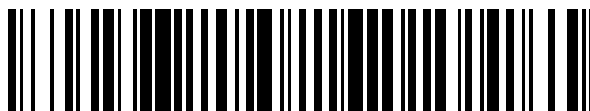


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 687 029**

51 Int. Cl.:

**C12N 15/86** (2006.01)  
**C12N 15/12** (2006.01)  
**C12N 5/10** (2006.01)  
**C12N 7/00** (2006.01)  
**A61K 38/00** (2006.01)  
**C07K 14/705** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **18.07.2002 PCT/EP2002/08047**

87 Fecha y número de publicación internacional: **06.02.2003 WO03010200**

96 Fecha de presentación y número de la solicitud europea: **18.07.2002 E 02767233 (6)**

97 Fecha y número de publicación de la concesión europea: **04.07.2018 EP 1414984**

54 Título: **Ácido nucleico que codifica un polipéptido implicado en la entrada celular del virus del PRRS**

30 Prioridad:

**24.07.2001 EP 01202824**  
**31.10.2001 EP 01204220**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**23.10.2018**

73 Titular/es:

**UNIVERSITEIT GENT (100.0%)**  
**Sint-Pietersnieuwstraat 25**  
**9000 Gent, BE**

72 Inventor/es:

**PENSAERT, MAURICE;**  
**NAUWYNCK, HANS y**  
**VANDERHEIJDEN, NATHALIE**

74 Agente/Representante:

**LEHMANN NOVO, María Isabel**

ES 2 687 029 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

Ácido nucleico que codifica un polipéptido implicado en la entrada celular del virus del PRRS

5 La invención se refiere a un nuevo polinucleótido que codifica un polipéptido implicado en la entrada celular del virus del PRRS (PRRSV), a un vector recombinante que comprende dicho polinucleótido, a una célula capaz de expresar dicho polipéptido, a un método para producir dicho polipéptido, así como a un cultivo celular y a un nuevo método de producción del virus PRRSV.

10 El síndrome reproductivo y respiratorio porcino (PRRS) se reconoció inicialmente en 1987 en los EE. UU. como una nueva enfermedad del cerdo que provoca insuficiencia reproductiva en cerdas gestantes e insuficiencia respiratoria en cerdos recién nacidos. La enfermedad fue descrita posteriormente en Europa y Canadá. Hasta que se identificó un agente causal, el síndrome era conocido por varios nombres, tales como la misteriosa enfermedad de los cerdos, la infertilidad porcina y el síndrome respiratorio, la enfermedad del cerdo azul, el aborto azul, etc.

Un virus ARN de cadena sencilla con envoltura, el virus PRRS, provoca el síndrome. El PRRSV es un miembro de la familia Arterivirus que incluye el virus elevador de la lactato deshidrogenasa (LDV) de ratones, el virus de la arteritis equina (EAV) y el virus de la fiebre hemorrágica de simios (SHFV).

15 El tipo celular predominante infectado por PRRSV es el macrófago. El virus se aisló por primera vez en macrófagos alveolares porcinos (PAM), pero también se ha informado que se replica en monocitos y en células microgliales. El PRRSV tiene un tropismo restringido para células monocíticas/macrófagas tanto in vitro como in vivo.

Actualmente la replicación del PRRSV solo ocurre en unas pocas líneas celulares establecidas.

20 Se ha demostrado la replicación del virus in vitro en líneas celulares establecidas tales como la línea celular de riñón de mono MA-104, y su clon permisivo MARC-145 (Kim et al., Arch. Virol., 133, 477-483, 1993).

Un procedimiento para cultivar el PRRSV en una línea celular clonada derivada de MA-104, 9009B (designado como ATCC CRL 11302) se describe en el documento US5510258.

Se ha informado de la falla de replicación del PRRSV en varias otras líneas celulares.

25 Se ha investigado el mecanismo que restringe la replicación del PRRSV en una diversidad de líneas celulares. Se descubrió que el PRRSV no se podía unir a la mayoría de los tipos celulares testados. Se sugirió que la ausencia de unión del PRRSV a las células era uno de los factores determinantes del tropismo celular de PRRSV. Se postuló que la ausencia de un componente celular específico podría ser crucial para la entrada del virus, y que la entrada del virus se produjo por endocitosis mediada por receptor (Kreutz, Virus Research, 53, 121-128, 1998).

30 La endocitosis mediada por receptor es conocida por otros virus con envuelta. Se ha informado de receptores para algunos virus. Por ejemplo, los receptores para los diferentes subgrupos de sarcoma aviar y virus de la leucosis (ASLV) han sido identificados y clonados. (Balliet et al., J. Virol., 73(4), 3054-3061, 1999.) El receptor celular para el virus de la hepatitis del ratón (MHV) también se ha identificado y clonado. Un vector en el que se había subclonado el ADNc para el receptor de MHV se utilizó para la expresión del receptor de MHV en células que normalmente eran resistentes a la infección por MHV (línea BHK de fibroblastos de hámster y la línea RD de células de rhabdomyosarcoma humano). Esto fue suficiente para permitir la infección de estas células con MHV (Dveksler et al., J. Virology, 65(12), 6881-6891, 1991).

40 También se ha informado de algunos esfuerzos para identificar un receptor del PRRSV. En un intento de identificar el receptor PAM que puede determinar la susceptibilidad de los macrófagos al PRRSV, se produjeron dos anticuerpos monoclonales (Mabs), 41 D3 y 41 D5, que bloquearon la infección por el PRRSV de PAM. (Duan et al., en Coronaviruses and Arteriviruses, editado por Enjuanes et al., Plenum Press, Nueva York, 81-87, 1998).

Con la ayuda de estos MAb se intentó identificar las moléculas que se utilizan para la unión del virus en la superficie de PAM.

45 Utilizando un kit de biotilación de proteínas, se marcaron todas las proteínas de membrana, y se inmunoprecipitaron proteínas del lisado celular biotinilado utilizando los MAb. Por lo tanto, una proteína marcada con biotina con un peso molecular de aproximadamente 210 kDa podría visualizarse después de la inmunoprecipitación (Duan et al., J. Virol., 4520-4523, 1998); Nauwynck et al., Vte. Res. 31 (2000) 22-22).

Sin embargo, hasta ahora no se ha informado de una caracterización del receptor potencial para el PRRSV y su papel real en el tropismo celular no se ha elucidado.

50 Como ya se explicó anteriormente, el PRRSV solo puede cultivarse actualmente en un número limitado de líneas celulares.

Con fines de vacuna, sería conveniente que el virus pudiera cultivarse en otras células, además del número limitado de líneas celulares disponibles en la actualidad para el cultivo del PRRSV. Solo cambiando el procedimiento de producción de las vacunas contra el PRRSV a otras líneas celulares se puede mejorar la calidad, la cantidad y el precio de las vacunas contra el PRRSV.

5 Por lo tanto, existe una necesidad de métodos de cultivo más eficientes para el virus.

Si de hecho un receptor del PRRSV es el responsable de la especificidad celular única del virus, la expresión del receptor en células que normalmente no son permisivas para el virus puede abrir rutas hacia formas más eficientes de cultivo del virus.

10 Los autores de la presente invención han logrado aislar una proteína de las membranas de PAM que parece jugar un papel crucial en la entrada del virus en la célula. Los autores de la presente invención tuvieron éxito por primera vez en identificar y aislar la proteína pura en cantidades suficientemente grandes como para que se analizara. Por medio de un proceso único para la identificación y el aislamiento de la proteína, los autores de la presente invención fueron, por lo tanto, los primeros en caracterizar la proteína y dilucidar la secuencia de aminoácidos de la proteína y la secuencia génica correspondiente.

15 La secuencia de nucleótidos dilucidada que codifica la proteína, así como la secuencia de aminoácidos de la proteína, se compararon con secuencias almacenadas en bases de datos de secuencias.

Sorprendentemente, el supuesto receptor del PRRSV proporcionado por la presente invención mostró una gran cantidad de homología con determinadas proteínas que pertenecen a la familia Siglec. La familia siglec es una familia de lectinas tipo inmunoglobulina (Ig) de unión al ácido siálico.

20 El polipéptido identificado mostró un 69% de identidad en el nivel de aminoácidos y un 75% de identidad en el nivel de ácidos nucleicos con una proteína sialoadhesina de ratón de 185 kDa y su correspondiente secuencia génica, respectivamente (Código de Acceso a Genbank: Z36293, registro de proteína anotado SwissProt: Q62230) según se describe por Crocker et al., EMBO J., 13(19), 4490-4503, 1994. El polipéptido mostró, además, un 78% de identidad en el nivel de aminoácidos y un 82% de identidad en el nivel de ácidos nucleicos con una proteína sialoadhesina humana de 200 kDa y su correspondiente secuencia génica, respectivamente (Código de Acceso a Genbank: AF230073), según se describe por Hartnell A. et al., Blood, 97(1), 288-296, 2001.

