <400> 11

ES 2 374 620 T3

```

ctgcagggtg gcgtggatga gcgtcaccgc ggggcaggcc gagctgaccg ccgcccaggt 60
ccgggttgct gcggcggcct acgagacggc gtatgggctg acggtgcccc cgccggatgat 120
cgccgagaac cgtgctgaac tgatgattct gatagcgacc aacctcttgg ggcaaaacac 180
cccggcgatc gcggtcaacg aggccgaata cggcgagatg tgggcccaag acgccgccgc 240
gatgtttggc tacgccgcgg cgacggcgac ggcgacggcg acggtgctgc cgttcgagga 300
ggcggccggag atgaccagcg cgggtgggct cctcgagcag gccgcccggg tcgaggaggc 360
ctccgacacc gccgcggcga accagttgat gaacaatgtg ccccaggcgc tgaaacagtt 420
ggcccagccc acgcagggca ccacgccttc ttccaagctg ggtggcctgt ggaagacggt 480
ctcggccgat cggtcgccga tcagcaacat ggtgtcgatg gccacaacc acatgtcgat 540
gaccaactcg ggtgtgtcga tgaccaacac cttgagctcg atgttgaagg gctttgctcc 600
ggcggcggcc gcccaggccg tgcaaaccgc ggcgcaaaac ggggtccggg cgatgagctc 660
gctgggcagc tcgctgggtt cttcgggtct gggcggtggg gtggccgcca acttgggtcg 720
ggcggcctcg gtacggtatg gtcaccggga tggcgaaaaa tatgcanagt ctggtcggcg 780
gaacggtggt ccggcgtaag gtttaccccc gttttctgga tgcggtgaac ttcgtcaacg 840
gaaacagtta c

```

<210> 12

<211> 263

<212> PRT

5 <213> *Mycobacterium tuberculosis*

<220>

<223> MTB39 (TbH9)

<220>

10 <221> MOD_RES

<222> (254)

<223> Xaa= cualquier aminoácido

<400> 12

ES 2 374 620 T3

Val Ala Trp Met Ser Val Thr Ala Gly Gln Ala Glu Leu Thr Ala Ala
 1 5 10 15
 Gln Val Arg Val Ala Ala Ala Ala Tyr Glu Thr Ala Tyr Gly Leu Thr
 20 25 30
 Val Pro Pro Pro Val Ile Ala Glu Asn Arg Ala Glu Leu Met Ile Leu
 35 40 45
 Ile Ala Thr Asn Leu Leu Gly Gln Asn Thr Pro Ala Ile Ala Val Asn
 50 55 60
 Glu Ala Glu Tyr Gly Glu Met Trp Ala Gln Asp Ala Ala Ala Met Phe
 65 70 75 80
 Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr Ala Thr Leu Leu Pro Phe
 85 90 95
 Glu Glu Ala Pro Glu Met Thr Ser Ala Gly Gly Leu Leu Glu Gln Ala
 100 105 110
 Ala Ala Val Glu Glu Ala Ser Asp Thr Ala Ala Ala Asn Gln Leu Met
 115 120 125
 Asn Asn Val Pro Gln Ala Leu Lys Gln Leu Ala Gln Pro Thr Gln Gly
 130 135 140
 Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu Trp Lys Thr Val Ser Pro
 145 150 155 160
 His Arg Ser Pro Ile Ser Asn Met Val Ser Met Ala Asn Asn His Met
 165 170 175
 Ser Met Thr Asn Ser Gly Val Ser Met Thr Asn Thr Leu Ser Ser Met
 180 185 190
 Leu Lys Gly Phe Ala Pro Ala Ala Ala Ala Gln Ala Val Gln Thr Ala
 195 200 205
 Ala Gln Asn Gly Val Arg Ala Met Ser Ser Leu Gly Ser Ser Leu Gly
 210 215 220
 Ser Ser Gly Leu Gly Gly Val Ala Ala Asn Leu Gly Arg Ala Ala
 225 230 235 240
 Ser Val Arg Tyr Gly His Arg Asp Gly Gly Lys Tyr Ala Xaa Ser Gly
 245 250 255
 Arg Arg Asn Gly Gly Pro Ala
 260

<210> 13

<211> 3058

<212> ADN

5 <213> *Mycobacterium tuberculosis*

<220>

<223> MTB39 (TbH9FL)

10 <400> 13

```

gatcgtaccc gtgcgagtgc tcgggccggt tgaggatgga gtgcacgtgt ctttcgtgat 60
ggcataccca gagatgttgg cggcggcggc tgacaccctg cagagcatcg gtgctaccac 120
tgtggctagc aatgccgctg cggcggcccc gacgactggg gtgggtgcccc ccgctgccga 180
tgagggtgctg gcgctgactg cggcgcactt cgccgcacat gcggcgatgt atcagtcctg 240
gagcgtcctg gctgctgca ttcattgacca gttcgtggcc acccttgcca gcagcggcag 300
ctcgtatgcg gccactgaag tcgccaatgc ggcggcggcc agctaagcca ggaacagtcg 360
gcacgagaaa ccacgagaaa tagggacacg taatggtgga tttcggggcg ttaccaccgg 420
agatcaactc cgcgaggatg tacgccggcc cgggttcggc ctcgctgggt gcccgggctc 480
agatgtggga cagcgtggcg agtgacctgt tttcggccgc gtcggcgttt cagtcgggtg 540
tctggggctc gacggtgggg tcgtggatag gttcgtcggc gggctctgat gtggcggcgg 600
cctcggccta tgtggcgtgg atgagcgtca ccgcggggca ggccgagctg acccggccc 660
aggtcgggtg tgctgcggcg gcctacgaga cggcgtatgg gctgacggtg cccccggcg 720
tgatcgccga gaaccgtgct gaactgatga ttctgatagc gaccaacctc ttggggcaaa 780
acacccggc gatcgcggtc aacgaggccg aatacggcga gatgtgggcc caagacggcg 840
ccgcgatgtt tggctacgcc gcggcgacgg cgacggcgac ggcgacggtg ctgccgttcg 900
aggaggcggc ggagatgacc agcgcgggtg ggctcctcga gcaggccgcc gcggtcgagg 960
agcctccga caccgcgcg gcgaaccagt tgtagaaca tgtgccccag cgctgcaac 1020
agctggccca gcccacgag ggcaccacgc cttcttcaa gctgggtggc ctgtggaaga 1080
cggctcggc gcacggtcg ccgatcagca acatggtgtc gatggccaac aaccacatgt 1140
cgatgaccaa ctcgggtgtg tcgatgacca acacctgag ctcgatgttg aagggtttg 1200
ctccggcggc ggccgcccag gccgtgcaaa ccgcggcgca aaacggggtc cgggcgatga 1260
gctcgtggg cagctcgtg gttcttcgg gtctggggc tgggggtggc gccaacttg 1320
gtcggcggc ctcggtcgg tcgtgtcgg tgccgcagg cggggcggc gccaaccagg 1380
cagtcacccc ggcggcggc gcgctgccc tgaccagcct gaccagcgc gcgaaagag 1440
ggcccggca gatgctggg gggctgccc tggggcagat gggcgccagg gccggtggg 1500
ggctcagtg tgtgctgct gttcgcggc gaccctatgt gatgccgat tctccggcg 1560
ccggctagga gaggggggc agactgtcgt tatttgacca gtgatcggcg gtctcgggt 1620
ttccgcgcc ggctatgaca acagtcaatg tgatgacaa gttacaggt ttaggtccag 1680
gttcaacaag gagacaggca acatggcctc acgttttatg acggatccgc acgcatgag 1740
ggacatggc ggccgttttg aggtgcacgc ccagacggtg gaggacgagg ctgccggat 1800
gtggcgctc gcgcaaaa tttccggtg gggctggagt ggcattggcc agggacctc 1860
gctagacacc atggcccaga tgaatcaggc gtttcgcaac atcgtgaaca tgctgcacg 1920
ggtgcgtgac gggctgggtc gcgacgcaa caactacgag cagcaagagc aggcctcca 1980
gcagatcctc agcagctaac gtcagccgct gcagcaaat acttttcaa gcaaggaga 2040
acaggttcga tgaccatcaa ctatcaattc ggggatgtc acgctcacg cgccatgat 2100
cgcgctcagg ccgggttgct ggaggccgag catcaggcca tcattcgtg tgtgttgac 2160
gcgagtgact tttggggcgg cgccggttc gcggcctgcc aggggttcat taccagttg 2220
ggccgtaact tccaggtgat ctacgagcag gccaacgccc acgggcagaa ggtgcaggc 2280
gccggcaaca acatggcgca aaccgacagc gccgtcggct ccagctggc ctgacaccag 2340

gccaaggcca gggacgtggt gtacgagtga agttcctcgc gtgatccttc ggggtggcagt 2400
ctaagtggc agtgctgggg tgttgggtgt ttgctgctg gcgggttctt cgggtgctggt 2460
cagtgctgct cgggctcggg tgaggacct gagggccagg tagcggcctc cttcgatcca 2520
ttcgtcgtgt tgttcggcga ggacggctcc gacgaggcgg atgatcgagg cgcggtcggg 2580
gaagatgccc acgacgtcgg ttcggcctcg tacctctcgg ttgaggcgtt cctgggggtt 2640
ggtggaccag atttggcgc agatctgctt ggggaaggcg gtgaacgcca gcaggctggt 2700
gcgggcggtg tcgaggtgct cggccaccgc ggggagttg tcggtcagag cgtcgagtac 2760
ccgatcatat tgggcaacaa ctgattcggc gtcgggctgg tcgtagatgg agtgcagcag 2820
ggtgcgcacc cacggccagg agggcttcgg ggtggctgcc atcagattgg ctgctagtg 2880
ggttctgcag cgctgccagg ccgctgccc cagggtggcg ccgatcgcgg ccaccaggcc 2940
ggcgtgggcg tcgctggtga ccagcgcgac cccggacagg ccgcggcgga ccaggtcgcg 3000
gaagaacgcc agccagccgg ccccgctcct ggcggaggtg acctggatgc ccaggatc 3058

```

<210> 14

<211> 391

5 <212> PRT

<213> *Mycobacterium tuberculosis*

<220>

<223> MTB39 (TbH9FL)

ES 2 374 620 T3

<400>

14

Met Val Asp Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met
 1 5 10 15
 Tyr Ala Gly Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Gln Met Trp
 20 25 30
 Asp Ser Val Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser
 35 40 45
 Val Val Trp Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly
 50 55 60
 Leu Met Val Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr
 65 70 75 80
 Ala Gly Gln Ala Glu Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala
 85 90 95
 Ala Tyr Glu Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala
 100 105 110

Glu Asn Arg Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly
 115 120 125
 Gln Asn Thr Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met
 130 135 140
 Trp Ala Gln Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala
 145 150 155 160
 Thr Ala Thr Ala Thr Leu Leu Pro Phe Glu Glu Ala Pro Glu Met Thr
 165 170 175
 Ser Ala Gly Gly Leu Leu Glu Gln Ala Ala Ala Val Glu Glu Ala Ser
 180 185 190
 Asp Thr Ala Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu
 195 200 205
 Gln Gln Leu Ala Gln Pro Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu
 210 215 220
 Gly Gly Leu Trp Lys Thr Val Ser Pro His Arg Ser Pro Ile Ser Asn
 225 230 235 240
 Met Val Ser Met Ala Asn Asn His Met Ser Met Thr Asn Ser Gly Val
 245 250 255
 Ser Met Thr Asn Thr Leu Ser Ser Met Leu Lys Gly Phe Ala Pro Ala
 260 265 270
 Ala Ala Ala Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Arg Ala
 275 280 285
 Met Ser Ser Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Gly Gly
 290 295 300
 Val Ala Ala Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser Val
 305 310 315 320
 Pro Gln Ala Trp Ala Ala Ala Asn Gln Ala Val Thr Pro Ala Ala Arg
 325 330 335
 Ala Leu Pro Leu Thr Ser Leu Thr Ser Ala Ala Glu Arg Gly Pro Gly
 340 345 350
 Gln Met Leu Gly Gly Leu Pro Val Gly Gln Met Gly Ala Arg Ala Gly
 355 360 365
 Gly Gly Leu Ser Gly Val Leu Arg Val Pro Pro Arg Pro Tyr Val Met
 370 375 380
 Pro His Ser Pro Ala Ala Gly
 385 390

<210> 15
 <211> 2287
 <212> ADN

ES 2 374 620 T3

gga	ttc	gcc	att	ccg	atc	ggg	cag	gcg	atg	gcg	atc	gcg	ggc	cag	atc	152
Gly	Phe	Ala	Ile	Pro	Ile	Gly	Gln	Ala	Met	Ala	Ile	Ala	Gly	Gln	Ile	
			25					30					35			
cga	tcg	ggt	ggg	ggg	tca	ccc	acc	ggt	cat	atc	ggg	cct	acc	gcc	ttc	200
Arg	Ser	Gly	Gly	Gly	Ser	Pro	Thr	Val	His	Ile	Gly	Pro	Thr	Ala	Phe	
		40					45					50				
ctc	ggc	ttg	ggt	ggt	gtc	gac	aac	aac	ggc	aac	ggc	gca	cga	gtc	caa	248
Leu	Gly	Leu	Gly	Val	Val	Asp	Asn	Asn	Gly	Asn	Gly	Ala	Arg	Val	Gln	
	55					60					65					
cgc	gtg	gtc	ggg	agc	gct	ccg	gcg	gca	agt	ctc	ggc	atc	tcc	acc	ggc	296
Arg	Val	Val	Gly	Ser	Ala	Pro	Ala	Ala	Ser	Leu	Gly	Ile	Ser	Thr	Gly	
	70				75					80					85	
gac	gtg	atc	acc	gcg	gtc	gac	ggc	gct	ccg	atc	aac	tcg	gcc	acc	gcg	344
Asp	Val	Ile	Thr	Ala	Val	Asp	Gly	Ala	Pro	Ile	Asn	Ser	Ala	Thr	Ala	
				90					95					100		
atg	gcg	gac	gcg	ctt	aac	ggg	cat	cat	ccc	ggt	gac	gtc	atc	tcg	gtg	392
Met	Ala	Asp	Ala	Leu	Asn	Gly	His	His	Pro	Gly	Asp	Val	Ile	Ser	Val	
			105				110						115			
acc	tgg	caa	acc	aag	tcg	ggc	ggc	acg	cg	aca	ggg	aac	gtg	aca	ttg	440
Thr	Trp	Gln	Thr	Lys	Ser	Gly	Gly	Thr	Arg	Thr	Gly	Asn	Val	Thr	Leu	
		120				125						130				
gcc	gag	gga	ccc	ccg	gcc	gaa	ttc	atg	gtg	gat	ttc	ggg	gcg	tta	cca	488
Ala	Glu	Gly	Pro	Pro	Ala	Glu	Phe	Met	Val	Asp	Phe	Gly	Ala	Leu	Pro	
	135					140				145						
ccg	gag	atc	aac	tcc	gcg	agg	atg	tac	gcc	ggc	ccg	ggt	tcg	gcc	tcg	536
Pro	Glu	Ile	Asn	Ser	Ala	Arg	Met	Tyr	Ala	Gly	Pro	Gly	Ser	Ala	Ser	
	150				155					160					165	
ctg	gtg	gcc	gcg	gct	cag	atg	tgg	gac	agc	gtg	gcg	agt	gac	ctg	ttt	584
Leu	Val	Ala	Ala	Ala	Gln	Met	Trp	Asp	Ser	Val	Ala	Ser	Asp	Leu	Phe	
				170				175						180		
tcg	gcc	gcg	tcg	gcg	ttt	cag	tcg	gtg	gtc	tgg	ggt	ctg	acg	gtg	ggg	632
Ser	Ala	Ala	Ser	Ala	Phe	Gln	Ser	Val	Val	Trp	Gly	Leu	Thr	Val	Gly	
			185					190					195			
tcg	tgg	ata	ggt	tcg	tcg	gcg	ggt	ctg	atg	gtg	gcg	gcg	gcc	tcg	ccg	680
Ser	Trp	Ile	Gly	Ser	Ser	Ala	Gly	Leu	Met	Val	Ala	Ala	Ala	Ser	Pro	
		200					205					210				
tat	gtg	gcg	tgg	atg	agc	gtc	acc	gcg	ggg	cag	gcc	gag	ctg	acc	gcc	728
Tyr	Val	Ala	Trp	Met	Ser	Val	Thr	Ala	Gly	Gln	Ala	Glu	Leu	Thr	Ala	
	215					220					225					
gcc	cag	gtc	ccg	ggt	gct	gcg	gcg	gcc	tac	gag	acg	gcg	tat	ggg	ctg	776
Ala	Gln	Val	Arg	Val	Ala	Ala	Ala	Ala	Tyr	Glu	Thr	Ala	Tyr	Gly	Leu	
	230			235						240					245	
acg	gtg	ccc	ccg	ccg	gtg	atc	gcc	gag	aac	cg	gct	gaa	ctg	atg	att	824
Thr	Val	Pro	Pro	Pro	Val	Ile	Ala	Glu	Asn	Arg	Ala	Glu	Leu	Met	Ile	
				250					255					260		
ctg	ata	gcg	acc	aac	ctc	ttg	ggg	caa	aac	acc	ccg	gcg	atc	gcg	gtc	872
Leu	Ile	Ala	Thr	Asn	Leu	Leu	Gly	Gln	Asn	Thr	Pro	Ala	Ile	Ala	Val	
			265				270						275			
aac	gag	gcc	gaa	tac	ggc	gag	atg	tgg	gcc	caa	gac	gcc	gcc	gcg	atg	920
Asn	Glu	Ala	Glu	Tyr	Gly	Glu	Met	Trp	Ala	Gln	Asp	Ala	Ala	Ala	Met	
		280					285					290				
ttt	ggc	tac	gcc	gcg	gcg	acg	gcg	acg	gcg	acg	gcg	acg	ttg	ctg	ccg	968

