

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 627 402**

51 Int. Cl.:

C12N 7/02 (2006.01)

C07K 14/11 (2006.01)

A61K 39/145 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **15.06.2010 PCT/EP2010/058360**

87 Fecha y número de publicación internacional: **23.12.2010 WO10146043**

96 Fecha de presentación y número de la solicitud europea: **15.06.2010 E 10724856 (9)**

97 Fecha y número de publicación de la concesión europea: **01.03.2017 EP 2443232**

54 Título: **Sistema para la expresión heteróloga de una proteína viral en una célula huésped de ciliado**

30 Prioridad:

17.06.2009 GB 0910357

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

28.07.2017

73 Titular/es:

**CILIAN AG (100.0%)
Johann-Krane-Weg 42
48149 Münster, DE**

72 Inventor/es:

**HARTMANN, MARCUS;
SACHSE, CHRISTINE;
APELT, JENNY y
BOCKAU, ULRIKE**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 627 402 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

Sistema para la expresión heteróloga de una proteína viral en una célula huésped de ciliado

5 La presente invención se refiere a un sistema para la expresión heteróloga de una proteína de la superficie viral en una célula huésped de ciliado.

10 Como muchos virus desempeñan una función considerable como patógenos humanos o animales, se ha prestado atención al uso de proteínas virales como vacunas contra patógenos virales. Los virus (especialmente virus de ARN) son altamente variables y muchas infecciones virales son debidas a virus con múltiples serotipos (por ejemplo, virus de la gripe, virus de FMD). Como consecuencia, muchas de las vacunas virales existentes son frecuentemente incapaces de arreglárselas con las cepas prevalentes en el campo y tienen que generarse nuevas a partir de las cepas de campo con nuevos brotes.

15 A pesar del hecho de que, gracias a las modernas técnicas de biotecnología, están disponibles métodos que permiten la expresión heteróloga de casi cualquier proteína concebible, la expresión de proteínas virales no es una cosa tan común hoy en día, ya que las proteínas virales son, en la mayoría de los casos, proteínas estructurales, es decir, ni enzimas ni hormonas de proteína ni anticuerpos, y así no son tan interesantes para fines farmacéuticos o industriales.

20 Sin embargo, como muchos virus desempeñan una función considerable como patógenos humanos o animales, se ha prestado atención al uso de proteínas virales como vacunas contra patógenos virales.

25 Hoy en día, el procedimiento convencional de producción de vacunas contra patógenos virales es cultivar el virus respectivo en un sistema dado, recoger e inactivar las partículas de virus así obtenidas, y fabricar una composición farmacéutica que comprende dichas partículas inactivadas como vacuna.

30 Dicho principio se pone en práctica, por ejemplo, en la producción de vacuna contra la gripe, que está siendo producida en huevos de pollo fecundados. Once días después de la fecundación, las cepas del virus de la gripe se inyectan en la albúmina de los huevos individuales y entonces se infectan los pulmones del embrión en desarrollo. Después de varios días de incubación, los virus se recogen y se purifican, se inactivan químicamente y se usan para producir una vacuna. En promedio, se necesitan aproximadamente uno y dos huevos para producir una dosis de vacuna. El proceso de producción entero dura al menos seis meses.

35 Esta forma de producción de vacuna está muy establecida y es rentable, pero tiene desventajas como la obtención de muchos millones de huevos, el mayor alcance temporal, la tediosa manipulación de los huevos y la flexibilidad limitada en caso de demanda rápidamente creciente o la repentina aparición de nuevas cepas de virus. Además, no pueden usarse huevos de pollo fecundados para producir vacunas contra varios virus patógenos, entre ellos el virus de la gripe aviar A (H5N1).

40 De cualquier forma, los huevos de pollo fecundados pueden usarse al menos principalmente para la producción de todos los virus que son compatibles con los embriones de pollo como huésped, o para la producción de vacunas que comprenden elementos de los anteriores.

45 Una forma alternativa de producir virus, o vacunas que comprenden elementos de los anteriores, se basa en cultivos de célula o de tejido. En un enfoque, el virus se inyecta en células de riñón de mamífero. Después de la propagación del virus, las células se lisan y entonces se recogen, purifican e inactivan partículas de virus. Sin embargo, estos sistemas requieren el uso de complejos medios de crecimiento, y su manipulación es laboriosa y requiere tiempo. Además, dicho enfoque de infección es difícil de llevar a cabo y no es completamente reproducible.

50 En otro enfoque desvelado en el documento WO03048348 se desvela un método de producción de vacunas de virus inactivados en la línea celular humana PER.C6®. Debido a la presencia de receptores que contienen Sia2-6Gal y Sia2-3Gal sobre su superficie, esta línea celular es altamente infectable con diferentes virus como tipos de virus de la gripe, virus paragripal, virus adeno-asociado o de virus del polio. Esto podría producir problemas de seguridad con respecto al proceso de producción.

55 Por una parte, los sistemas basados en cultivo celular de mamífero pueden, a diferencia de los sistemas basados en huevos anteriores, ser rápidamente expandidos y aumentarse de escala en tiempos de emergencia. Por otra parte, los costes directos para la rapidez operacional de tales instalaciones de producción (con sus enormes biorreactores) son muy superiores a los costes para los sistemas basados en huevos, y el rendimiento puede ser ligeramente más lento. Otra vez, estos sistemas requieren el uso de complejos medios de crecimiento, y su manipulación es laboriosa y requiere tiempo.

65 Por tanto, tiene alta prioridad el desarrollo de métodos alternativos para la producción de proteínas virales que puedan usarse como vacunas.

Un enfoque es la producción de aquellas proteínas usando técnicas de ADN recombinante. Una ventaja obvia es una seguridad enormemente mejorada de la vacuna, debido a la oportunidad de purificar proteína viral expresada recombinante a diferencia de la producción de vacuna usando huevos. Además, la flexibilidad para adaptar a diferentes subtipos de virus estacionales es altamente elevada.

Sin embargo, como las proteínas virales son, en la mayoría de los casos, proteínas estructurales, la producción para fines farmacéuticos e industriales es compleja y laboriosa. Al principio se intentó expresar polipéptidos correspondientes a proteínas virales como hemaglutinina en *Escherichia coli*. Es, sin embargo, digno de mención que la expresión heteróloga de proteínas que van a usarse como vacunas en procariotas no tiene sentido, ya que las últimas no tienen un aparato de modificación postraducciona. Por tanto, las proteínas expresadas en procariotas como *Escherichia coli* carecen de modificaciones postraduccionales, como patrón de glucosilación, que parece afectar la estabilidad y contribuir a un gran grado al potencial inmunogénico de un antígeno (Nayak et al. 1984).

La expresión de proteínas virales en el organismo eucariota *Saccharomyces cerevisiae* (documento US4752473 para hemaglutinina) causó problemas por la hiper-glucosilación de la proteína recombinante con la amplia adición de mananos de cadena externa de alto peso molecular (Jabbar y Nayak 1987, Jabbar et al. 1985) y no ha demostrado ser satisfactoria.

El documento US5858368 desvela un método de expresión de hemaglutinina del virus de la gripe A (HA) en líneas de células de insecto infectadas con baculovirus recombinantes. Las proteínas se purifican cromatográficamente después de extraerse de las membranas periféricas de las células infectadas con un detergente no iónico no desnaturizante. Sin embargo, el proceso de purificación requiere bastante tiempo. Además, es cuestionable si, debido a su patrón de glucosilación, las proteínas producidas en un sistema de expresión de baculovirus/célula de insecto son adecuadas o no para uso terapéutico en mamíferos (Kulakosky et al. 1998). Hay además indicios de que ciertos restos en N-glicanos de proteínas expresadas en células de insecto representan posibles epítopes alergénicos (Tomiya et al. 2004).

Definiciones

El término "expresión heteróloga", como se usa en el presente documento, debe referirse a la expresión de proteínas de un gen, un ácido nucleico o un ADNc, que es extraña para el organismo en el que se produce la expresión.

El término "ciliado", como se usa en el presente documento, debe referirse al filo científico Ciliophora, que son eucariotas unicelulares ("protozoos" o "protistas") caracterizados, entre otros, por su tamaño relativamente grande (algunas especies tienen hasta 2 mm de longitud), su superficie celular ciliada y por dos tipos diferentes de núcleos, es decir, un micronúcleo diploide pequeño y un macronúcleo poliploide grande (activo para la transcripción). El último se genera a partir del micronúcleo por amplificación del genoma y edición pesada.

El término "ADNc", como se usa en el presente documento, debe referirse a una molécula de ADN que codifica una proteína que va a expresarse, y carece de cualquier parte no codificante, como intrones. En muchos casos, un ADNc ha sido directamente sintetizado a partir de un molde de ARN usando transcriptasa inversa, y un cebador de oligo dT. Sin embargo, el término también debe comprender genes sintéticos y que codifican ADN obtenidos de otro modo.

El término "promotor", como se usa en el presente documento, debe referirse a una región reguladora de ADN generalmente localizada en la dirección 5' (hacia la región 5' de la hebra codificante) de un gen o un ADNc, que contiene elementos genéticos esenciales que permiten o incluso potencian la transcripción del gen, o el ADNc.

El término "fragmento", como se usa en el presente documento, debe referirse a una parte de una proteína que carece de algunas partes, o dominios, de la proteína nativa, o natural, mientras que retiene alguna actividad en términos de actividad enzimática, inmunogenicidad, unión a diana o similares.

El término "secuencia señal", como se usa en el presente documento, debe referirse a una secuencia de ácidos nucleicos que codifica un oligopéptido ("péptido señal" o "péptido de tránsito") que dirige el transporte de una proteína a ciertos orgánulos tales como el núcleo, matriz mitocondrial, retículo endoplásmico, cloroplasto, apoplasto y peroxisoma. Casi todas las proteínas que son transportadas al retículo endoplásmico tienen una secuencia que consiste en 5-10 aminoácidos hidrófobos en el extremo N. Aquellos péptidos señal se escinden de la proteína por una peptidasa señal después de la inserción co-traducciona de la proteína en la luz del ER. La mayoría de las proteínas son entonces transportadas mediante el aparato de Golgi en la dirección 3' en la vía secretora.

El término "operativamente unidas", como se usa en el presente documento, significa que una secuencia de nucleótidos, que puede codificar un producto génico, está unida a un promotor de forma que el promotor regula la expresión del producto génico bajo condiciones apropiadas.

El término "dominio hidrófobo" está siendo usado sinónimamente con los términos "dominio lipófilo" y/o "dominio no polar". Como se usa en el presente documento, dichos términos deben referirse a un dominio de proteína que tiene,

al menos en su periferia, propiedades hidrófobas (también llamadas "hidropatía") principalmente gobernadas por la abundancia de restos de aminoácidos hidrófobos/no polares, y que tiene así algo de afinidad por fracciones de lípido como, por ejemplo, bicapas de fosfolípidos como se encuentran en membranas celulares.

- 5 Dicho dominio hidrófobo puede, por ejemplo, cualificarse como un llamado "dominio transmembranario", es decir, una estructura de proteína tridimensional que es termodinámicamente estable en una membrana. Tal dominio transmembranario puede comprender una única hélice alfa, un complejo estable de varias hélices alfa transmembranarias, un barril beta transmembranario, una hélice beta, un anillo externo de aminoácidos hidrófobos que rodean los restos de aminoácidos posiblemente hidrófilos, y similares.

10 Los inventores se han dado cuenta de que los dominios hidrófobos pueden dificultar la secreción de una proteína en el entorno. Ejemplos de aminoácidos hidrófobos se dan en la siguiente tabla:

Tabla 1

Aminoácido	Índice de hidropatía (Kyte y Doolittle 1982)
Isoleucina	4,5
Valina	4,2
Leucina	3,8
Fenilalanina	2,8
Cisteína	2,5
Metionina	1,9
Alanina	1,8
Glicina	-0,4
Treonina	-0,7
Serina	-0,8

15 El término "molécula de ácido nucleico" está previsto para indicar cualquier molécula de ácido nucleico mono- o bicatenaria que comprenda ADN (ADNc y/o ADN genómico), ARN (preferentemente ARNm), PNA, LNA y/o morfolino.

20 El término "condiciones rigurosas" se refiere a condiciones bajo las que una sonda se hibridará preferentemente con su subsecuencia diana y mucho menos con otras secuencias. Las condiciones rigurosas son dependientes de secuencia y serán diferentes en diferentes circunstancias. Secuencias más largas se hibridan específicamente a temperaturas más altas. Generalmente, las condiciones rigurosas están seleccionadas para ser aproximadamente 5 °C más bajas que el punto de fusión térmico (T_m) para la secuencia específica a una fuerza iónica y pH definidos.

25 La T_m es la temperatura (bajo fuerza iónica, pH y concentración de ácido nucleico definidos) a la que el 50 % de las sondas complementarias a la secuencia diana se hibridan con la secuencia diana en equilibrio. (Como las secuencias diana están generalmente presentes en exceso, a T_m, el 50 % de las sondas están ocupadas en equilibrio). Normalmente, condiciones rigurosas serán aquellas en las que la concentración de sales es ión Na inferior a aproximadamente 1,0 M, normalmente ión Na aproximadamente 0,01 a 1,0 M (u otras sales) a pH 7,0 a 8,3

30 y la temperatura es al menos aproximadamente 30 °C para sondas cortas (por ejemplo, 10 a 50 nucleótidos) y al menos aproximadamente 60 °C para sondas más largas. También pueden lograrse condiciones rigurosas con la adición de agentes desestabilizantes, tales como formamida y similares.

35 El término "código de PDB" se refiere a un código de cuatro letras único que sirve de identificador primario por el que pueden recuperarse entradas del Protein Data Bank, que es el depósito mundial de datos estructurales macromoleculares biológicos.

El término "fragmento de la molécula de ácido nucleico" está previsto para indicar un ácido nucleico que comprende un subconjunto de una molécula de ácido nucleico según una de las secuencias reivindicadas.

40 Lo mismo es aplicable al término "fracción de la molécula de ácido nucleico".

El término "variante de la molécula de ácido nucleico" se refiere en el presente documento a una molécula de ácido nucleico que es sustancialmente similar en estructura y actividad biológica a una molécula de ácido nucleico según una de las secuencias reivindicadas.

45 El término "homólogo de la molécula de ácido nucleico" se refiere a una molécula de ácido nucleico cuya secuencia tiene uno o más nucleótidos añadidos, delecionados, sustituidos o químicamente modificados de otro modo en comparación con una molécula de ácido nucleico según una de las secuencias reivindicadas, siempre que el homólogo retenga sustancialmente las mismas propiedades que las últimas.

50

El término "de codón optimizado", como se usa en el presente documento, debe referirse a un proceso en el que el ADNc que codifica la proteína heteróloga que va a expresarse se adapta a un uso de codones específico de huésped que se deriva del esquema del código genético universal. Los ciliados tienen un genoma rico en AT, con ADN de *Tetrahymena* que consiste en aproximadamente el 75 % de AT (véase la Fig. 7). El uso de codones se diferencia de aquél en otros organismos particularmente en cómo de frecuentemente se usa un codón para codificar un aminoácido dado ("preferencia codónica"). Si el ADNc no optimizado que codifica una proteína heteróloga usa codones que se usan raramente en ciliados, esto podría afectar fuertemente la eficiencia de expresión de proteínas. Esto significa, a su vez, que la expresión heteróloga de proteínas puede mejorar espectacularmente cuando la frecuencia de codones del gen en estudio coincide con la del sistema de expresión del ciliado. Además, muchos ciliados, entre ellos *Tetrahymena*, utilizan códigos de nucleótidos no canónicos con tripletes UAA y UAG que codifican glutamina, mientras que en la mayoría de los otros organismos estos codones se usan como codón de terminación que terminan la traducción. Esto puede conducir al hecho de que genes extraños (no ciliados) que llevan tripletes UAA y UAG como codón de terminación no se expresan correctamente. Para este fin, antes de transformar la célula huésped de ciliado, el ADNc que codifica una proteína heteróloga debe ser de codón optimizado de tal forma que los tripletes UAA y UAG estén modificados en UAA. La optimización del código puede, por ejemplo, llevarse a cabo por mutagénesis dirigida al sitio, o por la síntesis de ADNc de *de novo*.

El término "derivado", como se usa en el presente documento, se refiere a una molécula de ácido nucleico relacionada que tiene características similares a una secuencia de ácidos nucleicos diana como molécula de ácido nucleico según una de las secuencias reivindicadas.

El término "identidad de secuencia de al menos el % de X", como se usa en el presente documento, se refiere a una identidad de secuencia como se determina después de un alineamiento de secuencias llevado a cabo con la familia de algoritmos BLAST (particularmente megablast, discontinuous megablast, blastn, blastp, PSI-BLAST, PHI-BLAST, blastx, tblastn y tblastx), como están accesibles en el dominio de internet respectivo proporcionado por NCBI.

El término "célula huésped", como se usa en el presente documento, tiene dos significados diferentes que pueden entenderse según el contexto respectivo. En el contexto de la expresión heteróloga de proteínas, el término "célula huésped" se refiere a una célula transgénica que se usa como huésped de expresión. Dicha célula, o su progenitor, ha sido así transfectada con un vector adecuado que comprende el ADNc de la proteína que va a expresarse. En el contexto del ciclo de vida viral, el término "célula huésped" se refiere a una célula que se infecta por un virus que usa la célula para la replicación.

El término "vector", como se usa en el presente documento, se refiere a un vehículo molecular usado para transferir material genético extraño en otra célula. El propio vector es generalmente una secuencia de ADN que consiste en una inserción (secuencia de interés) y una secuencia más grande que sirve de "esqueleto" del vector. El fin de un vector para transferir información genética a otra célula normalmente es aislar, multiplicar o expresar la inserción en la célula diana.