25 Las secuencias se compararon con secuencias en las bases de datos utilizando un programa BLAST (BLASTF 2.1.2 [19 oct. 2000]) (Altschul, Stephen F., Thomas, L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller y David J. Lipman (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs" Nucleic Acids Res. 25, 3389 - 3402). El programa se utilizó para buscar alineamientos de secuencias.

30 Las sialoadhesinas de ratón y humanas muestran aproximadamente un 72% de identidad en el nivel de aminoácidos y aproximadamente un 77% de identidad en el nivel de ácidos nucleicos, con la mayor identidad en la región extracelular, que comprende 17 dominios de Ig en ambas especies. El patrón de expresión de sialoadhesina humana y sialoadhesina de ratón es similar. Están ausentes de monocitos y otros leucocitos de la sangre periférica, pero se expresan fuertemente por los macrófagos tisulares en el bazo, los ganglios linfáticos, la médula ósea, el hígado, el colon y los pulmones.

La sialoadhesina humana (Sn) es una molécula de interacción celular restringida a macrófagos, que probablemente esté implicada en interacciones entre macrófagos y células hematopoyéticas en la médula ósea y posiblemente también en otras interacciones célula-célula.

40 La Sn humana es un miembro prototípico de la familia de moléculas de interacción celular Siglec (lectina tipo Ig de unión a ácido siálico) y se denominó Siglec-I.

45 Además de Sn, los miembros de la familia Siglec incluyen CD22 (Siglec-2) en células B, CD33 (Siglec-3) en células mieloides inmaduras y monocitos, y glicoproteína asociada a mielina (MAG) (Siglec-4A) en células de Schwann y oligodendrocitos, y Siglecs-5, -6, -7, -8 y -9 expresadas en diversos subconjuntos hemopoyéticos. Las funciones biológicas de Sn son poco conocidas.

Basándose en su % de identidad con las sialoadhesinas de ratón y humana mencionadas anteriormente, el polipéptido identificado, que sirve como el receptor para el PRRSV en macrófagos alveolares porcinos, puede ser una sialoadhesina (Sn) porcina.

50 La presente invención, en un aspecto, proporciona un polinucleótido aislado que comprende una secuencia de ácidos nucleicos que codifica un polipéptido porcino que comprende una secuencia de aminoácidos que tiene al menos un 90% de identidad en toda su longitud con una secuencia de aminoácidos tal como se representa en SEQ ID NO: 2, o un fragmento funcional de dicho polipéptido de al menos 50 aminoácidos, que tiene las siguientes características:

dicho polipéptido o fragmento de polipéptido, cuando se expresa en la superficie de una célula, es capaz de hacer que la célula sea receptiva para el PRRSV,

dicho polipéptido tiene un peso molecular aparente de aproximadamente 210 kD y

dicho polipéptido es reactivo con Mab 41 D3.

- 5 El hibridoma 41 D3 que produce este Mab está depositado en la CNCM del Instituto Pasteur, París, Francia, bajo nº de acceso I-2719.

El polinucleótido de acuerdo con la invención codifica un polipéptido que juega un papel crucial en la entrada del PRRSV en las células. Debido a que el polipéptido media en la entrada del virus en las células, se postula que es el receptor del PRRSV de macrófagos alveolares porcinos.

- 10 De este modo, con la presente invención, se proporcionan polinucleótidos que codifican una proteína de tipo Sn, que es un miembro de la superfamilia de moléculas de tipo inmunoglobulina con la propiedad de que facilita la entrada del PRRSV en una célula.

- 15 El Mab 41 D3 reconoce específicamente el polipéptido identificado y bloquea la entrada de PRRSV en una célula. No se observó tinción de inmunofluorescencia de membrana con el MAb 41 D3 en células mononucleares de la sangre periférica porcina (PBMC), macrófagos peritoneales porcinos (PPM), ST (testículos porcinos), SK (riñón porcino), PK-15 y células MARC-145. De estas células, solo se observó una leve tinción intracelular en algunas PBMC y PPM.

- 20 El polipéptido identificado tiene un peso molecular aparente de aproximadamente 210 kD. El peso molecular se determinó en SDS-PAGE utilizando condiciones reductoras. El término "aproximadamente" debe interpretarse en el sentido de que el peso molecular es superior a 200 kD (marcador más cercano en el gel) y en el intervalo entre 205 y 225 kD.

- 25 La secuencia del polinucleótido, tal como fue dilucidada, por los autores de la presente invención se representa en SEQ ID NO: 1. La secuencia del polipéptido codificado por este polinucleótido se representa en SEQ ID NO: 2. La información de secuencia, tal como se proporciona en esta memoria, no debe interpretarse de forma tan restrictiva como para requerir la exclusión de bases erróneamente identificadas o variaciones que se producen de forma natural de la secuencia de nucleótidos en la población porcina.

- 30 Fragmentos de la secuencia de ácidos nucleicos proporcionada que codifica un fragmento funcional del polipéptido son asimismo parte de la presente invención. Un fragmento funcional del polipéptido es un fragmento que al menos representa la parte del polipéptido, que es esencial para que el polipéptido sea capaz de servir como un receptor para el PRRSV, y puede cumplir esta función, por ejemplo, cuando se utiliza solo o fusionado a secuencias heterólogas.

Por lo tanto, tales polinucleótidos, que codifican fragmentos funcionales, pueden codificar polipéptidos que son funcionales per se, o los fragmentos pueden ser funcionales cuando se unen a otros polipéptidos, para obtener proteínas quiméricas.

- 35 Por ejemplo, un polinucleótido que codifica un fragmento funcional de este tipo del polipéptido se puede fusionar a polinucleótidos que codifican regiones de transmembrana y/o secuencias señal. Al crear tales proteínas quiméricas, el sitio funcional puede transferirse a proteínas de la superficie celular más abundantes, creando células con una sensibilidad más alta para el PRRSV. Alternativamente, para potenciar la sensibilidad de las células para el PRRSV, pueden crearse dominios funcionales múltiples en una sola molécula o pueden colocarse espaciadores de longitud óptima entre la región de transmembrana y el sitio de unión del PRRSV funcional.

- 40 Los polinucleótidos que codifican fragmentos de los polipéptidos completos, preferiblemente codifican fragmentos que tienen al menos 50 aminoácidos, más preferiblemente al menos 100 aminoácidos o al menos 200 aminoácidos de longitud.

- 45 Los polinucleótidos de acuerdo con la invención también abarcan aquellos polinucleótidos que codifican variantes del polipéptido identificado. Con las variantes del polipéptido identificado, se quiere dar a entender que los polipéptidos tienen una actividad de receptor del PRRSV y comprenden variaciones de la secuencia de aminoácidos identificada, al tiempo que siguen manteniendo las características funcionales, ya que todavía juegan un papel en la entrada viral del PRRSV en las células. Por lo tanto, estas variantes son funcionalmente equivalentes a un polipéptido con la secuencia de aminoácidos identificada. Estos polipéptidos no abarcan necesariamente la
- 50 secuencia de aminoácidos de longitud completa tal como se proporciona en esta memoria.

Las variaciones que pueden ocurrir en una secuencia de aminoácidos pueden demostrarse mediante una o más diferencias de aminoácidos en la secuencia global o mediante delecciones, sustituciones, inserciones, inversiones o adiciones de uno o más aminoácidos en dicha secuencia. Se han descrito sustituciones de aminoácidos que se

espera que no alteren esencialmente las actividades biológicas e inmunológicas. Sustituciones de aminoácidos entre aminoácidos relacionados o reemplazos que se han producido con frecuencia en la evolución son, entre otros, Ser/Ala, Ser/Gly, Asp/Gly, Asp/Asn, Ile/Val (véase Dayhof, M.D., Atlas of protein sequence and structure, Nat. Biomed. Res. Found., Washington D.C., 1978, volumen 5, supl. 3). Basándose en esta información, Lipman y Pearson desarrollaron un método para la comparación de proteínas rápida y sensible (Science, 1985 227, 1435-1441) y determinaron la similitud funcional entre polipéptidos homólogos. Resultará claro que los polinucleótidos que codifican tales variantes son asimismo parte de la invención.

Polinucleótidos como se definen con la presente invención también incluyen polinucleótidos que tienen variaciones en la secuencia de ácido nucleico cuando se comparan con la secuencia de ácido nucleico identificada o que tienen sitios polimórficos. Con polinucleótidos de "variantes" se entiende que difieren de la secuencia de ácido nucleico identificada, pero aún codifican un polipéptido que tiene la misma actividad de receptor de PRRSV que el polipéptido identificado.