ES 2 374 620 T3

Phe	Gly	Tyr	Ala	Ala	Ala	Thr	Ala	Thr	Ala	Thr	Ala	Thr	Leu	Leu	Pro		
	295					300				305							
ttc	gag	gag	gcg	ccg	gag	atg	acc	agc	gcg	ggt	ggg	ctc	ctc	gag	cag	1016	
Phe	Glu	Glu	Ala	Pro	Glu	Met	Thr	Ser	Ala	Gly	Gly	Leu	Leu	Glu	Gln	310	325
gcc	gcc	gcg	gtc	gag	gag	gcc	tcc	gac	acc	gcc	gcg	gcg	aac	cag	ttg	1064	
Ala	Ala	Ala	Val	Glu	Glu	Ala	Ser	Asp	Thr	Ala	Ala	Ala	Asn	Gln	Leu	330	340
atg	aac	aat	gtg	ccc	cag	gcg	ctg	caa	cag	ctg	gcc	cag	ccc	acg	cag	1112	
Met	Asn	Asn	Val	Pro	Gln	Ala	Leu	Gln	Gln	Leu	Ala	Gln	Pro	Thr	Gln	345	355
ggc	acc	acg	cct	tct	tcc	aag	ctg	ggt	ggc	ctg	tgg	aag	acg	gtc	tcg	1160	
Gly	Thr	Thr	Pro	Ser	Ser	Lys	Leu	Gly	Gly	Leu	Trp	Lys	Thr	Val	Ser	360	370
ccg	cat	cgg	tcg	ccg	atc	agc	aac	atg	gtg	tcg	atg	gcc	aac	aac	cac	1208	
Pro	His	Arg	Ser	Pro	Ile	Ser	Asn	Met	Val	Ser	Met	Ala	Asn	Asn	His	375	380
atg	tcg	atg	acc	aac	tcg	ggt	gtg	tcg	atg	acc	aac	acc	ttg	agc	tcg	1256	
Met	Ser	Met	Thr	Asn	Ser	Gly	Val	Ser	Met	Thr	Asn	Thr	Leu	Ser	Ser	390	405
atg	ttg	aag	ggc	ttt	gct	ccg	gcg	gcg	gcc	cgc	cag	gcc	gtg	caa	acc	1304	
Met	Leu	Lys	Gly	Phe	Ala	Pro	Ala	Ala	Ala	Arg	Gln	Ala	Val	Gln	Thr	410	420
gcg	gcg	caa	aac	ggg	gtc	cgg	gcg	atg	agc	tcg	ctg	ggc	agc	tcg	ctg	1352	
Ala	Ala	Gln	Asn	Gly	Val	Arg	Ala	Met	Ser	Ser	Leu	Gly	Ser	Ser	Leu	425	435
ggt	tct	tcg	ggt	ctg	ggc	ggt	ggg	gtg	gcc	gcc	aac	ttg	ggt	cgg	gcg	1400	
Gly	Ser	Ser	Gly	Leu	Gly	Gly	Gly	Val	Ala	Ala	Asn	Leu	Gly	Arg	Ala	440	450
gcc	tcg	gtc	ggt	tcg	ttg	tcg	gtg	ccg	cag	gcc	tgg	gcc	gcg	gcc	aac	1448	
Ala	Ser	Val	Gly	Ser	Leu	Ser	Val	Pro	Gln	Ala	Trp	Ala	Ala	Ala	Asn	455	465
cag	gca	gtc	acc	ccg	gcg	gcg	cgg	gcg	ctg	ccg	ctg	acc	agc	ctg	acc	1496	
Gln	Ala	Val	Thr	Pro	Ala	Ala	Arg	Ala	Leu	Pro	Leu	Thr	Ser	Leu	Thr	470	485
agc	gcc	gcg	gaa	aga	ggg	ccc	ggg	cag	atg	ctg	ggc	ggg	ctg	ccg	gtg	1544	
Ser	Ala	Ala	Glu	Arg	Gly	Pro	Gly	Gln	Met	Leu	Gly	Gly	Leu	Pro	Val	490	500
ggg	cag	atg	ggc	gcc	agg	gcc	ggt	ggt	ggg	ctc	agt	ggt	gtg	ctg	cgt	1592	
Gly	Gln	Met	Gly	Ala	Arg	Ala	Gly	Gly	Gly	Leu	Ser	Gly	Val	Leu	Arg	505	515
gtt	ccg	ccg	cga	ccc	tat	gtg	atg	ccg	cat	tct	ccg	gca	gcc	ggc	gat	1640	
Val	Pro	Pro	Arg	Pro	Tyr	Val	Met	Pro	His	Ser	Pro	Ala	Ala	Gly	Asp	520	530
atc	gcc	ccg	ccg	gcc	ttg	tcg	cag	gac	cgg	ttc	gcc	gac	ttc	ccc	gcg	1688	
Ile	Ala	Pro	Pro	Ala	Leu	Ser	Gln	Asp	Arg	Phe	Ala	Asp	Phe	Pro	Ala	535	545
ctg	ccc	ctc	gac	ccg	tcc	gcg	atg	gtc	gcc	caa	gtg	ggg	cca	cag	gtg	1736	
Leu	Pro	Leu	Asp	Pro	Ser	Ala	Met	Val	Ala	Gln	Val	Gly	Pro	Gln	Val	550	565
gtc	aac	atc	aac	acc	aaa	ctg	ggc	tac	aac	aac	gcc	gtg	ggc	gcc	ggg	1784	
Val	Asn	Ile	Asn	Thr	Lys	Leu	Gly	Tyr	Asn	Asn	Ala	Val	Gly	Ala	Gly	585	595

ES 2 374 620 T3

				570					575					580		
acc	ggc	atc	gtc	atc	gat	ccc	aac	ggt	gtc	gtg	ctg	acc	aac	aac	cac	1832
Thr	Gly	Ile	Val	Ile	Asp	Pro	Asn	Gly	Val	Val	Leu	Thr	Asn	Asn	His	
			585					590					595			
gtg	atc	gcg	ggc	gcc	acc	gac	atc	aat	gcg	ttc	agc	gtc	ggc	tcc	ggc	1880
Val	Ile	Ala	Gly	Ala	Thr	Asp	Ile	Asn	Ala	Phe	Ser	Val	Gly	Ser	Gly	
		600					605					610				
caa	acc	tac	ggc	gtc	gat	gtg	gtc	ggg	tat	gac	cgc	acc	cag	gat	gtc	1928
Gln	Thr	Tyr	Gly	Val	Asp	Val	Val	Gly	Tyr	Asp	Arg	Thr	Gln	Asp	Val	
		615				620					625					
gcg	gtg	ctg	cag	ctg	cgc	ggt	gcc	ggt	ggc	ctg	ccg	tcg	gcg	gcg	atc	1976
Ala	Val	Leu	Gln	Leu	Arg	Gly	Ala	Gly	Gly	Leu	Pro	Ser	Ala	Ala	Ile	
630					635					640					645	
ggt	ggc	ggc	gtc	gcg	ggt	ggt	gag	ccc	gtc	gtc	gcg	atg	ggc	aac	agc	2024
Gly	Gly	Gly	Val	Ala	Val	Gly	Glu	Pro	Val	Val	Ala	Met	Gly	Asn	Ser	
			650						655					660		
ggt	ggg	cag	ggc	gga	acg	ccc	cgt	gcg	gtg	cct	ggc	agg	gtg	gtc	gcg	2072
Gly	Gly	Gln	Gly	Gly	Thr	Pro	Arg	Ala	Val	Pro	Gly	Arg	Val	Val	Ala	
			665					670					675			
ctc	ggc	caa	acc	gtg	cag	gcg	tcg	gat	tcg	ctg	acc	ggt	gcc	gaa	gag	2120
Leu	Gly	Gln	Thr	Val	Gln	Ala	Ser	Asp	Ser	Leu	Thr	Gly	Ala	Glu	Glu	
		680					685					690				
aca	ttg	aac	ggg	ttg	atc	cag	ttc	gat	gcc	gcg	atc	cag	ccc	ggt	gat	2168
Thr	Leu	Asn	Gly	Leu	Ile	Gln	Phe	Asp	Ala	Ala	Ile	Gln	Pro	Gly	Asp	
	695					700					705					
tcg	ggc	ggg	ccc	gtc	gtc	aac	ggc	cta	gga	cag	gtg	gtc	ggt	atg	aac	2216
Ser	Gly	Gly	Pro	Val	Val	Asn	Gly	Leu	Gly	Gln	Val	Val	Gly	Met	Asn	
710				715						720				725		
acg	gcc	gcg	tcc	tag	gatatccatc			aactggcgg			ccgctcgagc			agatccggnt		2271
Thr	Ala	Ala	Ser		730											

gtaacaaagc ccgaaa 2287

<210> 16

<211> 729

<212> PRT

5 <213> Secuencia Artificial

<223> Descripción de secuencia artificial: proteína de fusión triple MTB72F (Ra12-TbH9-Ra35 o fusión MTB32-MTB39)

10 <400> 16

ES 2 374 620 T3

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

Gly Ile Ser Thr Gly₈₅ Asp Val Ile Thr Ala₉₀ Val Asp Gly Ala Pro Ile
 Asn Ser Ala Thr₁₀₀ Ala Met Ala Asp Ala₁₀₅ Leu Asn Gly His His₁₁₀ Pro Gly
 Asp Val Ile₁₁₅ Ser Val Thr Trp Gln Thr Lys Ser Gly Gly₁₂₅ Thr Arg Thr
 Gly Asn Val Thr Leu Ala Glu₁₃₅ Gly Pro Pro Ala Glu₁₄₀ Phe Met Val Asp
 Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala₁₅₅ Arg Met Tyr Ala Gly₁₆₀
 Pro Gly Ser Ala Ser₁₆₅ Leu Val Ala Ala Ala₁₇₀ Gln Met Trp Asp Ser₁₇₅ Val
 Ala Ser Asp Leu₁₈₀ Phe Ser Ala Ala Ser₁₈₅ Ala Phe Gln Ser Val₁₉₀ Val Trp
 Gly Leu Thr₁₉₅ Val Gly Ser Trp Ile Gly Ser Ser Ala Gly₂₀₅ Leu Met Val
 Ala Ala Ala Ser Pro Tyr Val₂₁₅ Ala Trp Met Ser Val₂₂₀ Thr Ala Gly Gln
 Ala Glu Leu Thr Ala Ala Gln Val Arg Val Ala₂₃₅ Ala Ala Tyr Glu₂₄₀
 Thr Ala Tyr Gly Leu₂₄₅ Thr Val Pro Pro Pro Val Ile Ala Glu Asn₂₅₅ Arg
 Ala Glu Leu Met₂₆₀ Ile Leu Ile Ala Thr₂₆₅ Asn Leu Leu Gly Gln₂₇₀ Asn Thr
 Pro Ala Ile Ala Val Asn Glu Ala₂₈₀ Glu Tyr Gly Glu Met₂₈₅ Trp Ala Gln
 Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr₃₀₀ Ala Thr Ala Thr
 Ala Thr Leu Leu Pro Phe Glu Glu Ala Pro Glu₃₁₅ Met Thr Ser Ala Gly₃₂₀
 Gly Leu Leu Glu Gln Ala Ala Ala Val Glu Glu Ala Ser Asp Thr Ala₃₃₅
 Ala Ala Asn Gln Leu Met Asn Asn Val₃₄₅ Pro Gln Ala Leu Gln₃₅₀ Gln Leu
 Ala Gln Pro Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu₃₆₅ Gly Gly Leu
 Trp Lys Thr Val Ser Pro His Arg Ser Pro Ile Ser₃₈₀ Asn Met Val Ser
 Met Ala Asn Asn His Met Ser Met Thr Asn Ser₃₉₅ Gly Val Ser Met Thr₄₀₀
 Asn Thr Leu Ser Ser₄₀₅ Met Leu Lys Gly Phe Ala Pro Ala Ala Ala Arg
 Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Arg Ala Met₄₃₀ Ser Ser
 Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Gly Gly₄₄₅ Val Ala Ala

ES 2 374 620 T3

Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser Val Pro Gln Ala
 450 455 460
 Trp Ala Ala Ala Asn Gln Ala Val Thr Pro Ala Ala Arg Ala Leu Pro
 465 470 475 480
 Leu Thr Ser Leu Thr Ser Ala Ala Glu Arg Gly Pro Gly Gln Met Leu
 485 490 495
 Gly Gly Leu Pro Val Gly Gln Met Gly Ala Arg Ala Gly Gly Gly Leu
 500 505 510
 Ser Gly Val Leu Arg Val Pro Pro Arg Pro Tyr Val Met Pro His Ser
 515 520 525
 Pro Ala Ala Gly Asp Ile Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe
 530 535 540
 Ala Asp Phe Pro Ala Leu Pro Leu Asp Pro Ser Ala Met Val Ala Gln
 545 550 555 560
 Val Gly Pro Gln Val Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn
 565 570 575
 Ala Val Gly Ala Gly Thr Gly Ile Val Ile Asp Pro Asn Gly Val Val
 580 585 590
 Leu Thr Asn Asn His Val Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe
 595 600 605
 Ser Val Gly Ser Gly Gln Thr Tyr Gly Val Asp Val Val Gly Tyr Asp
 610 615 620
 Arg Thr Gln Asp Val Ala Val Leu Gln Leu Arg Gly Ala Gly Gly Leu
 625 630 635 640
 Pro Ser Ala Ala Ile Gly Gly Gly Val Ala Val Gly Glu Pro Val Val
 645 650 655
 Ala Met Gly Asn Ser Gly Gly Gln Gly Gly Thr Pro Arg Ala Val Pro
 660 665 670
 Gly Arg Val Val Ala Leu Gly Gln Thr Val Gln Ala Ser Asp Ser Leu
 675 680 685
 Thr Gly Ala Glu Glu Thr Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala
 690 695 700
 Ile Gln Pro Gly Asp Ser Gly Gly Pro Val Val Asn Gly Leu Gly Gln
 705 710 715 720
 Val Val Gly Met Asn Thr Ala Ala Ser
 725

<210> 17

<211> 2190

<212> ADN

5 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: MTB72FmutSA (Ra12-TbHp-Ra35MutSA), ADNc

<400>

17

```

atgcatcacc atcaccatca cacggccgcg tccgataact tccagctgtc ccaggggtggg 60
cagggattcg ccattccgat cgggcaggcg atggcgatcg cgggccagat ccgatcgggt 120
gggggggtcac ccaccgttca tatcgggcct accgccttcc tcggcttggg tgttgtcgac 180
aacaacggca acggcgcacg agtccaacgc gtggtcggga gcgctccggc ggcaagtctc 240
ggcatctcca ccggcgacgt gatcaccgcg gtcgacggcg ctccgatcaa ctcgccacc 300
gcgatggcgg acgcgcttaa cgggcatcat cccggtgacg tcatctcggg gacctggcaa 360
accaagtctg gcggcacgcg tacagggaac gtgacattgg ccgagggacc cccggccgaa 420
ttcatggtgg atttcggggc gttaccaccg gagatcaact ccgagggat gtacgccggc 480
ccgggttcgg cctcgctggt ggcccgcgct cagatgtggg acagcgtggc gagtgaacctg 540
ttttcggcgg cgteggcggt tcagtcggtg gtctggggtc tgacgggtggg gtcgtggata 600
ggttcgtcgg cgggtctgat ggtggcggcg gcctcgccgt atgtggcgtg gatgagcgtc 660
accgcggggc aggccgagct gaccgccgcc caggtccggg ttgctgcggc ggcctacgag 720
acggcgtatg ggctgacggg gcccccgccg gtgatcgccg agaaccgtgc tgaactgatg 780
attctgatag cgaccaacct ctggggcaa aacaccccg cgatcgcggg caacgaggcc 840
gaatacggcg agatgtgggc ccaagacgcc gccgcgatgt ttggctacgc cgcggcgacg 900
gcgacggcga cggcgacgtt gctgccgttc gaggaggcgc cggagatgac cagcgcgggt 960
gggctcctcg agcaggccgc cgcggctcag gaggcctccg acaccgccgc ggcgaaccag 1020
ttgatgaaca atgtgcccc a ggcgctgcaa cagctggccc agcccacgca ggcgaccacg 1080
ccttcttcca agctgggtgg cctgtggaag acggtctcgc cgcacggtc cgcgatcagc 1140
aacatggtgt cgatggccaa caaccacatg tcgatgacca actcgggtgt gtcgatgacc 1200
aacaccttga gctcgatggt gaagggcctt gctccggcgg cggccgccca ggcggtgcaa 1260
accgcggcgc aaaacggggg cggggcgatg agctcgctgg gcagctcgtt gggttcttcg 1320
ggctctggcg gtgggggtgg cggccaactt ggtcgggcgg cctcggtcgg ttcgttgtcg 1380
gtgccgcagg cctgggcccgc ggccaaccag gcagtcaccc cggcggcgcg ggcgctgccg 1440
ctgaccagcc tgaccagcgc cgcggaaaga gggcccgggc agatgctggg cgggctgccg 1500
gtggggcaga tgggcgccag ggccggtggt gggctcagtg gtgtgctgcg tgttccgccg 1560
cgaccctatg tgatgcgcga ttctccggca gccggcgata tcgccccgcc ggccttgtcg 1620
caggaccggt tcgcccactt ccccgcgctg cccctcgacc cgtcccgat ggtcgcccaa 1680
gtggggccac aggtggtcaa catcaacacc aaactgggct acaacaacgc cgtgggcgcc 1740
gggaccggca tcgtcatcga tcccaacggt gtcgtgctga ccaacaacca cgtgatcgcg 1800
ggcggcaccg acatcaatgc gttcagcgtc ggctccggcc aaacctacgg cgtcgatgtg 1860
gtcgggtatg accgcaccca ggatgtcgcg gtgctgcagc tgcgcggtgc cggtggcctg 1920
ccgtcggcgg cgatcgggtg cggcgtcgcg gttggtgag cgtcgtcgc gatgggcaac 1980
agcgtggggc agggcggaac gccccgtcgc gtcctggca ggggtggtcgc gctcggccaa 2040
accgtgcagg cgtcggatc gctgaccggt gccgaagaga cattgaacgg gttgatccag 2100
ttcgatgccc cgatccagcc cgggtgatgcg ggcgggcccg tcgtcaacgg cctaggacag 2160
gtggtcggta tgaacaaggc cgcgtcctag . 2190

```

5 <210> 18

<211> 729

<212> PRT

<213> Secuencia artificial

10 <220>

<223> Descripción de la secuencia artificial: MTB72FmutSA (Ra12-TbHp-Ra35MutSA)

<400> 18

ES 2 374 620 T3

Met His His His His His His Thr Ala Ala Ser Asp Asn Phe Gln Leu
 1 5 10 15
 Ser Gln Gly Gly Gln Gly Phe Ala Ile Pro Ile Gly Gln Ala Met Ala
 20 25 30
 Ile Ala Gly Gln Ile Arg Ser Gly Gly Gly Ser Pro Thr Val His Ile
 35 40 45
 Gly Pro Thr Ala Phe Leu Gly Leu Gly Val Val Asp Asn Asn Gly Asn
 50 55 60
 Gly Ala Arg Val Gln Arg Val Val Gly Ser Ala Pro Ala Ala Ser Leu
 65 70 75 80
 Gly Ile Ser Thr Gly Asp Val Ile Thr Ala Val Asp Gly Ala Pro Ile
 85 90 95
 Asn Ser Ala Thr Ala Met Ala Asp Ala Leu Asn Gly His His Pro Gly
 100 105 110
 Asp Val Ile Ser Val Thr Trp Gln Thr Lys Ser Gly Gly Thr Arg Thr
 115 120 125