El término "plásmido", como se usa en el presente documento, se refiere a vectores plasmídicos, es decir, secuencias de ADN circular que son capaces de replicación autónoma dentro de un huésped adecuado debido a un origen de replicación ("ORI"). Además, un plásmido puede comprender un marcador de selección para indicar el éxito de la transformación u otros procedimientos que pretenden introducir ADN extraño en una célula y un sitio de clonación múltiple que incluye múltiples sitios consenso de enzima de restricción para permitir la inserción de una inserción. Los vectores plasmídicos llamados vectores de clonación o donantes se usan para facilitar la clonación y para amplificar una secuencia de interés. Vectores plasmídicos llamados vectores de expresión o aceptores son específicamente para la expresión de un gen de interés en una célula diana definida. Aquellos vectores plasmídicos generalmente muestran un casete de expresión, que consiste en un promotor, el transgén y una secuencia terminadora. Los plásmidos de expresión pueden ser plásmidos lanzadera que contienen elementos que permiten la propagación y selección en diferentes células huésped.

El término "proteína viral", como se usa en el presente documento, se refiere a proteínas generadas por un virus. Dichas proteínas pueden tanto formar la envoltura viral y/o la cápside, o pueden ser proteínas no estructurales, reguladoras y accesorias.

El término "proteína de la superficie", como se usa en el presente documento, debe referirse a proteínas de virus que forman parte de la capa externa del virus. Tales proteínas son, por ejemplo, (i) proteínas ancladas en los recubrimientos de bicapa lipídica de virus envueltos, (ii) proteínas de la cápside tanto en virus envueltos como no envueltos y/o (iii) proteínas de la espícula como se encuentran en muchos virus, como bacteriófagos y virus que infectan seres humanos, que incluyen mamíferos que incluyen aves. Dichos recubrimientos de bicapa lipídica de virus envueltos son adquiridos en la mayoría de los casos por el virus durante su ciclo de replicación, concretamente tras la exocitosis o lisis celular, de membranas intracelulares en la célula huésped (por ejemplo, la membrana nuclear interna, o la membrana de Golgi), o de la membrana externa de células huésped.

El término "proteína de fusión viral", como se usa en el presente documento, se refiere a glucoproteínas de virus envueltos, que facilitan la infección de células huésped (inserción de material genético de virus en la célula

huésped). Los virus envueltos, con una bicapa lipídica como parte integral de su estructura, entran en las células que infectan por fusión de la membrana viral y de la célula huésped. Dichas proteínas de fusión virales combinan dos características principales: contienen una función de unión a receptor, que une el virus a la célula huésped, e incluyen una función de fusión que puede activarse para mediar en la fusión de la membrana viral y de célula huésped. Algunas proteínas de fusión virales importantes son las proteínas de fusión virales de clase I (véase más adelante).

El término "péptido de fusión", como se usa en el presente documento, se refiere a una región hidrófoba conservada distinta dentro de una proteína de fusión viral que se inserta en la membrana de la célula huésped durante el proceso de fusión. Los péptidos de fusión tienden a ser regiones apolares, relativamente ricas en restos de glicina y alanina, y que contienen varios restos hidrófobos voluminosos. Para las proteínas de fusión virales de clase I el péptido de fusión está localizado en el extremo N de la región transmembranaria de la proteína.

Objetivo de la presente invención

El objetivo de la presente invención es proporcionar un método y/o una realización que venza las desventajas como se exponen anteriormente. Tal objetivo se logra mediante las realizaciones expuestas en las reivindicaciones independientes.

Descripción

Según la presente invención, se proporciona un sistema para la expresión heteróloga de una proteína de la superficie viral, comprendiendo dicho sistema una célula huésped ciliada, al menos un ácido nucleico que codifica una proteína de la superficie viral y un promotor operativamente unido a dicho ácido nucleico.

En una realización preferida, dicho ácido nucleico es un ADNc.

Básicamente, la expresión heteróloga de proteínas virales no es una cosa común hoy en día, ya que las proteínas virales son, en la mayoría de los casos, proteínas estructurales, es decir, ni enzimas ni hormonas de proteína ni anticuerpos, y así no son tan interesantes para fines farmacéuticos o industriales.

Además, la expresión de proteínas virales en ciliados todavía no ha sido sugerida en las referencias anteriores. Los inventores se han dado cuenta de que para los ciliados, a diferencia de las bacterias o metazoos, hasta la fecha no se conocen virus específicos. Esto podría ser debido al dimorfismo nuclear que es común a los ciliados. Otro motivo para esto podría ser el uso de codones poco usual y el genoma rico en AT en ciliados. Los inventores asumen así que los virus patógenos de organismos superiores no pueden amplificarse en la mayoría de los ciliados, por ejemplo, *Tetrahymena*.

Los argumentos anteriores dan razones para el hecho de que, hasta la fecha, no se conozcan intentos para producir proteínas virales en ciliados.

El hecho de que, como se sabe hasta la fecha, los ciliados no son susceptibles a los virus, surge como una ventaja sorprendente. Esto significa que en los procesos de producción basados en ciliados, no se produce la amplificación o crecimiento de virus inesperados. Esto significa, además, que en el caso de una proteína que se produce para uso terapéutico, pueden saltarse los costosos procedimientos de agotamiento de virus según que son necesarios en procesos industriales con cultivos de células humanas y animales.

Además, se desvelan promotores adecuados para la expresión de proteínas virales en ciliados, por ejemplo, en el documento WO2007006812A1 que también está registrado por el solicitante de la presente invención, cuyo contenido debe incorporarse con la presente por referencia. Las Seq ID No 12 y 13 del listado de secuencias adjunto dan las secuencias de dos promotores específicos de ciliado particularmente preferidos en el contexto de la presente invención, concretamente un promotor inducible por calor y un promotor de la metalotioneína.

Merece la pena mencionar que Gaertig et al. (1999) (véase también el documento EP1151118) han demostrado el uso del protozoo *Tetrahymena thermophila* como sistema de expresión de proteína para una proteína de antígeno de superficie de otro protozoo, es decir, *Ichthyophthirius multifiliis*, que es un parásito de pescado. Informan que la proteína así producida está siendo presentada sobre la superficie de células de *Tetrahymena* transformadas, donde puede recogerse y ser usada para la fabricación de una vacuna contra *Ichthyophthirius multifiliis*. Otra idea es usar directamente las células de *Tetrahymena* transformadas que presentan el antígeno sobre su superficie como vacuna viva.

Sin embargo, los autores sugieren que *Tetrahymena* solo puede usarse como huésped para la clonación y expresión de genes de organismos con genomas ricos en AT, es decir, protozoos como *Plasmodium*, *Mycoplasma* o *Leishmania*, que también son patógenos humanos. Debido a la inestabilidad inherente de los tramos de AT en sistemas convencionales tales como *Escherichia coli*, ha demostrado ser difícil clonar genes de estos organismos. El ADN de *Tetrahymena* consiste, por ejemplo, en aproximadamente el 75 % de AT.

Además, se informa que la expresión de genes extraños en *Tetrahymena* está impedida por problemas del uso de codones. *T. thermophila* utiliza los tripletes UAA y UAG para glutamina, mientras que en la mayoría de los otros organismos estos codones se usan como codones de terminación que terminan la traducción. Esto puede conducir al hecho de que genes extraños que llevan un triplete UAA y UAG como codón de terminación no se expresen correctamente.

En una realización preferida según la invención, el sistema comprende además una secuencia señal operativamente unida a dicho ácido nucleico, secuencia señal que representa la secreción de una proteína viral codificada por dicho ácido nucleico en el medio extracelular.

Las proteínas virales no son, cuando se expresan en la naturaleza (es decir, por una célula huésped bacteriana infectada por un bacteriófago, o por una célula huésped de metazoo infectada por un virus), objeto de secreción, ya que un virus usa la maquinaria y el metabolismo de su huésped para producir simplemente sus propias proteínas, que entonces se ensamblan con virus de segunda generación antes de la lisis de la célula huésped. Esto significa a su vez que las proteínas de virus nunca han estado bajo la presión evolutiva para optimizar la capacidad de ser secretadas.

Éste es un motivo por el hecho de que la producción recombinante de proteínas de virus solubles y la posterior secreción de estas proteínas virales solubles es un problema complicado. Secuencias señal adecuadas se desvelan, por ejemplo, en el documento WO03078566A1, que también está registrado por el solicitante de la presente invención.

Las Seq ID Nos 8 y 10 del listado de secuencias adjunto dan las secuencias de ácidos nucleicos de dos péptidos señal particularmente preferidos en el contexto de la presente invención, concretamente el péptido señal endógeno del gen HA, y el péptido señal de fosfolipasa A1 de ciliado.

En otra realización preferida más de la invención, dicha proteína viral es una proteína de la superficie viral, como se ha definido anteriormente.

Tras la infección de un huésped, dichas proteínas virales de la superficie se ponen en contacto con el huésped y, en algunos casos, provocan respuestas inmunitarias, si dicho huésped permite tal respuesta (por ejemplo, en mamíferos, que tienen un sistema inmunitario bien desarrollado). Estas proteínas tienen así el potencial de actuar, en forma aislada o a propósito de un adyuvante (virus desactivado, o un fragmento del mismo), de vacuna.

En otra realización preferida más de la presente invención, se proporciona que dicho ciliado transgénico es un miembro de la familia *Tetrahymenidae*.

En una realización particularmente preferida, dicho ciliado transgénico es *Tetrahymena sp.* (particularmente *Tetrahymena thermophila*). *Tetrahymena* es un microorganismo eucariota unicelular no patógeno que se ha establecido en algunos laboratorios como huésped de expresión. Presenta varias ventajas que le hacen adecuado para la expresión heteróloga de proteínas. *Tetrahymena* es un organismo modelo ampliamente examinado y, en más de 50 años de investigación básica, no se observaron virus o endoparásitos. Exámenes con líneas celulares indicadoras no revelaron agentes infecciosos endógenos como virus o micoplasma, que pueden infectar animales superiores.

Primero de todo, las consideraciones anteriores como se relacionan con el uso de codones en ciliados también se aplican para *Tetrahymena*. Además, están disponibles plásmidos de altos números de copias para *Tetrahymena*, que contiene un origen de replicación (ori) de un ADNr minicromosómico. Este ADNr minicromosómico está presente en hasta 9.000 copias por célula. Por encima, puede tener lugar la integración estable en el ADN macronuclear, en el que todos los genes están presentes en número de copias de 45 veces. La alta dosis de gen es la condición previa ideal para una eficiente biosíntesis de proteínas y así para una alta productividad. A diferencia de las levaduras y bacterias, ciliados del género *Tetrahymena* secretan biológicamente proteínas muy eficientemente al sobrenadante de fermentación.

Tetrahymena es capaz de unir modificaciones postraduccionales a proteínas, como puentes de disulfuro, anclaje de GPI, fosforilación, acetilación y glucosilación, que son más similares a aquellas en células de mamífero que aquellas detectadas en levadura u otros sistemas de expresión eucariotas.

A diferencia de las células de mamífero, *Tetrahymena* combina la facilidad de crecimiento con tiempos de generación cortos (1,5 - 3 h), y la reducción de costes, ya que pueden usarse medios químicamente definidos y no existe necesidad de péptidos o componentes de suero, como factores de crecimiento.

Se establecen fermentación discontinua, de lotes alimentados y continua de *Tetrahymena* con densidades celulares hasta 2×10^7 células/ml y pesos secos de hasta 80 g/l, y podrían demostrarse aumentos de la producción (aumento de escala) hasta 1000 l sin ningún problema. En estudios de factibilidad con proteínas indicadoras ya podrían lograrse rendimientos de espacio-tiempo de 50 - 90 pg/célula al día. Los primeros experimentos con expresión

homóloga produjeron un rendimiento de más de 200 mg/l al día para proteínas secretadas. Puede fermentarse *Tetrahymena* en instalaciones de producción convencionales para sistemas de expresión microbiológicos (bacterias o levaduras). Esto significa que no son necesarias costosas modificaciones en las plantas de producción existentes o una nueva construcción de las instalaciones de producción.

A pesar de dichas ventajas, los sistemas de expresión en ciliados, particularmente *Tetrahymena*, son todavía relativamente desconocidos, y el experto en la materia, cuando es preguntado sobre los posibles sistemas de expresión heteróloga de proteína, pensaría más bien en sistemas basados en *E. coli*, levadura, baculovirus y líneas celulares de mamífero inmortalizadas.

Además, el uso de un sistema de expresión en ciliados, particularmente *Tetrahymena*, tiene otra ventaja significativa que no era previsible en el presente contexto. Como la auto/no auto-discriminación de sistemas inmunitarios de mamífero se realiza basándose en la composición de hidratos de carbono de las glucoproteínas, el patrón de glucosilación de un antígeno contribuye, a un gran grado, a su potencial inmunogénico.

Reading et al. (2000) han informado que, en células de mamífero, el receptor de manosa desempeña una función como un principal receptor endocítico en la entrada infecciosa del virus de la gripe, y quizás otros virus envueltos, en macrófagos.

El receptor de manosa puede clasificarse como una proteína de lectina unida a membrana (también llamada "lectina de unión a manosa", "MBL") que media en la captación de glucoproteínas que terminan en manosa, fucosa o N-acetilglucosamina y lectinas de tipo C que contienen dominios de reconocimiento de hidratos de carbono conservados (CRD).

El CRD de la lectina de unión a manosa (MBL) une hexosas tales como manosa y N-acetilglucosamina con grupos hidroxilo ecuatoriales en la posición C3 y C4 del anillo de piranosa. Así, no tienen afinidad por oligosacáridos comúnmente encontrados en las glucoproteínas de mamífero.

Los macrófagos desempeñan una función clave en los sistemas inmunitarios de mamíferos, ya que captan patógenos por endocitosis y, después de digerir los últimos, presentan antígenos relacionados con patógenos a los linfocitos T cooperadores correspondientes. La presentación se hace integrando el antígeno en la membrana celular y presentándola unida a una molécula de clase II de MHC a un linfocito T cooperador correspondiente, que indica a otros glóbulos blancos que el macrófago no es un patógeno, a pesar de tener antígenos sobre su superficie.

Los inventores han deducido, a partir de la investigación de Reading et al. (2000), que, con el fin de potenciar la inmunogenicidad de una vacuna, podría ser útil aumentar la cantidad de manosa en los patrones de glucosilación de proteínas comprendidas en la vacuna (es decir, para potenciar la manosilación). La glucosilación de proteínas se lleva a cabo, sin embargo, en el aparato de modificación postraducciona, que es exclusivo de eucariotas, y así difícil de modificar por diseño.

Sin embargo, los taxones eucariotas muestran diferencias en sus esquemas de glucosilación, particularmente en los esquemas de N-glucosilación. Dicha glucosilación se produce principalmente en eucariotas y arqueas, pero no en bacterias. Generalmente, el término N-glucosilación se refiere a la glucosilación de los restos de aminoácidos asparagina (N). Aquí, una cadena de oligosacárido está unida por la enzima oligosacariltransferasa a aquellos restos de asparagina que se producen en las secuencias de tripéptidos Asn-X-Ser o Asn-X-Thr, donde X puede ser cualquier aminoácido, aunque Pro y Asp son raramente encontrados.

Los inventores de la presente invención se han dado cuenta sorprendentemente de que los ciliados tienen un patrón de N-glucosilación único, que lleva un alto grado de restos de manosa (véase Fig. 1). Es además el mérito de los inventores de la presente invención que encontraron que proteínas que van a usarse para la vacunación en mamíferos podrían beneficiarse de los restos de manosa terminales, concretamente debido al aumento de captura de estas proteínas en macrófagos mediante los receptores de manosa que podrían producir un aumento de la respuesta inmunitaria.

Los inventores han así deducido que las proteínas expresadas en ciliados, particularmente en *Tetrahymena*, tienen un potencial inmunogénico más alto debido a un alto grado de restos de manosa, y que los ciliados transgénicos son así candidatos prometedores para la producción de vacunas, particularmente vacunas de virus, para su uso en mamíferos.

Los inventores se han dado cuenta además de que los macrófagos demostraron un aumento de la fagocitosis de patógenos que tienen un patrón de glucosilación rico en manosa. Por este motivo, hay evidencia de que la expresión de vacunas en ciliados conducirá, debido al patrón de glucosilación rico en manosa, a una reacción inmunitaria potenciada una vez las vacunas así producidas se administran a un sujeto mamífero.

Este hallazgo no se espera por el trabajo de Gaertig et al. (1999, véase anteriormente), ya que los últimos solo han descrito el uso de una proteína de ciliado producida en un *Tetrahymena* transgénico para su uso como una vacuna

en pescado. Sin embargo, el efecto anteriormente descrito del receptor de manosa en macrófagos es según el actual conocimiento solo aplicable para mamíferos, no para peces. Por este motivo, es bastante poco probable que una vacuna de pez sacara beneficio similar de una elevada cantidad de manosa en el patrón de glucosilación de la vacuna.

En una realización particularmente preferida, dicha proteína de la superficie viral es una proteína de fusión viral (véase la definición anteriormente).

En otra realización preferida más, dicha proteína de la superficie viral es al menos una proteína de fusión viral seleccionada del grupo que consiste en proteínas de fusión virales de clase I, clase II y/o clase III.

Dichas proteínas de fusión virales son estructuralmente bastante diversas, pero sus subunidades de fusión se pliegan por último lugar en un trímero de horquillas, en el que tres regiones del extremo C se acumulan en el exterior de un núcleo trímero del extremo N central.