Las variantes pueden ser variantes naturales o no naturales. La variante que se produce de forma no natural se puede introducir mediante mutagénesis. Las variantes naturales pueden ser variantes alélicas. Una variante alélica es una de varias formas alternativas de un gen que ocupa un locus en un cromosoma de un organismo. A veces, un gen se expresa en un determinado tejido como una variante de corte y empalme, dando como resultado un ARNm 5' o 3' alterado o la inclusión o exclusión de una o más secuencias de exón. Se espera que todas estas secuencias, así como también las proteínas codificadas por estas secuencias, realicen las mismas funciones o funciones similares y también formen parte de la invención.

Una secuencia de ADNc aislada puede ser incompleta debido a la transcripción incompleta a partir del ARNm correspondiente, o pueden obtenerse clones que contienen fragmentos del ADNc completo. En la técnica se conocen diversas técnicas para completar dichas secuencias de ADNc, tales como RACE (siglas inglesas de amplificación rápida de extremos de ADNc).

Los polinucleótidos que tienen una secuencia de ácidos nucleicos que es una variante de la secuencia de ácidos nucleicos identificada se pueden aislar mediante un método que comprende las etapas de: a) hibridar un ADN que comprende toda o parte de la secuencia identificada tal como se refleja en SEQ ID NO: 1, en condiciones rigurosas contra ácidos nucleicos que son ADN (genómico) o ADNc aislado de células porcinas (preferiblemente PAMs) que expresan altamente el polinucleótido de interés; y b) aislar dichos ácidos nucleicos por métodos conocidos por una persona experta en la técnica.

Las condiciones de hibridación son preferiblemente muy rigurosas.

El término "riguroso" significa condiciones de lavado de 1 x SSC, SDS al 0,1% a una temperatura de 65 °C; condiciones altamente rigurosas se refieren a una reducción en SSC a 0,3 x SSC, más preferiblemente a 0,1 x SSC. Preferiblemente, los primeros dos lavados se llevan a cabo dos veces cada uno durante 15-30 minutos. Si existe la necesidad de lavar bajo condiciones altamente rigurosas, se realiza un lavado adicional con 0,1 x SSC una vez durante 15 minutos. La hibridación se puede realizar, p. ej., durante la noche en tampón fosfato 0,5 M, pH 7,5/SDS al 7% a 65 °C. Dichos métodos de hibridación se describen en cualquier libro de texto estándar sobre clonación molecular, por ejemplo: Molecular Cloning: a laboratory manual, 3ª ed.; eds: Sambrook et al., CSHL press, 2001.

Como alternativa, el método de aislamiento podría comprender una metodología de amplificación de ácido nucleico utilizando cebadores y/o sondas derivados de la secuencia de ácidos nucleicos proporcionada con la presente invención. Dichos cebadores y/o sondas son oligonucleótidos que tienen al menos 15 nucleótidos de longitud, los oligonucleótidos preferidos tienen aproximadamente 25-50 nucleótidos.

Para testar si los polinucleótidos, cuya secuencia de ácidos nucleicos representa una variante de la secuencia de ácidos nucleicos identificada, codifican polipéptidos que están funcionalmente relacionados con la secuencia identificada, se pueden utilizar métodos tales como los ejemplificados en los ejemplos. En el ejemplo 6 se describe cómo se puede expresar una secuencia en células y cómo las células transfectadas posteriormente pueden testarse para determinar su susceptibilidad frente a la infección por el PRRSV.

Por lo tanto, en un aspecto adicional, la presente invención proporciona polinucleótidos que comprenden una secuencia de ácidos nucleicos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos un 90% de identidad con la secuencia de aminoácidos tal como se representa en SEQ ID NO: 2. Se prefieren polinucleótidos que codifican polipéptidos que tienen al menos un 95% de identidad con SEQ ID NO: 2 y más preferidos son aquellos polinucleótidos que codifican poliproteínas que tienen al menos un 97% de identidad con SEQ ID NO: 2, en donde aquellos que codifican proteínas que tienen al menos un 98 o 99% son más preferidos. Los más preferidos son polinucleótidos que codifican el polipéptido de SEQ ID NO: 2. Debido a la degeneración del código genético, los polinucleótidos que codifican una secuencia de aminoácidos idéntica o sustancialmente idéntica pueden utilizar diferentes codones específicos. Todos los polinucleótidos que codifican los polipéptidos según se definen arriba se consideran parte de la invención.

- En particular, polinucleótidos preferidos de acuerdo con la invención son polinucleótidos aislados que tienen al menos un 90% de identidad con la secuencia de ácidos nucleicos completa de SEQ ID NO: 1. Más preferidos son aquellos polinucleótidos que tienen al menos un 95% de identidad, y aún más preferidos al menos 97, preferiblemente 98% o 99% de identidad con la secuencia completa de SEQ ID NO: 1. Dichos polinucleótidos incluyen polinucleótidos que comprenden la secuencia de ácidos nucleicos representada en SEQ ID NO: 1. Un polinucleótido que codifica un polipéptido con una secuencia tal como se representa en SEQ ID NO: 2 puede comprender la secuencia de ácidos nucleicos tal como se representa en SEQ ID NO: 1. En una realización preferida adicional de la invención, el polinucleótido consiste en la secuencia de ácidos nucleicos tal como se representa en SEQ ID NO: 1.
- Los polinucleótidos de acuerdo con la invención pueden ser ADN o ARN, preferiblemente ADN. El ADN de acuerdo con la invención se puede obtener a partir de ADNc. Alternativamente, la secuencia codificante podría ser ADN genómico, o podría prepararse utilizando técnicas de síntesis de ADN. Si el polinucleótido es ADN, éste puede estar en forma de cadena sencilla o de doble cadena. La cadena sencilla podría ser la cadena codificante o la cadena no codificante (anti-sentido).
- También se incluyen dentro de la definición de polinucleótidos ARNs o ADNs modificados. Se pueden hacer modificaciones en las bases del ácido nucleico y se pueden incorporar bases tales como inosina. Otra modificación puede implicar, por ejemplo, modificaciones de la cadena principal.
- Con "aislado" se quiere dar a entender que el polinucleótido se aísla del estado natural, es decir, se ha cambiado o movido desde su entorno natural o ambos. La molécula está separada y es discreta de todo el organismo con el que se encuentra la molécula en la naturaleza.
- "% de identidad" define la relación entre dos o más polinucleótidos o polipéptidos sobre la base de una comparación entre sus secuencias alineadas.
- La identidad se puede calcular por métodos conocidos. Los porcentajes de identidad, tal como se menciona en esta memoria, son los que se pueden calcular con el programa GAP, que se ejecuta bajo GCG (Genetics Computer Group Inc., Madison WI).
- Los parámetros para la comparación de la secuencia polipeptídica incluyen los siguientes:
- Algoritmo: Needleman y Wunch, J.Mol.Biol., 48, 443-453, 1970.
- Como una matriz de comparación para los alineamientos de aminoácidos, se utiliza la matriz BLOSUM62 (Hentikoff y Hentikoff, P.N.A.S. USA, 89, 10915-10919, 1992) utilizando los siguientes parámetros:
- Penalización del hueco: 8
- Penalización de longitud de hueco: 2
- Sin penalización para los huecos finales.
- Parámetros para la comparación de nucleótidos utilizados:
- Algoritmo: Needleman y Wunch (supra).
- Matriz de comparación: coincide = +10, desajuste = 0.
- Penalización del hueco: 50.
- Penalización de longitud de hueco: 3.
- El ADN de acuerdo con la invención será muy útil para la expresión in vivo o in vitro del polipéptido codificado en cantidades suficientes y en forma sustancialmente pura. Cuando los polinucleótidos de acuerdo con la invención se utilizan para la expresión del polipéptido codificado, los polinucleótidos pueden incluir, además de la secuencia codificante del polipéptido o fragmento funcional del mismo, otras secuencias codificantes, por ejemplo, secuencias conductoras o porciones de fusión, tales como secuencias de marcador y similares.
- Se puede emplear de manera útil una amplia diversidad de combinaciones de células huésped y vehículos de clonación en la clonación de la secuencia de ácidos nucleicos de acuerdo con la invención. Un polinucleótido de acuerdo con la invención puede clonarse en un sistema de expresión apropiado, tal como un sistema de expresión bacteriano (p. ej., Ecoli DH5 $\alpha$ ), un sistema de expresión viral (p. ej., Baculovirus), un sistema de levadura (p. ej., S. Cerevisiae, Pichia) o células eucarióticas (p. ej., células Cos, BHK, MDCK, MDBK, HeLa, PK15). En todos los sistemas, el polinucleótido se clona primero en un vector apropiado bajo el control de un promotor constitutivo o inducible adecuado.
- En otro aspecto, la presente invención se refiere, por lo tanto, a un vector recombinante que comprende un polinucleótido de acuerdo con la invención. Vectores adecuados son, por ejemplo, cósmidos, plásmidos bacterianos

o de levadura, plásmidos de amplia gama de huéspedes y vectores derivados de combinaciones de plásmidos y fagos o ADN de virus. Los vectores derivados de ADN cromosómico también están incluidos. Además, puede estar presente un origen de replicación y/o un marcador de selección dominante en el vector de acuerdo con la invención. Los vectores de acuerdo con la invención son adecuados para transformar una célula huésped. Ejemplos de vectores de clonación adecuados son vectores plasmídicos tales como pBR322, los diversos plásmidos pUC, pEMBL y Bluescript.