ES 2 374 620 T3

Gly Asn Val Thr Leu Ala Glu Gly Pro Pro Ala Glu Phe Met Val Asp
 130 135 140
 Phe Gly Ala Leu Pro Pro Glu Ile Asn Ser Ala Arg Met Tyr Ala Gly
 145 150 155
 Pro Gly Ser Ala Ser Leu Val Ala Ala Ala Gln Met Trp Asp Ser Val
 165 170 175
 Ala Ser Asp Leu Phe Ser Ala Ala Ser Ala Phe Gln Ser Val Val Trp
 180 185
 Gly Leu Thr Val Gly Ser Trp Ile Gly Ser Ser Ala Gly Leu Met Val
 195 200 205
 Ala Ala Ala Ser Pro Tyr Val Ala Trp Met Ser Val Thr Ala Gly Gln
 210 215 220
 Ala Glu Leu Thr Ala Ala Gln Val Arg Val Ala Ala Ala Ala Tyr Glu
 225 230 235 240
 Thr Ala Tyr Gly Leu Thr Val Pro Pro Pro Val Ile Ala Glu Asn Arg
 245 250 255
 Ala Glu Leu Met Ile Leu Ile Ala Thr Asn Leu Leu Gly Gln Asn Thr
 260 265 270
 Pro Ala Ile Ala Val Asn Glu Ala Glu Tyr Gly Glu Met Trp Ala Gln
 275 280 285
 Asp Ala Ala Ala Met Phe Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr
 290 295 300
 Ala Thr Leu Leu Pro Phe Glu Glu Ala Pro Glu Met Thr Ser Ala Gly
 305 310 315 320
 Gly Leu Leu Glu Gln Ala Ala Ala Val Glu Glu Ala Ser Asp Thr Ala
 325 330 335
 Ala Ala Asn Gln Leu Met Asn Asn Val Pro Gln Ala Leu Gln Gln Leu
 340 345 350
 Ala Gln Pro Thr Gln Gly Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu
 355 360 365
 Trp Lys Thr Val Ser Pro His Arg Ser Pro Ile Ser Asn Met Val Ser
 370 375 380
 Met Ala Asn Asn His Met Ser Met Thr Asn Ser Gly Val Ser Met Thr
 385 390 395 400
 Asn Thr Leu Ser Ser Met Leu Lys Gly Phe Ala Pro Ala Ala Ala Ala
 405 410 415
 Gln Ala Val Gln Thr Ala Ala Gln Asn Gly Val Arg Ala Met Ser Ser
 420 425 430
 Leu Gly Ser Ser Leu Gly Ser Ser Gly Leu Gly Gly Gly Val Ala Ala
 435 440 445
 Asn Leu Gly Arg Ala Ala Ser Val Gly Ser Leu Ser Val Pro Gln Ala
 450 455 460
 Trp Ala Ala Ala Asn Gln Ala Val Thr Pro Ala Ala Arg Ala Leu Pro
 465 470 475 480
 Leu Thr Ser Leu Thr Ser Ala Ala Glu Arg Gly Pro Gly Gln Met Leu
 485 490 495

Gly Gly Leu Pro Val Gly Gln Met Gly Ala Arg Ala Gly Gly Gly Leu
 500 505
 Ser Gly Val Leu Arg Val Pro Pro Arg Pro Tyr Val Met Pro His Ser
 515 520 525
 Pro Ala Ala Gly Asp Ile Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe
 530 535 540
 Ala Asp Phe Pro Ala Leu Pro Leu Asp Pro Ser Ala Met Val Ala Gln
 545 550 555
 Val Gly Pro Gln Val Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn
 565 570
 Ala Val Gly Ala Gly Thr Gly Ile Val Ile Asp Pro Asn Gly Val Val
 580 585 590
 Leu Thr Asn Asn His Val Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe
 595 600 605
 Ser Val Gly Ser Gly Gln Thr Tyr Gly Val Asp Val Val Gly Tyr Asp
 610 615 620
 Arg Thr Gln Asp Val Ala Val Leu Gln Leu Arg Gly Ala Gly Gly Leu
 625 630 635 640
 Pro Ser Ala Ala Ile Gly Gly Gly Val Ala Val Gly Glu Pro Val Val
 645 650 655
 Ala Met Gly Asn Ser Gly Gly Gln Gly Gly Thr Pro Arg Ala Val Pro
 660 665 670
 Gly Arg Val Val Ala Leu Gly Gln Thr Val Gln Ala Ser Asp Ser Leu
 675 680 685
 Thr Gly Ala Glu Glu Thr Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala
 690 695 700
 Ile Gln Pro Gly Asp Ala Gly Gly Pro Val Val Asn Gly Leu Gly Gln
 705 710 715 720
 Val Val Gly Met Asn Thr Ala Ala Ser
 725

<210> 19

<211> 1797

<212> ADN

5 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: proteína fusión doble TbH9-Ra35 (designada MTB59F)

<220>

10 <221> CDS

<222> (1)..(1791)

<223> MTB59F

ES 2 374 620 T3

<400>

19

cat atg cat cac cat cac cat cac atg gtg gat ttc ggg gcg tta cca 48
 His Met His His His His His His Met Val Asp Phe Gly Ala Leu Pro
 1 5 10 15

ccg gag atc aac tcc gcg agg atg tac gcc ggc ccg ggt tcg gcc tcg 96
 Pro Glu Ile Asn Ser Ala Arg Met Tyr Ala Gly Pro Gly Ser Ala Ser
 20 25 30

ctg gtg gcc gcg gct cag atg tgg gac agc gtg gcg agt gac ctg ttt 144
 Leu Val Ala Ala Ala Gln Met Trp Asp Ser Val Ala Ser Asp Leu Phe
 35 40 45

tcg gcc gcg tcg gcg ttt cag tcg gtg gtc tgg ggt ctg acg gtg ggg 192
 Ser Ala Ala Ser Ala Phe Gln Ser Val Val Trp Gly Leu Thr Val Gly
 50 55 60

ES 2 374 620 T3

tcg tgg ata ggt tcg tcg gcg ggt ctg atg gtg gcg gcg gcc tcg ccg Ser Trp Ile Gly Ser Ser Ala Gly Leu Met Val Ala Ala Ala Ser Pro 65 70 75 80	240
tat gtg gcg tgg atg agc gtc acc gcg ggg cag gcc gag ctg acc gcc Tyr Val Ala Trp Met Ser Val Thr Ala Gly Gln Ala Glu Leu Thr Ala 85 90 95	288
gcc cag gtc cgg gtt gct gcg gcg gcc tac gag acg gcg tat ggg ctg Ala Gln Val Arg Val Ala Ala Ala Tyr Glu Thr Ala Tyr Gly Leu 100 105 110	336
acg gtg ccc ccg ccg gtg atc gcc gag aac cgt gct gaa ctg atg att Thr Val Pro Pro Pro Val Ile Ala Glu Asn Arg Ala Glu Leu Met Ile 115 120 125	384
ctg ata gcg acc aac ctc ttg ggg caa aac acc ccg gcg atc gcg gtc Leu Ile Ala Thr Asn Leu Leu Gly Gln Asn Thr Pro Ala Ile Ala Val 130 135 140	432
aac gag gcc gaa tac ggc gag atg tgg gcc caa gac gcc gcc gcg atg Asn Glu Ala Glu Tyr Gly Glu Met Trp Ala Gln Asp Ala Ala Ala Met 145 150 155 160	480
ttt ggc tac gcc gcg gcg acg gcg acg gcg acg gcg acg ttg ctg ccg Phe Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr Ala Thr Leu Leu Pro 165 170 175	528
ttc gag gag gcg ccg gag atg acc agc gcg ggt ggg ctc ctc gag cag Phe Glu Glu Ala Pro Glu Met Thr Ser Ala Gly Gly Leu Leu Glu Gln 180 185 190	576
gcc gcc gcg gtc gag gag gcc tcc gac acc gcc gcg gcg aac cag ttg Ala Ala Ala Val Glu Glu Ala Ser Asp Thr Ala Ala Ala Asn Gln Leu 195 200 205	624
atg aac aat gtg ccc cag gcg ctg caa cag ctg gcc cag ccc acg cag Met Asn Asn Val Pro Gln Ala Leu Gln Gln Leu Ala Gln Pro Thr Gln 210 215 220	672
ggc acc acg cct tct tcc aag ctg ggt ggc ctg tgg aag acg gtc tcg Gly Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu Trp Lys Thr Val Ser 225 230 235 240	720
ccg cat cgg tcg ccg atc agc aac atg gtg tcg atg gcc aac aac cac Pro His Arg Ser Pro Ile Ser Asn Met Val Ser Met Ala Asn Asn His 245 250 255	768
atg tcg atg acc aac tcg ggt gtg tcg atg acc aac acc ttg agc tcg Met Ser Met Thr Asn Ser Gly Val Ser Met Thr Asn Thr Leu Ser Ser 260 265 270	816
atg ttg aag ggc ttt gct ccg gcg gcg gcc gcc cag gcc gtg caa acc Met Leu Lys Gly Phe Ala Pro Ala Ala Ala Ala Gln Ala Val Gln Thr 275 280 285	864
gcg gcg caa aac ggg gtc cgg gcg atg agc tcg ctg ggc agc tcg ctg Ala Ala Gln Asn Gly Val Arg Ala Met Ser Ser Leu Gly Ser Ser Leu 290 295 300	912
ggt tct tcg ggt ctg ggc ggt ggg gtg gcc gcc aac ttg ggt cgg gcg Gly Ser Ser Gly Leu Gly Gly Gly Val Ala Ala Asn Leu Gly Arg Ala 305 310 315 320	960
gcc tcg gtc ggt tcg ttg tcg gtg ccg cag gcc tgg gcc gcg gcc aac Ala Ser Val Gly Ser Leu Ser Val Pro Gln Ala Trp Ala Ala Ala Asn 325 330 335	1008

ES 2 374 620 T3

cag gca gtc acc ccg gcg gcg cgg gcg ctg ccg ctg acc agc ctg acc	1056
Gln Ala Val Thr Pro Ala Ala Arg Ala Leu Pro Leu Thr Ser Leu Thr	
340 345 350	
agc gcc gcg gaa aga ggg ccc ggg cag atg ctg ggc ggg ctg ccg gtg	1104
Ser Ala Ala Glu Arg Gly Pro Gly Gln Met Leu Gly Gly Leu Pro Val	
355 360 365	
ggg cag atg ggc gcc agg gcc ggt ggt ggg ctc agt ggt gtg ctg cgt	1152
Gly Gln Met Gly Ala Arg Ala Gly Gly Gly Leu Ser Gly Val Leu Arg	
370 375 380	
gtt ccg ccg cga ccc tat gtg atg ccg cat tct ccg gca gcc ggc gat	1200
Val Pro Pro Arg Pro Tyr Val Met Pro His Ser Pro Ala Ala Gly Asp	
385 390 395 400	
atc gcc ccg ccg gcc ttg tcg cag gac cgg ttc gcc gac ttc ccc gcg	1248
Ile Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala	
405 410 415	
ctg ccc ctc gac ccg tcc gcg atg gtc gcc caa gtg ggg cca cag gtg	1296
Leu Pro Leu Asp Pro Ser Ala Met Val Ala Gln Val Gly Pro Gln Val	
420 425 430	
gtc aac atc aac acc aaa ctg ggc tac aac aac gcc gtg ggc gcc ggg	1344
Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val Gly Ala Gly	
435 440 445	
acc ggc atc gtc atc gat ccc aac ggt gtc gtg ctg acc aac aac cac	1392
Thr Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His	
450 455 460	
gtg atc gcg ggc gcc acc gac atc aat gcg ttc agc gtc ggc tcc ggc	1440
Val Ile Ala Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly	
465 470 475 480	
caa acc tac ggc gtc gat gtg gtc ggg tat gac cgc acc cag gat gtc	1488
Gln Thr Tyr Gly Val Asp Val Val Gly Tyr Asp Arg Thr Gln Asp Val	
485 490 495	
gcg gtg ctg cag ctg cgc ggt gcc ggt ggc ctg ccg tcg gcg gcg atc	1536
Ala Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala Ile	
500 505 510	
ggt ggc ggc gtc gcg gtt ggt gag ccc gtc gtc gcg atg ggc aac agc	1584
Gly Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser	
515 520 525	
ggt ggg cag ggc gga acg ccc cgt gcg gtg cct ggc agg gtg gtc gcg	1632
Gly Gly Gln Gly Gly Thr Pro Arg Ala Val Pro Gly Arg Val Val Ala	
530 535 540	
ctc ggc caa acc gtg cag gcg tcg gat tcg ctg acc ggt gcc gaa gag	1680
Leu Gly Gln Thr Val Gln Ala Ser Asp Ser Leu Thr Gly Ala Glu Glu	
545 550 555 560	
aca ttg aac ggg ttg atc cag ttc gat gcc gcg atc cag ccc ggt gat	1728
Thr Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp	
565 570 575	
tcg ggc ggg ccc gtc gtc aac ggc cta gga cag gtg gtc ggt atg aac	1776
Ser Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn	
580 585 590	
acg gcc gcg tcc taggatatc	1797
Thr Ala Ala Ser	
595	

ES 2 374 620 T3

<210> 20

<211> 596

<212> PRT

<213> Secuencia artificial

5

<223> Descripción de la secuencia artificial: proteína fusión doble TbH9-Ra35 (designada MTB59F)

<400> 20

ES 2 374 620 T3

His Met His His His His His His Met Val Asp Phe Gly Ala Leu Pro
 1 5 10 15
 Pro Glu Ile Asn Ser Ala Arg Met Tyr Ala Gly Pro Gly Ser Ala Ser
 20 25 30
 Leu Val Ala Ala Ala Gln Met Trp Asp Ser Val Ala Ser Asp Leu Phe
 35 40 45
 Ser Ala Ala Ser Ala Phe Gln Ser Val Val Trp Gly Leu Thr Val Gly
 50 55 60
 Ser Trp Ile Gly Ser Ser Ala Gly Leu Met Val Ala Ala Ala Ser Pro
 65 70 75 80
 Tyr Val Ala Trp Met Ser Val Thr Ala Gly Gln Ala Glu Leu Thr Ala
 85 90 95
 Ala Gln Val Arg Val Ala Ala Ala Ala Tyr Glu Thr Ala Tyr Gly Leu
 100 105 110
 Thr Val Pro Pro Pro Val Ile Ala Glu Asn Arg Ala Glu Leu Met Ile
 115 120 125
 Leu Ile Ala Thr Asn Leu Leu Gly Gln Asn Thr Pro Ala Ile Ala Val
 130 135 140
 Asn Glu Ala Glu Tyr Gly Glu Met Trp Ala Gln Asp Ala Ala Ala Met
 145 150 155 160
 Phe Gly Tyr Ala Ala Ala Thr Ala Thr Ala Thr Ala Thr Leu Leu Pro
 165 170 175
 Phe Glu Glu Ala Pro Glu Met Thr Ser Ala Gly Gly Leu Leu Glu Gln
 180 185 190
 Ala Ala Ala Val Glu Glu Ala Ser Asp Thr Ala Ala Ala Asn Gln Leu
 195 200 205
 Met Asn Asn Val Pro Gln Ala Leu Gln Gln Leu Ala Gln Pro Thr Gln
 210 215 220
 Gly Thr Thr Pro Ser Ser Lys Leu Gly Gly Leu Trp Lys Thr Val Ser
 225 230 235 240
 Pro His Arg Ser Pro Ile Ser Asn Met Val Ser Met Ala Asn Asn His
 245 250 255
 Met Ser Met Thr Asn Ser Gly Val Ser Met Thr Asn Thr Leu Ser Ser
 260 265 270
 Met Leu Lys Gly Phe Ala Pro Ala Ala Ala Ala Gln Ala Val Gln Thr
 275 280 285
 Ala Ala Gln Asn Gly Val Arg Ala Met Ser Ser Leu Gly Ser Ser Leu
 290 295 300
 Gly Ser Ser Gly Leu Gly Gly Gly Val Ala Ala Asn Leu Gly Arg Ala
 305 310 315 320
 Ala Ser Val Gly Ser Leu Ser Val Pro Gln Ala Trp Ala Ala Ala Asn

ES 2 374 620 T3

<223> MTB8.4 (DPV), ADNc

<400> 21

```

cgtggcaatg tcgttgaccg tcggggccgg ggtcgctcc gcagatcccg tggacgcggt 60
cattaacacc acctgcaatt acgggcaggt agtagctgcg ctcaacgcga cggatccggg 120
ggctgccgca cagttcaacg cctcaccggg ggcgcagtcc tatttgcgca atttcctcgc 180
cgcaccgcca cctcagcgcg ctgccatggc cgcgcaattg caagctgtgc cgggggcggc 240
acagtacatc ggccttgctg agtcggttgc cggctcctgc aacaactatt aagcccatgc 300
gggccccatc ccgcgacccg gcatcgtcgc cggggctagg ccagattgcc ccgctcctca 360
acggggccgca tcccgcgacc cggcatcgtc gccggggcta ggccagattg ccccgtcct 420
caacgggccc catctcgtgc cgaattcctg cagcccgggg gatccactag ttctagagcg 480
gccgcccaccg cgggtggagct                                     500
    
```

5 <210> 22

<211> 96

<212> PRT

<213> *Mycobacterium tuberculosis*

10 <220>

<223> MTB8.4 (DPV)

<400> 22

```

Val Ala Met Ser Leu Thr Val Gly Ala Gly Val Ala Ser Ala Asp Pro
1          5          10          15
Val Asp Ala Val Ile Asn Thr Thr Cys Asn Tyr Gly Gln Val Val Ala
20          25          30
Ala Leu Asn Ala Thr Asp Pro Gly Ala Ala Ala Gln Phe Asn Ala Ser
35          40          45
Pro Val Ala Gln Ser Tyr Leu Arg Asn Phe Leu Ala Ala Pro Pro Pro
50          55          60
Gln Arg Ala Ala Met Ala Ala Gln Leu Gln Ala Val Pro Gly Ala Ala
65          70          75          80
Gln Tyr Ile Gly Leu Val Glu Ser Val Ala Gly Ser Cys Asn Asn Tyr
85          90          95
    
```

15 <210> 23

<211> 585

<212> ADN

<213> *Mycobacterium tuberculosis*

20 <220>

<223> MTB9.8 (MSL)

<400> 23

ES 2 374 620 T3

```
tggattccga tagcggtttc ggcccctcga cgggcgacca cggcgcgcag gcctccgaac 60
ggggggccgg gacgctggga ttcgccggga ccgcaaccaa agaacgccgg gtccgggagg 120
tcgggctgac cgcactggcc ggtgatgagt tcggcaacgg cccccggatg ccgatggtgc 180
cggggacctg ggagcagggc agcaacgagc ccgaggcgcc cgacggatcg gggagagggg 240
gaggcgacgg cttaccgcac gacagcaagt aaccgaattc cgaatcacgt ggaccctgac 300
gggtcgaag gagagatggt atgagccttt tggatgctca tatcccacag ttggtggcct 360
cccagtcggc gtttgccgcc aaggcggggc tgatgcggca cacgatcggg caggccgagc 420
aggcggcgat gtcggctcag gcgtttcacc agggggagtc gtcggcggcg tttcaggccg 480
cccatgcccg gtttggtggc gcggccgcca aagtcaacac cttgttggat gtcgcgcagg 540
cgaatctggg tgaggccgcc ggtacctatg tggccgccga tgctg 585
```