Como se indica anteriormente, las proteínas de fusión virales se subdividen en proteínas de fusión virales de clase I, clase II y clase III. La Tabla 2 da una visión general de algunas proteínas de fusión virales de virus envueltos que cumplen la definición anterior.

En el presente documento, "-S-S-" indica un puente disulfuro, mientras que "/" indica que las subunidades indicadas están asociadas entre sí, pero no unidas por disulfuro. "GPX" o "gpX" representa "glucoproteína X", mientras que "FX" representa "proteína de fusión X".

Tabla 2

Tabla 2				
familia de virus	Especies de virus de ejemplo	Proteínas de ejemplo	Código PDB	Proteína de fusión viral/ subunidades de fusión
Clase I				
Orthomyxoviridae	Virus de la gripe A (HA) Virus de la gripe C (HEF)	HA1, HA2, HEF1, HEF2,	1HA0, 1FLC	HA1-S-S-HA2
Retroviridae	Virus de la leucemia murina de Moloney (TM) virus de la inmunodeficiencia humana	TM, SU HIV-1 gp41, gp120 gp21	1AOL ENV, 1AIK 1MGI	SU-S-S-TM gp120-S-S-gp41
Paramyxo-viridae	Virus paragripal simio 5 (F) virus paragripal humano F virus de la enfermedad de Newcastle F respiratorio sincitial F	F1, F2	2B9B, ISVF IZTM 1G5G 1G2C	F2-S-S-F1
Corona-viridae	Virus de la hepatitis de ratón S2 SARS corona virus E2	S1, S2	1WDG 2BEQ	S1/S2
Filoviridae	Virus del Ébola gp2	GP1, GP2	1EBO, 2EBO	GP1-S-S-GP2
Clase II				
Arena-viridae	Virus Junín	GP1, GP2		GP1/GP2/SSP
Toga-viridae	Virus del bosque de Semliki E1	E1, E2	1E9W, 1RER	E1/E2
Flavi-viridae	Virus de la encefalitis transmitida por garrapatas (E) virus E del dengue 2 y 3	E	1URZ, 1SVB 1OKB, 1UZG	E, E1/E2f
Bunya-viridae	Fiebre hemorrágica de Crimea-Congo (CCHF)	GN, GC		GN/GC
	Virus de Hantaan (HTNV)	G1, G2		
Clase III				
Rhabdo-viridae	Virus de la rabia virus de la estomatitis vesicular	Proteínas G	2GUM	G
Herpes-viridae	Virus del herpes simple gB	gB, gD, gH/L	2CMZ	gB, gH/L

Muy frecuentemente, dos o más subunidades de una proteína de fusión viral de clase 1 se unen por un puente disulfuro, o por otros medios. Ejemplos de tales dímeros son HA1-S-S-HA2, SU-S-S-TM, HA1-S-S-HA2, SU-S-S-TM, SU/TM, F2-S-S-F1, S1/S2, GP1-S-S-GP2, GP1/GP2/SSP. En el presente documento, al menos una subunidad está

dominada por dominios hidrófobos, como, por ejemplo, la subunidad HA2 en la construcción HA1-S-S-HA2.

En una realización preferida, dicha proteína es una proteína de fusión viral de clase I. Véase la Fig. 4 para un esquema de una proteína de fusión viral de clase I.

Una proteína viral particularmente preferida es la hemaglutinina. La hemaglutinina (HA) es una glucoproteína antigénica que consiste en dos subunidades de proteína de fusión de membrana de clase 1, concretamente HA2 y HA1, que se encuentra sobre la superficie del virus de la gripe (Orthomyxoviridae), además de muchas otras bacterias y virus.

La hemaglutinina es responsable de la unión del virus a la célula que está siendo infectada. Hoy en día se conocen al menos 16 subtipos de HA diferentes, marcadas H1 - H16. Las tres primeras hemaglutininas, H1, H2 y H3, se encuentran en el virus de la gripe humano, mientras que el virus de la gripe aviar tiene, por ejemplo, una hemaglutinina H5.

La principal función de la HA es el reconocimiento de células de vertebrado diana, realizado mediante los receptores que contienen ácido siálico respectivos sobre la membrana de célula diana, y la entrada del genoma viral en las células diana produciendo la fusión de la membrana de célula diana con la membrana viral. En este contexto, merece la pena mencionar que los Orthomyxoviridae tienen una membrana de fosfolípido que encierra la nucleocápside, siendo dicha membrana adquirida de la célula huésped tras la liberación de la última.

HA es una glucoproteína de membrana integral homotrimérica. Tiene forma de un cilindro, y tiene aproximadamente 135 Å de longitud. Los tres monómeros idénticos que constituyen la HA están contruidos en una bobina de hélice α central; tres cabezas esféricas contienen los sitios de unión a ácido siálico. Los monómeros de HA se sintetizan como precursores que entonces se glucosilan y se escinden en dos polipéptidos más pequeños, es decir, las subunidades HA1 y HA2. Cada monómero de HA consiste en una cadena helicoidal larga anclada en la membrana por HA2 y coronada por un gran glóbulo de HA1 (véase la Fig. 3).

El mecanismo de unión es del siguiente modo: HA se une al ácido siálico de monosacárido que está presente en la superficie de sus células diana. Esto hace que las partículas virales se peguen a la superficie de la célula. La membrana celular engulle entonces el virus y se cierra para formar un nuevo compartimento unido a membrana dentro de la célula, llamado el endosoma, que contiene el virus engullido. La célula digiere entonces el contenido del endosoma acidificando su interior y transformándolo en un lisosoma. Sin embargo, un pH bajo (<6,0) desencadena cambios conformacionales de la molécula de HA, en el que HA1 se separa de HA2 y entonces se repliega en una forma completamente diferente. Al final de este proceso, el llamado "péptido de fusión" actúa como un gancho molecular insertándose él mismo en la membrana endosómica y bloqueándose. Tan pronto como el resto de la molécula de HA se repliega en una nueva estructura (que es más estable al pH más bajo), retrae el "garfio" y tira de la membrana endosómica hasta a continuación de la propia membrana de la partícula de virus, haciendo que los dos se fusionen entre sí. Una vez ha ocurrido esto, el contenido del virus, que incluye su genoma de ARN, está libre para inundar el citoplasma de la célula diana.

Los inventores encontraron que resultó difícil la expresión como proteína soluble (expresión intracelular), además de la secreción (expresión extracelular) de una hemaglutinina completa en *Tetrahymena*. Aunque en experimentos preliminares la expresión de proteínas en la membrana celular como tal fue satisfactoria, la expresión de proteína soluble y la secreción de la misma planteó problemas, ya que pareció que las moléculas de hemaglutinina expresadas satisfactoriamente en *Tetrahymena*, y equipadas con un péptido señal, seguían unidas a estructuras de membrana intracelular, además de a la membrana celular de la célula huésped.

Este hallazgo fue a pesar del uso de péptidos señal que han resultado ser útiles en la expresión y secreción de otras proteínas, como se muestra, por ejemplo, en el documento WO03078566, que se incorpora en el presente documento por referencia.

Los inventores han atribuido este fenómeno a los dominios altamente hidrófobos, particularmente de HA2, que pueden ser responsables de retener las proteínas satisfactoriamente expresadas en la membrana celular.

Se cree que este fenómeno es similar al proceso de replicación natural de partículas de virus en una célula huésped infectada, en el que la mayoría de las proteínas virales expresadas por el huésped infectado se usan para ensamblar el virión dentro del citoplasma, mientras que la hemaglutinina es llevada mediante el retículo endoplásmico y el aparato de Golgi a la superficie celular, donde queda anclada en la membrana celular, hasta que el virus maduro se desprenda de la célula en una esfera de membrana de fosfolípido huésped (primera membrana de célula huésped), adquiriendo así su envoltura de hemaglutinina.

En otra realización preferida más de la presente invención, se proporciona que la proteína de fusión viral codificada por el ácido nucleico carece de al menos un dominio hidrófobo, o una fracción del mismo.

En el contexto de la presente invención, los inventores han descubierto sorprendentemente que una proteína viral que carece de al menos el dominio de péptido de fusión hidrófobo, o una fracción del mismo, facilita la secreción de proteína.

En la expresión heteróloga de proteínas, la secreción de proteína está soportada por una secuencia señal (véase anteriormente). Un gran dominio hidrófobo puede producir problemas cuando se llega a la secreción de proteína, ya que puede ejercer alguna afinidad por la membrana celular, o por membranas intracelulares, del huésped de expresión. Básicamente, la secreción de proteína en ciliados transgénicos como *Tetrahymena* se obtiene bastante fácilmente, ya que los últimos secretan grandes cantidades de enzimas hidrolíticas para la digestión extracelular de alimentos. Sin embargo, resultó que la secreción de proteínas de la superficie viral de escala completa en ciliados puede llegar a ser incómoda.

Los inventores han encontrado ahora que la eliminación de un dominio hidrófobo ayuda en tales casos a facilitar la secreción de proteína, y a obtener altos rendimientos de la proteína secretada, sin afectar el potencial inmunogénico de la proteína así modificada.

En proteínas virales, los dominios antigénicos son frecuentemente hidrófilos, ya que se extienden desde la superficie del virus en el medio circundante, que es frecuentemente un medio acuoso (saliva, sangre, moco y otros líquidos corporales, aguas residuales, etc.). Esto significa que una eliminación de dominios hidrófobos no afecta, en estos casos, la inmunogenicidad de la proteína.

Dicho principio de eliminación de un dominio hidrófobo en una proteína viral con el fin de facilitar la secreción de proteína o facilitar la purificación de proteínas después de la lisis celular, mientras que retiene la inmunogenicidad de proteína, se demostrará con hemaglutinina (HA) a continuación:

En la expresión de la proteína HA mencionada, HA0 es una proteína precursora, que consiste en dos subunidades HA1 y HA2 codificadas por un ADNc respectivo. HA0 es escindida por enzimas proteolíticas (por ejemplo, una triptasa como se encuentra en el tejido de pulmón de muchos vertebrados) en sus subunidades, que siguen unidas entre sí por un puente disulfuro (proteína HA1-HA2, véanse las Fig. 3 y 4).

Aunque HA1 es responsable del reconocimiento de receptor, y comprende los dominios con el potencial inmunogénico más alto, particularmente un dominio esférico ("cabeza") que comprende sitios de unión para ácidos siálicos localizados en el glucocáliz de la membrana de célula huésped, HA2 (que también ancla el complejo HA1-HA2 completo en la membrana viral) es responsable de la fusión de membranas. HA2 comprende un péptido de fusión de extremo N (FP) que consiste en una secuencia relativamente hidrófoba (véase la Tabla 1).

Este péptido de fusión consiste en aproximadamente 24 restos altamente conservados del extremo N, de los que al menos seis son glicinas (véase la Fig. 4, donde los péptidos de fusión comprenden ocho glicinas). Siguen tres hélices α (α -H1, α -H2 y α -H3). Sigue un dominio transmembranario altamente hidrófobo (TMD), que ancla la proteína en la bicapa lipídica de la envoltura viral (véase la Fig. 3).

La siguiente tabla, en la que "TMD" representa "dominio transmembranario" y "FP" representa "péptido de fusión", da una visión general sobre las posibles hemaglutininas modificadas que carecen de al menos un dominio hidrófobo, según la presente invención. Es bastante claro de lo anterior que dicho principio se aplica también a otras proteínas de fusión virales de clase I, clase II y/o clase III.

Tabla 3

Ejemplo	HA1	HA2	También conocida como/ Seq ID No
1	no truncada	truncada en el extremo C	"HA Larga", Seq ID No 2 y 3
2	no truncada	TMD eliminado	"HA_Corta", Seq ID No 4 y 5
3	no truncada	HA2 completa (incluido FP) eliminada	"HA_1", Seq ID No 6 y 7
4	Reducida al dominio esférico ("cabeza")	HA2 completa (incluido FP) eliminada	

Merece la pena mencionar que, como una alternativa a la secreción de proteína, los inventores encontraron que la expresión intracelular de una proteína viral en una forma que es soluble en el citoplasma es una opción prometedora. Los inventores han encontrado que las proteínas de fusión virales de clase I completamente caracterizadas expresadas intracelularmente tenderán a unirse a membranas intracelulares del huésped de ciliado, de las que hay muchas (retículo endoplásmico, membrana de golgi, nuclear, mitocondrias, etc.), además de a la membrana celular, y en este caso pueden no aislarse después de la lisis de las células huésped. Los inventores han encontrado además que las proteínas de fusión virales presentarán tendencia reducida a unirse a membranas

intracelulares si carecen de al menos un dominio hidrófobo. Las proteínas modificadas de tal forma (por ejemplo, como se muestra en la Tabla 3) son así más fáciles de aislar a partir de lisados de célula completa.

Además, según la divulgación, se proporciona al menos una molécula de ácido nucleico, la última seleccionada del grupo que consiste en

- a) una molécula de ácido nucleico que comprende una secuencia de nucleótidos presentada como SEQ ID NO: 2, 4 y/o 6,
- b) una molécula de ácido nucleico que codifica un polipéptido que comprende la secuencia de aminoácidos presentada como SEQ ID NO: 3, 5 y/o 7, en la que dicho polipéptido es una proteína viral, o un fragmento de la misma,
- c) una molécula de ácido nucleico que codifica una forma truncada de una proteína viral que carece de al menos un dominio hidrófobo, preferentemente de una proteína de fusión viral, y más preferentemente de una proteína de fusión viral de clase I, clase II y/o clase III,
- d) una molécula de ácido nucleico que codifica una forma truncada de una proteína de fusión viral de clase I que carece del dominio transmembranario (TMD), el péptido de fusión (FP) y/o la subunidad HA2,
- e) una molécula de ácido nucleico que es una fracción, variante, homólogo o derivado de las moléculas de ácidos nucleicos de a) - d),
- f) una molécula de ácido nucleico que es un complemento a cualquiera de las moléculas de ácidos nucleicos de a) - e),
- g) una molécula de ácido nucleico que es capaz de hibridarse con cualquiera de las moléculas de ácidos nucleicos de a) - f) bajo condiciones rigurosas,
- h) una molécula de ácido nucleico que comprende, en comparación con cualquiera de las moléculas de ácidos nucleicos de a) - g), al menos una sustitución silenciosa de un solo nucleótido,
- i) una molécula de ácido nucleico según a) y c) - h) que es de codón optimizado para un huésped de expresión en protozoos, y/o
- j) una molécula de ácido nucleico que tiene una identidad de secuencia de al menos el 70 %, preferentemente el 95 % con cualquiera de las moléculas de ácidos nucleicos de a) - i).

Posiciones de último recurso particularmente preferidas para dicha molécula de ácido nucleico tienen una identidad de secuencia del 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 o el 99 % con cualquiera de las moléculas de ácidos nucleicos de a) - i).

Además, se proporciona un vector para la transfección de una célula huésped de ciliado, comprendiendo dicho vector

- a) un ácido nucleico que codifican una proteína viral, o un fragmento de la misma, y
- b) un promotor operativamente unido a dicho ácido nucleico.

Dicho ácido nucleico es preferentemente un ADNc.

La invención proporciona además una célula huésped de ciliado transfectada con un vector según la invención. Dicho ciliado es preferible un miembro de la familia Tetrahymenidae.

Además, la invención proporcionó un proceso para la producción de al menos una proteína viral, o un fragmento de la misma, en una célula huésped de ciliado, comprendiendo dicho proceso las etapas de

- a) transfectar la célula huésped de ciliado con el vector según la invención,
- b) cultivar la célula huésped en condiciones que permiten la expresión de una proteína.

Dicho proceso comprende además, en una realización preferida, una etapa de

- c) cultivar la célula huésped en condiciones que permiten la secreción de la proteína expresada de una proteína.

Además, la divulgación proporciona una proteína seleccionada del grupo que consiste en

- a) una proteína codificada por un ácido nucleico según la divulgación,
- b) una proteína que comprende la secuencia de aminoácidos según la divulgación,
- c) una proteína según a) o b) que comprende al menos una sustitución de aminoácidos conservativa, y/o
- d) una proteína obtenible por el proceso según la invención.

Además, la invención proporciona un método para la producción de una composición farmacéutica, comprendiendo dicho método las etapas de

- a) expresar una proteína según la invención en un sistema de expresión de ciliado según la invención, y
- b) aislar y/o purificar la proteína así obtenida.

En una realización alternativa, la invención proporciona un método para la producción de una composición farmacéutica, comprendiendo dicho método el proceso según la invención. Además, la invención proporciona una composición farmacéutica que comprende una proteína según la invención, o que se produce con un método según la invención. Dicha composición es, en una realización preferida, una vacuna.

Descargo de responsabilidad

Como reconocerán aquellos expertos en la materia, pueden ocurrírsele variaciones, modificaciones y otras implementaciones de lo que se describe en el presente documento, a aquellos expertos habituales en la materia sin apartarse del alcance de la invención como se reivindica. Por consiguiente, la descripción anterior es a modo de ejemplo solo y no está prevista como limitante. El alcance de la invención se define en las siguientes reivindicaciones.

Además, los signos de referencia usados en la descripción y reivindicaciones no limitan el alcance de la invención como se reivindica.

Breve descripción de los ejemplos y dibujos

Detalles adicionales, rasgos, características y ventajas del objetivo de la invención se desvelan en las reivindicaciones dependientes, y la siguiente descripción de las figuras y ejemplos respectivos, que, en un modo a modo de ejemplo, muestran realizaciones preferidas de la presente invención. Sin embargo, estos dibujos ni mucho menos deben entenderse como que limitan el alcance de la invención.