Cuando se utiliza en la expresión del polipéptido o fragmentos funcionales del mismo, un vector recombinante de acuerdo con la presente invención, puede comprender, además, una secuencia de control de la expresión enlazada operativamente a la secuencia de ácidos nucleicos que codifica la proteína. Enlazada operativamente se refiere a una disposición en la que las secuencias de control están configuradas con el fin de realizar su función habitual, al efectuar la expresión del polinucleótido. Generalmente, dichas secuencias de control de la expresión comprenden una secuencia promotora y secuencias que regulan la transcripción y la traducción y/o potencian los niveles de expresión. No todas estas secuencias de control necesitan estar presentes en un vector recombinante, siempre que el polinucleótido deseado sea capaz de transcribirse y traducirse. Por supuesto, el control de la expresión y otras secuencias pueden variar dependiendo de la célula huésped seleccionada o pueden hacerse inducibles. Dichas secuencias de control de la expresión son bien conocidas en la técnica y se extienden a cualquier promotor eucariota, procariota o viral o señal de poli-A capaz de dirigir la transcripción génica.

Ejemplos de promotores útiles son el promotor SV-40 (Science 222, 524-527, 1983), el promotor de metalotioneína (Nature, 296, 39-42, 1982), el promotor de choque térmico (Voellmy et al., P.N.A.S. USA, 82, 4949-4953, 1985), el promotor gX del PRV (Mettenleiter y Rauh, J. Virol. Methods, 30, 55-66, 1990), el promotor IE humano del CMV (documento US 5.168.062), el promotor LTR del virus del sarcoma Rous (Gorman et al., PNAS, 79, 6777-6781, 1982) o factor de elongación humano 1 alfa o promotor de ubiquitina, etc.

Después de que el polinucleótido se haya clonado en un vector apropiado, la construcción puede transferirse a la célula, a las bacterias o a la levadura sola por medio de un método apropiado, tal como electroporación, transfección con CaCl<sub>2</sub> o lipofectinas. Cuando se utiliza un sistema de expresión de baculovirus, el vector de transferencia que contiene el polinucleótido puede transfectarse junto con un genoma de baculo completo.

Un virus recombinante, que comprende un polinucleótido de acuerdo con la invención, es asimismo parte de la presente invención.

Todas estas técnicas son bien conocidas en la técnica y se describen ampliamente en protocolos proporcionados por fabricantes de materiales biológicos moleculares (tales como Promega, Stratagene, Clonetech y/o Invitrogen) y en la bibliografía o libros de texto de referencia, por ejemplo en Rodríguez, R.L. y D.T. Denhardt, edit., Vectors: A survey of molecular cloning vectors and their, Butterworths, 1988; Current protocols in Molecular Biology, eds.: F.M. Ausbel et al., Wiley N.Y., 1995; Molecular Cloning: a laboratory manual, 3<sup>a</sup> ed.; eds: Sambrook et al., CSHL press, 2001 y DNA Cloning, Vol. 1-4, 2<sup>a</sup> edición 1995, eds.: Glover y Hames, Oxford University Press).

Las células transformadas con un polinucleótido o un vector de acuerdo con la invención son asimismo parte de la presente invención. Por lo tanto, en otro aspecto, la presente invención proporciona una célula capaz de expresar un polipéptido recombinante, caracterizada porque la célula comprende un polinucleótido de acuerdo con la invención que codifica el polipéptido recombinante expresado. El término "recombinante" en este contexto se refiere a un polipéptido que no se expresa en la célula en la naturaleza. Si la célula huésped es de origen porcino, la secuencia de polinucleótidos puede estar presente en el material genómico de la célula, pero no se expresa en el tipo particular de célula porcina en un cerdo. Por lo tanto, una célula huésped que comprende el ADN o el vector de expresión de acuerdo con la invención también está dentro del alcance de la invención. Las células huésped modificadas genéticamente se pueden cultivar en medios nutrientes convencionales que pueden modificarse, p. ej., para la selección, amplificación o inducción de la transcripción apropiadas. Las condiciones de cultivo tales como temperatura, pH, nutrientes, etc., son bien conocidas por los expertos ordinarios en la técnica.

Las células que se transforman con un vector de acuerdo con la invención pueden ser de origen porcino o no porcino. Las células que son de origen porcino pueden ser células PK 15, células SK o células ST. Las células de origen no porcino que pueden transformarse para expresar un polipéptido de acuerdo con la invención son, por ejemplo, células MDCK, células BHK, células MDBK, células de insectos, células HeLa o células COS.

Una célula transformada de acuerdo con la invención puede comprender un polinucleótido de acuerdo con la invención integrado de manera estable en el material genómico o como parte de un vector de replicación autónoma.

Un cultivo celular que comprende una multitud de células de acuerdo con la invención es asimismo parte de la presente invención. En una realización preferida, dicho cultivo celular está infectado con un PRRSV. El PRRSV puede ser un virus de tipo salvaje o un virus atenuado. Este último es particularmente adecuado para la preparación de vacunas. Las células de acuerdo con la invención se pueden utilizar para expresar el polipéptido y el polipéptido se puede aislar del cultivo celular.

En otro aspecto, la presente invención proporciona, por lo tanto, un método para producir un polipéptido que, cuando se expresa en la superficie de una célula, es capaz de hacer que la célula sea receptiva al PRRSV, tiene un peso molecular aparente de aproximadamente 210 kD y es reactivo con MAB 41 D3, que comprende las etapas de:

- cultivar una célula de acuerdo con la invención, y
- 5 - aislar material que contiene polipéptido del cultivo celular.

Células preferidas de acuerdo con la invención son células que presentan el polipéptido en su superficie. Dado que el polipéptido juega un papel en la entrada del PRRSV en las células, las células que son capaces de expresar un polipéptido de acuerdo con la invención en su superficie se vuelven accesibles al virus PRRSV. Por lo tanto, la presente invención permite la infección con PRRSV de células que normalmente no son receptivas para el PRRSV. De esta manera, se proporcionan nuevas rutas para cultivar el PRRSV.

La presente invención proporciona, por lo tanto, un método para producir PRRSV, en el que un cultivo celular que comprende células de acuerdo con la invención se infecta con un PRRSV y se cultiva, después de lo cual el virus se puede recoger del cultivo celular. Mientras que el virus, debido a su tropismo celular limitado, podría cultivarse solamente en células muy específicas, con la presente invención están disponibles nuevas rutas de cultivo del virus. Con la presente invención, es posible ahora desarrollar PRRSV en células que se pueden cultivar en medios libres de compuestos animales, una característica que mejora la calidad de los productos basados en el virus cultivado, por ejemplo vacuna. También las células que pueden crecer en suspensión pueden ahora hacerse susceptibles al PRRSV, lo que facilita un procedimiento de producción más óptimo en los fermentadores. Las células pueden hacerse más receptivas para el virus y pueden seleccionarse para títulos de producción de PRRSV más altos. Una vez que el virus ha crecido a títulos elevados, puede procesarse según el uso previsto por medios conocidos en la técnica. Por ejemplo, los fluidos virales pueden inactivarse, por ejemplo con, p. ej., formalina, BPL, BEA o irradiación gamma, para su uso en vacunas. En la alternativa, la cepa viral utilizada en la infección puede ser una cepa atenuada para su uso en la producción de vacunas vivas atenuadas. Las vacunas se pueden formular por medios conocidos en la técnica. Habitualmente, esto puede implicar la adición de un adyuvante y/o un soporte adecuado.

Las líneas celulares que se han vuelto más susceptibles al virus se pueden utilizar en los tests de detección de virus, mientras que, con los métodos de la técnica anterior, se necesitan PAMs. Las células PAMs son células primarias no crecientes que deben ser recogidas de cerdos. Se evita el tedioso proceso de recogida de PAMs y se puede obtener una mayor sensibilidad en dichos sistemas celulares.