<210> 24

<211> 97

<212> PRT

5 <213> *Mycobacterium tuberculosis*

<220>

<223> MTB9.8 (MSL)

10 <400> 24

```
Met Ser Leu Leu Asp Ala His Ile Pro Gln Leu Val Ala Ser Gln Ser
 1          5          10          15
Ala Phe Ala Ala Lys Ala Gly Leu Met Arg His Thr Ile Gly Gln Ala
          20          25          30
Glu Gln Ala Ala Met Ser Ala Gln Ala Phe His Gln Gly Glu Ser Ser
          35          40          45
Ala Ala Phe Gln Ala Ala His Ala Arg Phe Val Ala Ala Ala Ala Lys
          50          55          60
Val Asn Thr Leu Leu Asp Val Ala Gln Ala Asn Leu Gly Glu Ala Ala
          65          70          75          80
Gly Thr Tyr Val Ala Ala Asp Ala Ala Ala Ala Ser Thr Tyr Thr Gly
          85          90          95
```

Phe

<210> 25

<211> 1742

<212> ADN

15 <213> *Mycobacterium tuberculosis*

<220>

<223> MTB9.9 (MTI, también conocido como MTI-A)

<220>

20 <221> base_modificada

<222> (1)...(1742)

ES 2 374 620 T3

<223> n= g, a, c o t

<400> 25

```

ccgctctctt tcaacgtcat aagttcgggt ggccagtcgg ccgcgcgctgc atatggcacc 60
aataacgcgt gtcccatgga taccgggacc gcacgacggt agagcggatc agcgcagccg 120
gtgccgaaca ctaccgcgtc cacgctcagc cctgccgctg tgcggaagat cgagcccagg 180
ttctcatggg cgttaacgcc ttccaacact gcgacgggtg gcgccccggc gaccacctga 240
gcaacgctcg gctccggcac ccggcgcgcg gctgccaaca cccacgatt gagatggaag 300
ccgatcacc gtgccatgac atcagccgac gctcgatagt acggcgcgcc gacaccggcc 360
agatcctct tgagctcggc cagccggcgg tcggtgccga acagcggcag cggcgtgaac 420
cgtgaggcca gcatgcgctg caccaccagc acaccctcgg cgatcaccaa cgccttgccg 480
gtcggcagat cgggacnacn gtcgatgctg ttcaggtcac ggaaatcgtc gagccgtggg 540
tcgtcgggat cgcagacgtc ctgaacatcg aggcgctcgg ggtgctgggc acaacggcct 600
tcggtcacgg gctttcgtcg accagagcca gcatcagatc ggcgcgctg cgaggatgt 660
cacgctcgtc gcggttcagc gtcgcgagcc gctcagccag ccactcttgc agagagccgt 720
tgctgggatt aattgggaga ggaagacagc atgtcgttcg tgaccacaca gccggaagcc 780
ctggcagctg cggcggcgaa cctacagggg attggcacga caatgaacgc ccagaacgcg 840
gccgcggctg ctccaaccac cggagttagt cccgcagccg ccgatgaagt atcagcgtg 900
accgcggctc agtttgctg gcacgcgcag atgtacaaa cggtcagcgc ccaggccctc 960
gccattcacg aaatgttcgt gaacacgctg gtggccagtt ctggctcata cgcgccacc 1020
gaggcggcca acgcagccgc tgccggctga acgggctcgc acgaacctgc tgaaggagag 1080
ggggaacatc cggagtctc gggtcagggg ttgcgccagc gccagccga ttcagntatc 1140
ggcgtccata acagcagacg atctaggcat tcagtactaa ggagacaggc aacatggcct 1200
cacgttttat gacggatccg catgcgatgc gggacatggc gggccgtttt gaggtgcacg 1260
cccagacggg ggaggacgag gctcgcggga tgtggcgctc cgcgcaaac atttccggtg 1320
cgggctggag tggcatggcc gaggcgacct cgtagacac catgacctag atgaatcagg 1380
cgtttcgcaa catcgtgaac atgctgcacg ggggtcgtga cgggctggtt cgcgacgcca 1440
acaantacga acagcaagag caggcctccc agcagatcct gagcagntag cgcgaaagc 1500
cacagctgng tacgntttct cacattagga gaacaccaat atgacgatta attaccagtt 1560
cggggacgct gacgctcatg gcgccatgat ccgcgctcag gcggcgtcgc ttgaggcggg 1620
gcatcaggcc atcgttcgtg atgtgttggc cgcggtgac ttttggggcg gcgccggttc 1680
ggtggcctgc caggagtcca ttaccagtt gggccgtaac ttccaggtga tctacgagca 1740
gg

```

5 <210> 26

<211> 2836

<212> ADN

<213> *Mycobacterium tuberculosis*

10 <220>

<223> MTB9.9A (MTI, también conocido como MTI-A)

<220>

<221> base_modificada

<222> (104)

15 <223> n= g, a, c o t

<400> 26

ES 2 374 620 T3

```

gttgattccg ttcgcggcgc cgccgaagac caccaactcc gctgggggtg tgcacagggc 60
ggttgctgctc gtcagctggc cgaatcccaa tgattggtgg ctngtgctgg ttgctgggct 120
cgattacccc cacggaaagg acgacgatcg ttcgtttgct cggtcagtcg tacttgccga 180
cgggcatggc gcggtttctt acctcgatcg cacagcagct gaccttcggc ccagggggca 240
caacggctgg ctccggcggg gcctggtacc caacgccaca attcggcggc ctgggtgcag 300
gcccggcggg gtcggcggg ttggcgcggg cggagccggg cgggaggttg tcggtgccg 360
caagtggggc cgtcgcggct ccggccttcg cggagaagcc tgaggcgggc acgccgatgt 420
ccgtcatcgg cgaagcgtcc agctgcggtc agggaggcct gcttcgaggc ataccgctgg 480
cgagagcggg gcggcgatca ggcgcttcg ctcaccgata cgggttcggc cacagcgtga 540
ttacccggtc tccgtcggcg ggatagcttt cgatccggtc tgcgcggccg ccggaaatgc 600
tgcagatagc gatcgaccgc gccggtcggg aaacgccgca cacggcacta tcaatggcga 660
cggcgggagc tgatgccaaa ttgaccgtcc cgacggggct ttatctgctg caagatttca 720
tccccagccc ggtcgggtgg ccgataaata cgctggctcag cgcgactctt ccggctgaat 780
tcgatgctct gggcgcccgc tcgacgccga gtatctcgag tgggccgcaa acccggtcaa 840
acgctgttac tgtggcgtta ccacaggtga atttgcgggtg ccaactggtg aacacttgctg 900
aacgggtggc atcgaatca acttgttgcg ttgcagtgat ctactctctt gcagagagcc 960
gttgcctggg ttaattggga gaggaagaca gcatgtcgtt cgtgaccaca cagccggaag 1020
ccctggcagc tgcggcggcg aacctacagg gtattggcac gacaatgaac gcccagaacg 1080
cggccgcggc tgctccaacc accggagtag tgcccgcagc cggcgatgaa gtatcagcgc 1140
tgaccgcggc tcagtttgct gcgcacgcgc agatgtacca aacggtcagc gcccaggccg 1200
cggccattca cgaaatgttc gtgaacacgc tggtggccag ttctggctca tacgcggcca 1260
ccgaggcggc caacgcagcc gctgccggct gaacgggctc gcacgaacct gctgaaggag 1320
agggggaaca tccggagttc tcgggtcagg ggttgcggca gcgcccagcc gattcagcta 1380
tcggcgcca taacagcaga cgatctaggc attcagtagt aaggagacag gcaacatggc 1440
ctcacgtttt atgacggatc cgcacgcgat gcgggacatg gcgggcccgt ttgaggtgca 1500
cgcccagacg gtggaggacg aggtcgcggc gatgtggggc tccgcgcaaa acatttccgg 1560
tgccggctgg agtggcatgg ccgagggcgc ctcgctagac accatgacct agatgaatca 1620
ggcgtttcgc aacatcgtga acatgctgca cggggtcgtg gacgggctgg ttcgcgagc 1680
caacaactac gaacagcaag agcaggcctc ccagcagatc ctgagcagct agcgcggaaa 1740
gccacagctg cgtacgcttt ctacattag gagaacacca atatgacgat taattaccag 1800
ttcggggagc tcgacgctca tggcgccatg atccgcgctc aggcggcgtc gcttgaggcg 1860
gagcatcagg ccatcgttcg tgatgtgttg gccgcgggtg acttttgggg cggcgccggg 1920
tcggtggctt gccaggagt cattaccag ttgggccgta acttccaggt gatctacgag 1980
caggccaacg cccacgggca gaaggtgcag gctgccggca acaacatggc gcaaaccgac 2040
agcgcgctc gctccagctg ggctaaaac tgaacttcag tcgcggcagc acaccaacca 2100
gccggtgtgc tgctgtgtcc tgcagttaac tagcactcga ccgctgaggt agcgatggat 2160
caacagagta cccgcaccga catcaccgtc aacgtcgacg gcttctggat gcttcaggcg 2220
ctactggata tccgccacgt tgcgcctgag ttacgttgcc ggccgtacgt ctccaccgat 2280
tccaatgact ggctaaacga gcacccgggg atggcggtca tgcgcgagca gggcattgtc 2340
gtcaacgacg cggctaacga acaggtcgtc gcccggatga aggtgcttgc cgcacctgat 2400
cttgaagtgc tcgccctgct gtcacgcggc aagttgctgt acggggtcat agacgacgag 2460
aaccagccgc cgggttcgcg tgacatccct gacaatgagt tccgggtggg gttggcccgg 2520
cgaggccagc actgggtgtc ggcggtacgg gttggcaatg acatcaccgt cgatgacgtg 2580
acggtctcgg atagcgcctc gatcgccgca ctggtaatgg acggtctgga gtcgattcac 2640
cacgccgacc cagcccgcat caacgcggtc aacgtgccaa tggaggagat ctcgtgccga 2700
attcggcagc aggcacgagg cgggtgcggg gacgacggga tcgatcacga tcatcgaccg 2760
gccgggatcc ttggcgatct cgttgagcac gaccggggcc cgcgggaagc tctgcgacat 2820
ccatgggttc ttcccc 2836

```

<210> 27

<211> 94

<212> PRT

5 <213> *Mycobacterium tuberculosis*

<220>

<223> MTB9.9A (MTI, también conocido como MTI-A), péptido ORF

10 <400> 27

ES 2 374 620 T3

Met Thr Ile Asn Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met
 1 5 10 15
 Ile Arg Ala Leu Ala Gly Leu Leu Glu Ala Glu His Gln Ala Ile Ile
 20 25 30
 Ser Asp Val Leu Thr Ala Ser Asp Phe Trp Gly Gly Ala Gly Ser Ala
 35 40 45
 Ala Cys Gln Gly Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile
 50 55 60
 Tyr Glu Gln Ala Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn
 65 70 75 80
 Asn Met Ala Gln Thr Asp Ser Ala Val Gly Ser Ser Trp Ala
 85 90

<210> 28

<211> 1200

<212> ADN

5 <213> *Mycobacterium tuberculosis*

<220>

<223> MTB40 (HTCC n° 1), ADNc

10 <400> 28

```

caggcatgag cagagcgttc atcatcgatc caacgatcag tgccattgac ggcttgtacg 60
accttctggg gattggaata cccaaccaag ggggtatcct ttactcctca cttaggtact 120
tcgaaaaagc cctggaggag ctggcagcag cgtttccggg tgatggctgg ttaggttcgg 180
ccgcggaaca atacgccggc aaaaaccgca accacgtgaa tttttccag gaactggcag 240
acctcgatcg tcagctcatc agcctgatcc acgaccaggc caacgcggtc cagacgacct 300
gcgacatcct ggagggcgcc aagaaaggtc tcgagttcgt gcgcccgggt gctgtggacc 360
tgacctacat cccggtcgtc gggcacgccc ttcggccgc ctccaggcg ccgttttgcg 420
cgggcgcgat ggccgtagtg ggcggcgcgc ttgcctactt ggtcgtgaaa acgctgatca 480
acgcgactca actcctcaaa ttgcttgcca aattggcgga gttggtcgcy gccgccattg 540
cggacatcat ttcggatgtg gcggacatca tcaagggcac cctcggagaa gtgtgggagt 600
tcatcacaana cgcgctcaac ggcctgaaa agctttggga caagctcacg ggggtgggtga 660
ccggactgtt ctctcgaggg tggtcgaacc tggagtcctt ctttgcgggc gtccccggct 720
tgaccggcgc gaccagcggc ttgtcgcaag tgactggctt gttcgggtgc gccggctctgt 780
ccgcatcgtc gggcttggct cacgcggata gcctggcgag ctacgccagc ttgccgccc 840
tggccggcat tgggggcggg tccggttttg ggggcttgcc gagcctggct cagggtccatg 900
ccgctcaac tcggcaggcg ctacggcccc gagctgatgg cccggtcggc gccgctgccg 960
agcaggtcgg cgggcagtcg cagctggtct ccgcgcaggg ttcccaaggt atgggcggac 1020
ccgtaggcat gggcggcagc caccctctt cgggggcgct gaaagggacg acgacgaaga 1080
agtactcggg aggcgcggcg gcgggcactg aagacgccga gcgcgcgcca gtcgaagctg 1140
acgcgggcgg tgggcaaaaag gtgctggtac gaaacgtcgt ctaacggcat ggcgagccaa 1200

```

<210> 29

<211> 392

<212> PRT

15 <213> *Mycobacterium tuberculosis*

<220>

<223> MTB40 (HTCC n° 1)

<400>

29

Met Ser Arg Ala Phe Ile Ile Asp Pro Thr Ile Ser Ala Ile Asp Gly
 1 5 10 15
 Leu Tyr Asp Leu Leu Gly Ile Gly Ile Pro Asn Gln Gly Gly Ile Leu
 20 25 30
 Tyr Ser Ser Leu Glu Tyr Phe Glu Lys Ala Leu Glu Glu Leu Ala Ala
 35 40 45
 Ala Phe Pro Gly Asp Gly Trp Leu Gly Ser Ala Ala Asp Lys Tyr Ala
 50 55 60
 Gly Lys Asn Arg Asn His Val Asn Phe Phe Gln Glu Leu Ala Asp Leu
 65 70 75 80
 Asp Arg Gln Leu Ile Ser Leu Ile His Asp Gln Ala Asn Ala Val Gln
 85 90 95
 Thr Thr Arg Asp Ile Leu Glu Gly Ala Lys Lys Gly Leu Glu Phe Val
 100 105 110
 Arg Pro Val Ala Val Asp Leu Thr Tyr Ile Pro Val Val Gly His Ala
 115 120 125
 Leu Ser Ala Ala Phe Gln Ala Pro Phe Cys Ala Gly Ala Met Ala Val
 130 135 140
 Val Gly Gly Ala Leu Ala Tyr Leu Val Val Lys Thr Leu Ile Asn Ala
 145 150 155 160
 Thr Gln Leu Leu Lys Leu Leu Ala Lys Leu Ala Glu Leu Val Ala Ala
 165 170 175
 Ala Ile Ala Asp Ile Ile Ser Asp Val Ala Asp Ile Ile Lys Gly Thr
 180 185 190
 Leu Gly Glu Val Trp Glu Phe Ile Thr Asn Ala Leu Asn Gly Leu Lys
 195 200 205
 Glu Leu Trp Asp Lys Leu Thr Gly Trp Val Thr Gly Leu Phe Ser Arg
 210 215 220
 Gly Trp Ser Asn Leu Glu Ser Phe Phe Ala Gly Val Pro Gly Leu Thr
 225 230 235 240
 Gly Ala Thr Ser Gly Leu Ser Gln Val Thr Gly Leu Phe Gly Ala Ala
 245 250 255
 Gly Leu Ser Ala Ser Ser Gly Leu Ala His Ala Asp Ser Leu Ala Ser
 260 265 270
 Ser Ala Ser Leu Pro Ala Leu Ala Gly Ile Gly Gly Gly Ser Gly Phe
 275 280 285
 Gly Gly Leu Pro Ser Leu Ala Gln Val His Ala Ala Ser Thr Arg Gln
 290 295 300
 Ala Leu Arg Pro Arg Ala Asp Gly Pro Val Gly Ala Ala Ala Glu Gln
 305 310 315 320
 Val Gly Gly Gln Ser Gln Leu Val Ser Ala Gln Gly Ser Gln Gly Met
 325 330 335
 Gly Gly Pro Val Gly Met Gly Gly Met His Pro Ser Ser Gly Ala Ser
 340 345 350
 Lys Gly Thr Thr Thr Lys Lys Tyr Ser Glu Gly Ala Ala Ala Gly Thr
 355 360 365
 Glu Asp Ala Glu Arg Ala Pro Val Glu Ala Asp Ala Gly Gly Gly Gln
 370 375 380
 Lys Val Leu Val Arg Asn Val Val
 385 390

5 <210> 30

<211> 1441

ES 2 374 620 T3

<212> ADN

<213> *Mycobacterium tuberculosis*

<220>

5 <223> MTB41 (MTCC n° 2), ADNc

<400> 30

```

gaggttgctg gcaatggatt tcgggctttt acctccggaa gtgaattcaa gccgaatgta 60
ttccggtccg gggccggagt cgatgctagc cgcccgggcc gcctgggacg gtgtggccgc 120
ggagttgact tccgccgcgg tctcgtatgg atcggtggtg tcgacgctga tcgttgagcc 180
gtggatgggg ccggcggcgg ccgcgatggc ggcccgggca acgccgtatg tggggtggct 240
ggccgccacg gcggcgtg cgaaggagac ggccacacag gcgagggcag cggcggaagc 300
gtttgggacg gcgttcgcga tgacggtgcc accatccctc gtcgcgggca accgcagccg 360
gttgatgtcg ctggctcgcg cgaacattct ggggcaaaac agtgcggcga tcgcggtac 420
ccaggccgag tatgccgaaa tgtgggcccc agacgctgcc gtgatgtaca gctatgaggg 480
ggcatctcgc gccgcgtcgc cgttgccgcc gttcactcca cccgtgcaag gcaccggccc 540
ggccgggccc gcggccgcag ccgcccgcac ccaagccgcc ggtgcgggcg ccgttgcgga 600
tgacagggcg aactggccc agctgcccc ggggatcctg agcgacattc tgtccgcatt 660
ggccgccaac gctgatccgc tgacatcggg actggtgggg atcgcgtcga ccctcaacc 720
gcaagtcgga tccgctcagc cgatagtgat cccaccccc ataggggaat tggacgtgat 780
cgcgctctac attgcatcca tcgcgaccgg cagcattgcg ctgcgatca cgaacacggc 840
cagaccctgg cacatcggcc tatacgggaa cgccggcggg ctgggaccga cgcagggcca 900
tccactgagt tcggcgaccg acgagccgga gccgcactgg ggccccttcg gggcgcgcc 960
gccggtgtcc gcgggcgtcg gccacgcagc attagtcgga gcggtgtcgg tgccgcacag 1020
ctggaccacg gccgccccgg agatccagct cgccgttcag gcaacacca ccttcagctc 1080
cagcgcgggc gccgacccga cggccctaaa cgggatgccg gcaggcctgc tcagcgggat 1140
ggctttggcg agcctggccg cacgcggcac gacgggaggg ggccggcacc gttagcggcac 1200
cagcactgac ggccaagagg acggccgcaa acccccggta gttgtgatta gagagcagcc 1260
gccgcccgga aacccccgc ggtaaaagtc cggcaaccgt tcgtcgccgc gcggaatg 1320
cctggtgagc gtggctatcc gacgggcccgt tcacaccgct tgtagtagcg tacggctatg 1380
gacgacggtg tctggattct cggcggctat cagagcgatt ttgctcgcaa cctcagcaaa 1441
g