Ejemplos y figuras

1. Construcción de vectores de expresión

Se clonaron los genes sintéticos para los diferentes fragmentos de hemaglutinina (véase SEQ ID No. 2, 4 y 6) en el vector donante (véase Fig. 2A). Los casetes de expresión de todos los vectores donantes se transfirieron en el vector aceptor (véase Fig. 2B) usando un sistema de recombinasa dependiente de Cre.

2. Cultivo de *Tetrahymena* no mutante y transformación de plásmidos de expresión

Se cultivaron cepas de *Tetrahymena thermophila* no mutantes (por ejemplo, B 1868/4, B 1868/7 y B 2068/1) en medio de leche desnatada, en proteosa peptona complementada (SPP) o en medio químicamente definido (CDM). La transformación de las células de *T. thermophila* se realizó como se describe previamente en Cassidy-Hanley et al. 1997.

3. Detección de hemaglutinina recombinante

Se cultivaron células de *Tetrahymena* transformadas en medio SPP bajo presión de selección a 30 °C en un agitador a 80 rpm. Se indujo la expresión del gen diana por choque térmico a 41 °C (HSP-P) o mediante la adición de Cd²⁺ 20 nM (MTT1-P) de cultivos de crecimiento logarítmico.

Se recogieron alícuotas de células transformadas y de sobrenadantes de SPP libres de célula 24 h después de la inducción de la expresión del gen diana. Las células recogidas se solubilizaron en tampón RIPA helado (5000 células/μl en NaCl 150 mM, Tris-HCl 10mM, EDTA 5 mM, 0,1 % de SDS, 0,1 % de DOC, 1 % de Triton X100, E64 2,5 μg/ml) y se rompieron durante 15 minutos en un sonicador. Se hicieron SDS-PAGE y análisis de transferencia Western según la materia. Brevemente, se resuspendieron alícuotas de tanto células rotas (es decir, 1000 células) como sobrenadante libre de células en tampón de muestra de Laemmli (Tris-HCl 125 mM a pH 6,8, 10 % de glicerol, 4 % de SDS) y se separaron en 12 % de SDS-PAGE. Los geles se transfirieron a membranas de nitrocelulosa y se bloquearon en PBS que contenía 0,05 % de Tween 20 y 5 % de leche desnatada o 3 % de albúmina de suero bovino. La expresión de hemaglutinina recombinante en ciliados transformados se detectó usando anticuerpos primarios específicos de la cepa de virus. Después de lavar con PBS/T y aplicar el anticuerpo secundario conjugado con HRP, las transferencias se desarrollaron usando Super Signal West Pico Chemoluminiscent Substrate (Perbio, Fischer Scientific) en combinación con revelado de película de rayos X convencional. La Figura 8 y las Figuras 9A - C muestran transferencias Western representativas de lisados celulares y sobrenadantes de células de *Tetrahymena* transformadas después de la inducción de la expresión del gen diana. El control no mutante (Fig. 9B, carril 3 sedimento de células y 4 sobrenadante) es todo blanco. En la Figura 8 se muestra la secreción de un antígeno de HA en el sobrenadante durante el proceso de fermentación. En la Figura 9 se muestra la detección de hemaglutininas truncadas (eliminación de dominio intracelular (Fig. 9A, "HA_larga"), eliminación adicional de región transmembranaria (Fig. 9B, "HA_corta") y eliminación de la subunidad HA2 completa (Fig. 9C, "HA_1") en el sobrenadante de ciliados transformados.

4. Producción de hemaglutinina

Para las fermentaciones se usaron un Braun UD50 (50 litros) y un Infors Sixfors (0,5 litros) equipados con impulsores Rushton estándar. La velocidad del agitador se limitó a 300 y 400 rpm; pO₂ se estableció al 20 % y el pH se estableció a 7,0 respectivamente. Las fermentaciones se llevaron a cabo en medio estándar.

5. Purificación de hemaglutinina recombinante

La recogida de las células del proceso de fermentación de 50 l se logró usando un módulo de fibra hueca (0,3 m², 3 l/min, diámetro de tubo 11 mm) para separar las células del caldo de fermentación. El sedimento de células se lavó 3 a 4 veces usando tampón fosfato de sodio helado a pH 7,4 (10 mM, PB) y se sedimentó por etapas de centrifugación a 10 °C y 2400 x g durante 8 min (Sorvall Evolution, rotor SLA-1500) para eliminar el contenido de micocistos. El sedimento de células obtenido se resuspendió en PB complementado con el inhibidor de proteasa de cisteína E-64 (70 µM) y 3 % de Tween®20. Las células se rompieron usando un Ultraturrax (IKA UT T25 + S25N-25G) a 10.000 rpm durante 5 min sobre hielo. El lisado se cargó hasta 1,8 l usando PB (pH 7,4) y se solubilizó agitando a 4 °C durante 17 horas. La filtración del lisado celular tuvo lugar usando un módulo de fibra hueca (0,45 µm con 850 cm² de superficie) que se lavó tres veces con 2,3 l de PB (pH 7,4).

Se realizó purificación cromatográfica usando un proceso de purificación cromatográfica de tres etapas que produjo un antígeno HA recombinante altamente purificado que no se desnaturaliza y es adecuado como componente de una vacuna.

Toda la cromatografía se realizó a 4 °C. El filtrado que contenía HA preparado como se ha descrito anteriormente se cargó primero a 15 ml/min a una columna de intercambio aniónico de amonio (columna Capto™ Q, una columna de perlas de agarosa con amonio cuaternario fuerte) equilibrada con PB que contenía 5 % de glicerol y 3 % de Tween®20 a pH 7,4. La columna se lavó primero con tampón de carga que contenía 3 % de Tween®20, seguido de una etapa de lavado con tampón de carga sin Tween®20 para eliminar el detergente. La elución de la HA parcialmente purificada se hizo usando PB que contenía 5 % de glicerol y NaCl 150 mM en una primera etapa y NaCl 1 M en una segunda etapa (pH 7,4). Las muestras de todas las fracciones recogidas se analizaron por SDS-PAGE, transferencia Western y ensayo Bradford, y las fracciones positivas para HA se reunieron para la siguiente etapa de cromatografía.

Se usó una columna de hidroxiapatita cerámica (CHT) para eliminar proteínas contaminantes, debido a que la HA recombinante pueden no unirse en esta columna. Se concentraron las fracciones reunidas obtenidas de la columna Capto™ Q, y se realizó un intercambio de tampón para eliminar el NaCl usando un módulo TFF de escala de laboratorio (30 kDa) con PB que contenía 5 % de glicerol. El pH de la muestra se ajustó a 7,5 antes de la carga a la columna CHT. El caudal de carga era 7 ml/min y la columna se lavó posteriormente con PB que contenía 5 % de glicerol. La elución se realizó aplicando PB 150 mM que contenía 5 % de glicerol a pH 7,5 por una elución en etapas. El eluato y el flujo a través se probaron por SDS-PAGE, análisis de transferencia Western y ensayo de Bradford, y se tomó el flujo que contenía HA a través de 200 ml para carga a la tercera columna.

Para la tercera etapa de purificación se usó una columna de Con A Sepharose 4B (un medio de afinidad con concanavalina A acoplada a Sepharose 4B por el método de bromuro de cianógeno). El flujo a través de la columna de hidroxiapatita se complementó con NaCl 150 mM y se aplicó con un caudal de 5 ml/min a la columna. La columna se lavó con solución salina tamponada con fosfato (PBS, pH 7,4) y la HA recombinante purificada se eluyó usando PBS que contenía metil-α-D-manopiranosido 0,5 M (pH 7,4). Se analizaron las fracciones de la elución por SDS-PAGE, transferencia Western y ensayo de Bradford.

Figuras

La Fig. 1A muestra una visión general de patrones de N-glucosilación de diferentes taxones. Generalmente, el término "N-glucosilación" se refiere a la glucosilación del resto de aminoácido asparagina (N). Aquí, una cadena de oligosacárido está unida por oligosacariltransferasa a aquellos restos de asparagina que se producen en las secuencias de tripéptidos Asn-X-Ser o Asn-X-Thr, donde X puede ser cualquier aminoácido excepto Pro.

Es obvio que, mientras que los procariotas no tienen N-glucosilación en absoluto, los eucariotas presentan patrones de N-glucosilación que son ricos en manosa, y así potencian el potencial inmunogénico de una proteína (véase el texto). La Fig. 1b muestra variaciones del potencial en dicho patrón en algunas especies de ciliados.

La Fig. 2A muestra el vector donante que codifica el gen HA sintético. El vector donante consiste en un esqueleto bacteriano para la amplificación en *E. coli* (pUC ori), una kanamicina (kanR), cloranfenicol (cmR) casete de selección y gen sucrasa (sacB) para la selección en *E. coli* y un marco de lectura abierto del gen diana (gen HA) bajo el control de un promotor inducible y seguido de secuencia terminadora de [beta]-tubulina 2 de *T. thermophila* (btu2).

La Fig. 2B muestra un vector de expresión para el uso en el ciliado *Tetrahymena thermophila*. El vector contiene un gen de resistencia a ampicilina (ampR) y cloranfenicol (cmR) para la selección en *E.coli*, un origen específico de *T. thermophila* (ori de ADNr) para la propagación de plásmido en *T. thermophila*, un casete de selección basado en neomicina (neoR) para la identificación de ciliados transformados y un marco de lectura abierto del gen diana (gen HA) bajo el control de un promotor inducible y seguido de secuencia terminadora de [beta]-tubulina 2 de *T. thermophila* (btu2).

La Fig. 3A muestra una hemaglutinina con sus subunidades HA1 (negro) y HA2 (gris). Se indican hojas de hélices alfa y beta.

La Fig. 3b muestra un dibujo esquemático de hemaglutinina con subunidades HA1 (negro) y HA2. El péptido de fusión (FP, véase anterior) se muestra además del dominio transmembranario (TMD) que ancla la proteína en la membrana de envuelta de virus. HA1 y HA2 están unidas por un puente disulfuro (DSB). La sección de HA2 que conecta las dos hélices α (α -H1 y α -H2) se llama el "dominio de bucle" (LD). No se muestra la cola citoplasmática del extremo C de HA-2.

HA1 comprende un dominio esférico ("cabeza") que está estabilizado por puentes disulfuro no mostrados. Dicha cabeza comprende sitios de unión para ácidos siálicos localizados en el glucocáliz de la membrana de célula huésped. Asimismo, la cabeza tiene un potencial inmunogénico alto y es así la diana para la mayoría de los anticuerpos generados contra HA después de la inmunización.

Para provocar la fusión de la bicapa de envoltura del virus con la membrana de célula huésped, HA1 y HA2 se separan entre sí provocado por una disminución de pH, que produce una carga positiva de las subunidades de HA1 en el homotrímero de HA.

Esto conduce a fuerzas de repulsión y las subunidades de HA1 se alejan las unas de las otras, retirándose así de las subunidades de HA2, que entonces se activan. La HA2 activada representa la fusión de las membranas. HA2 puede activarse solo una vez, y después queda inactiva. Así, es importante que HA1 no se retire demasiado pronto, de otro modo el virus pierde su potencial infectivo.

HA2 tiene, con respecto a estiramientos grandes, una estructura de hélice alfa y contiene una gran región de bucle. Además, HA2 contiene el dominio transmembranario y el péptido de fusión. El péptido de fusión se libera por el desprendimiento de subunidades de HA1.

Con el fin de liberar la fusión, el dominio esférico ("cabeza") de HA1 debe separarse de HA2. Así, HA2 puede cambiar su conformación, de manera que se desdoble y el péptido de fusión pueda penetrar en la membrana huésped. Así, el virus llega a conectarse directamente a la membrana del huésped.

Debido al proceso de despliegue, HA2 está estirada en un lado. Otras áreas están enrolladas, de manera que no hay ampliación neta de la proteína. Entonces, HA cambia su conformación y pone el virus cerca de la membrana del huésped, de manera que tiene lugar la fusión de membrana.

La Fig. 4A muestra la estructura general de una proteína de fusión viral de clase I como se define en el texto, como se ejemplifica por hemaglutinina. HA1 y HA2 son subunidades de una proteína de la clase I de HA0 de la gripe. Las marcas especifican las posiciones del péptido señal (SP), el péptido de fusión (FP), sitios de procesamiento (flechas), el dominio transmembranario (TMD), que son subunidades de una proteína de fusión viral de clase I general.

En la Fig. 4B (véase también SEQ ID 3), las subunidades de la hemaglutinina de la cepa A del virus de la gripe A/Nueva Caledonia/20/1999 H1N1 (Número de acceso de GenBank: AAP34324.1) se indican a modo de ejemplo. Fragmentos subrayados indican el péptido señal, (SP), el péptido de fusión HA2 (FP) y el dominio transmembranario de HA2 (TMD).

La proteína mostrada consiste en 565 restos de aminoácidos ("AA"). Características de secuencia específicas se indican en la Tabla 4. Obsérvese que el péptido de fusión de HA2 y el dominio transmembranario de HA2 tienen una alta participación de aminoácidos hidrófobos (datos calculados basándose en los datos de hidropatía a partir de la Tabla 1), en la que, por ejemplo, el péptido de fusión comprende 8 restos de glicina. Los inventores han mostrado que para fines de expresión heteróloga en ciliados, una truncación de estos dominios, como se proporciona por la presente invención, facilita la secreción de proteína y/o facilita la purificación de proteínas después de la lisis celular.

Tabla 4

Subunidades	Dominios funcionales	AA	Participación de AA hidrófobos
Péptido señal		1 - 17	
Cadena de HA1		18 - 343	42 %

Cadena de HA2		344 - 565	41 %
	Péptido de fusión de HA2	344 - 368	71 %
	Dominio extracelular de HA2	368 - 529	
	Dominio transmembranario de HA2	530 - 550	66,6 %
	Dominio intracelular de HA2	551 - 565	
Sitios de N-glucosilación		28, 40, 104, 142, 176, 303, 497	
Enlace disulfuro intercatenario		21 - 480	

La Tabla 4 aclara que el dominio transmembranario de HA2 y el péptido de fusión de HA2 tienen la participación más alta de aminoácidos hidrófobos. Los inventores han deducido de este hallazgo que la truncación de al menos uno de estos dos dominios facilitará la secreción de la proteína en el medio.

Es digno de atención que las características mostradas para la proteína en la Fig. 4 son aplicables a otras hemaglutininas también, como para la mayoría de las proteínas de fusión virales de clase I mostradas en la Tabla 2.

La Fig. 5 muestra una comparación entre el uso de codones en *Tetrahymena thermophila* y *Homo sapiens*. El último es aplicable para proteínas virales que se expresan en un huésped humano, como hemaglutinina expresada de una célula huésped humana. Véase el texto para explicaciones adicionales.

La Fig. 6 muestra el código genético como se usa en ciliados, particularmente en *Tetrahymena*. Los códigos de nucleótidos no canónicos UAA y UAG, que codifican glutamina, están en negrita. Según el código genético general, estos tripletes son, sin embargo, codones de terminación (véanse los tripletes tachados). "1LC" representa un "código de una letra", mientras que "3LC" representa "código de tres letras".

La Fig. 7 muestra la comparación a modo de ejemplo entre el uso de codones entre la secuencia de hemaglutinina viral (véase SEQ ID No. 1) y la secuencia de hemaglutinina de codón optimizado (véase SEQ ID No. 2) para una expresión en *Tetrahymena thermophila*. Diferencias en el uso de codones se indican usando recuadros grises.

En la Fig. 8 se muestran inmunotransferencias de ciliados transformados cultivados en un proceso de fermentación de lotes alimentados (0,5 l) que expresa HA recombinante. Se aplicaron muestras del sobrenadante de diferentes momentos de tiempo de recogida (SN1- 47,5 h, SN2- 66 h, SN3- 70 h, SN4- 90 h) a SDS-PAGE bajo condiciones no reductoras seguido de transferencia Western. La tinción tuvo lugar usando un suero anti-B/Florida/4/2006 de gripe específico. El antígeno NIBSC-HA aplicado (control positivo) mostró una señal correspondiente a aproximadamente 90 kDa (A, NIBSC, 45 ng). Después de 47,5 h de cultivo apareció una banda específica de HA a aproximadamente 72 kDa y 90 kDa en el sobrenadante. Dicha proteína se corresponde con Seq ID No 9.

En la Fig. 9 se muestran inmunotransferencias de fragmentos de hemaglutinina de proteínas de diferentes longitudes (truncadas según la enseñanza de la presente invención) recombinantemente expresadas en *Tetrahymena*. La tinción tuvo lugar usando un anticuerpo anti-H1N1 de cobaya específico. P es sedimento de células de *Tetrahymena* y SN es sobrenadante de cultivo celular. El carril 1 y 2 se refiere a clones de *Tetrahymena* mientras que el carril 3 y 4 son el control no mutante.

La Fig. 9A muestra una proteína llamada "HA_larga", que es un dímero de HA1-HA2 completo que carece del dominio intracelular (CT, véase la Fig. 4A) de HA2 (peso molecular predicho 61,4 kDa). Dicha proteína se corresponde con Seq ID No 3.

La Fig. 9B muestra una proteína llamada "HA_corta" que es un dímero de HA1-HA2 que carece del dominio intracelular (CT) y el dominio hidrófobo transmembranario (TMD) de HA2 (peso molecular predicho 57,6 kDa). Dicha proteína se corresponde con Seq ID No 5.

La Fig. 9C muestra HA1 solo (llamada "HA_1", peso molecular predicho 34,2 kDa), es decir, HA2 con sus al menos dos dominios hidrófobos (péptido de fusión y dominio transmembranario) truncados de la misma. Dicha proteína se corresponde con Seq ID No 7.