Cuando se utilizan células porcinas, estas células pueden transformarse, al igual que otros tipos de células, con un vector que comprende la información genética para el polipéptido. Sin embargo, las células porcinas tienen, en su material genómico, ya la secuencia de ADN genómico correspondiente al ácido nucleico de la presente invención. Pero el gen no se expresa en todas las células porcinas, ya que está en los macrófagos alveolares porcinos. Con la presente invención, el gen, en forma aislada, y su secuencia, se han elucidado. Ahora que está disponible esta información, proporcionada con la presente invención, también se han puesto a disposición otras rutas (además de la transformación de la célula con un vector que contiene el gen) para la expresión del gen en células porcinas. La expresión del gen en células porcinas también se puede lograr "conectando" la expresión del gen ya presente en células porcinas. Con la presente información, se proporciona la secuencia del gen y se puede insertar un promotor de elección delante del codón de inicio mediante recombinación homóloga, seguida de la selección de las células que expresan el gen del receptor. Este promotor puede ser un promotor constitutivo de mamífero (p. ej., HCMV, SV40 LTR) o un promotor inducible (p. ej., sistema Tet).

Las células que expresan el polipéptido en su superficie también se pueden utilizar en un ensayo de exploración utilizado en el rastreo de bancos de compuestos para compuestos que se unen específicamente al polipéptido. Dado que el polipéptido juega un papel en la entrada de PRRSV en las células, compuestos de este tipo se pueden utilizar en el tratamiento o la prevención de una infección por el PRRSV.

Compuestos que pueden identificarse con el método de exploración de la invención pueden derivarse de una diversidad de fuentes que incluyen bancos de compuestos químicos o mezclas de compuestos (naturales). El método de exploración puede implicar medir la unión de un compuesto candidato al polipéptido, o a células o membranas que portan el polipéptido, o a una proteína de fusión que porta solamente el dominio de unión del polipéptido de longitud completa. La unión puede medirse directa o indirectamente. La unión puede medirse directamente, por ejemplo, por medio de una etiqueta asociada con el compuesto. La unión también puede medirse indirectamente. Por ejemplo, medir indirectamente la unión de un compuesto candidato puede implicar competencia con un competidor (etiquetado). La medición de la unión de un compuesto candidato puede, por ejemplo, determinarse también en un ensayo basado en células, en el que se puede determinar si un compuesto candidato es capaz de bloquear el virus PRRSV para que no entre en una célula. En ese caso, se puede determinar si, en presencia del compuesto, las células pueden seguir siendo infestadas con el PRRSV.

Métodos para tratar o prevenir una infección por el PRRSV en cerdos inhibiendo la unión del virus PRRSV se pueden lograr administrando un inhibidor del polipéptido, que bloqueará la entrada del virus a las células. Un inhibidor de este tipo puede ser un compuesto químico, que puede ser (un derivado de) un compuesto identificado

con un método de exploración. Un inhibidor de este tipo puede ser asimismo otra molécula capaz de unirse al polipéptido, p. ej., un anticuerpo o fragmento de anticuerpo.

- 5 Otro método para inhibir la captación de virus es cubriendo el virus mediante receptores solubles. De esta forma, se bloqueará lo adjunto y la posterior entrada del virus en las células. Por lo tanto, el uso del polipéptido en la fabricación de un medicamento para el tratamiento o la prevención de la infección por el PRRSV en cerdos también forma parte de la presente invención.

La composición utilizada para administrar el inhibidor debe formularse con un soporte farmacéuticamente aceptable, adaptado a la vía de administración elegida.

FIGURAS:

- 10 Figura 1. Expresión de p210 en células PK-15 recombinantes transfectadas de manera estable con el plásmido pcDNA3.1D/V5-His-TOPO que contiene el ADNc de p210 (clon 80.2). Se realizó una tinción de inmunofluorescencia indirecta utilizando el MAb 41 D3 (dirigido contra p210) e Ig anti-ratón de suero de cabra marcada con FITC en células fijadas con metanol (A) o no fijadas (B). Los macrófagos alveolares porcinos fijados con metanol se tiñeron usando el mismo protocolo y se incluyeron como un control positivo (C).

- 15 Figura 2. Co-localización de la nucleocápside de PRRSV y p210 en células PK-15 recombinantes que expresan p210 e infectadas con el virus Lelystad. Una tinción doble de inmunofluorescencia indirecta utilizando el MAb A27 (dirigido contra la nucleocápside de PRRSV) e Ig anti-ratón en suero de cabra marcada con FITC seguido de tinción con el MAb 41 D3 biotinilado (dirigido contra p210) y estreptavidina marcada con rojo de Texas se realizó con células fijadas en metanol. La imagen (A) muestra la señal de FITC, la imagen (B) muestra la señal de rojo Texas y la imagen (3), obtenida mediante la fusión de (A) y (B) da una señal amarilla cuando (A) y (B) se localizan conjuntamente.
- 20

Figura 3: Secuencia de ácidos nucleicos genómicos (exones) identificada del gen p210 y la secuencia de aminoácidos codificada.

- 25 Figura 4: Secuencia de ácidos nucleicos de ADNc identificada de un clon que expresa un receptor de PRRSV p210 funcional y la secuencia de aminoácidos codificada.

EJEMPLOS:

Ejemplo 1: identificación del receptor de PRRSV

1) Producción de anticuerpos monoclonales:

- 30 Para identificar el receptor que puede determinar el tropismo macrófago del PRRSV, se produjo un anticuerpo monoclonal (41 D3). Este Mab bloqueó completamente la infección de PAM por el PRRSV (Duan et al., J. Virol, 72, 4520-4523, 1998; Duan et al., Adv.Exp. Med.Biol., 440, 81-88, 1998). El Mab 41 D3 redujo la unión del PRRSV a PAM e inmunoprecipitó una proteína de 210 kDa de PAM. Esta proteína se detectó en la membrana celular de PAM mediante citometría de flujo y microscopía de fluorescencia. La tinción de fluorescencia estaba ausente en las membranas de células no permisivas para el PRRSV, que incluyen PBMC, macrófagos peritoneales porcinos, células ST, SK y PK15. Las células de una línea celular permisiva para el PRRSV, MARC-145 también se tiñeron negativamente con Mab 41 D3.
- 35

Ejemplo 2: Purificación de p210

MATERIALES Y MÉTODOS:

1) Purificación de la membrana de PAM

- 40 Se obtuvieron PAM de cerdos Landrace belgas convencionales de 4 a 6 semanas de edad de una piara negativa al PRRSV de acuerdo con el método previamente descrito por Wensvoort et al. (VetQ., 13, 121-130, 1991).

- PAM se mantuvieron en medio RPMI complementado con suero bovino fetal al 10%, 1% de penicilina/estreptomina, y 1% de kanamicina. Veinticuatro horas después de la siembra a una densidad de 50x10<sup>6</sup> células/175 cm<sup>2</sup>, medio reciente complementado con 50 U/ml de interferón  $\alpha$  porcino recombinante (rPo IFN- $\alpha$ 1, Lefevre et al., J.Gen.Virol., 71, 1057-1063, 1990) y PAM se mantuvieron en este medio durante 48-72 horas adicionales. PAM no adherentes se recogieron en medio por centrifugación a baja velocidad y PAM adherentes se desalojaron en PBS mediante agitación manual. Las células se agruparon, se lavaron tres veces con PBS y se recogieron por centrifugación (7 min 700 g 4°C). El sedimento celular se resuspendió en tampón A (Tris-HCl 10 mM pH 7,4, cóctel de inhibidor de proteasa Complete™ (Boehringer)) y se lisó utilizando un homogeneizador Douncer B1. Los núcleos de las células se sedimentaron por centrifugación (10 min 800 g 4°C) y el sobrenadante se
- 50

centrifugó adicionalmente (1 h 100.000 g 4°C) para sedimentar la fracción de membrana. Esta fracción se resuspendió posteriormente en sacarosa al 70% en tampón A y se sometió a un gradiente por etapas de sacarosa del 60%-30%-0% (1 h 100.000 g 4°C). La interfase de 60-30% se recogió, se diluyó 3 veces en tampón A y se sedimentó por centrifugación (1 h 100.000 g 4°C). La fracción de membrana purificada correspondiente a 3x175 cm<sup>2</sup> de PAM se resuspendió en 1 ml de tampón A y se congeló a -70°C.