```

<210> 31

10 <211> 423

<212> PRT

<213> *Mycobacterium tuberculosis*

<220>

15 <223> MTB41 (MTCC n° 2)

<400> 31

Met Asp Phe Gly Leu Leu Pro Pro Glu Val Asn Ser Ser Arg Met Tyr
1 5 10 15
Ser Gly Pro Gly Pro Glu Ser Met Leu Ala Ala Ala Ala Ala Trp Asp
20 25 30
Gly Val Ala Ala Glu Leu Thr Ser Ala Ala Val Ser Tyr Gly Ser Val
35 40 45
Val Ser Thr Leu Ile Val Glu Pro Trp Met Gly Pro Ala Ala Ala Ala
50 55 60
Met Ala Ala Ala Ala Thr Pro Tyr Val Gly Trp Leu Ala Ala Thr Ala
65 70 75 80
Ala Leu Ala Lys Glu Thr Ala Thr Gln Ala Arg Ala Ala Ala Glu Ala
85 90 95
Phe Gly Thr Ala Phe Ala Met Thr Val Pro Pro Ser Leu Val Ala Ala
100 105 110
Asn Arg Ser Arg Leu Met Ser Leu Val Ala Ala Asn Ile Leu Gly Gln
115 120 125
Asn Ser Ala Ala Ile Ala Ala Thr Gln Ala Glu Tyr Ala Glu Met Trp
130 135 140
Ala Gln Asp Ala Ala Val Met Tyr Ser Tyr Glu Gly Ala Ser Ala Ala
145 150 155 160
Ala Ser Ala Leu Pro Pro Phe Thr Pro Pro Val Gln Gly Thr Gly Pro
165 170 175
Ala Gly Pro Ala Ala Ala Ala Ala Thr Gln Ala Ala Gly Ala Gly
180 185 190
Ala Val Ala Asp Ala Gln Ala Thr Leu Ala Gln Leu Pro Pro Gly Ile
195 200 205
Leu Ser Asp Ile Leu Ser Ala Leu Ala Ala Asn Ala Asp Pro Leu Thr
210 215 220
Ser Gly Leu Leu Gly Ile Ala Ser Thr Leu Asn Pro Gln Val Gly Ser
225 230 235 240
Ala Gln Pro Ile Val Ile Pro Thr Pro Ile Gly Glu Leu Asp Val Ile
245 250 255
Ala Leu Tyr Ile Ala Ser Ile Ala Thr Gly Ser Ile Ala Leu Ala Ile
260 265 270
Thr Asn Thr Ala Arg Pro Trp His Ile Gly Leu Tyr Gly Asn Ala Gly
275 280 285
Gly Leu Gly Pro Thr Gln Gly His Pro Leu Ser Ser Ala Thr Asp Glu
290 295 300
Pro Glu Pro His Trp Gly Pro Phe Gly Gly Ala Ala Pro Val Ser Ala
305 310 315 320
Gly Val Gly His Ala Ala Leu Val Gly Ala Leu Ser Val Pro His Ser
325 330 335
Trp Thr Thr Ala Ala Pro Glu Ile Gln Leu Ala Val Gln Ala Thr Pro
340 345 350
Thr Phe Ser Ser Ser Ala Gly Ala Asp Pro Thr Ala Leu Asn Gly Met
355 360 365
Pro Ala Gly Leu Leu Ser Gly Met Ala Leu Ala Ser Leu Ala Ala Arg
370 375 380
Gly Thr Thr Gly Gly Gly Gly Thr Arg Ser Gly Thr Ser Thr Asp Gly
385 390 395 400
Gln Glu Asp Gly Arg Lys Pro Pro Val Val Val Ile Arg Glu Gln Pro
405 410 415
Pro Pro Gly Asn Pro Pro Arg
420

ES 2 374 620 T3

<211> 154
 <212> ADN
 <213> *Mycobacterium tuberculosis*

5 <220>
 <223> ESAT-6

<400> 32

atgacagagc agcagtggaa ttcgcggggt atcgaggccg cggcaagcgc aatccagggga 60
 aatgtcacgt ccattcattc cctccttgac gaggggaagc agtccctgac caagctcgca 120

10 gcggcctggg gcggtagcgg ttcggaagcg tacc 154

<210> 33
 <211> 51
 <212> PRT
 <213> *Mycobacterium tuberculosis*

15 <220>
 <223> ESAT-6

<400> 33

Met Thr Glu Gln Gln Trp Asn Phe Ala Gly Ile Glu Ala Ala Ala Ser
 1 5 10 15
 Ala Ile Gln Gly Asn Val Thr Ser Ile His Ser Leu Leu Asp Glu Gly
 20 25 30
 Lys Gln Ser Leu Thr Lys Leu Ala Ala Ala Trp Gly Gly Ser Gly Ser
 35 40 45
 Glu Ala Tyr
 50

20 <210> 34
 <211> 327
 <212> ADN
 <213> *Mycobacterium tuberculosis*

25 <220>
 <223> Tb38-1 o 38-1 (MTb11)

<400> 34

ES 2 374 620 T3

```

cggcacgaga gaccgatgcc gctaccctcg cgcaggaggc aggtaatttc gagcggatct 60
ccggcgacct gaaaaccag atcgaccagg tggagtcgac ggcaggttcg ttgcagggcc 120
agtggcgcgg cgcggcgggg acggccgcc aggccgcggt ggtgcgcttc caagaagcag 180
ccaataagca gaagcaggaa ctcgacgaga tctcgacgaa tattcgtcag gccggcgctcc 240
aatactcgag ggccgacgag gaggcagcag aggcgctgtc ctcgcaaatg ggcttctgac 300
ccgctaatac gaaaagaaac ggagcaa 327

```

<210> 35

<211> 95

<212> PRT

5 <213> *Mycobacterium tuberculosis*

<220>

<223> Tb38-1 o 38-1(MTb11)

10 <400> 35

```

  Thr Asp Ala Ala Thr Leu Ala Gln Glu Ala Gly Asn Phe Glu Arg Ile
  1      5      10      15
  Ser Gly Asp Leu Lys Thr Gln Ile Asp Gln Val Glu Ser Thr Ala Gly
      20      25      30
  Ser Leu Gln Gly Gln Trp Arg Gly Ala Ala Gly Thr Ala Ala Gln Ala
      35      40      45
  Ala Val Val Arg Phe Gln Glu Ala Ala Asn Lys Gln Lys Gln Glu Leu
      50      55      60
  Asp Glu Ile Ser Thr Asn Ile Arg Gln Ala Gly Val Gln Tyr Ser Arg
  65      70      75      80
  Ala Asp Glu Glu Gln Gln Gln Ala Leu Ser Ser Gln Met Gly Phe
      85      90      95

```

<210> 36

<211> 542

<212> ADN

15 <213> *Mycobacterium tuberculosis*

<220>

<223> TbRa3

<220>

20 <221> base_modificada

<222> (406)

<223> n= g, a, c o t

<400> 36

ES 2 374 620 T3

```

gaattcggca cgagaggtga tgcacatcat cgggaccagc cccacatcct gggaacaggc 60
ggcggcggag gcggtccagc gggcgcggga tagcgtcgat gacatccgcg tcgctcgggt 120
cattgagcag gacatggccg tggacagcgc cggcaagatc acctaccgca tcaagctcga 180
agtgtcgttc aagatgaggc cggcgcaacc gcgctagcac gggccggcga gcaagacgca 240
aaatcgcacg gtttgcgggt gattcgtgcg attttgtgtc tgctcgccga gccctaccag 300
gcgcggccca ggtccgcgtg ctgccgtatc caggcgtgca tcgcgattcc ggcggccacg 360
ccggagtaa tgcttcgctg cgacccgaac tgggcgatcc gccggngagc tgatcgatga 420
ccgtggccag cccgtcgatg cccgagttgc ccgaggaaac gtgctgccag gccggtagga 480
agcgtccgta ggcggcggtg ctgaccggct ctgcctgctc cctcagtgcg gccagcgagc 540
gg

```

<210> 37

<211> 66

<212> PRT

5 <213> *Mycobacterium tuberculosis*

<220>

<223> TbRa3

10 <400> 37

```

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

```

<210> 38

<211> 1993

<212> ADN

15 <213> *Mycobacterium tuberculosis*

<220>

<223> 38 kDa

20 <400> 38

ES 2 374 620 T3

```

tgttcttcga cggcaggctg gtggaggaag ggcccaccga acagctgttc tcctcgccga 60
agcatgcgga aaccgccccg tacgtcgccg gactgtcggg ggacgtcaag gacgccaagc 120
gcggaaaattg aagagcacag aaaggtatgg cgtgaaaatt cgtttgata cgctgttggc 180
cgtgttgacc gctgccccg tgctgctagc agcggcggg tgtggctcga aaccaccgag 240
cggttcgctt gaaacgggcg ccggcgccgg tactgtcgg actacccccg cgtcgtcggc 300
ggtgacgttg gcggagaccg gtagcacgct gctctaccgg ctgttcaacc tgtgggggtcc 360
ggcctttcac gagaggatc cgaacgtcac gatcaccgct cagggcaccg gttctggtgc 420
cgggatcgcg caggccgccc ccgggacggt caacattggg gcctccgacg cctatctgtc 480

ggaaggtgat atggccgcgc acaaggggct gatgaacatc gcgctagcca tctccgctca 540
gcaggtcaac tacaacctgc ccggagtgaag cgagcacctc aagctgaacg gaaaagtctt 600
ggcggccatg taccagggca ccatcaaaa ctgggacgac cgcagatcg ctgcgtcaa 660
ccccggcgtg aacctgcccc gcaccgcggt agttccgctg caccgctccg acgggtccgg 720
tgacaccttc ttgttcaccc agtacctgtc caagcaagat cccgagggct ggggcaagtc 780
gccccgcttc ggcaccaccg tcgacttccc ggcggtgccc ggtgcgctgg gtgagaacgg 840
caacggcggc atggtgaccg gttgcgccga gacaccgggc tgcgtggcct atatcgcat 900
cagcttcttc gaccaggcca gtcaacgggg actcggcgag gcccaactag gcaatagctc 960
tggcaatttc ttgttgcggc acgcgcaaag cttcaggcc gcggcggtg gcttcgcatc 1020
gaaaaccccc gcgaaccagg cgatttcgat gatcgacggg cccgccccgg acggctacct 1080
gatcatcaac tacgagtacg ccatcgtcaa caaccggcaa aaggacgccg ccaccgcgca 1140
gaccttgacg gcatttctgc actgggcatg caccgacggc aacaaggcct cgttcctcga 1200
ccaggttcat ttccagcccg tgccgcccgc ggtgggtaag ttgtctgacg cgttgatcgc 1260
gacgatttcc agctagcctc gttgaccacc acgcgacagc aacctccgct gggccatcgg 1320
gctgcttttg ggagcatgct ggcccgtgcc ggtgaagtcg gccgcgctgg cccggccatc 1380
cggtggttgg gtgggatagg tgcggtgatc ccgctgctt cgctggtctt ggtgctggtg 1440
gtgctggtca tcgaggcgat ggggtgcgat aggtcaacg ggttgcattt cttaccgcc 1500
accgaatgga atccaggcaa cacctacggc gaaaccgtt tcaccgacgc gtcgcccac 1560
cggtcggcgc ctactacggg gcgttgcccg tgatcgtcgg gacgctggcg acctcgcaa 1620
tcgccctgat catcgcggtg ccggtctctg taggagcggc gctggtgatc gtggaacggc 1680
tgccgaaacg gttggccgag gctgtgggaa tagtctgga attgctcgcc ggaatcccca 1740
gcgtggtcgt cggtttgtgg ggggcaatga cgttcgggcc gttcatcgct catcacatcg 1800
ctccggtgat cgctcacaac gctcccgatg tgccggtgct gaactactg cgcggcgacc 1860
cgggcaacgg ggagggcatg ttggtgtccg gctggtggt ggcggtgatg gtcgttccca 1920
ttatcgccac caccactcat gacctgttcc ggcaggtgcc ggtgttggcc cgggagggcg 1980
cgatcgggaa ttc 1993

```

<210> 39

<211> 374

5 <212> PRT

<213> *Mycobacterium tuberculosis*

<220>

<223> 38 kDa

10

<400> 39

ES 2 374 620 T3

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

ES 2 374 620 T3

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

<210> 40

<211> 999

<212> ADN

5 <213> *Mycobacterium tuberculosis*

<220>

<223> DPEP

10 <400> 40

ES 2 374 620 T3

```

atgcatcacc atcaccatca catgcatcag gtggaccca acttgacacg tcgcaagggg 60
cgattggcgg cactggctat cgcggcgatg gccagcgcca gcctggtgac cgttgcggtg 120
cccgcgaccg ccaacgccga tccggagcca gcgcccccg taccacaac ggccgcctcg 180
ccgccgtcga ccgctgcagc gccaccgca ccggcgacac ctggtgccc cccaccaccg 240
gccgcccga acacgccgaa tgcccagccg ggcgatcca acgcagcacc tccgccggcc 300
gacccgaacg caccgccgc acctgtcatt gccccaaacg caccacaacc tgtccggatc 360
gacaaccccg ttggaggatt cagcttcgct ctgcctgctg gctgggtgga gtctgacgcc 420
gcccacttcg actacggttc agcactcctc agcaaaacca ccggggacc gccatttccc 480
ggacagccgc cgccggtggc caatgacacc cgtatcgtgc tcggccggct agaccaaaag 540
ctttacgcca gcgccgaagc caccgactcc aaggccgctg cccggttggg ctcgacatg 600
ggtgagttct atatgcccta cccgggcacc cggatcaacc aggaaaccgt ctgctcgac 660
gccaacgggg tgtctggaag cgcgtcgtat tacgaagtca agttcagcga tccgagtaag 720
ccgaacggcc agatctggac gggcgtaatc ggctcgccc cggcgaacgc accggacgcc 780
gggccccctc agcgctggtt tgtggtatgg ctcgggaccg ccaacaacc ggtggacaag 840
ggcgcggcca aggcgctggc cgaatcgatc cggcctttgg tcgccccgcc gccggcgccg 900
gcaccggctc ctgcagagcc cgctccggcg ccggcgccgg ccggggaagt cgctcctacc 960
ccgacgacac cgacaccgca gcggaccta ccggcctga 999

```

<210> 41

<211> 332

<212> PRT

5 <213> *Mycobacterium tuberculosis*

<220>

<223> DPEP

10 <400> 41

Met His His His His His His Met His Gln Val Asp Pro Asn Leu Thr
 1 5 10
 Arg Arg Lys Gly Arg⁴ Leu Ala Ala Leu Ala Ile Ala Ala Met Ala Ser
 20 25 30
 Ala Ser Leu Val Thr Val Ala Val Pro Ala Thr Ala Asn Ala Asp Pro
 35 40 45
 Glu Pro Ala Pro Pro Val Pro Thr Thr Ala Ala Ser Pro Pro Ser Thr
 50 55 60
 Ala Ala Ala Pro Pro Ala Pro Ala Thr Pro Val Ala Pro Pro Pro Pro
 65 70 75 80
 Ala Ala Ala Asn Thr Pro Asn Ala Gln Pro Gly Asp Pro Asn Ala Ala
 85 90 95
 Pro Pro Pro Ala Asp Pro Asn Ala Pro Pro Pro Pro Val Ile Ala Pro
 100 105 110
 Asn Ala Pro Gln Pro Val Arg Ile Asp Asn Pro Val Gly Phe Ser
 115 120 125
 Phe Ala Leu Pro Ala Gly Trp Val Glu Ser Asp Ala Ala His Phe Asp
 130 135 140
 Tyr Gly Ser Ala Leu Leu Ser Lys Thr Thr Gly Asp Pro Pro Phe Pro
 145 150 155 160
 Gly Gln Pro Pro Pro Val Ala Asn Asp Thr Arg Ile Val Leu Gly Arg
 165 170 175
 Leu Asp Gln Lys Leu Tyr Ala Ser Ala Glu Ala Thr Asp Ser Lys Ala
 180 185 190
 Ala Ala Arg Leu Gly Ser Asp Met Gly Glu Phe Tyr Met Pro Tyr Pro
 195 200 205
 Gly Thr Arg Ile Asn Gln Glu Thr Val Ser Leu Asp Ala Asn Gly Val
 210 215 220
 Ser Gly Ser Ala Ser Tyr Tyr Glu Val Lys Phe Ser Asp Pro Ser Lys
 225 230 235 240
 Pro Asn Gly Gln Ile Trp Thr Gly Val Ile Gly Ser Pro Ala Ala Asn
 245 250 255
 Ala Pro Asp Ala Gly Pro Pro Gln Arg Trp Phe Val Val Trp Leu Gly
 260 265 270
 Thr Ala Asn Asn Pro Val Asp Lys Gly Ala Ala Lys Ala Leu Ala Glu
 275 280 285
 Ser Ile Arg Pro Leu Val Ala Pro Pro Ala Pro Ala Pro Ala Pro
 290 295 300
 Ala Glu Pro Ala Pro Ala Pro Ala Pro Ala Gly Glu Val Ala Pro Thr
 305 310 315 320
 Pro Thr Thr Pro Thr Pro Gln Arg Thr Leu Pro Ala
 325 330

<210> 42

<211> 702

<212> ADN

5 <213> *Mycobacterium tuberculosis*

<220>

<223> TbH4

<220>

10 <221> base_modificada

<222> (1)...(702)