Es evidente de las inmunotransferencias que HA_corta y HA_1, que carecen de al menos un dominio hidrófobo, son más abundantes en el sobrenadante que HA_larga, que soporta la conclusión que la secreción de proteína de proteínas de fusión virales en ciliados se promueve por eliminación de al menos un dominio hidrófobo, debido a que se reduce la tendencia de unirse a membranas celulares. Además, como los dominios inmunogénicos de proteínas de fusión virales, particularmente hemaglutinina, están localizados en la subunidad HA1, debe esperarse que una vacuna que comprende solo la subunidad HA1, o una subunidad HA1 más una subunidad HA2 truncada, todavía provocará una respuesta inmunitaria, y es así útil para la producción de una vacuna.

La Fig. 10A muestra una representación de hidrofobicidad de la proteína según Seq ID No 3 (es decir, "HA_larga") de la presente invención, que ha sido representada según la escala de Kyte-Doolittle. Se determina fácilmente que el péptido de fusión (AA 344 - 368) y el dominio transmembranario (AA 530 - 550) están entre las principales regiones hidrófobas de la proteína (véanse las flechas), y así impiden una secreción eficiente de la proteína en el medio.

La Fig. 10B muestra una representación de hidrofobicidad de una proteína secretada importante de ciliado (SEQ ID NO 2 del documento WO03078566 por los inventores de la presente invención), que es secretada por *Tetrahymena* en su entorno natural en grandes cantidades. Es bastante obvio que esta proteína no tiene dominios altamente hidrófobos y es en promedio menos hidrófoba que HA_larga. Los inventores hacen este fenómeno responsable de la facilidad de secreción de esta proteína a diferencia de HA_larga.

Listado de secuencias

En el apéndice, se añade un listado completo de secuencias, que se refiere a las siguientes secuencias (NA: ácido nucleico; AA; aminoácido).

Tabla 5

Seq ID NO	Descripción	Composición	También conocido como
1	Secuencia de NA de un gen HA completo (cepa A/Nueva Caledonia/20/99 H1N1)	HA1-HA2, que carece de dominio intracelular de HA2	
2	Versión de codón optimizado de Seq ID NO 1		
3	Secuencia de AA codificada por Seq ID 3		HA_larga
4	Secuencia de NA (de codón optimizado) de un fragmento truncado de Seq ID NO 2	HA1-HA2, que carece de dominio intracelular de HA2 y dominio transmembranario	
5	Secuencia de AA codificada por Seq ID 6		HA corta
6	Secuencia de NA (de codón optimizado) de otro fragmento truncado Seq ID NO 2	HA1 solo, es decir, que carece de HA2	
7	Secuencia de AA codificada por Seq ID 6		HA 1
8	Secuencia de NA de péptido señal de fosfolipasa A1 de <i>Tetrahymena thermophila</i> .		
9	Secuencia de AA codificada por Seq ID 8		
10	Secuencia de NA del péptido señal endógeno del gen HA		
11	Secuencia de AA codificada por Seq ID 10		
12	Secuencia de NA de un promotor inducible por calor de <i>Tetrahymena thermophila</i>		PLA_P
13	Secuencia de NA de un promotor de metalotioneína de <i>Tetrahymena thermophila</i>		MTT1_P
14	Secuencia de NA de un gen HA completo (cepa B/Florida/4/2006)	HA completa	
15	Secuencia de AA codificada por Seq ID 14		HA

Referencias:

Cassidy-Hanley, D.; Bowen, J.; Lee, J.; Cole, E.; VerPlank, L.; Gaertig, J.; Gorovsky, M. & Bruns, P. Germline and somatic transformation of mating *Tetrahymena thermophila* by particle bombardment. *Genetics*, 1997, 146, 135-47.

Gaertig, J.; Gao, Y.; Tishgarten, T.; Clark, T. & Dickerson, H. Surface display of a parasite antigen in the ciliate *Tetrahymena thermophila*. *Nat Biotechnol*, 1999, 17, 462-5.

Kulakosky, P.; Hughes, P. & Wood, H. N-Linked glycosylation of a baculovirus-expressed recombinant glycoprotein in insect larvae and tissue culture cells. *Glycobiology*, 1998, 8, 741-5.

Reading, P.; Miller, J. & Anders, E. Involvement of the mannose receptor in infection of macrophages by influenza virus. J Virol, 2000, 74, 5190-7.

5 Tomiya, N.; Narang, S.; Lee, Y.C. & Betenbaugh, M.J. Comparing N-glycan processing in mammalian cell lines to native and engineered lepidopteran insect cell lines. Glycoconj J, 2004, 21, 343-60.

Kyte, J.; Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 157, 105, 1982.

10 LISTADO DE SECUENCIAS

<110> Cilian AG

15 <120> Método para la expresión de una proteína viral en un protozoo

<130> CD 40333

<150> EP 0910357.3

20 <151> 17-06-2009

<160> 15

<170> PatentIn versión 3.5

25 <210> 1

<211> 1697

<212> ADN

<213> Virus de la gripe

30 <400> 1

atgaaagcaa aactactggt cctggttatgt acattttacag ctacatatgc agacacaata	60
tgtataggtt accatgccaa caactcaacc gacactgttg acacagtact tgagaagaat	120
gtgacagtga cacactctgt caacctactt gaggacagtc acaatggaaa actatgtcta	180
ctaaaaggaa tagccccact acaattgggt aattgcagcg ttgccggatg gatcttagga	240
aaccagaat gcgaattact gatttccaag gaatcatggt cctacattgt agaaacacca	300
aatcctgaga atggaacatg ttaccagggg tatttcgccg actatgagga actgagggag	360
caattgagtt cagtatcttc atttgagaga ttcgaaatat tccccaaaga aagctcatgg	420
cccaaccaca ccgtaaccgg agtatcagca tcatgctccc ataatgggaa aagcagtttt	480
tacagaaatt tgctatggct gacggggaag aatgggtttgt acccaaacct gagcaagtcc	540
tatgtaaaca acaaagagaa agaagtcctt gtactatggg gtgttcatca cccgcctaac	600
ataggggaacc aaagggccct ctatcataca gaaaatgctt atgtctctgt agtgtcttca	660
cattatagca gaagattcac ccagaaata gccaaaagac ccaaagtaag agatcaggaa	720
ggaagaatca actactactg gactctgctg gaacctgggg atacaataat atttgaggca	780
aatggaaatc taatagcgcc atgggtatgct tttgcaactga gtagaggctt tggatcagga	840
atcatcacct caaatgcacc aatggatgaa tgtgatgcga agtgtcaaac acctcagggg	900
gctataaaca gcagtcttcc ttccagaat gtacaccag tcacaatagg agagtgtcca	960
aagtatgtca ggagtgc aaaattaaggatg gttacaggac taaggaaacat cccatccatt	1020
caatccagag gtttgtttgg agccattgcc ggtttcattg aaggggggtg gactggaatg	1080
gtagatgggt ggtatggtta tcatcatcag aatgagcaag gatctggcta tgctgcagat	1140
caaaaaagta cacaaatgc cattaacggg attacaaaca aggtgaattc tgtaattgag	1200
aaaatgaaca ctcaattcac agctgtgggc aaagaattca acaaattgga aagaaggatg	1260
gaaaacttaa ataaaaaagt tgatgatggg ttcttagaca ttggacaat aatgcagaat	1320
tggttggttct actggaaaat gaaaggactt tggatttcca tgactccaat gtgaagaatc	1380
tgtatgagaa agtaaaaagc caattaaaga ataatgccaa agaaatagga aacgggtgtt	1440
ttgaattcta tcacaagtgt aacaatgaat gcatggagag tgtgaaaaat ggaacttatg	1500
actatccaaa atattccgaa gaatcaaagt taaacaggga gaaaattgat ggagtgaat	1560
tggaatcaat gggagtctat cagattctgg cgatctactc aactgtcgcc agttccctgg	1620
ttcttttgggt ctcctgggg gcaatcagct tctggatgtg ttccaatggg tctttgcagt	1680
gtagaatatg catctga	1697

<210> 2
 <211> 1698
 <212> ADN
 <213> Virus de la gripe

<400> 2

ES 2 627 402 T3

```

atgaaagcta agttattagt tttactttgt actttcactg ctacttatgc tgatactatt      60
tgtattgggt atcacgctaa taattctact gatactgttg atactgtttt agaaaaaat      120
gttactgtta ctcaactctgt taatctttta gaagattctc acaatggtaa attatgttta      180
cttaaaggta ttgctccttt ataacttggg aattgttctg ttgctgggtg gattcttggg      240
aatcctgaat gtgaattact tatttctaaa gaatcttggg cttatattgt tgaaactcct      300
aatcctgaaa atgggtacttg ttatcctggg tatttcgctg attatgaaga attaagagaa      360
taactttctt ctgtttcttc atttgaaaga ttcgaaattt tccctaaaga atcttcttgg      420
cctaatacaca ctgttactgg tgtttctgct tcttgttctc ataatggtaa atcttcattc      480
tatagaaatc ttttatgggt aactggtaaa aatggtttat atcctaactt ttctaaatct      540
tatgttaata ataaggaaaa agaagtttta gttctttggg gtgttcacca cctcctaact      600
attggtaatt aaagagcttt atatcacact gaaaatgctt atgtttctgt tgtttcttct      660
cactattcta gaagattcac tcctgaaatt gctaaaagac ctaaagttag agattaagaa      720
ggtagaatta actattattg gactttatta gaacctgggt atactattat ttctgaagct      780
aatggtaatc ttattgctcc ttggtacgct ttcgctttat ctagagggtt cggttctggt      840
attattactt ctaatgctcc tatggatgaa tgtgatgcta aatgttaaac tccttaaggt      900
gctattaatt cttctttacc tttctaaaac gttcacctg ttactattgg tgaatgtcct      960
aaatatgtta gatcagctaa attaagaatg gttactgggt taagaaatat tccttctatc     1020
taatctagag gacttttcgg tgctattgct ggtttcattg agggaggatg gacaggtatg     1080
gttgatgggt ggtacggta tcaccaccaa aatgaataag gttctgggta tgctgctgat     1140
taaaagtcta ctcaaacgc tattaatggg attactaaca aagttaattc tgttattgaa     1200
aaaaatgaata ctttaattcac tgctgttggg aaagaattca ataagttaga aagaagaatg     1260
gaaaatctta ataagaaagt tgatgatggg ttcttagata tttggactta taatgctgaa     1320

ttattagttc ttttagaaaa tgaaagaact ttagatttcc acgattctaa tgttaaaaat     1380
ctttatgaaa aagttaaatc ttaacttaag aataatgcta aagaaattgg taatggttgt     1440
ttcgaattct atcacaaatg taataatgaa tgtatggaat ctgtaagaa tgggtacttat     1500
gattatccta aatattctga agaactctag ttaaataagag aaaaaattga tgggtgttaa     1560
ttagaatcta tgggagtta ctaaattcct gctatctatt ctactgttgc ttcttcttta     1620
gttttattag tttctttagg tgctatttct ttctggatgt gttctaattg ttctttataa     1680
tgtagaattt gtatttga                                     1698

```

<210> 3
 <211> 550
 <212> PRT
 <213> Virus de la gripe

<400> 3

ES 2 627 402 T3

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

Leu Ser Lys Ser Tyr Val Asn Asn Lys Glu Lys Glu Val Leu Val Leu
 180 185 190

Trp Gly Val His His Pro Pro Asn Ile Gly Asp Gln Arg Ala Leu Tyr
 195 200 205

His Thr Glu Asn Ala Tyr Val Ser Val Val Ser Ser His Tyr Ser Arg
 210 215 220

Arg Phe Thr Pro Glu Ile Ala Lys Arg Pro Lys Val Arg Asp Gln Glu
 225 230 235 240

Gly Arg Ile Asn Tyr Tyr Trp Thr Leu Leu Glu Pro Gly Asp Thr Ile
 245 250 255

Ile Phe Glu Ala Asn Gly Asn Leu Ile Ala Pro Trp Tyr Ala Phe Ala
 260 265 270

Leu Ser Arg Gly Phe Gly Ser Gly Ile Ile Thr Ser Asn Ala Pro Met
 275 280 285

Asp Glu Cys Asp Ala Lys Cys Gln Thr Pro Gln Gly Ala Ile Asn Ser
 290 295 300

Ser Leu Pro Phe Gln Asn Val His Pro Val Thr Ile Gly Glu Cys Pro
 305 310 315 320

Lys Tyr Val Arg Ser Ala Lys Leu Arg Met Val Thr Gly Leu Arg Asn
 325 330 335

Ile Pro Ser Ile Gln Ser Arg Gly Leu Phe Gly Ala Ile Ala Gly Phe
 340 345 350

Ile Glu Gly Gly Trp Thr Gly Met Val Asp Gly Trp Tyr Gly Tyr His
 355 360 365

His Gln Asn Glu Gln Gly Ser Gly Tyr Ala Ala Asp Gln Lys Ser Thr
 370 375 380

Gln Asn Ala Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu
 385 390 395 400

Lys Met Asn Thr Gln Phe Thr Ala Val Gly Lys Glu Phe Asn Lys Leu
 405 410 415

Glu Arg Arg Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu
 420 425 430

Asp Ile Trp Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu
435 440 445

Arg Thr Leu Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys
450 455 460

Val Lys Ser Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys
465 470 475 480

Phe Glu Phe Tyr His Lys Cys Asn Asn Glu Cys Met Glu Ser Val Lys
485 490 495

Asn Gly Thr Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn
500 505 510

Arg Glu Lys Ile Asp Gly Val Lys Leu Glu Ser Met Gly Val Tyr Gln
515 520 525

Ile Leu Ala Ile Tyr Ser Thr Val Ala Ser Ser Leu Val Leu Leu Val
530 535 540

Ser Leu Gly Ala Ile Ser
545 550

<210> 4

<211> 1604

<212> ADN

<213> Virus de la gripe

<400> 4

```

ggcgcgccat ggatatcaaa gctaagttat tagttttact ttgtactttc actgctactt      60
atgctgatac tatttgtatt gggtatcacg ctaataattc tactgatact gttgatactg      120
ttttagaaaa aaatgttact gttactcact ctgttaatct tttagaagat tctcacaatg      180
gtaaattatg tttacttaaa ggtattgctc ctttataact tggtaattgt tctgttgctg      240
gttggattct tggtaatcct gaatgtgaat tacttatttc taaagaatct tgggtctata      300
ttgttgaaac tcctaactct gaaaatggta cttgttatcc tggttatttc gctgattatg      360
aagaattaag agaataactt tcttctgttt cttcatttga aagattcgaa attttccta      420
aagaatcttc ttggcctaata cacactgtta ctgggtgttc tgcttcttgt tctcataatg      480
gtaaatcttc attctataga aatcttttat ggttaactgg taaaaatggg ttatatccta      540
atctttctaa atcttatgtt aataataagg aaaaagaagt tttagttctt tgggggtgtc      600
accacccctc taatattggg aattaaagag ctttatatca cactgaaaat gcttatgttt      660
ctgttggttc ttctcactat tctagaagat tcactcctga aattgctaaa agacctaaag      720
ttagagatta agaaggtaga attaactatt attggacttt attagaacct ggtgatacta      780
ttattttcga agctaattgg aatcttattg ctcttggtga cgctttcgct ttatctagag      840

gtttcgggtc tggattattt acttctaatt ctctatgga tgaatgtgat gctaaatgtt      900
aaactcctta aggtgctatt aattcttctt tacctttcta aaacgttcac cctgttacta      960
ttggtgaatg tcctaaatat gttagatcag ctaaattaag aatggttact ggtttaagaa     1020
atattccttc tatctaactc agaggacttt tgggtgctat tgctgggttc attgaggagag     1080
gatggacagg tatggttgat ggttggtacg gttatcacca ccaaatgaa taagggtctg      1140
gttatgctgc tgattaaaag tctactcaaa acgctattaa tggattactt aacaaagtta      1200
attctgttat tgaaaaaatg aatacttaat tcactgctgt tggtaaagaa ttcaataagt      1260
tagaaagaag aatggaaaat ctttaataaga aagttgatga tggtttctta gatatttgga      1320
cttataatgc tgaattatta gttcttttag aaaatgaaag aacttttagat ttccacgatt      1380
ctaattgtaa aaatctttat gaaaaagtta aatcttaact taagaataat gctaaagaaa      1440
ttggtaatgg ttgtttcgaa ttctatcaca aatgtaataa tgaatgtatg gaatctgtta      1500
agaatggtag ttatgattat cctaaatatt ctgaagaatc taagttaaag agagaaaaaa      1560
ttgatgggtg taaattagaa tctatgggag tttactaaat ttga                          1604