## 2) Purificación de p210 mediante inmunoprecipitación y SDS-PAGE.

La proteína Sepharose-G (100 microlitros, Pharmacia) se incubó durante 1 h a 20°C con 20 µl de suero policlonal anti-ratón de conejo (Dako), se lavó con PBS y se incubó adicionalmente durante 1 h a 20°C con 20 µl de MAb 41 D3. Después de lavar con PBS, se añadieron 400 µl de fracción de membrana solubilizada y se incubaron durante 16 h a 4°C. La fracción de membrana solubilizada se obtuvo tratando la fracción de membrana purificada de PAM con Triton X-100 al 0,1% en Tris-HCl 25 mM pH 7,4 durante 1 h a 37°C y centrifugando (30 min, 13000 g, 4°C) para obtener productos no solubilizados en el sedimento. Los inmunocomplejos se lavaron cuatro veces con PBS-Tween20 al 0,1% y una vez con agua. Para permitir la recuperación eficiente de p210 y permitir que su concentración posterior proporcione un volumen final compatible con SDS-PAGE, se añadió un tampón Laemmli modificado (Tris-HCl 15,5 mM pH 6,8, SDS al 1%, 2-mercaptoetanol al 1,25%) a los inmunocomplejos antes de hervir durante 3 min. Después de centrifugar durante 1 min a 13000 g, el sobrenadante se concentró 8 veces utilizando una centrifuga de vacío (Jouan RC 10.10), se añadió glicerol al 10%, y las proteínas se separaron mediante SDS-PAGE utilizando gel de acrilamida al 7% y condiciones reductoras. La banda de 210 kDa correspondió a p210. Tres preparaciones de p210 purificadas a partir de 8x175 cm<sup>2</sup> de PAM (correspondientes aproximadamente a la cantidad de PAM extraída de un cerdo) se agruparon y se analizaron mediante espectrometría de masas.

## RESULTADOS:

Se observó que el tratamiento de PAM durante 72 h con interferón porcino dio como resultado una expresión en superficie aumentada de p210 en un factor de 4 aproximadamente, según se determinó mediante análisis FACS utilizando MAb 41 D3. Se encontró que MAb41D3 es dependiente de la conformación, porque no se produjo el reconocimiento de p210 en la transferencia Western en condiciones reductoras. Se encontraron muchas dificultades con respecto al reconocimiento del antígeno por MAb 41 D3 después de la solubilización del detergente. En consecuencia, las condiciones de solubilización e inmunoprecipitación que permiten una recuperación eficiente de p210 se determinaron empíricamente. El tipo de detergente (iónico, no iónico), su concentración, la temperatura de incubación, el tampón de solubilización fueron parámetros críticos que afectan el reconocimiento de p210 por MAb 41 D3. Otro problema crítico fue provocado por la migración de los anticuerpos en el gel, en condiciones no reductoras, justo debajo de p210, evitando una separación eficiente. Después de marcar p210 (véase más adelante) e inmunoprecipitación con MAb 41 D3, los autores de la invención pudieron determinar el peso molecular aparente de la glucoproteína reducida y utilizar condiciones reductoras para separarlo de las moléculas de anticuerpo reducidas. Utilizando las condiciones descritas, pudieron purificar 2-4 pmoles de p210 que fueron necesarios para realizar la secuenciación interna.

### Ejemplo 3: Análisis de la secuencia de proteínas.

Se realizó la digestión trípica en gel durante la noche en la banda de proteína de 210 kDa escindida en presencia de urea 2M, Tris-HCl 0,1 M pH 8,2. Los péptidos resultantes se extrajeron y se prepararon para cargarlos en una columna corta de fase inversa (1 mm D.I. x 100 mm - C18) en un sistema de fase móvil de TFA agua/acetonitrilo al 0,1% (v/v). Los péptidos eluidos se recogieron automáticamente en 18 fracciones de 50 µl. Se tomaron partes alícuotas de 0,5 µl de cada una de las fracciones y se mezclaron con 0,5 µl de solución de matriz (una relación de ácido ciano-4-hidroxi cinámico/2,5-dimetilbenzoico <sup>^</sup>A en peso) y se analizaron por el modo reflectrón. Los péptidos se pudieron detectar solo en las fracciones 16 y 17. La espectrometría de masas Maldi-Tof reveló en la fracción 16 péptidos con masas moleculares monoisotópicas de 956,1 Da, 1676,34 Da, 1853,51 Da, 1870,57 Da y en la fracción 17, 1790,47 Da, respectivamente. Se agruparon las fracciones 7 a 14 y las fracciones 15, 17 y 18. Las fracciones agrupadas y la fracción 16 se liofilizaron. La cromatografía capilar se realizó en una columna de fase inversa PepMap C18 (0,3 mm D.I. x 250 mm) en un sistema de fase móvil de agua ácido fórmico/acetonitrilo al 0,05% (v/v) a un caudal de aproximadamente 3 microlitros/min. Los péptidos eluidos se recogieron manualmente en fracciones de 2 a 5 microlitros. Se cargaron partes alícuotas de picos relevantes en agujas recubiertas de oro y se sometieron a nanoproyección ESI-TOF-MS. Los iones más intensivos fueron seleccionados para la fragmentación.

## RESULTADOS:

Se determinó la secuencia de aminoácidos de 7 péptidos (Tabla 1). Después de búsquedas en bases de datos de proteínas (BLAST 2.0.9., Altschul et al., Nucleic Acids Res., 25, 3389-3402, 1997) con la secuencia de aminoácidos de estos péptidos, las identidades de secuencia que varían desde 75% a 91% a sialoadhesina de ratón (Crocker et al., EMBO J., 13, 4490-4503, 1994) se encontraron con los péptidos 1-5. No se observó homología significativa con ninguna proteína para los péptidos 6 y 7. Claramente, p210 es el ortólogo porcino de la sialoadhesina de ratón.

Tabla 1: Secuencia de los péptidos derivados de p210 y comparación con sialoadhesina de ratón.

Péptido	Secuencia de los péptidos de p210	Secuencia deducida de ADN	Secuencia de sialoadhesina de ratón
PEP1	FSWYR	FSWYL	FSWYL
PEP2	PPAQLQLIHR	PPAQLRLLHG	PPAQLQLFHR
PEP3	ASSTAASVP	ASSTAASVP	ASSTEASVP
PEP4	WLQEGSAASLSF	WLQEGSAASLSF	WLQEGPASSLQF
PEP5	DAVLBSFWDSR	DAVLSSFWDSDR	DAVLSSFRDSR
PEP6	ALLLGQVEQR	ALLLGQVEQR	/
PEP7	QATLTTIMOSGLGR	QATLTTLMDSGLGR	/

Ejemplo 4: Caracterización bioquímica de p210.

## MATERIALES Y MÉTODOS

5 Para caracterizar p210, las proteínas de superficie de  $7 \times 10^6$  células PAM se marcaron por biotilación utilizando el protocolo recomendado (protocolo D 2, módulo de biotilación de proteína ECL, RPN2202, Amersham). El sedimento celular se lisó en 100  $\mu$ l de Tris 25 mM pH 7,4, Triton-X-100 al 0,1% y cóctel de inhibidor de proteasa Complete™ durante 1 h a 37°C. Después de la centrifugación (30 min, 13000 g, 4°C), se añadieron 30  $\mu$ l de sobrenadante a 20  $\mu$ l de proteína G-sefarosa preincubados con 2  $\mu$ l de Mab 41 D3 y se incubaron durante 2 h a 20°C. Los inmunocomplejos se lavaron cuatro veces con PBS-Tween20 al 0,1% y una vez con agua.

10 Posteriormente, se realizaron tratamientos con endoglicosidasas H y F, neuraminidasa, O-glicosidasa (Boehringer) y heparinasa I (Sigma). Se realizaron tratamientos con EndoF (16 h a 37°C) después de la desnaturalización utilizando el kit de desglucosilación endoF (Boehringer, nº de cat. 1836552) y 2,5-4 U de endoF (Boehringer, nº de cat. 1 365 169) o sin pre-desnaturalización, en PBS. El tratamiento con endoH (16 h a 37°C) se realizó en NaAc 50 mM pH 5,5, SDS al 0,02%, 2-mercaptoetanol 0,1M y 10mU de endoH (Boehringer 1088726). El tratamiento con neuraminidasa (16 h a 37°C) en NaAc 50 mM, pH 5,5, CaCl<sub>2</sub> 4 mM, BSA 100  $\mu$ g/ml, neuraminidasa 2mU (Boehringer 1080725), seguido de incubación de 2mU de O-glucosidasa (Boehringer 1347101) en PBS durante 16 h a 37°C. También se realizó una digestión combinada de endoglicosidasas utilizando endoF (2,5U), neuraminidasa (2mU) y O-glucosidasa (2mU) en PBS durante 16 h a 37°C. La digestión con heparinasa I (Sigma H2519) se realizó a 37°C durante 16 h en PBS utilizando 5 U de la enzima, como en Keil et al., 1996, J. Virol. 70, 3032-3038. La transferencia Western se llevó a cabo según lo recomendado (Amersham), excepto que la etapa de detección se realizó utilizando reactivo DAB (Sigma) (Sambrook J. et al., Molecular Cloning: a Laboratory manual, 2ª ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York, 1989).