<223> n= g, a, c o t

ES 2 374 620 T3

```

<400>          42
cggcacgagg atcgggtacc cgcggcatcg gcagctgccg attcgccggg tttccccacc 60
cgaggaaagc cgctaccaga tggcgctgcc gaagtagggc gatccgttcg cgatgccggc 120
atgaacgggc ggcatcaaat tagtgcagga acctttcagt ttagcgacga taatggctat 180
agcactaagg aggatgatcc gatatgacgc agtcgcagac cgtgacggtg gatcagcaag 240
agatTTTgaa cagggccaac gaggtggagg ccccgatggc ggaccaccg actgatgtcc 300
ccatcacacc gtgcgaactc acgnggnta aaaacgccgc ccaacagntg gtnttgccg 360
ccgacaacat gcgggaatac ctggcgcccg gtgccaaaga gcggcagcgt ctggcgacct 420
cgctgcgcaa cgcgccaag gngtatggcg aggttgatga ggaggctgcg accgcgctgg 480
acaacgacgg cgaaggaact gtgcaggcag aatcggccgg ggccgtcgga ggggacagtt 540
cggccgaact aaccgatacg ccgagggtgg ccacggccgg tgaacccaac ttcattggatc 600
tcaaagaagc ggcaaggaag ctcgaaacgg gcgaccaagg cgcattcgctc gcgcactgng 660
gggatgggtg gaacacttnc accctgacgc tgcaaggcga cg                               702

```

<210> 43

5 <211> 286

<212> PRT

<213> *Mycobacterium tuberculosis*

<220>

10 <223> TbH4

<220>

<221> MOD_RES

<222> (1)...(286)

<223> Xaa= cualquier aminoácido

15

<400> 43

Gly Asp Ser Phe Trp Ala Ala Ala Asp Gln Met Ala Arg Gly Phe Val
 1 5 10 15
 Leu Gly Ala Thr Ala Gly Arg Thr Thr Leu Thr Gly Glu Gly Leu Gln
 20 25 30
 His Ala Asp Gly His Ser Leu Leu Leu Asp Ala Thr Asn Pro Ala Val
 35 40 45
 Val Ala Tyr Asp Pro Ala Phe Ala Tyr Glu Ile Gly Tyr Ile Xaa Glu
 50 55 60
 Ser Gly Leu Ala Arg Met Cys Gly Glu Asn Pro Glu Asn Ile Phe Phe
 65 70 75 80
 Tyr Ile Thr Val Tyr Asn Glu Pro Tyr Val Gln Pro Pro Glu Pro Glu
 85 90 95
 Asn Phe Asp Pro Glu Gly Val Leu Gly Gly Ile Tyr Arg Tyr His Ala
 100 105 110
 Ala Thr Glu Gln Arg Thr Asn Lys Xaa Gln Ile Leu Ala Ser Gly Val
 115 120 125
 Ala Met Pro Ala Ala Leu Arg Ala Ala Gln Met Leu Ala Ala Glu Trp
 130 135 140
 Asp Val Ala Ala Asp Val Trp Ser Val Thr Ser Trp Gly Glu Leu Asn
 145 150 155 160
 Arg Asp Gly Val Val Ile Glu Thr Glu Lys Leu Arg His Pro Asp Arg
 165 170 175
 Pro Ala Gly Val Pro Tyr Val Thr Arg Ala Leu Glu Asn Ala Arg Gly
 180 185 190
 Pro Val Ile Ala Val Ser Asp Trp Met Arg Ala Val Pro Glu Gln Ile
 195 200 205
 Arg Pro Trp Val Pro Gly Thr Tyr Leu Thr Leu Gly Thr Asp Gly Phe
 210 215 220
 Gly Phe Ser Asp Thr Arg Pro Ala Gly Arg Arg Tyr Phe Asn Thr Asp
 225 230 235 240
 Ala Glu Ser Gln Val Gly Arg Gly Phe Gly Arg Gly Trp Pro Gly Arg
 245 250 255
 Arg Val Asn Ile Asp Pro Phe Gly Ala Gly Arg Gly Pro Pro Ala Gln
 260 265 270
 Leu Pro Gly Phe Asp Glu Gly Gly Gly Leu Arg Pro Xaa Lys
 275 280 285

<210> 44

<211> 339

<212> ADN

5 <213> *Mycobacterium tuberculosis*

<220>

<223> DPPD, ADN genómico

10 <400> 44

atgaagtga agtttgctc cctgagtact gcgatactgg gttgtgcagc ggcgcttgtg 60
 ttccctgcct cggttgccag cgcagatcca cctgacccgc atcagccgga catgacgaaa 120
 ggctattgcc cgggtggccg atggggtttt ggcgacttgg ccgtgtgcga cggcgagaag 180
 taccctgacg gctcgtttg gcaccagtgg atgcaaacgt ggtttaccgg cccacagttt 240
 tacttcgatt gtgtcagcgg cggtagccc ctccccggcc cgccgccacc gggtagttgc 300
 ggtggggcaa ttccgtccga gcagcccaac gctccctga 339

<210> 45

ES 2 374 620 T3

<211> 112
 <212> PRT
 <213> *Mycobacterium tuberculosis*

5 <220>
 <223> DPPD

<400> 45

```

Met Lys Leu Lys Phe Ala Arg Leu Ser Thr Ala Ile Leu Gly Cys Ala
 1           5           10           15
Ala Ala Leu Val Phe Pro Ala Ser Val Ala Ser Ala Asp Pro Pro Asp
           20           25           30
Pro His Gln Pro Asp Met Thr Lys Gly Tyr Cys Pro Gly Gly Arg Trp
           35           40           45
Gly Phe Gly Asp Leu Ala Val Cys Asp Gly Glu Lys Tyr Pro Asp Gly
           50           55           60
Ser Phe Trp His Gln Trp Met Gln Thr Trp Phe Thr Gly Pro Gln Phe
           65           70           75           80
Tyr Phe Asp Cys Val Ser Gly Gly Glu Pro Leu Pro Gly Pro Pro Pro
           85           90           95
Pro Gly Gly Cys Gly Gly Ala Ile Pro Ser Glu Gln Pro Asn Ala Pro
           100           105           110
  
```

10 <210> 46
 <211> 921
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> Descripción de la secuencia artificial: proteína de fusión triple DPV-MTI-MSL (designada MTb31F),
 ADNc

<220>
 <221> CDS

20 <222> (1)..(900)
 <223> MTb31F

<400> 4 6

ES 2 374 620 T3

cat His 1	atg Met	cat His	cac His	cat His 5	cac His	cat His	cac His	gat Asp 10	ccc Pro	gtg Val	gac Asp	gcg Ala	gtc Val	att Ile 15	aac Asn	48
acc Thr	acc Thr	tgc Cys	aat Asn 20	tac Tyr	ggg Gly	cag Gln	gta Val	gta Val 25	gct Ala	gcg Ala	ctc Leu	aac Asn	gcg Ala 30	acg Thr	gat Asp	96
ccg Pro	ggg Gly	gct Ala 35	gcc Ala	gca Ala	cag Gln	ttc Phe	aac Asn 40	gcc Ala	tca Ser	ccg Pro	gtg Val	gcg Ala 45	cag Gln	tcc Ser	tat Tyr	144
ttg Leu	cgc Arg 50	aat Asn	ttc Phe	ctc Leu	gcc Ala	gca Ala 55	ccg Pro	cca Pro	cct Pro	cag Gln	cgc Arg 60	gct Ala	gcc Ala	atg Met	gcc Ala	192
gcg Ala 65	caa Gln	ttg Leu	caa Gln	gct Ala	gtg Val 70	ccg Pro	ggg Gly	gcg Ala	gca Ala	cag Gln 75	tac Tyr	atc Ile	ggc Gly	ctt Leu	gtc Val 80	240
gag Glu	tcg Ser	gtt Val	gcc Ala	ggc Gly 85	tcc Ser	tgc Cys	aac Asn	aac Asn	tat Tyr 90	gag Glu	ctc Leu	atg Met	acg Thr	att Ile 95	aat Asn	288
tac Tyr	cag Gln	ttc Phe	ggg Gly 100	gac Asp	gtc Val	gac Asp	gct Ala	cat His 105	ggc Gly	gcc Ala	atg Met	atc Ile	cgc Arg 110	gct Ala	cag Gln	336
gcg Ala	gcg Ala	tcg Ser 115	ctt Leu	gag Glu	gcg Ala	gag Glu	cat His 120	cag Gln	gcc Ala	atc Ile	gtt Val	cg Arg 125	gat Asp	gtg Val	ttg Leu	384
gcc Ala 130	gcg Ala	ggt Gly	gac Asp	ttt Phe	tgg Trp	ggc Gly 135	ggc Gly	gcc Ala	ggt Gly	tcg Ser	gtg Val 140	gct Ala	tgc Cys	cag Gln	gag Glu	432
ttc Phe 145	att Ile	acc Thr	cag Gln	ttg Leu	ggc Gly 150	cg Arg	aac Asn	ttc Phe	cag Gln	gtg Val 155	atc Ile	tac Tyr	gag Glu	cag Gln	gcc Ala 160	480
aac Asn	gcc Ala	cac His	ggg Gly	cag Gln 165	aag Lys	gtg Val	cag Gln	gct Ala	gcc Ala 170	ggc Gly	aac Asn	aac Asn	atg Met	gcg Ala 175	caa Gln	528

ES 2 374 620 T3

acc gac agc gcc gtc ggc tcc agc tgg gcc act agt atg agc ctt ttg 576
 Thr Asp Ser Ala Val Gly Ser Ser Trp Ala Thr Ser Met Ser Leu Leu
 180 185 190

gat gct cat atc cca cag ttg gtg gcc tcc cag tcg gcg ttt gcc gcc 624
 Asp Ala His Ile Pro Gln Leu Val Ala Ser Gln Ser Ala Phe Ala Ala
 195 200 205

aag gcg ggg ctg atg cgg cac acg atc ggt cag gcc gag cag gcg gcg 672
 Lys Ala Gly Leu Met Arg His Thr Ile Gly Gln Ala Glu Gln Ala Ala
 210 215 220

atg tcg gct cag gcg ttt cac cag ggg gag tcg tcg gcg gcg ttt cag 720
 Met Ser Ala Gln Ala Phe His Gln Gly Glu Ser Ser Ala Ala Phe Gln
 225 230 235 240

gcc gcc cat gcc cgg ttt gtg gcg gcg gcc gcc aaa gtc aac acc ttg 768
 Ala Ala His Ala Arg Phe Val Ala Ala Ala Lys Val Asn Thr Leu
 245 250 255

ttg gat gtc gcg cag gcg aat ctg ggt gag gcc gcc ggt acc tat gtg 816
 Leu Asp Val Ala Gln Ala Asn Leu Gly Glu Ala Ala Gly Thr Tyr Val
 260 265 270

gcc gcc gat gct gcg gcc gcg tcg acc tat acc ggg ttc gat atc cat 864
 Ala Ala Asp Ala Ala Ala Ala Ser Thr Tyr Thr Gly Phe Asp Ile His
 275 280 285

cac act ggc ggc cgc tcg agc aga tcc ggc tgc taa caaagcccga 910
 His Thr Gly Gly Arg Ser Ser Arg Ser Gly Cys 300

aaggaagctg a 921

<210> 47

<211> 299

<212> PRT

5 <213> Secuencia Artificial

<223> Descripción de la secuencia artificial: proteína de fusión triple DPV-MTI-MSL (designada MTb31F), ADNc

10 <400> 47

His Met His His His His His His Asp Pro Val Asp Ala Val Ile Asn
 1 5 10 15

Thr Thr Cys Asn Tyr Gly Gln Val Val Ala Ala Leu Asn Ala Thr Asp
 20 25 30

Pro Gly Ala Ala Ala Gln Phe Asn Ala Ser Pro Val Ala Gln Ser Tyr
 35 40 45

Leu Arg Asn Phe Leu Ala Ala Pro Pro Pro Gln Arg Ala Ala Met Ala
 50 55 60
 Ala Gln Leu Gln Ala Val Pro Gly Ala Ala Gln Tyr Ile Gly Leu Val
 65 70 75 80
 Glu Ser Val Ala Gly Ser Cys Asn Asn Tyr Glu Leu Met Thr Ile Asn
 85 90 95
 Tyr Gln Phe Gly Asp Val Asp Ala His Gly Ala Met Ile Arg Ala Gln
 100 105 110
 Ala Ala Ser Leu Glu Ala Glu His Gln Ala Ile Val Arg Asp Val Leu
 115 120 125
 Ala Ala Gly Asp Phe Trp Gly Gly Ala Gly Ser Val Ala Cys Gln Glu
 130 135 140
 Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile Tyr Glu Gln Ala
 145 150 155 160
 Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn Asn Met Ala Gln
 165 170 175
 Thr Asp Ser Ala Val Gly Ser Ser Trp Ala Thr Ser Met Ser Leu Leu
 180 185 190
 Asp Ala His Ile Pro Gln Leu Val Ala Ser Gln Ser Ala Phe Ala Ala
 195 200 205
 Lys Ala Gly Leu Met Arg His Thr Ile Gly Gln Ala Glu Gln Ala Ala
 210 215 220
 Met Ser Ala Gln Ala Phe His Gln Gly Glu Ser Ser Ala Ala Phe Gln
 225 230 235 240
 Ala Ala His Ala Arg Phe Val Ala Ala Ala Ala Lys Val Asn Thr Leu
 245 250 255
 Leu Asp Val Ala Gln Ala Asn Leu Gly Glu Ala Ala Gly Thr Tyr Val
 260 265 270
 Ala Ala Asp Ala Ala Ala Ala Ser Thr Tyr Thr Gly Phe Asp Ile His
 275 280 285
 His Thr Gly Gly Arg Ser Ser Arg Ser Gly Cys
 290 295

<210> 48

<211> 2168

<212> ADN

5 <213> Secuencia artificial

<220>

ES 2 374 620 T3

<223> Descripción de la secuencia artificial: proteína de fusión cuádruple DPV-MTI-MSL-MTCC nº2 (designada MTb71F)

<220>

<221> CDS

5 <222> (1)..(2133)

<223> MTb71F

<400> 48

cat	atg	cat	cac	cat	cac	cat	cac	gat	ccc	gtg	gac	gcg	gtc	att	aac	48
His	Met	His	His	His	His	His	His	Asp	Pro	Val	Asp	Ala	Val	Ile	Asn	
1				5					10					15		
acc	acc	tgc	aat	tac	ggg	cag	gta	gta	gct	gcg	ctc	aac	gcg	acg	gat	96
Thr	Thr	Cys	Asn	Tyr	Gly	Gln	Val	Val	Ala	Ala	Leu	Asn	Ala	Thr	Asp	
			20					25					30			
ccg	ggg	gct	gcc	gca	cag	ttc	aac	gcc	tca	ccg	gtg	gcg	cag	tcc	tat	144
Pro	Gly	Ala	Ala	Ala	Gln	Phe	Asn	Ala	Ser	Pro	Val	Ala	Gln	Ser	Tyr	
		35					40					45				
ttg	cgc	aat	ttc	ctc	gcc	gca	ccg	cca	cct	cag	cgc	gct	gcc	atg	gcc	192
Leu	Arg	Asn	Phe	Leu	Ala	Ala	Pro	Pro	Pro	Gln	Arg	Ala	Ala	Met	Ala	
	50				55						60					
gcg	caa	ttg	caa	gct	gtg	ccg	ggg	gcg	gca	cag	tac	atc	ggc	ctt	gtc	240
Ala	Gln	Leu	Gln	Ala	Val	Pro	Gly	Ala	Ala	Gln	Tyr	Ile	Gly	Leu	Val	
65					70					75					80	
gag	tcg	gtt	gcc	ggc	tcc	tgc	aac	aac	tat	gag	ctc	atg	acg	att	aat	288

ES 2 374 620 T3

Glu	Ser	Val	Ala	Gly 85	Ser	Cys	Asn	Asn	Tyr 90	Glu	Leu	Met	Thr	Ile 95	Asn		
tac	cag	ttc	ggg	gac	gtc	gac	gct	cat	ggc	gcc	atg	atc	cgc	gct	cag		336
Tyr	Gln	Phe	Gly 100	Asp	Val	Asp	Ala	His 105	Gly	Ala	Met	Ile	Arg 110	Ala	Gln		
gcg	gcg	tcg	ctt	gag	gcg	gag	cat	cag	gcc	atc	gtt	cgt	gat	gtg	ttg		384
Ala	Ala	Ser 115	Leu	Glu	Ala	Glu	His 120	Gln	Ala	Ile	Val	Arg 125	Asp	Val	Leu		
gcc	gcg	ggt	gac	ttt	tgg	ggc	ggc	gcc	ggt	tcg	gtg	gct	tgc	cag	gag		432
Ala	Ala	Gly 130	Asp	Phe	Trp	Gly 135	Gly	Ala	Gly	Ser	Val	Ala	Cys	Gln	Glu		
ttc	att	acc	cag	ttg	ggc	cgt	aac	ttc	cag	gtg	atc	tac	gag	cag	gcc		480
Phe	Ile	Thr	Gln	Leu	Gly 150	Arg	Asn	Phe	Gln	Val 155	Ile	Tyr	Glu	Gln	Ala 160		
aac	gcc	cac	ggg	cag	aag	gtg	cag	gct	gcc	ggc	aac	aac	atg	gcg	caa		528
Asn	Ala	His	Gly 165	Gln	Lys	Val	Gln	Ala	Ala 170	Gly	Asn	Asn	Met	Ala 175	Gln		
acc	gac	agc	gcc	gtc	ggc	tcc	agc	tgg	gcc	act	agt	atg	agc	ctt	ttg		576
Thr	Asp	Ser	Ala 180	Val	Gly	Ser	Ser	Trp 185	Ala	Thr	Ser	Met	Ser 190	Leu	Leu		
gat	gct	cat	atc	cca	cag	ttg	gtg	gcc	tcc	cag	tcg	gcg	ttt	gcc	gcc		624
Asp	Ala	His 195	Ile	Pro	Gln	Leu	Val 200	Ala	Ser	Gln	Ser	Ala 205	Phe	Ala	Ala		
aag	gcg	ggg	ctg	atg	cgg	cac	acg	atc	ggt	cag	gcc	gag	cag	gcg	gcg		672
Lys	Ala 210	Gly	Leu	Met	Arg	His 215	Thr	Ile	Gly	Gln	Ala 220	Glu	Gln	Ala	Ala		
atg	tcg	gct	cag	gcg	ttt	cac	cag	ggg	gag	tcg	tcg	gcg	gcg	ttt	cag		720
Met	Ser	Ala	Gln	Ala	Phe 230	His	Gln	Gly	Glu	Ser 235	Ser	Ala	Ala	Phe	Gln 240		
gcc	gcc	cat	gcc	cgg	ttt	gtg	gcg	gcg	gcc	gcc	aaa	gtc	aac	acc	ttg		768
Ala	Ala	His	Ala 245	Arg	Phe	Val	Ala	Ala	Ala 250	Ala	Lys	Val	Asn	Thr 255	Leu		
ttg	gat	gtc	gcg	cag	gcg	aat	ctg	ggt	gag	gcc	gcc	ggt	acc	tat	gtg		816
Leu	Asp	Val 260	Ala	Gln	Ala	Asn	Leu	Gly 265	Glu	Ala	Ala	Gly	Thr 270	Tyr	Val		
gcc	gcc	gat	gct	gcg	gcc	gcg	tcg	acc	tat	acc	ggg	ttc	gat	atc	atg		864
Ala	Ala	Asp 275	Ala	Ala	Ala	Ala	Ser 280	Thr	Tyr	Thr	Gly	Phe 285	Asp	Ile	Met		
gat	ttc	ggg	ctt	tta	cct	ccg	gaa	gtg	aat	tca	agc	cga	atg	tat	tcc		912
Asp	Phe 290	Gly	Leu	Leu	Pro	Pro 295	Glu	Val	Asn	Ser	Ser 300	Arg	Met	Tyr	Ser		
ggt	ccg	ggg	ccg	gag	tcg	atg	cta	gcc	gcc	gcg	gcc	gcc	tgg	gac	ggt		960
Gly	Pro	Gly	Pro	Glu 310	Ser	Met	Leu	Ala	Ala 315	Ala	Ala	Ala	Trp	Asp 320	Gly		
gtg	gcc	gcg	gag	ttg	act	tcc	gcc	gcg	gtc	tcg	tat	gga	tcg	gtg	gtg		1008
Val	Ala	Ala	Glu 325	Leu	Thr	Ser	Ala	Ala	Val 330	Ser	Tyr	Gly	Ser	Val 335	Val		
tcg	acg	ctg	atc	gtt	gag	ccg	tgg	atg	ggg	ccg	gcg	gcg	gcc	gcg	atg		1056
Ser	Thr	Leu 340	Ile	Val	Glu	Pro	Trp	Met 345	Gly	Pro	Ala	Ala	Ala 350	Ala	Met		
gcg	gcc	gcg	gca	acg	ccg	tat	gtg	ggg	tgg	ctg	gcc	gcc	acg	gcg	gcg		1104
Ala	Ala	Ala	Ala	Thr	Pro	Tyr	Val	Gly	Trp	Leu	Ala	Ala	Thr	Ala	Ala		