```

<210> 5
 <211> 538
 <212> PRT
 <213> Virus de la gripe

 <400> 5

ES 2 627 402 T3

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

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

Gln Asn Ala Ile Asn Gly Ile Thr Asn Lys Val Asn Ser Val Ile Glu
385 390 395 400

Lys Met Asn Thr Gln Phe Thr Ala Val Gly Lys Glu Phe Asn Lys Leu
405 410 415

Glu Arg Arg Met Glu Asn Leu Asn Lys Lys Val Asp Asp Gly Phe Leu
420 425 430

Asp Ile Trp Thr Tyr Asn Ala Glu Leu Leu Val Leu Leu Glu Asn Glu
435 440 445

Arg Thr Leu Asp Phe His Asp Ser Asn Val Lys Asn Leu Tyr Glu Lys
450 455 460

Val Lys Ser Gln Leu Lys Asn Asn Ala Lys Glu Ile Gly Asn Gly Cys
465 470 475 480

Phe Glu Phe Tyr His Lys Cys Asn Asn Glu Cys Met Glu Ser Val Lys
485 490 495

Asn Gly Thr Tyr Asp Tyr Pro Lys Tyr Ser Glu Glu Ser Lys Leu Asn
500 505 510

Arg Glu Lys Ile Asp Gly Val Lys Leu Glu Ser Met Gly Val Tyr Gln
515 520 525

Ile Leu Ala Ile Tyr Ser Thr Val Ala Ser
530 535

<210> 6

<211> 1032

<212> ADN

<213> Virus de la grippe

<400> 6

atcaaagcta agttattagt tttactttgt actttcactg ctacttatgc tgatactatt	60
tgtattgggt atcacgctaa taattctact gatactgttg atactgtttt agaaaaaaat	120
gttactgtta ctcaactctgt taatctttta gaagattctc acaatggtaa attatgttta	180
cttaaaggta ttgctccttt ataacttggg aattgttctg ttgctgggtg gattcttggg	240
aatcctgaat gtgaattact tatttctaaa gaatcttggg cttatatattg tgaaactcct	300
aatcctgaaa atgggtacttg ttatcctggg tatttcgctg attatgaaga attaagagaa	360
taactttctt ctgtttcttc atttgaaaga ttogaaattt tccctaaaga atcttcttgg	420
cctaatacaca ctgttactgg tgtttctgct tcttgttctc ataatggtaa atcttcattc	480
tatagaaatc ttttatgggt aactggtaaa aatggtttat atcctaattc ttctaaatct	540
tatgttaata ataaggaaaa agaagtttta gttctttggg gtgttcacca ccctccta	600
attggtaatt aaagagcttt atatcacact gaaaatgctt atgtttctgt tgtttcttct	660
cactattcta gaagattcac tcctgaaatt gctaaaagac ctaaagtag agattaagaa	720
ggtagaatta actattattg gactttatta gaacctgggtg atactattat ttctgaagct	780
aatggtaatc ttattgctcc ttggtacgct ttogctttat ctagagggtt cggttctggt	840
attattactt ctaatgctcc tatggatgaa tgtgatgcta aatgttaaac tccttaaggt	900
gctattaatt cttctttacc tttctaaaac gttcacctg ttactattgg tgaatgtcct	960
aaatatgtta gatcagctaa attaagaatg gttactgggt taagaaatat tccttctatc	1020
taatctagat ga	1032

<210> 7

<211> 343

<212> PRT

<213> Virus de la gripe

<400> 7

ES 2 627 402 T3

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

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

<210> 8

<211> 328

<212> ADN

<213> *Tetrahymena thermophila*

<400> 8

```

atgaacaaga ctctcatctt agcttttagtt ggtggttttgg ctttaactgc caccaccttg      60
gttgcttttcc acaaccactc tcacaacatc agagttgact aagaccccgcc cactctcttc      120
aagcaattca agcaaaactta caataagaag tatgctgatc ctactttcga aacctacaga      180
ttcgggtgtct tcacccaaaa cttagaaatc gtcaagactg actctacttt cgggtgtcacc      240
taattcatgg acttaactcc tgctgaattc gctcaacaat tcttcacttt acacgaaaag      300

gttaacagca ccgaagttaa cagagctc      328

```

<210> 9
 <211> 113
 <212> PRT
 <213> *Tetrahymena thermophila*
 <400> 9

```

Met Asn Lys Thr Leu Ile Leu Ala Leu Val Val Val Leu Ala Leu Thr
 1              5              10              15

Ala Thr Thr Leu Val Ala Phe His Asn His Ser His Asn Ile Arg Val
      20              25              30

Asp Gln Asp Pro Ala Thr Leu Phe Lys Gln Phe Lys Gln Thr Tyr Asn
      35              40              45

Lys Lys Tyr Ala Asp Ala Thr Phe Glu Thr Tyr Arg Phe Gly Val Phe
      50              55              60

Thr Gln Asn Leu Glu Ile Val Lys Thr Asp Ser Thr Phe Gly Val Thr
65              70              75              80

Gln Phe Met Asp Leu Thr Pro Ala Glu Phe Ala Gln Gln Phe Leu Thr
      85              90              95

Leu His Glu Lys Val Asn Ser Thr Glu Val Tyr Arg Ala Gln Gly Glu
      100             105             110

```

Ala

<210> 10
 <211> 65
 <212> ADN
 <213> Virus de la gripe
 <400> 10

```

ggcgcgccat ggatatcaaa gctaagttat tagtttttact ttgtactttc actgctactt      60
atgct      65

```

<210> 11

<211> 17
 <212> PRT
 <213> Virus de la gripe

5 <400> 11

Met Lys Ala Lys Leu Leu Val Leu Leu Cys Thr Phe Thr Ala Thr Tyr
 1 5 10 15

Ala

10 <210> 12
 <211> 980
 <212> ADN
 <213> *Tetrahymena thermophila*

15 <400> 12

```

agcatgcttt ttcatgtact attcctaact atagcttaat gctttatata tcaaagtatg      60
aagaagacag gttttaaaaa accaaaaaat atgtttttta ttgaaaagt attatatagg      120
aataataaat gctgcataaa tactctaagt agttgataaa aaatctattt atgcaaagaa      180
aaatttagaa aaatttagaa aaacaaagaa aataattggt attataaagc attttgttta      240
ctaagcaaaa gatataaaat ttgagataga aatatacata caagattaac gttttgtttg      300
tgctttgaaa attgaaatat ttaattatta aatctgctaa cttttttaga tatttatttg      360
ctttatttta ttttttaaat ttttgcaaag tggagaaaaa tgaacaatc aatctttttt      420
ttataaataa tataaatggt attaagccta atttttattg ctggagagtg ttattcaaatt      480
atattgctga atgtggctag atggaattcg ctttggaagg aaagtgttta taaaataagt      540
gatgtattat agcacattgc taattattat aaagaatgta ttggatattt aaaaattaga      600
aaataaaatt tagctgaaaa gtaagaaagc aagcaaagat atatatatat atatggagat      660
gataaaaaaga taaattcgaa aaaagaaaat ttctaaagtg aaaagaatta tggatttgat      720
taaataaaaa tattttttaga atgggctgat taaagaggga tcttcgagaa tgaaatgatt      780
tagaaaaaaa gaaagaaaga ttataacatc tacaagagt tgaagattct agaagaggag      840
gaataattag ctatcagtct tattaaaaga tatcgcaaaa caagaaatat ttttgaaatt      900
aattaaaaaa tttaaaaaac aaaagataaa aattttgcac aaaaaagcaa ttaattaaaa      960
aaaaagatat atcataagaa                                     980
    
```

20 <210> 13
 <211> 998
 <212> ADN
 <213> *Tetrahymena thermophila*
 <400> 13

togaggataa gtaatatatt tagtgcacaa tgtttgaatg tttttttgat aaatttgtaa	60
attaaatcta ttttaagtcag tagaatttta gatattttca tagatttttt aataagtaga	120
caattttatt ctaaaaaata tttaaaaata aaaaataata agggttttga ataactcctt	180
taattttaat acacattttt aaattttttt tagctcttta aatattcata aaaataaaaa	240
ataactaact aaaaataaat aaaaagataa taatgattaa aggtataata ctgaataaga	300
aaaaacataa tagagtactt atttttttata tcactatttt taatatcttg aaagcaaaac	360
ttttttatat atcttaaaat atattgtatc gtttattcaa ttattttctt taaatttcaa	420
atatattgat aaaaagatg acatgttttt taaagaaaac atgaaatata aaatagataa	480
atatcaatta ttttatttat taaatatata agctgctcaa aacatagctc attcatcaat	540
tataatatgt gaatcattaa ttttcaaaat attactcatt atttaggcta tcattttattt	600
tttattttca attatccgtt tctatttatat tttaatatta agttgtgatt cttgaattttt	660
gtgtcatgaa ttatttgtaa atctttttat ttctgataaa aaatataaat tgattgactc	720
atgatttaaa tcatgagtca acctaaactaa ttttcaaaat tcttctattc taaaatatag	780
atgtgattct tgaatctctc ttgaatataa agtaattttt tataatttctg atataattct	840
tagctacgtg attcacgatt tatgcaatga tccatataaa ataatgtaa tagtgtatat	900
atatatattc gtctttttta ttcttttatat aatttaaaaa aatttaaaaa atttaataaa	960
gctctaataa aataaataat aataactaaac ttaaaata	998

<210> 14
 <211> 1722
 <212> ADN
 <213> Virus de la gripe

 <400> 14

atgaaggcaa taattgtact actcatggta gtaacatcca atgcagatcg aatctgcact	60
ggaataacat cttcaaactc acctcatgtg gtcaaaacag ccactcaagg ggaggtcaat	120
gtgactgggtg tgataccact aacaacaaca ccaacaaaat cttattttgc aaatctcaaa	180
ggaacaagga ccagagggaa actatgccca gactgtctca actgcacaga tctggatgtg	240
gctttgggca gaccaatgtg tgtggggacc acaccttcgg cgaaagcttc aatactccac	300
gaagtcaaac ctgttacatc cgggtgcttt cctataatgc acgacagaac aaaaatcagg	360
caactacca atcttctcag aggatatgaa aatatcaggc tatcaacca aaacgtcatc	420
gatgcggaaa aggcaccagg aggacctac agacttggaa cctcaggatc ttgccctaac	480
gctaccagta agagcggatt ttctgcaaca atggcttggg ctgtcccaaa ggacaacaac	540
aaaaatgcaa cgaaccctc aacagtagaa gtaccataca tttgtacaga aggggaagac	600
caaatcactg tttgggggtt ccattcagat gacaaaaccc aatgaagaa cctctatgga	660
gactcaaac ctcaaaagtt cacctcatct gctaattggag taaccacaca ctatgtttct	720
cagattggca gcttcccaga tcaaacagaa gacggaggac taccacaaag cggcaggatt	780
gttgttgatt acatgatgca aaaacctggg aaaacaggaa caattgtcta ccaaagaggt	840
gttttgttgc ctcaaaaggt gtggtgcgcg agtggcagga gcaaagtaat aaaagggccc	900
ttgcctttaa ttggtgaagc agattgcctt catgaaaaat acggtggatt aaacaaaagc	960
aagccttact acacaggaga acatgcaaaa gccataggaa attgcccaat atgggtgaaa	1020
acacctttga agcttgccaa tggaacaaaa tatagacctc ctgcaaaact attaaaggaa	1080
aggggtttct tcggagctat tgctggttct ctagaaggag gatgggaagg aatgattgca	1140
ggctggcacg gatacacatc tcacggagca catggagtgg cagtggcggc ggaccttaag	1200
agtacgcaag aagctataaa caagataaca aaaaatctca attctttgag tgagctagaa	1260
gtaaagaatc ttcaaagact aagtgggtgc atggatgaac tccacaacga aatactcgag	1320
ctggatgaga aagtggatga tctcagagct gacactataa gctcgcaaat agaacttgca	1380
gtcttgcttt ccaacgaagg aataataaac agtgaagatg agcatctatt ggcacttgag	1440
agaaaactaa agaaaatgct gggccctct gctgtagaga taggaaatgg atgcttcgaa	1500
accaaacaca agtgcaacca gacctgctta gacaggatag ctgctggcac ctttaatgca	1560
ggagaatttt ctctccccac ttttgattca ctgaacatta ctgctgcac ttttaatgat	1620
gatggattgg ataaccatac tatactgctc tattactcaa ctgctgcttc tagtttggct	1680
gtaacattga tgctagctat ttttattgtt tatatggtct cc	1720

<210> 15
 <211> 574
 <212> PRT
 <213> Virus de la gripe

<400> 15

ES 2 627 402 T3

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

Ala Thr Ser Lys Ser Gly Phe Phe Ala Thr Met Ala Trp Ala Val Pro
 165 170 175
 Lys Asp Asn Asn Lys Asn Ala Thr Asn Pro Leu Thr Val Glu Val Pro
 180 185 190
 Tyr Ile Cys Thr Glu Gly Glu Asp Gln Ile Thr Val Trp Gly Phe His
 195 200 205
 Ser Asp Asp Lys Thr Gln Met Lys Asn Leu Tyr Gly Asp Ser Asn Pro
 210 215 220
 Gln Lys Phe Thr Ser Ser Ala Asn Gly Val Thr Thr His Tyr Val Ser
 225 230 235 240
 Gln Ile Gly Ser Phe Pro Asp Gln Thr Glu Asp Gly Gly Leu Pro Gln
 245 250 255
 Ser Gly Arg Ile Val Val Asp Tyr Met Met Gln Lys Pro Gly Lys Thr
 260 265 270
 Gly Thr Ile Val Tyr Gln Arg Gly Val Leu Leu Pro Gln Lys Val Trp
 275 280 285
 Cys Ala Ser Gly Arg Ser Lys Val Ile Lys Gly Ser Leu Pro Leu Ile
 290 295 300
 Gly Glu Ala Asp Cys Leu His Glu Lys Tyr Gly Gly Leu Asn Lys Ser
 305 310 315 320
 Lys Pro Tyr Tyr Thr Gly Glu His Ala Lys Ala Ile Gly Asn Cys Pro
 325 330 335
 Ile Trp Val Lys Thr Pro Leu Lys Leu Ala Asn Gly Thr Lys Tyr Arg
 340 345 350
 Pro Pro Ala Lys Leu Leu Lys Glu Arg Gly Phe Phe Gly Ala Ile Ala
 355 360 365
 Gly Phe Leu Glu Gly Gly Trp Glu Gly Met Ile Ala Gly Trp His Gly
 370 375 380
 Tyr Thr Ser His Gly Ala His Gly Val Ala Val Ala Ala Asp Leu Lys
 385 390 395 400
 Ser Thr Gln Glu Ala Ile Asn Lys Ile Thr Lys Asn Leu Asn Ser Leu
 405 410 415

Ser	Glu	Leu	Glu	Val	Lys	Asn	Leu	Gln	Arg	Leu	Ser	Gly	Ala	Met	Asp	
			420					425					430			
Glu	Leu	His	Asn	Glu	Ile	Leu	Glu	Leu	Asp	Glu	Lys	Val	Asp	Asp	Leu	
		435					440					445				
Arg	Ala	Asp	Thr	Ile	Ser	Ser	Gln	Ile	Glu	Leu	Ala	Val	Leu	Leu	Ser	
	450					455					460					
Asn	Glu	Gly	Ile	Ile	Asn	Ser	Glu	Asp	Glu	His	Leu	Leu	Ala	Leu	Glu	
465					470					475					480	
Arg	Lys	Leu	Lys	Lys	Met	Leu	Gly	Pro	Ser	Ala	Val	Glu	Ile	Gly	Asn	
				485					490					495		
Gly	Cys	Phe	Glu	Thr	Lys	His	Lys	Cys	Asn	Gln	Thr	Cys	Leu	Asp	Arg	
			500					505					510			
Ile	Ala	Ala	Gly	Thr	Phe	Asn	Ala	Gly	Glu	Phe	Ser	Leu	Pro	Thr	Phe	
		515					520					525				
Asp	Ser	Leu	Asn	Ile	Thr	Ala	Ala	Ser	Leu	Asn	Asp	Asp	Gly	Leu	Asp	
	530					535					540					
Asn	His	Thr	Ile	Leu	Leu	Tyr	Tyr	Ser	Thr	Ala	Ala	Ser	Ser	Leu	Ala	
545					550					555					560	
Val	Thr	Leu	Met	Leu	Ala	Ile	Phe	Ile	Val	Tyr	Met	Val	Ser			
				565					570							

REIVINDICACIONES

1. Un sistema para la expresión heteróloga de una proteína de la superficie viral para su uso en o como una vacuna, comprendiendo dicho sistema
 - a) una célula huésped de ciliado
 - b) al menos una molécula de ácido nucleico que comprende una secuencia de ácidos nucleicos que codifica la proteína de la superficie viral, y
 - c) un promotor operativamente unido a dicha molécula de ácido nucleico.
2. El sistema según la reivindicación 1, que comprende además
 - d) una secuencia señal operativamente unida a dicho ácido nucleico, secuencia señal que representa la secreción de la proteína de la superficie del virus, codificada por dicho ácido nucleico, en el medio extracelular.
3. El sistema según cualquiera de las reivindicaciones anteriormente mencionadas, caracterizado por que dicho ciliado transgénico es un miembro de la familia *Tetrahymenidae*.
4. El sistema según cualquiera de las reivindicaciones anteriormente mencionadas, caracterizado por que la proteína de la superficie viral es una proteína de fusión viral, preferentemente al menos una seleccionada del grupo que consiste en proteína de fusión viral de clase I, clase II o clase III.
5. El sistema según cualquiera de las reivindicaciones anteriormente mencionadas, caracterizado por que la proteína de la superficie viral, codificada por dicho ácido nucleico, carece de al menos un dominio hidrófobo, o de una fracción del mismo.
6. Una célula huésped de ciliado transfectada con un vector que comprende una molécula de ácido nucleico que codifica una proteína de la superficie viral para su uso en o como una vacuna.
7. Un proceso para la producción de al menos una proteína de la superficie del virus en una célula huésped de ciliado para su uso en o como una vacuna, comprendiendo dicho proceso las etapas de
 - a) transfectar la célula huésped de ciliado con el vector según la reivindicación 6, y
 - b) cultivar la célula huésped en condiciones que permiten la expresión de una proteína.
8. El proceso según la reivindicación 7, caracterizado porque comprende además una etapa de c) cultivar la célula huésped en condiciones que permiten la secreción de la proteína expresada de una proteína.
9. Un método para la producción de una composición farmacéutica, comprendiendo dicho método el proceso según la reivindicación 7.
10. Una composición farmacéutica que comprende una proteína producida con un método según la reivindicación 9, composición que es una vacuna.