20

## RESULTADOS:

25 MAb 41 D3 no se pudo unir a p210 en una transferencia Western en condiciones reductoras. Por lo tanto, los autores de la invención marcaron p210 con biotina, y después de la inmunoprecipitación con MAb 41 D3, se realizó un experimento de transferencia Western utilizando estreptavidina conjugada con peroxidasa de rábano picante. Se realizaron diferentes tratamientos en p210 inmunoprecipitado, marcado con biotina.

## 1) migración en condiciones reductoras y no reductoras:

30 En presencia de 2-mercaptoetanol, el p210 biotilado migró como una única banda de 220 kDa. En condiciones no reductoras, se observó una banda difusa principal que migra como una proteína de 180 kDa junto con una banda menor de peso molecular aparente mucho más alto. Como no se pudieron detectar otras proteínas, los autores de la invención sugirieron que p210 se produjo como un multímero (banda menor, probablemente un dímero) y un monómero (banda principal) que se enlazaron mediante puentes disulfuro intermoleculares, dando como resultado una única banda después de la reducción. Además, los datos también sugieren la presencia de puentes disulfuro intramoleculares, ya que la proteína reducida migró mucho más lentamente que el monómero no reducido.

35

## 2) Tratamiento con endoglicosidasa H, F y heparinasa I

Para analizar la glicosilación de p210, los autores de la invención trataron la proteína inmunoprecipitada con endoglicosidasas H (endoH) o F (endoF). La falta de sensibilidad a endoH sugirió que p210 no contenía glicanos híbridos o de manosa alta y que los n-glicanos presentes en p210 eran del tipo complejo. Se observó un

desplazamiento importante de aproximadamente 30 kDa en presencia de endoF, produciendo una proteína de 180 kDa en condiciones reductoras. En condiciones no reductoras, observaron un desplazamiento en el peso molecular tanto del dímero como del monómero, dando como resultado la aparición de dos bandas de alta movilidad. Estos resultados indicaron que p210 estaba muy modificado con glucanos enlazados a N y proporcionaban una indicación adicional de que el dímero era un complejo de p210.

Debido a que se demostró que la unión de PRRSV a células MARC se inhibió cuando estas células se trataron con heparinasa I, también se evaluó la sensibilidad a la heparinasa I de p210. No se observó desplazamiento en SDS-PAGE, lo que indica la ausencia de restos tipo heparina sensibles a heparinasa I en p210.

### 3) Tratamiento con O-glicosidasa y Neuraminidasa

p210 inmunoprecipitado se trató con Neuraminidasa para detectar la presencia de residuos de ácido siálico. No se observó desplazamiento en SDS-PAGE, lo que indica la falta de cantidades importantes de residuos de ácido siálico en p210. Además, el tratamiento posterior con O-glicosidasa no dio como resultado un desplazamiento del peso molecular aparente, lo que indica la falta de O-glicanos sensibles a la escisión por O-glicosidasa en p210.

#### Ejemplo 5: Clonación y secuenciación del ADNc de p210.

El ADN se preparó a partir de sangre entera de cerdo tal como se describe en Innis et al. (en PCR protocols. A guide to methods and applications., Academic Press Inc., Harcourt Brace Jovanovich, Publishers, 1990) y sirvió como objetivo inicial para experimentos de PCR utilizando oligonucleótidos no degenerados basados en los péptidos 3 y 4. Los códigos del nucleótido degenerado en los cebadores se eligieron de forma similar a la secuencia de nucleótidos correspondientes a los péptidos p210 en ratón (número de acceso: EMBL Z36293) o en ser humano (número de acceso: EMBL AL 109804). Dos cebadores derivados de ratón (directo 5' TCCTCAACTGCAGCCTCTGT e inverso 3' AGTGAGGCAGCCGTTCCCTC) amplificaron un fragmento de 340 nt correspondiente al extremo del exón 14, un intrón, y el exón 15 del gen de sialoadhesina de ratón.

Oligonucleótidos porcinos específicos se derivaron de esta primera secuencia y se utilizaron para rastrear un banco de BAG de cerdo (Rogel-Gaillard et al., Cytogenet Cell genet., 85, 205-211, 1999) mediante PCR. Se pudo seleccionar un clon (BAC634C10) y se utilizó para un análisis adicional de la secuencia del gen p210 de cerdo. Utilizando técnicas estándares, se obtuvo información de la secuencia del exón 4 al 18 del gen p210.

Para determinar la secuencia de los exones 1-4 y los exones 18-21 se realizó una RACE 5' y 3' utilizando ARN de macrófago alveolar de cerdo total de acuerdo con las instrucciones del fabricante (GibcoBRL, protocolos RACE para ADNc rico en GC).

Para el 5' RACE se sintetizó el ADNc de la primera cadena utilizando un cebador inverso derivado del exón4 (5' TCTGGTCTTTGAGCTTCGTC 3') y TdT con cola de dCTP. La síntesis de la segunda cadena se realizó utilizando el cebador de anclaje abreviado 5'-RACE suministrado y un cebador anidado derivado del exón4 (5' ACCTGAGGGTTGCTGCTATT 3'). Se realizó una PCR semianidada utilizando el cebador de amplificación universal abreviado (AUAP) suministrado y un cebador anidado derivado también del exón 4 (5' cacctggcagctgagggtgaccagatc 3').

Para el 3' RACE se utilizó el cebador adaptador suministrado para preparar el ADNc de la primera cadena y un cebador directo derivado del exón 18 (5' GACGCCACCATGACTGIIIIIG 3') junto con el cebador AUAP suministrado para la PCR semianidada.

La secuencia de ADN completa de los exones correspondientes a la secuencia codificante de p210 se ensambló a partir de los datos de PCR, RT-PCR, 5'RACE y 3'RACE (Figura 3 y SEQ ID 1-2).

Para la secuenciación del ADNc de p210 y la clonación en un vector de expresión eucariótico, se aisló en primer lugar el ARN total de los macrófagos.

Para aislar el ARN total y para obtener cantidades óptimas del ARN que codifica p210, se lisaron 5x10<sup>7</sup> macrófagos directamente en el recipiente de cultivo y se extrajo el ARN utilizando el kit Rneasy (Quiagen). La primera cadena se hizo utilizando RNasa H-Transcriptasa inversa Superscript II (GibcoBRL) y nonameros aleatorios 2,5 µM (Sigma) siguiendo las instrucciones del fabricante (GibcoBRL). Para la PCR, se mezclaron 0,4 µM de cebador directo 5' CACCATGGACTTCCTGCTCCTGCTCCTC 3' y 0,4 µM de cebador inverso 5' CTTGGGGTTTGAAGCTAGGTCATAA 3' con 200 µM de cada dNTP, 1 U de ThermalAce DNA polimerasa y tampón de reacción ThermalAce (Invitrogen). El ciclo térmico se realizó de la siguiente manera: 3 min a 95°C y luego 30 ciclos de {20 s a 95°C, 30 s a 65°C y 5,3 min a 74°C} y finalmente 10 min a 74°C. Utilizando estos parámetros, se amplificó el ADNc de p210 total de 5,2 kb. El ADN amplificado se purificó a partir de un gel de agarosa (GeneClean) y TOPO se clonó en el vector de expresión eucariótico (Invitrogen) de acuerdo con las instrucciones del fabricante. Después de la transformación (cepa TOP10 de E. coli, Invitrogen), se seleccionaron 9 colonias para la presencia del gen clonado en la orientación esperada. En

este vector, la secuencia codificante se expresa como una proteína sin fusión del promotor de CMV. Finalmente, se secuenció el clon que producía un gen p210 funcional y se determinó la secuencia de ADNc del gen p210 (Figura 4).

Ejemplo 6: Transfección de células PK-15 y detección de la expresión del receptor p210 de PRRSV y cultivo de PRRSV.

5 Se transfectaron células PK-15 al 50% de confluencia con plásmido pcDNA3.1D/V5-His-TOPO purificado que contenía el ADNc de p210 utilizando la técnica de fosfato de calcio (kit de transfección CellPfect, Pharmacia). Las células transfectadas se tripsinizaron 72 h después de la transfección y se cultivaron en presencia de 1 mg/ml de geneticina (GibcoBRL). Diez días después de la adición de la geneticina, se aislaron colonias de células resistentes a la geneticina utilizando papeles de filtro 3MM estériles empapados en tripsina. Después del cultivo de estas colonias durante 2 semanas adicionales en presencia de geneticina, se verificó la expresión de p210 recombinante.

10 Para verificar la expresión de p210, las células se fijaron con metanol durante 10 min a 4°C, se rehidrataron en PBS y se incubaron durante 1 h a 37°C con MAb p210 (ascitis diluida 1/300 en PBS suplementado con suero de cabra descomplementado al 10%). Después de lavar con PBS, las células se incubaron durante 1 h a 37°C con una Ig anti-ratón en suero de cabra marcado con FITC (Molecular Probes) diluido 1/100 en PBS complementado con suero de cabra descomplementado al 10%. Después de lavar con PBS y agua, las células se secaron y se observaron bajo un microscopio de fluorescencia.