ES 2 374 620 T3

355					360					365						
ctg	gcg	aag	gag	acg	gcc	aca	cag	gcg	agg	gca	gcg	gcg	gaa	gcg	ttt	1152
Leu	Ala	Lys	Glu	Thr	Ala	Thr	Gln	Ala	Arg	Ala	Ala	Ala	Glu	Ala	Phe	
	370					375					380					
ggg	acg	gcg	ttc	gcg	atg	acg	gtg	cca	cca	tcc	ctc	gtc	gcg	gcc	aac	1200
Gly	Thr	Ala	Phe	Ala	Met	Thr	Val	Pro	Pro	Ser	Leu	Val	Ala	Ala	Asn	
385					390					395					400	
cgc	agc	cgg	ttg	atg	tcg	ctg	gtc	gcg	gcg	aac	att	ctg	ggg	caa	aac	1248
Arg	Ser	Arg	Leu	Met	Ser	Leu	Val	Ala	Ala	Asn	Ile	Leu	Gly	Gln	Asn	
				405						410				415		
agt	gcg	gcg	atc	gcg	gct	acc	cag	gcc	gag	tat	gcc	gaa	atg	tgg	gcc	1296
Ser	Ala	Ala	Ile	Ala	Ala	Thr	Gln	Ala	Glu	Tyr	Ala	Glu	Met	Trp	Ala	
			420					425					430			
caa	gac	gct	gcc	gtg	atg	tac	agc	tat	gag	ggg	gca	tct	gcg	gcc	gcg	1344
Gln	Asp	Ala	Ala	Val	Met	Tyr	Ser	Tyr	Glu	Gly	Ala	Ser	Ala	Ala	Ala	
		435					440					445				
tcg	gcg	ttg	ccg	ccg	ttc	act	cca	ccc	gtg	caa	ggc	acc	ggc	ccg	gcc	1392
Ser	Ala	Leu	Pro	Pro	Phe	Thr	Pro	Pro	Val	Gln	Gly	Thr	Gly	Pro	Ala	
	450					455					460					
ggg	ccc	gcg	gcc	gca	gcc	gcg	gcg	acc	caa	gcc	gcc	ggt	gcg	ggc	gcc	1440
Gly	Pro	Ala	Ala	Ala	Ala	Ala	Ala	Thr	Gln	Ala	Ala	Gly	Ala	Gly	Ala	
465					470					475					480	
gtt	gcg	gat	gca	cag	gcg	aca	ctg	gcc	cag	ctg	ccc	ccg	ggg	atc	ctg	1488
Val	Ala	Asp	Ala	Gln	Ala	Thr	Leu	Ala	Gln	Leu	Pro	Pro	Gly	Ile	Leu	
				485					490					495		
agc	gac	att	ctg	tcc	gca	ttg	gcc	gcc	aac	gct	gat	ccg	ctg	aca	tcg	1536
Ser	Asp	Ile	Leu	Ser	Ala	Leu	Ala	Ala	Asn	Ala	Asp	Pro	Leu	Thr	Ser	
			500					505					510			
gga	ctg	ttg	ggg	atc	gcg	tcg	acc	ctc	aac	ccg	caa	gtc	gga	tcc	gct	1584
Gly	Leu	Leu	Gly	Ile	Ala	Ser	Thr	Leu	Asn	Pro	Gln	Val	Gly	Ser	Ala	
		515					520					525				
cag	ccg	ata	gtg	atc	ccc	acc	ccg	ata	ggg	gaa	ttg	gac	gtg	atc	gcg	1632
Gln	Pro	Ile	Val	Ile	Pro	Thr	Pro	Ile	Gly	Glu	Leu	Asp	Val	Ile	Ala	
	530					535					540					
ctc	tac	att	gca	tcc	atc	gcg	acc	ggc	agc	att	gcg	ctc	gcg	atc	acg	1680
Leu	Tyr	Ile	Ala	Ser	Ile	Ala	Thr	Gly	Ser	Ile	Ala	Leu	Ala	Ile	Thr	
	545				550					555					560	
aac	acg	gcc	aga	ccc	tgg	cac	atc	ggc	cta	tac	ggg	aac	gcc	ggc	ggg	1728
Asn	Thr	Ala	Arg	Pro	Trp	His	Ile	Gly	Leu	Tyr	Gly	Asn	Ala	Gly	Gly	
				565					570					575		
ctg	gga	ccg	acg	cag	ggc	cat	cca	ctg	agt	tcg	gcg	acc	gac	gag	ccg	1776
Leu	Gly	Pro	Thr	Gln	Gly	His	Pro	Leu	Ser	Ser	Ala	Thr	Asp	Glu	Pro	
			580					585					590			
gag	ccg	cac	tgg	ggc	ccc	ttc	ggg	ggc	gcg	gcg	ccg	gtg	tcc	gcg	ggc	1824
Glu	Pro	His	Trp	Gly	Pro	Phe	Gly	Gly	Ala	Ala	Pro	Val	Ser	Ala	Gly	
		595					600					605				
gtc	ggc	cac	gca	gca	tta	gtc	gga	gcg	ttg	tcg	gtg	ccg	cac	agc	tgg	1872
Val	Gly	His	Ala	Ala	Leu	Val	Gly	Ala	Leu	Ser	Val	Pro	His	Ser	Trp	
	610					615					620					
acc	acg	gcc	gcc	ccg	gag	atc	cag	ctc	gcc	gtt	cag	gca	aca	ccc	acc	1920
Thr	Thr	Ala	Ala	Pro	Glu	Ile	Gln	Leu	Ala	Val	Gln	Ala	Thr	Pro	Thr	
					630					635					640	

ES 2 374 620 T3

ttc agc tcc agc gcc ggc gcc gac ccg acg gcc cta aac ggg atg ccg 1968
 Phe Ser Ser Ser Ala Gly Ala Asp Pro Thr Ala Leu Asn Gly Met Pro
 645 650 655
 gca ggc ctg ctc agc ggg atg gct ttg gcg agc ctg gcc gca cgc ggc 2016
 Ala Gly Leu Leu Ser Gly Met Ala Leu Ala Ser Leu Ala Ala Arg Gly
 660 665 670
 acg acg ggc ggt ggc ggc acc cgt agc ggc acc agc act gac ggc caa 2064
 Thr Thr Gly Gly Gly Gly Thr Arg Ser Gly Thr Ser Thr Asp Gly Gln
 675 680 685
 gag gac ggc cgc aaa ccc ccg gta gtt gtg att aga gag cag ccg ccg 2112
 Glu Asp Gly Arg Lys Pro Pro Val Val Val Ile Arg Glu Gln Pro Pro
 690 695 700
 ccc gga aac ccc ccg cgg taa gatttctaaa tccatcacac tggcggccgc 2163
 Pro Gly Asn Pro Pro Arg
 705 710

tcgag 2168

<210> 49

<211> 710

<212> PRT

5 <213> Secuencia artificial

<223> Descripción de la secuencia artificial: proteína de fusión cuádruple DPV-MTI-MSL-MTCC n°2 (designada MTb71F)

10 <400> 49

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

ES 2 374 620 T3

Phe Ile Thr Gln Leu Gly Arg Asn Phe Gln Val Ile Tyr Glu Gln Ala
 145 150 155 160
 Asn Ala His Gly Gln Lys Val Gln Ala Ala Gly Asn Asn Met Ala Gln
 165 170 175
 Thr Asp Ser Ala Val Gly Ser Ser Trp Ala Thr Ser Met Ser Leu Leu
 180 185 190
 Asp Ala His Ile Pro Gln Leu Val Ala Ser Gln Ser Ala Phe Ala Ala
 195 200 205
 Lys Ala Gly Leu Met Arg His Thr Ile Gly Gln Ala Glu Gln Ala Ala
 210 215 220
 Met Ser Ala Gln Ala Phe His Gln Gly Glu Ser Ser Ala Ala Phe Gln
 225 230 235 240
 Ala Ala His Ala Arg Phe Val Ala Ala Ala Lys Val Asn Thr Leu
 245 250 255
 Leu Asp Val Ala Gln Ala Asn Leu Gly Glu Ala Ala Gly Thr Tyr Val
 260 265 270
 Ala Ala Asp Ala Ala Ala Ala Ser Thr Tyr Thr Gly Phe Asp Ile Met
 275 280 285
 Asp Phe Gly Leu Leu Pro Pro Glu Val Asn Ser Ser Arg Met Tyr Ser
 290 295 300
 Gly Pro Gly Pro Glu Ser Met Leu Ala Ala Ala Ala Trp Asp Gly
 305 310 315 320
 Val Ala Ala Glu Leu Thr Ser Ala Ala Val Ser Tyr Gly Ser Val Val
 325 330 335
 Ser Thr Leu Ile Val Glu Pro Trp Met Gly Pro Ala Ala Ala Ala Met
 340 345 350
 Ala Ala Ala Ala Thr Pro Tyr Val Gly Trp Leu Ala Ala Thr Ala Ala
 355 360 365
 Leu Ala Lys Glu Thr Ala Thr Gln Ala Arg Ala Ala Ala Glu Ala Phe
 370 375 380
 Gly Thr Ala Phe Ala Met Thr Val Pro Pro Ser Leu Val Ala Ala Asn
 385 390 395 400
 Arg Ser Arg Leu Met Ser Leu Val Ala Ala Asn Ile Leu Gly Gln Asn
 405 410 415
 Ser Ala Ala Ile Ala Ala Thr Gln Ala Glu Tyr Ala Glu Met Trp Ala
 420 425 430
 Gln Asp Ala Ala Val Met Tyr Ser Tyr Glu Gly Ala Ser Ala Ala Ala
 435 440 445
 Ser Ala Leu Pro Pro Phe Thr Pro Pro Val Gln Gly Thr Gly Pro Ala
 450 455 460
 Gly Pro Ala Ala Ala Ala Ala Thr Gln Ala Ala Gly Ala Gly Ala
 465 470 475 480
 Val Ala Asp Ala Gln Ala Thr Leu Ala Gln Leu Pro Pro Gly Ile Leu
 485 490 495
 Ser Asp Ile Leu Ser Ala Leu Ala Ala Asn Ala Asp Pro Leu Thr Ser
 500 505 510

ES 2 374 620 T3

Gly Leu Leu Gly Ile Ala Ser Thr Leu Asn Pro Gln Val Gly Ser Ala
 515 520 525
 Gln Pro Ile Val Ile Pro Thr Pro Ile Gly Glu Leu Asp Val Ile Ala
 530 535 540
 Leu Tyr Ile Ala Ser Ile Ala Thr Gly Ser Ile Ala Leu Ala Ile Thr
 545 550 555 560
 Asn Thr Ala Arg Pro Trp His Ile Gly Leu Tyr Gly Asn Ala Gly Gly
 565 570 575
 Leu Gly Pro Thr Gln Gly His Pro Leu Ser Ser Ala Thr Asp Glu Pro
 580 585 590
 Glu Pro His Trp Gly Pro Phe Gly Gly Ala Ala Pro Val Ser Ala Gly
 595 600 605
 Val Gly His Ala Ala Leu Val Gly Ala Leu Ser Val Pro His Ser Trp
 610 615 620
 Thr Thr Ala Ala Pro Glu Ile Gln Leu Ala Val Gln Ala Thr Pro Thr
 625 630 635 640
 Phe Ser Ser Ser Ala Gly Ala Asp Pro Thr Ala Leu Asn Gly Met Pro
 645 650 655
 Ala Gly Leu Leu Ser Gly Met Ala Leu Ala Ser Leu Ala Ala Arg Gly
 660 665 670
 Thr Thr Gly Gly Gly Thr Arg Ser Gly Thr Ser Thr Asp Gly Gln
 675 680 685
 Glu Asp Gly Arg Lys Pro Pro Val Val Val Ile Arg Glu Gln Pro Pro
 690 695 700
 Pro Gly Asn Pro Pro Arg
 705 710

REIVINDICACIONES

1. Un ácido nucleico que codifica un polipéptido que comprende un antígeno MTB32A de una especie de *Mycobacterium* del complejo de tuberculosis, en el que al menos un aminoácido de la triada del sitio activo del antígeno MTB32A (SEQ ID N° 2 o 4) se ha sustituido por un aminoácido diferente.
- 5 2. El ácido nucleico de la reivindicación 1, en el que un residuo de serina correspondiente a la posición aminoacídica 183 de la SEQ ID N° 4 o la posición 208 de la SEQ ID N° 2 se ha sustituido por otro aminoácido.
3. El ácido nucleico de la reivindicación 1, en el que se ha sustituido un residuo de alanina por un residuo de serina.
4. El ácido nucleico de la reivindicación 4, en el que el ácido nucleico comprende una secuencia nucleotídica de SEQ ID N° 5.
5. Una composición farmacéutica que comprende el ácido nucleico de la reivindicación 1.
- 10 6. Un ácido nucleico que codifica un polipéptido de fusión que comprende el ácido nucleico de la reivindicación 1.
7. Un polipéptido que comprende un polipéptido MTB32A de una especie de *Mycobacterium* del complejo de tuberculosis, en el que al menos un aminoácido de la triada del sitio activo del antígeno MTB32A (SEQ ID N° 2 o 4) se ha sustituido por un aminoácido diferente.
- 15 8. El polipéptido de la reivindicación 7, en el que un residuo de serina correspondiente a la posición aminoacídica 183 de la SEQ ID N° 4 o la posición 208 de la SEQ ID N° 2 se ha sustituido por otro aminoácido.
9. El polipéptido de la reivindicación 8, en el que se ha sustituido un residuo de alanina por un residuo de serina.
10. Un polipéptido de la reivindicación 9, en el que el polipéptido comprende una secuencia aminoacídica de SEQ ID N° 6.
11. Una composición farmacéutica que comprende el polipéptido de la reivindicación 7.
- 20 12. Un polipéptido de fusión que comprende el polipéptido de la reivindicación 7.
13. Un ácido nucleico que codifica un polipéptido de fusión que comprende un antígeno MTB39 (SEQ ID N° 12 o 14) de una especie de *Mycobacterium* del complejo de tuberculosis y un antígeno que comprende al menos 195 aminoácidos del extremo N de un antígeno MTB32A (SEQ ID N° 2 o 4) de una especie de *Mycobacterium* del complejo de tuberculosis, en el que un aminoácido de la triada del sitio activo del antígeno MTB32A (SEQ ID N° 2 o 4) se ha sustituido por un aminoácido diferente.
- 25 14. El ácido nucleico de la reivindicación 13, en el que se ha sustituido un residuo de serina correspondiente al aminoácido en la posición 183 de la SEQ ID N° 4 o la posición 208 de la SEQ ID N° 2 por otro aminoácido.
15. El ácido nucleico de la reivindicación 14, en el que se ha sustituido un residuo de alanina por un residuo de serina.
- 30 16. Una composición farmacéutica que comprende el ácido nucleico de la reivindicación 13.
17. Un ácido nucleico que codifica un polipéptido de fusión que comprende el ácido nucleico de la reivindicación 13.
18. Un ácido nucleico que codifica un polipéptido de fusión, en el que el ácido nucleico comprende una secuencia nucleotídica de SEQ ID N° 17.
- 35 19. Un polipéptido de fusión que comprende un antígeno MTB39 (SEQ ID N° 12 o 14) de una especie de *Mycobacterium* del complejo de tuberculosis y un antígeno que comprende al menos 195 aminoácidos del extremo N de un antígeno MTB32A (SEQ ID N° 2 o 4) de una especie de *Mycobacterium* del complejo de tuberculosis, en el que un aminoácido de la triada del sitio activo del antígeno MTB32A (SEQ ID N° 2 o 4) se ha sustituido por un aminoácido diferente.
- 40 20. El polipéptido de la reivindicación 19, en el que se ha sustituido un residuo de resina correspondiente a la posición aminoacídica 183 de la SEQ ID N° 4 o la posición aminoacídica 208 de la SEQ ID N° 2 por otro aminoácido.
21. El polipéptido de la reivindicación 19, en el que se ha sustituido un residuo de alanina por un residuo de serina.
22. Una composición farmacéutica que comprende el polipéptido de la reivindicación 19.

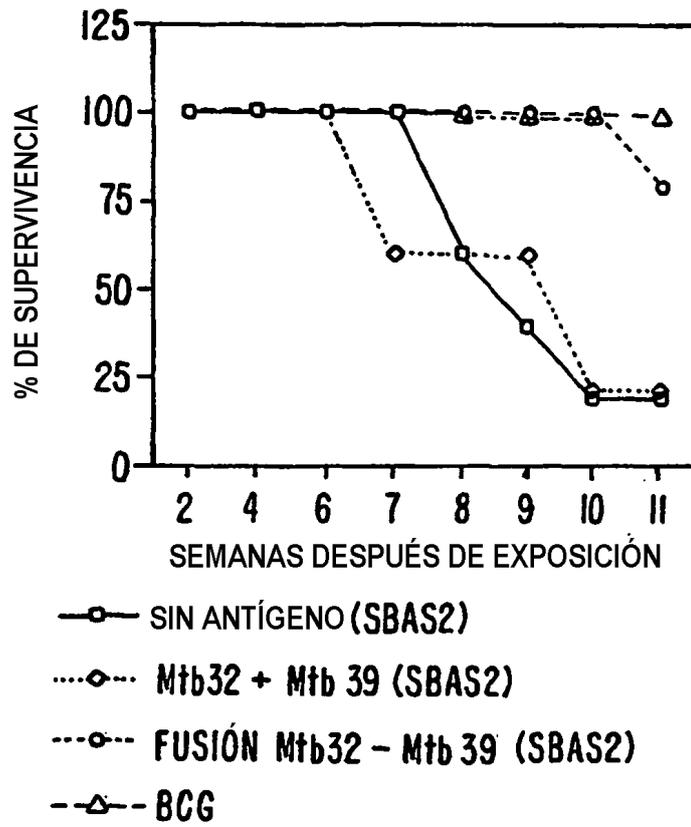


FIG. 1.