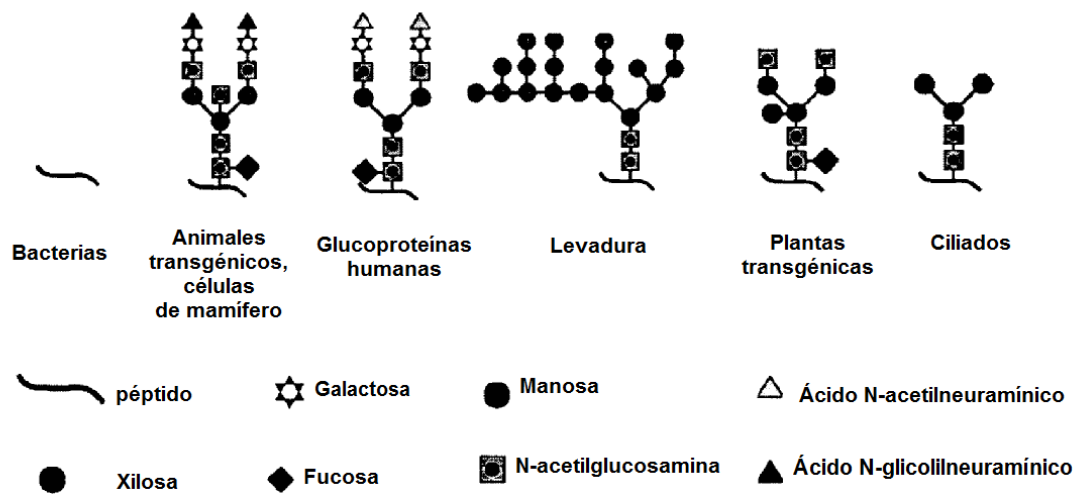


Fig. 1A

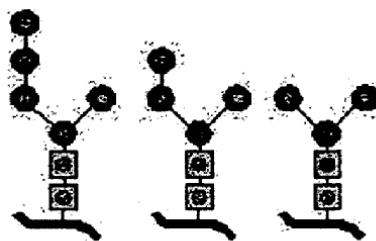


Fig. 1B

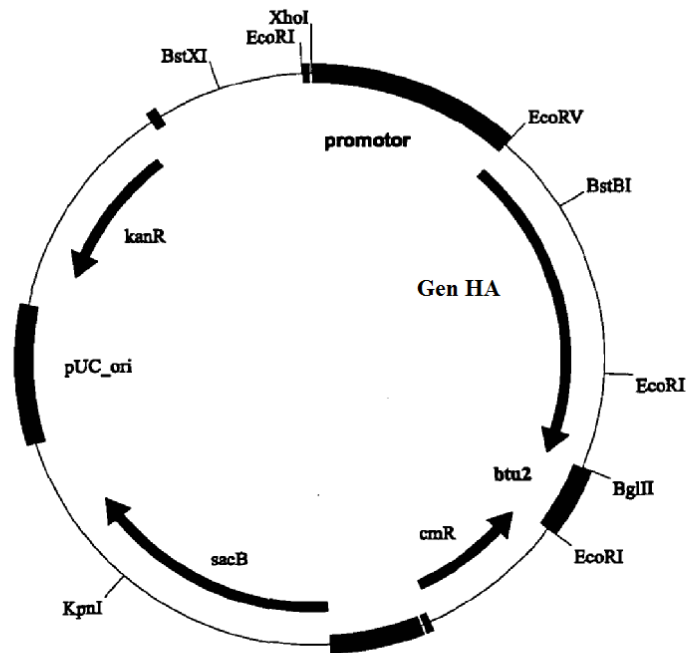


Fig. 2A

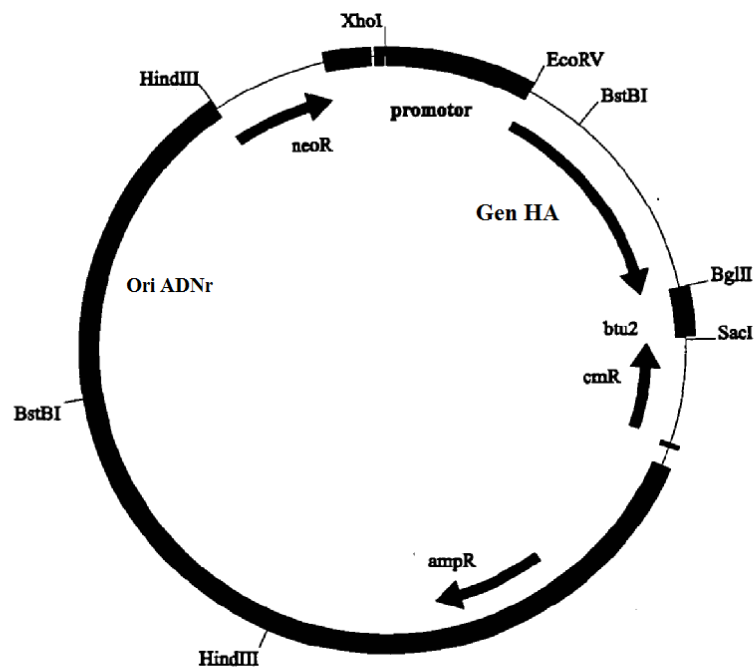


Fig. 2B

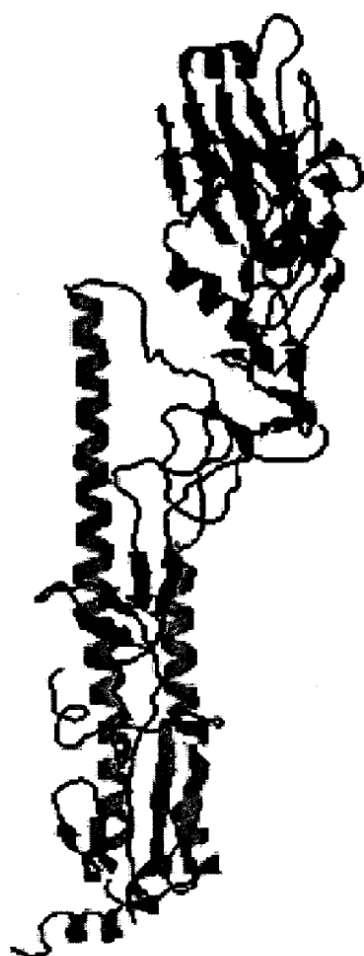


Fig. 3A

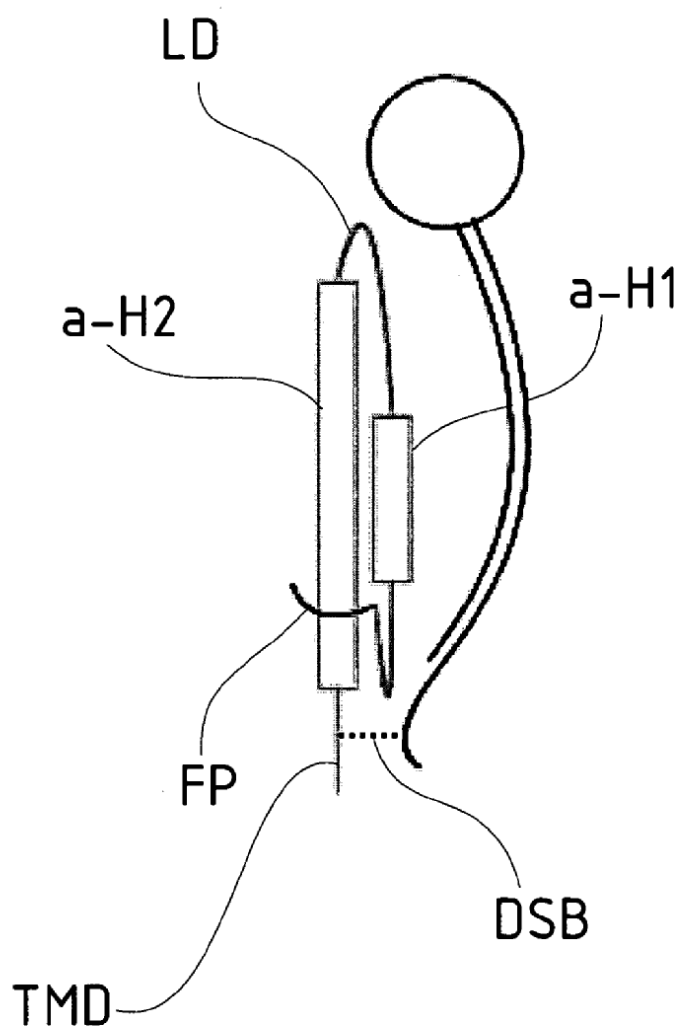


Fig. 3B

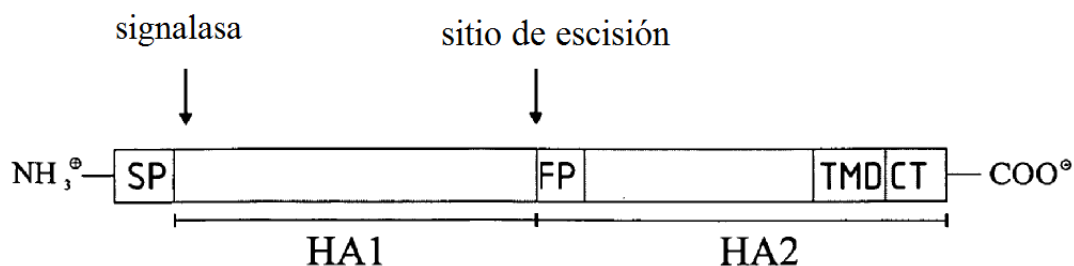


Fig. 4A

Dominio	AA	Símbolo
Péptido señal	1 - 17	SP
Dominio extracelular de HA1	18 - 343	
Péptido de fusión de HA2	344 - 368	FP
Dominio extracelular de HA2	368 - 529	
Dominio transmembranario de HA2	530 - 550	TMD
Dominio intracelular de HA2	551 - 565	CT

```

      10          20          30          40          50          60
MKAKLLVLLC TFTATYADTI CIGYHANNST DTVDTVLEKN VTVTHSVNLL EDSHNGKLCL

      70          80          90         100         110         120
LKGIAPLQLG NCSVAGWILG NPECELLISK ESWSYIVETP NPENGTCYPG YFADYEELRE

      130         140         150         160         170         180
QLSSVSSFER FEIFPKESSW PNHTVTGVSA SCSHNGKSSF YRNLLWLTGK NGLYPNLSKS

      190         200         210         220         230         240
YVNNKEKEVL VLWGVVHPPN IGNQRALYHT ENAYVSVVSS HYSRRFTPEI AKRPKVRDQE

      250         260         270         280         290         300
GRINYWTLL EPGDTIIFEA NGNLIAPWYA FALSRGFGSG IITSNAPMDE CDAKCQTPQG

      310         320         330         340         350         360
AINSSLPFQN VHPVTIGECF KYVRS AKLRM VTGLRNIPSI QSRGLFGAIA GFIEGGWTGM

      370         380         390         400         410         420
VDGWYGYHHQ NEQSGGYAAD QKSTQNAING ITNKVNSVIE KMNTQFTAVG KEFNKLERRM

      430         440         450         460         470         480
ENLNKKVDDG FLDIWTYNAE LLVLENERT LDFHDSNVKN LYEKVKSQK NNAKEIGNGC

      490         500         510         520         530         540
FEFYHKCNE CMESVKNGTY DYPKYSEESK LNREKIDGVK LESMGVYQIL AIYSTVASSL

      550         560
VLLVSLGAIS FWMCSNGSLQ CRICI

```

Fig. 4B

***Tetrahymena thermophila* [gbinv]: 283 CDS (145922 codones)**

campos: [triplete] [frecuencia: por mil] ([número])

UUU 26,1 (3815)	UCU 24,4 (3557)	UAU 23,3 (3407)	UGU 9,7 (1412)
UUC 19,4 (2827)	UCC 6,5 (948)	UAC 14,5 (2110)	UGC 8,8 (1282)
UUA 29,8 (4346)	UCA 16,8 (2453)	UAA 36,8 (5366)	UGA 2,0 (286)
UUG 14,1 (2054)	UCG 1,5 (222)	UAG 11,0 (1606)	UGG 7,4 (1080)
CUU 20,3 (2955)	CCU 17,6 (2574)	CAU 8,7 (1267)	CGU 4,6 (677)
CUC 10,3 (1497)	CCC 4,6 (676)	CAC 6,4 (930)	CGC 0,9 (136)
CUA 7,4 (1078)	CCA 8,2 (1202)	CAA 19,8 (2894)	CGA 0,5 (73)
CUG 2,6 (378)	CCG 0,5 (68)	CAG 3,3 (477)	CGG 0,1 (8)
AUU 39,3 (5733)	ACU 27,2 (3968)	AAU 48,0 (7002)	AGU 13,5 (1963)
AUC 16,2 (2367)	ACC 7,8 (1140)	AAC 24,2 (3530)	AGC 9,2 (1344)
AUA 19,1 (2783)	ACA 14,8 (2153)	AAA 58,7 (8562)	AGA 26,6 (3887)
AUG 19,3 (2811)	ACG 0,8 (111)	AAG 34,3 (5001)	AGG 2,8 (412)
GUU 25,8 (3763)	GCU 30,3 (4428)	GAU 42,5 (6208)	GGU 24,5 (3576)
GUC 10,1 (1469)	GCC 7,5 (1098)	GAC 12,4 (1815)	GGC 4,3 (629)
GUA 11,6 (1693)	GCA 11,8 (1726)	GAA 58,2 (8499)	GGA 15,1 (2205)
GUG 3,1 (451)	GCG 0,6 (88)	GAG 11,2 (1630)	GGG 1,5 (216)

Codificación GC 32,53 % 1ª letra GC 38,64 % 2ª letra GC 31,25 % 3ª letra GC 27,69 %

***Homo sapiens* [gbpri]: 93487 CDS (40662582 codones)**

campos: [triplete] [frecuencia: por mil] ([número])

UUU 17,6 (714298)	UCU 15,2 (618711)	UAU 12,2 (495699)	UGU 10,6 (430311)
UUC 20,3 (824692)	UCC 17,7 (718892)	UAC 15,3 (622407)	UGC 12,6 (513028)
UUA 7,7 (311881)	UCA 12,2 (496448)	UAA 1,0 (40285)	UGA 1,6 (63237)
UUG 12,9 (525688)	UCG 4,4 (179419)	UAG 0,8 (32109)	UGG 13,2 (535595)
CUU 13,2 (536515)	CCU 17,5 (713233)	CAU 10,9 (441711)	CGU 4,5 (184609)
CUC 19,6 (796638)	CCC 19,8 (804620)	CAC 15,1 (613713)	CGC 10,4 (423516)
CUA 7,2 (290751)	CCA 16,9 (688038)	CAA 12,3 (501911)	CGA 6,2 (250760)
CUG 39,6 (1611801)	CCG 6,9 (281570)	CAG 34,2 (1391973)	CGG 11,4 (464485)
AUU 16,0 (650473)	ACU 13,1 (533609)	AAU 17,0 (689701)	AGU 12,1 (493429)
AUC 20,8 (846466)	ACC 18,9 (768147)	AAC 19,1 (776603)	AGC 19,5 (791383)
AUA 7,5 (304565)	ACA 15,1 (614523)	AAA 24,4 (993621)	AGA 12,2 (494682)
AUG 22,0 (896005)	ACG 6,1 (246105)	AAG 31,9 (1295568)	AGG 12,0 (486463)
GUU 11,0 (448607)	GCU 18,4 (750096)	GAU 21,8 (885429)	GGU 10,8 (437126)
GUC 14,5 (588138)	GCC 27,7 (1127679)	GAC 25,1 (1020595)	GGC 22,2 (903565)
GUA 7,1 (287712)	GCA 15,8 (643471)	GAA 29,0 (1177632)	GGA 16,5 (669873)
GUG 28,1 (1143534)	GCG 7,4 (299495)	GAG 39,6 (1609975)	GGG 16,5 (669768)

Codificación GC 52,27 % 1ª letra GC 55,72 % 2ª letra GC 42,54 % 3ª letra GC 58,55 %

Fig. 5

1LC	3LC	aminoácido	Codones						
A	Ala	Alanina	GCU	GCA	GCC	GCG			
C	Cys	Cisteína	UGU	UGC					
D	Asp	Ácido aspártico	GAU	GAC					
E	Glu	Ácido glutámico	GAA	GAG					
F	Phe	Fenilalanina	UUU	UUC					
G	Gly	Glicina	GGU	GGA	GGC	GGG			
H	His	Histidina	CAU	CAC					
I	Ile	Isoleucina	AUU	AUA	AUC				
K	Lys	Lisina	AAA	AAG					
L	Leu	Leucina	UUA	CUU	UUG	CUC	CUA	CUG	
M	Met	Metionina / INICIO	AUG						
N	Asn	Asparagina	AAU	AAC					
P	Pro	Prolina	CCU	CCA	CCC	CCG			
Q	Gln	Glutamina	CAA	CAG	UAA	UAG			
R	Arg	Arginina	AGA	CGU	AGG	CGC	CGA	CGG	
S	Ser	Serina	UCU	UCA	AGU	AGC	UCC	UCG	
T	Thr	Treonina	ACU	ACA	ACC	ACG			
V	Val	Valina	GUU	GUA	GUC	GUG			
W	Trp	Triptófano	UGG						
Y	Tyr	Tirosina	UAU	UAC					
		FIN	UGA	UAA	UAG				

Fig. 6

Hemagglutinina
HA larga
1 at gaaagcaaaact act ggtctgtt at gtae attacagat ac at at gtagar aaaa at gt at aggtacac at gta caaaact caa accgact gtt g
1 at gaaagct aagtt at agttttactt gta ctttcaat gtt actt at gctgat act at gt attggtat cagctta ataat tct actgata cttgt g
M K A K L L V L L C I F T A T Y A D T I C I G Y H A N N S T D I V D

Hemagglutinina
HA larga
101 ac ac agt actt gaaagaaatgc gacagt gacacactct gtaacacttctt gaggacagc cac aat gyaaaactat gtt ctt aaaaagaaat agt cccact
101 at actgttt gaaaaaaatgc tactgtt actcactctgttaact ttttt agaaagtt cc cac aat gyaaaactat gtt ctt actt aaaaagat att gttctttt
I V L E K N V T V T H S V N L L E D S H N G K L C L L K G I A P L

Hemagglutinina
HA larga
201 ac aattt ggt aatt gtagcgtt gc cggat ggtatctt agyaaaccc aagaatgc gaatt act gatt ttt aaggaat cat ggtctacat t gtag aaaaact a
201 at aattt ggt aatt gtt ctt gtt gtt gtt gtt ctt gtt aatt ctt gaaatgc gaatt act gatt ttt aaggaat cat ggtctacat t gtag aaaaact a
Q L G N C S V A G W I L G N P R C L L I S K R S W S Y I V X T P

Hemagglutinina
HA larga
301 aatcct gagaat ggaac at gtt ac ctaggtt attt cccgact at gag aact gaggagac aat t gatt tcaat ctt cttt gtag at cgaactat
301 aatcct gagaat ggt act gtt act ctt ggtt attt tttt gtt at gaa aat aagaat aact ttt ttt ttt ctt ttt gtag at cgaactat
N P E N G I C Y P C Y F A D Y E E L H R W S S P I R F E I Y P K E S

Hemagglutinina
HA larga
401 tccc aagaagaaatggtat ggtccaaact ac act gtt aac cggagat cagc acc at gtt cccat aat gyyga aagaagtt ttt aagaat ttt gtt ctt ggtt
401 tccc aagaagaaatggtat ggtccaaact ac act gtt aac cggagat cagc acc at gtt cccat aat gyyga aagaagtt ttt aagaat ttt gtt ctt ggtt
S W P N H Q L S S T V I T G V S A S C S H N G K S S F Y R N L L W L

Hemagglutinina
HA larga
501 gacggg aagaat ggtt ttt acce aaaaact gagg aagtt ctt atgt aaaa aac aagaaga aagaagtt ctt ttt actat ggggt gtt tttt acccgtt aac
501 at tggta at taaagagttt at ac act gtaaat gtt ttt gtt ctt gtagt gtt ttt cact at agagaat cactt ctaaat gaaat tttt aagaat
T G K N G L Y P N L S K S Y V N N K E K E W L V L W G V H H P P N

Hemagglutinina
HA larga
601 at agggaaac aaaaaggtt cttt ac cact agaaaat gtt ttt gtt ctt gtagt gtt ttt cact at agagaat cactt ctaaat gaaat tttt aagaat
601 at tggta at taaagagttt at ac act gtaaat gtt ttt gtt ctt gtagt gtt ttt cact at agagaat cactt ctaaat gaaat tttt aagaat
I G N Q R A L Y H T I N A Y V S V V S S H Y S R R F I P H I A K R P

Hemagglutinina
HA larga
701 ccaaggt aagat caggaagaa gaaatc aactactt ggt ctt gtt ggaac ctt gggat ac aat aat ttt gaggc aat ggaat ctaact agcgt c
701 ccaaggt aagat caggaagaa gaaatc aactactt ggt ctt gtt ggaac ctt gggat ac aat aat ttt gaggc aat ggaat ctaact agcgt c
K V R D Q E C R I N Y Y W I L L R P C D T I I F E A N C N L I A P

Hemagglutinina
HA larga
801 ac ggtat gtt ttt gtagt agt ag aggtt ttt gtagt caggat cactacact caaatgc ac ccaat ggt gaatt gtt gtt ggtt ccaaac ctt cagggt a
801 at tggta at taaagagttt at ac act gtaaat gtt ttt gtt ctt gtagt gtt ttt cact at agagaat cactt ctaaat gaaat tttt aagaat
W Y A F A L S R G Y G S G I Y I S N A P H D B C D A K C Q Y P Q G

Hemagglutinina
HA larga
901 gttat aaaaact gtt ctt cttt cagaaatg ac act cagtt ac aat agggagtt gtt ccaagt at gtt aggt gtt aaaaat aaggt gtt ttt caggat
901 gttat aaaaact gtt ctt cttt cagaaatg ac act cagtt ac aat agggagtt gtt ccaagt at gtt aggt gtt aaaaat aaggt gtt ttt caggat
A I N S S L P F Q N V H P V T I C R C P K Y V R S A K L H M V I G L

Hemagglutinina
HA larga
1001 caaggaacat ccaact ccaact ccaaggtt gtt gtaggtt att gtt ggtt ttaatt gaaaggggtt ggtt ggtt at gtt agat ggtt gtt gtt a
1001 caaggaacat ccaact ccaact ccaaggtt gtt gtaggtt att gtt ggtt ttaatt gaaaggggtt ggtt ggtt at gtt agat ggtt gtt gtt a
R N I P S I Q S R G L Y C A I A G F I E G G W I C M V D G W Y C Y

Hemagglutinina
HA larga
1101 ccaact ccaaggtt gtaggtt ggtt at gtt gtagt ccaaaaggt ac caaaat gttatt aaggggtt ttt aac aaggtt gaaat ctt gttatt ggt
1101 ccaact ccaaggtt gtaggtt ggtt at gtt gtagt ccaaaaggt ac caaaat gttatt aaggggtt ttt aac aaggtt gaaat ctt gttatt ggt
H H Q N E Q G S C Y A A D Q K S T Q N A I N C I T N K V N S V I E

Hemagglutinina
HA larga
1201 aaaaatgaaact ccaact ccaact gtt ggtt ccaaggtt ac aac aaaaat ggaagaggtt ggaagaaact aaaaatgaaact gaaat ggtt ctt ggtt a
1201 aaaaatgaaact ccaact ccaact gtt ggtt ccaaggtt ac aac aaaaat ggaagaggtt ggaagaaact aaaaatgaaact gaaat ggtt ctt ggtt a
K M N T Q F I A V G K E F N K L E R R M E N L N K K V D D C G F L D I

Hemagglutinina
HA larga
1301 tttggat at aat gtagaatt gtt gttt ctaact ggaat gaaaggtt ctt ggtt ctaact ggtt ccaat gtt gtagaatt ctt gtagaatt aaaaatg
1301 tttggat at aat gtagaatt gtt gttt ctaact ggaat gaaaggtt ctt ggtt ctaact ggtt ccaat gtt gtagaatt ctt gtagaatt aaaaatg
W I Y N A E L L V L L E R N E R T L D F H D S N V K N L Y E K V K S

Hemagglutinina
HA larga
1401 tcaatc aagaat at gccaagaat aggaat ggtt gttt gttt tttt ccaact ccaact gaaat gtagaatt gtagaatt gtagaatt gtagaatt gtagaatt
1401 tcaatc aagaat at gccaagaat aggaat ggtt gttt gttt tttt ccaact ccaact gaaat gtagaatt gtagaatt gtagaatt gtagaatt gtagaatt
Q L K N N A K R I C N C C Y R F Y H K C N N E C H K S V K N C T Y

Hemagglutinina
HA larga
1501 ggtt ac ccaaat at ccaagaaat ccaaggtt aaaaaggggaa aat tggat gtagt gaaat ggtt ccaact ggaat gtagaatt ctt gtagaatt ctt gtagaatt
1501 ggtt ac ccaaat at ccaagaaat ccaaggtt aaaaaggggaa aat tggat gtagt gaaat ggtt ccaact ggaat gtagaatt ctt gtagaatt ctt gtagaatt
D Y P K Y S E E S K L N R E K I D C V K L E S H G V Y Q I L A I Y S

Hemagglutinina
HA larga
1601 caactgt cgtcagtt cttt ggtt tttt ggtt cttt gggggc aat agctt ctt ggtt gtt ttt ccaat ggtt cttt gtagt gtagaatt atgt ac cta
1601 caactgt cgtcagtt cttt ggtt tttt ggtt cttt gggggc aat agctt ctt ggtt gtt ttt ccaat ggtt cttt gtagt gtagaatt atgt ac cta
I V A S S L V L L V S L C A I S P W M C S N G S L Q C R I C I F I N

Fig. 7

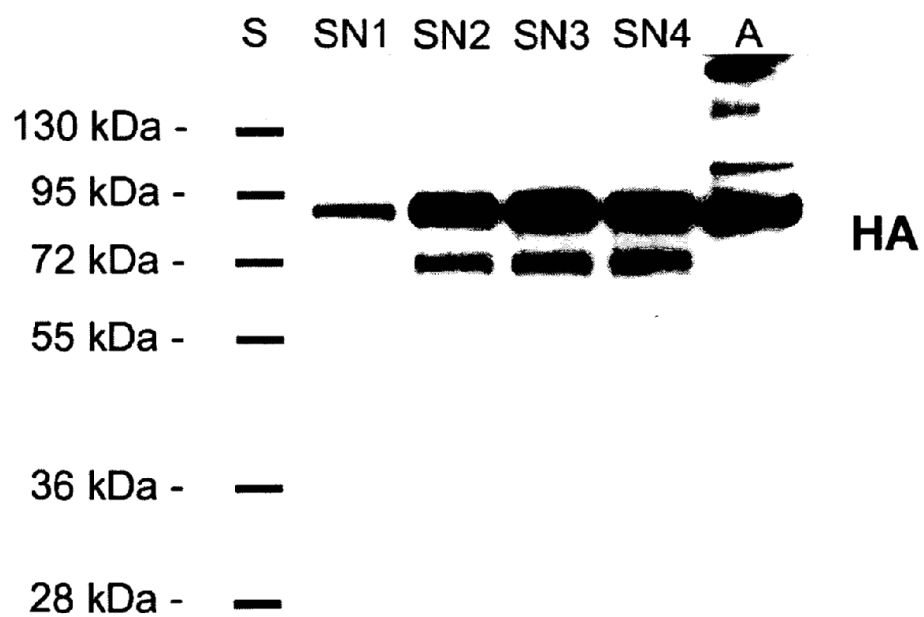


Fig. 8

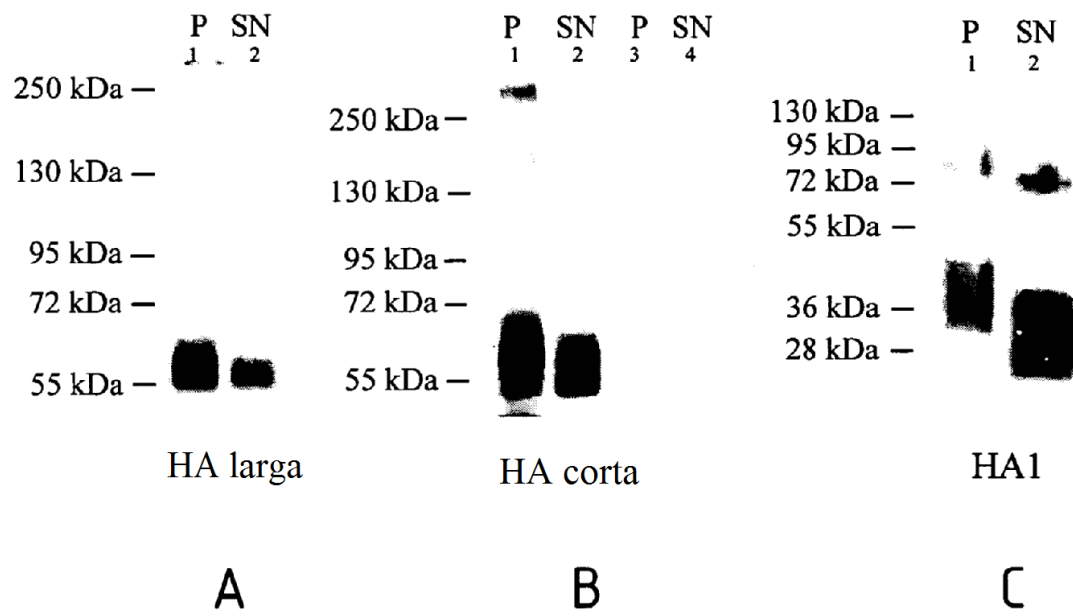


Fig. 9

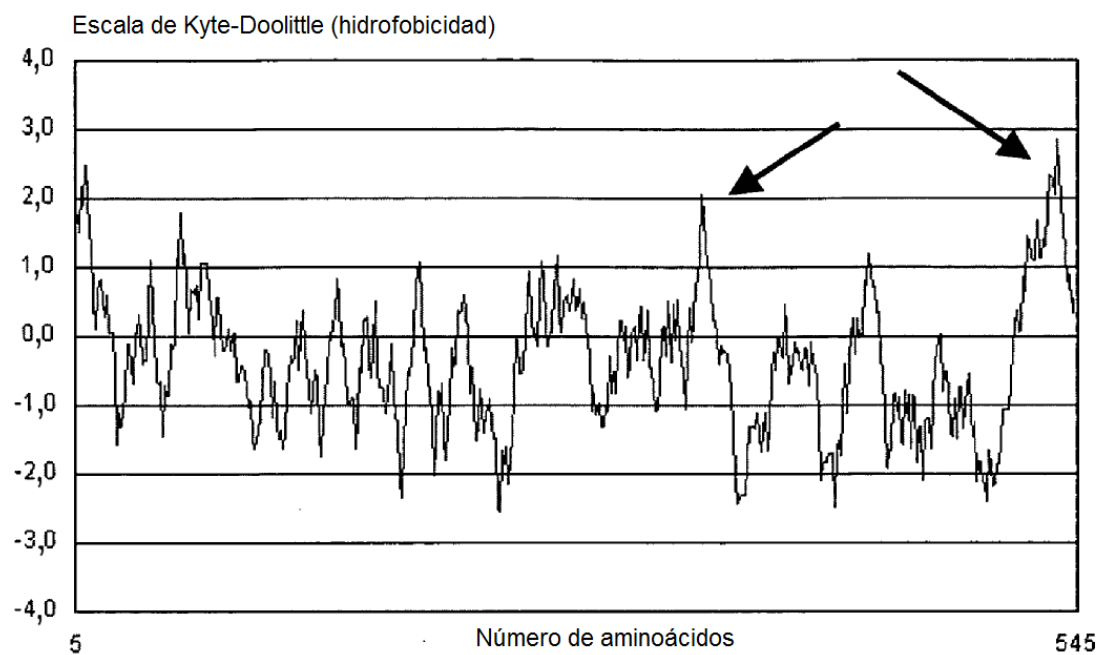


Fig. 10a

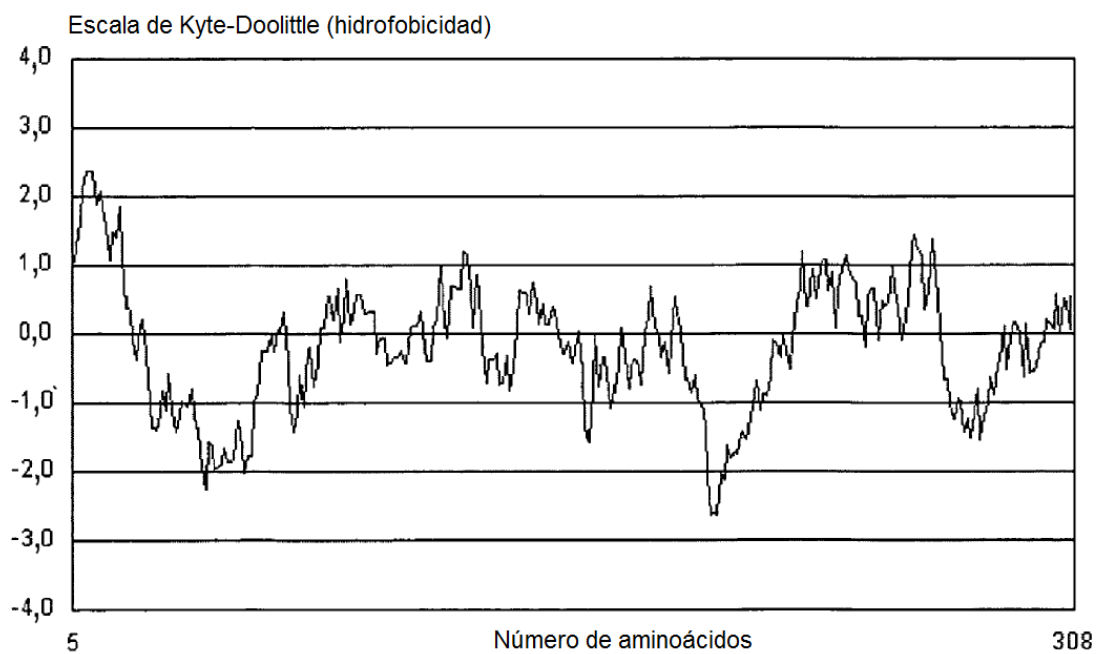


Fig. 10b