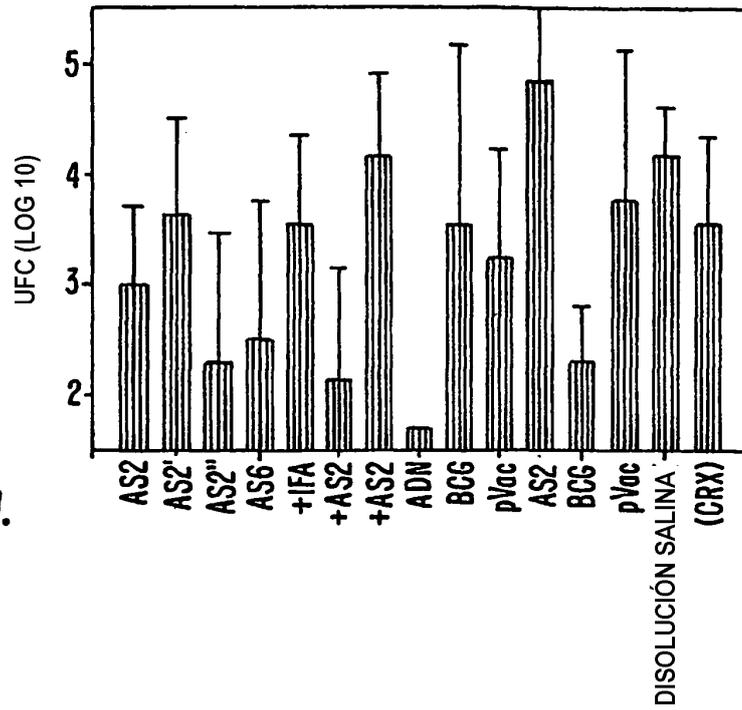


FIG. 2A.

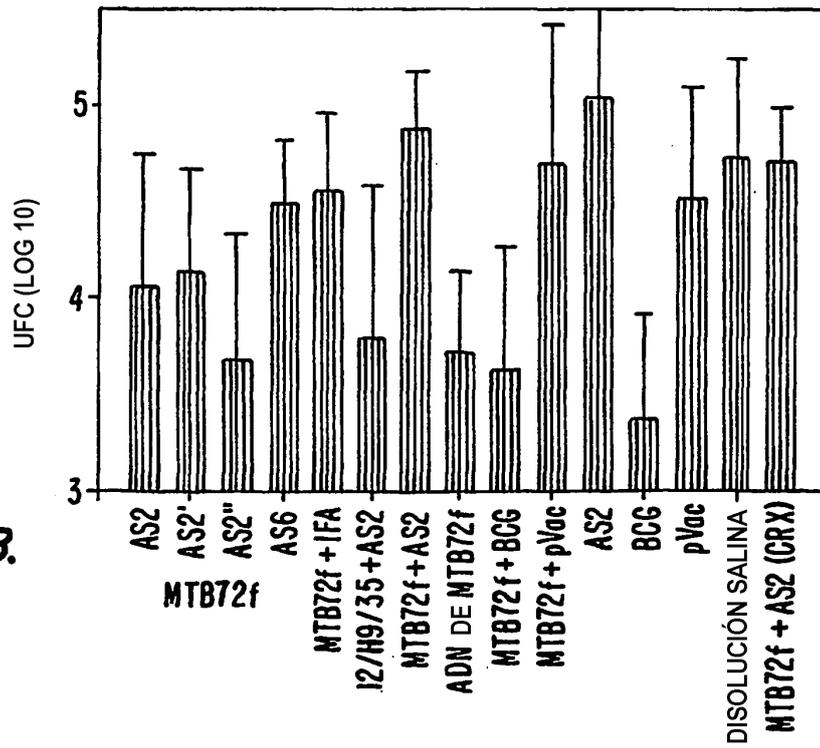


FIG. 2B.

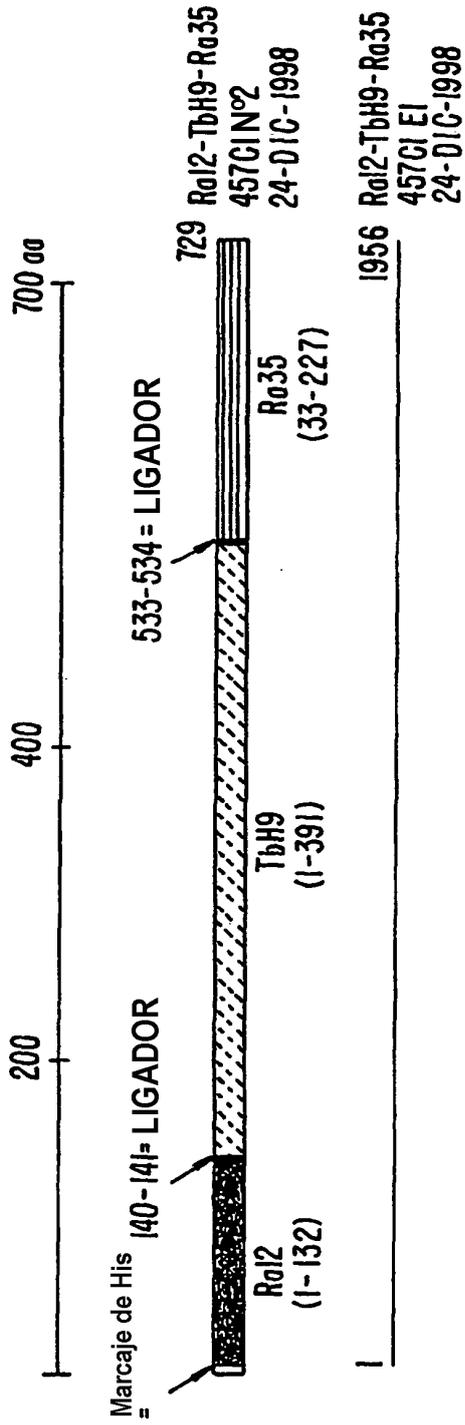


FIG. 3.

ADN N-terminal de Ra35

gccccgccg ccttgtcgca ggaccgggttc gccgacttcc ccgcgctgcc cctcgaccgg tccgcgatgg 70
 tcgccccagt ggggccacag gtggtcaaca tcaacaccaa actgggctac aacaacgccg tgggcgccgg 140
 gaccggcatc gtcatcgatc ccaacgggtg cgtgctgacc aacaaccacg tcatcgccgg cgccaccgac 210
 atcaatgctg tcagcgtcgg ctccggccaa acctaccggc tcgatgtggt cgggtatgac cgcaccagg 280
 atgtcgggtt gctgcagctg cgcggtgccc gtggcctacc atcggcggcg atcgggtggcg gcgtcgcggt 350
 tggtagcccc gtcgtcgcga tgggcaacag cggtagggcag ggcggaacgc cccgtgcccgt ccctggcagg 420
 gtggtcgcgc tcggccaaac cgtgcaggcg tcggattcgc tgaccgggtg cgaagagaca ttgaacgggt 490
 tgatccagtt cgatgcccgg atccagcccc gtgattcggg cgggcccgtc gtcaacggcc taggacaggt 560
 ggtcggtatg aacacggccc cgtccctag 588

Secuencia aminoacídica N-terminal de Ra35

Ala Pro Pro Ala Leu Ser Gln Asp Arg Phe Ala Asp Phe Pro Ala Leu Pro Leu Asp Pro Ser Ala
 5 10 15 20
 Met Val Ala Gln Val Gly Pro Gln Val Val Asn Ile Asn Thr Lys Leu Gly Tyr Asn Asn Ala Val
 25 30 35 40
 Gly Ala Gly Thr Gly Ile Val Ile Asp Pro Asn Gly Val Val Leu Thr Asn Asn His Val Ile Ala
 45 50 55 60 65
 Gly Ala Thr Asp Ile Asn Ala Phe Ser Val Gly Ser Gly Gln Thr Tyr Gly Val Asp Val Val Gly
 70 75 80 85
 Tyr Asp Arg Thr Gln Asp Val Ala Val Leu Gln Leu Arg Gly Ala Gly Gly Leu Pro Ser Ala Ala
 90 95 100 105 110

FIG. 4.

Ile Gly Gly Gly Val Ala Val Gly Glu Pro Val Val Ala Met Gly Asn Ser Gly Gly Gln Gly Gly
 115 120 125 130
 Thr Pro Arg Ala Val Pro Gly Arg Val Ala Leu Gly Gln Thr Val Gln Ala Ser Asp Ser Leu
 135 140 145 150
 Thr Gly Ala Glu Glu Thr Leu Asn Gly Leu Ile Gln Phe Asp Ala Ala Ile Gln Pro Gly Asp Ser
 155 160 165 170 175
 Gly Gly Pro Val Val Asn Gly Leu Gly Gln Val Val Gly Met Asn Thr Ala Ala Ser
 180 185 190 195

FIG. 4. (CONTINUACIÓN)

Ra12

1 MHHHHH[TAASDNFQLSQGGQGFPIPIQAMAIAGQIRSGGGSPVHIGPTAFLG Mtb72f
1 MHHHHH[TAASDNFQLSQGGQGFPIPIQAMAIAGQIRSGGGSPVHIGPTAFLG Mtb72f-mutSA

56 LGVVDNNGGARVQRVVGSAAPASLIGSTGDVITAVDGA PINSATAMADALNGHH Mtb72f
56 LGVVDNNGGARVQRVVGSAAPASLIGSTGDVITAVDGA PINSATAMADALNGHH Mtb72f-mutSA

111 PGDVISVTWQTKSFFTRTFNVTLAEGPPA[E]MVD[FL]Tb9FL
111 PGDVISVTWQTKSFFTRTFNVTLAEGPPA[E]MVD[FL]Tb9FL Mtb72f-mutSA

166 LVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSWIGSAGLMVAAAASPYVAVMSV Mtb72f
166 LVAAAQMWDSVASDLFSAASAFQSVVWGLTVGSWIGSAGLMVAAAASPYVAVMSV Mtb72f-mutSA

221 TAGQAE[LTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMILLIATNLLGQNTPAI Mtb72f
221 TAGQAE[LTAAQVRVAAAAYETAYGLTVPPVPIAENRAELMILLIATNLLGQNTPAI Mtb72f-mutSA

276 AVNEAEYGE[MWAQDAAMFGYAAATATATATLLPFE[EAPE]MTSAGLL[EQAAA]VE Mtb72f
276 AVNEAEYGE[MWAQDAAMFGYAAATATATATLLPFE[EAPE]MTSAGLL[EQAAA]VE Mtb72f-mutSA

331 EASDTAAANQ[LMNNV]PQALQ[LAQ]PTQ[GT]PSSKLG[GLW]KTVSPHRSPISNMVSM Mtb72f
331 EASDTAAANQ[LMNNV]PQALQ[LAQ]PTQ[GT]PSSKLG[GLW]KTVSPHRSPISNMVSM Mtb72f-mutSA

386 ANNHMSMTNSGV[SMTNT]LSSMLKGFAPAAAQAVQTA[AAQ]NGVRAMSS[LGSS]LGSS Mtb72f
386 ANNHMSMTNSGV[SMTNT]LSSMLKGFAPAAAQAVQTA[AAQ]NGVRAMSS[LGSS]LGSS Mtb72f-mutSA

441 GLGGVAANLGRAASV[GLSV]PQAWAAANQAVT[PAARAL]PLTSLTSA[ER]GPGQM Mtb72f
441 GLGGVAANLGRAASV[GLSV]PQAWAAANQAVT[PAARAL]PLTSLTSA[ER]GPGQM Mtb72f-mutSA

FIG. 5.

Ra35

496 LGGLPVGQMGARAGGGLSGVLRVPPRPYVMPHSPAAGDIAAPPALSQDRFADFPAL Mtb72f
496 LGGLPVGQMGARAGGGLSGVLRVPPRPYVMPHSPAAGDIAAPPALSQDRFADFPAL Mtb72f-mutSA

551 PLDPSAMVAQVGPQVNNINTKLGYNNNAVAGTGIVIDPNGVVLTNNNVIAGATDI Mtb72f
551 PLDPSAMVAQVGPQVNNINTKLGYNNNAVAGTGIVIDPNGVVLTNHHVIAGATDI Mtb72f-mutSA

606 NAFSVGSGQTYGVDVVGVDYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVAMGN Mtb72f
606 NAFSVGSGQTYGVDVVGVDYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVAMGN Mtb72f-mutSA

661 SGGGGTTPRAVPGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDSGGPVV Mtb72f
661 SGGGGTTPRAVPGRVVALGQTVQASDSLGTGAETLNGLIQFDAAIQPGDAAGGPVV Mtb72f-mutSA

716 NGLGQVVGMMNTAAS Mtb72f
716 NGLGQVVGMMNTAAS Mtb72f-mutSA

FIG. 5. (CONTINUACIÓN)

extremo N de Ra35

1 MHHHHH[A]PPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNA TBRa35_mat

1 MHHHHH[A]PPALSQDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNA TBRa35_mutSA

51 VGAGTGIVIDPENGVVLTNHHVIAGATDINAFSVGSGQTYGVDDVVGYDRTO TBRa35_mat

51 VGAGTGIVIDPENGVVLTNHHVIAGATDINAFSVGSGQTYGVDDVVGYDRTO TBRa35_mutSA

101 DVAVLQLRGAGGLPSAAIGGGVAVGEPVAMNGSQQGTPRAVPGRVVA TBRa35_mat

101 DVAVLQLRGAGGLPSAAIGGGVAVGEPVAMNGSQQGTPRAVPGRVVA TBRa35_mutSA

151 LGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLQVVGMMN[A] TBRa35_mat

151 LGQTVQASDSLTAETLNGLIQFDAAIQPGD[A]GGPVVNGLQVVGMMN[A] TBRa35_mutSA

201 AS[DNFQLSQGGGF]AIPIGQAMAIAQIRSGGSPVHIGPTAFLGLGVV TBRa35_mat

201 AS[DNFQLSQGGGF]AIPIGQAMAIAQIRSGGSPVHIGPTAFLGLGVV TBRa35_mutSA

251 DNNGGARVQRVVGSAAPAASLGISTGDVITAVDVGAPINSATAMADALNGH TBRa35_mat

251 DNNGGARVQRVVGSAAPAASLGISTGDVITAVDVGAPINSATAMADALNGH TBRa35_mutSA

301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA] final de Ra12 TBRa35_mat

301 HPGDVISVTWQTKSGGTRTGNVTLAEGPPA] TBRa35_mutSA

Ra12 Cterm

FIG. 6.

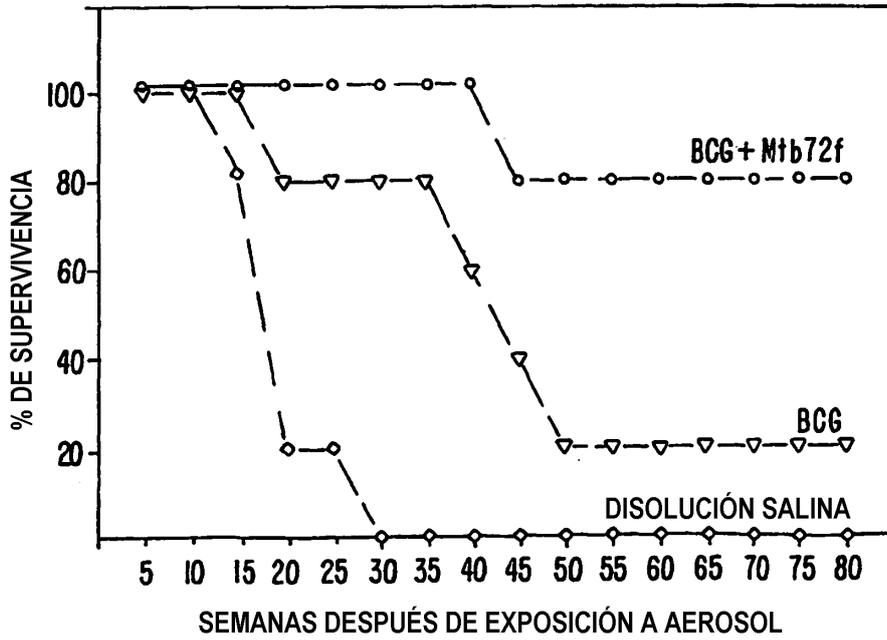
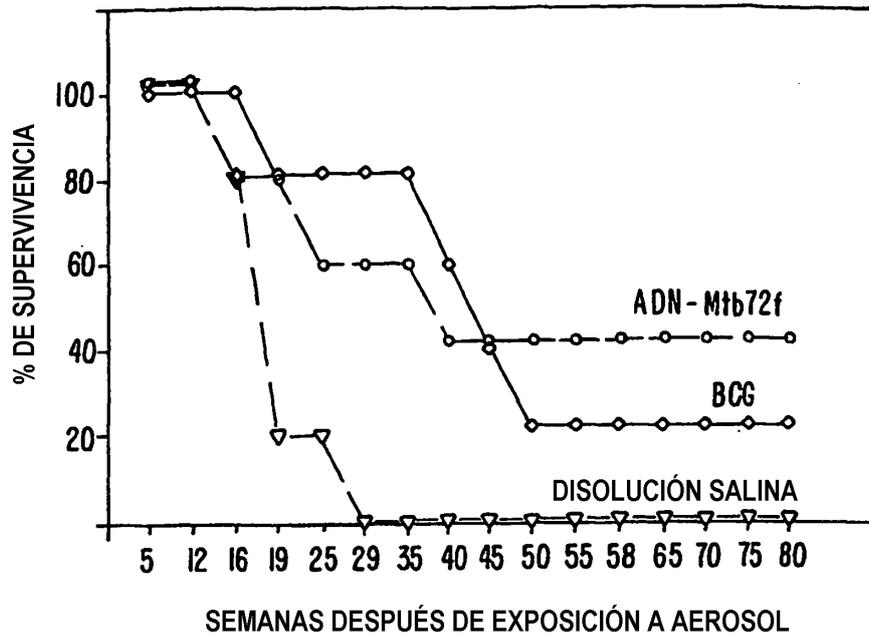


FIG. 7