15 Se detectó una tinción positiva de grupos de células en monocapas derivadas de las transfecciones realizadas con 3 de los 9 plásmidos seleccionados.

20 Para evaluar la expresión superficial de p210, se utilizó el mismo protocolo de tinción que el descrito anteriormente, excepto que las células no se fijaron, el PBS se reemplazó por medio de cultivo (MEM sin aditivos), se añadió azida sódica al 0,1% a los anticuerpos y se realizaron etapas de incubación a 4°C para bloquear la endocitosis de anticuerpos. De los 9 plásmidos utilizados, solo 1 mostró expresión en superficie de p210 en grupos de células.

25 Para verificar si la expresión de p210 haría que la célula fuera permisiva para el PRRS, las células se infectaron en un volumen de cultivo mínimo con diferentes cepas de PRRSV a una multiplicidad de infección de 0,1 TCID50/célula. Después de 1 h de incubación, se añadió medio de cultivo y las células se incubaron a 37°C durante 20 h más. Después de lavar con PBS, las células se fijaron con metanol, se lavaron con PBS y se incubaron durante 1 h a 37°C con MAb A27 específico para PRRSV (sobrenadante de hibridoma complementado con suero de cabra descomplementado al 10%). Después de lavar con PBS, las células se incubaron durante 1 h a 37°C con Ig anti-ratón en suero de cabra marcada con FITC (Molecular Probes) complementada con suero de cabra descomplementado al 10%. Después de lavar con PBS y agua, las células se secaron y se observaron bajo un microscopio de fluorescencia (filtro FITC). Después, las células se lavaron en PBS y se incubaron durante 1 h a 37°C con Mab 41 D3 biotinilado en PBS. Después de lavar con PBS, las células se incubaron durante 1 h a 37°C con estreptavidina marcada con rojo Texas (Molecular Probes) diluida en PBS. Después de lavar con PBS y agua, las células se secaron y se observaron bajo un microscopio de fluorescencia (filtro Texas-Red).

35 Células infectadas positivas se observaron solo en células transfectadas con p210, en donde p210 se mostraba en la superficie de las células. Además, todas las cepas de virus incubadas en esta célula (cepa LV, VR2332 y dos cepas de PRRSV de tipo salvaje de Bélgica) pudieron infectar la célula que expresa p210.

Ejemplo 7: Transfección de la línea celular humana HEK293T, detección de la expresión del receptor de PRRSV p210 y cultivo de PRRSV.

40 Células HEK293T se transfectaron al 25% de confluencia con el plásmido QiagenpcDNA3.1 D/V5-His-TOPO (Invitrogen) que contenía el ADNc de p210 utilizando la técnica de fosfato de calcio (kit de transfección CellPfect, Amersham Pharmacia Biotech). Dieciséis horas después de la transfección, las células se inocularon con el PRRSV de tipo americano VR-2332 (cultivadas en células Marc-145), con el PRRSV de tipo europeo virus Lelystad (cultivadas en macrófagos alveolares porcinos) y con el aislado belga 94V360 (cultivado en células Marc-145).

45 Ocho horas después de la infección, las células se fijaron con metanol durante 10 minutos a 4°C, se rehidrataron en PBS, y tanto las células infectadas como las células que expresaban p210 se identificaron utilizando una tinción de inmunofluorescencia doble. Para detectar la infección por PRRSV, las células se incubaron durante 1 hora a 37°C con Mab A27 (sobrenadante de cultivo diluido 1/100 en PBS) que reconoce la proteína de la nucleocápside de PRRSV. Las células se lavaron con PBS y se incubaron con suero anti-ratón de cabra marcado con FITC diluido 1/100 en PBS. Las células HEK293T que expresan la proteína p210 se detectaron incubando las células durante 1 hora a 37°C con Mab 41 D3 biotinilado (ascitis diluida 1/200 en PBS), que se dirige contra la proteína p210. Después de lavar con PBS, las células se incubaron adicionalmente durante 1 hora a 37°C con estreptavidina Texas Red, diluida 1/50 en PBS. Finalmente, las células se lavaron con PBS y se montaron en una solución de glicerina tamponada que contenía DABCO al 2,5% (Janssen Chimica). Las células infectadas con PRRSV (verde) y las

células que expresaban p210 (rojo) se visualizaron mediante microscopía de fluorescencia. Las células HEK293T transfectadas se podían infestar tanto con el tipo americano como con los dos tipos europeos del PRRSV.

**REIVINDICACIONES**

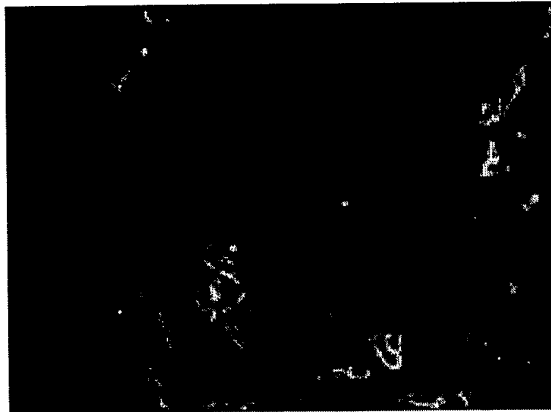
1. Un polinucleótido aislado que comprende una secuencia de ácidos nucleicos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos un 90% de identidad en toda su longitud con una secuencia de aminoácidos tal como se representa en SEQ ID NO: 2, o un fragmento funcional de dicho polipéptido de al menos 50 aminoácidos, teniendo dicho polipéptido las siguientes características:
- dicho polipéptido o fragmento de polipéptido, cuando se expresa en la superficie de una célula, es capaz de hacer que la célula sea receptiva para el PRRSV,
  - dicho polipéptido tiene un peso molecular aparente de aproximadamente 210 kD y
  - dicho polipéptido es reactivo con MAB 41 D3 depositado en la CNCM del Instituto Pasteur, bajo el nº de acceso I-2719.
2. Un polinucleótido de acuerdo con la reivindicación 1, que codifica una secuencia de aminoácidos que tiene un 100% de identidad a lo largo de toda su longitud con una secuencia de aminoácidos tal como se representa en SEQ ID NO: 2.
3. Un polinucleótido de acuerdo con cualquiera de las reivindicaciones 1 o 2, teniendo dicho polinucleótido al menos 90% de identidad a lo largo de toda su longitud con la secuencia de ácidos nucleicos tal como se representa en SEQ ID NO: 1.
4. Un polinucleótido de acuerdo con cualquiera de las reivindicaciones precedentes, teniendo dicho polinucleótido un 100% de identidad a lo largo de toda su longitud con la secuencia de ácidos nucleicos tal como se representa en SEQ ID NO: 1.
5. Un polinucleótido de acuerdo con cualquiera de las reivindicaciones 1-4, que comprende con la secuencia de ácidos nucleicos tal como se representa en SEQ ID NO: 1.
6. Un vector recombinante, que comprende un polinucleótido de acuerdo con cualquiera de las reivindicaciones 1-5.
7. Un vector recombinante de acuerdo con la reivindicación 6, que comprende un polinucleótido de acuerdo con cualquiera de las reivindicaciones 1-5, enlazado operativamente a una secuencia de control de la expresión.
8. Un virus recombinante, que comprende un polinucleótido de acuerdo con cualquiera de las reivindicaciones 1-5.
9. Una célula que comprende un vector de expresión, caracterizada por que dicho vector de expresión comprende un polinucleótido según se describe en cualquiera de las reivindicaciones 1-5.
10. Un método para producir un polipéptido, codificado por un polinucleótido de acuerdo con cualquiera de las reivindicaciones 1-5. Que comprende las etapas de:
- cultivar células de acuerdo con la reivindicación 9 o células infestadas con un virus recombinante de acuerdo con la reivindicación 8, y
  - aislar material que contiene polipéptido del cultivo celular.
11. Un cultivo celular, que comprende una multitud de células de acuerdo con la reivindicación 9.
12. Un cultivo celular de acuerdo con la reivindicación 11, infestado con un virus PRRS.
13. Un método para producir PRRSV, que comprende las etapas de
- infestar un cultivo celular de acuerdo con la reivindicación 11 con un PRRSV,
  - cultivar las células, y
  - recoger el virus del cultivo celular.
14. Uso de una célula transformada con el polinucleótido de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en la fabricación de una vacuna para la prevención de una infección por el PRRSV en cerdos.
15. Uso de un polipéptido codificado por un polinucleótido de acuerdo con cualquiera de las reivindicaciones 1-5, en la fabricación de un medicamento para el tratamiento o la prevención de una infección por el PRRSV en cerdos.

FIGURA 1

A.



B.



C.

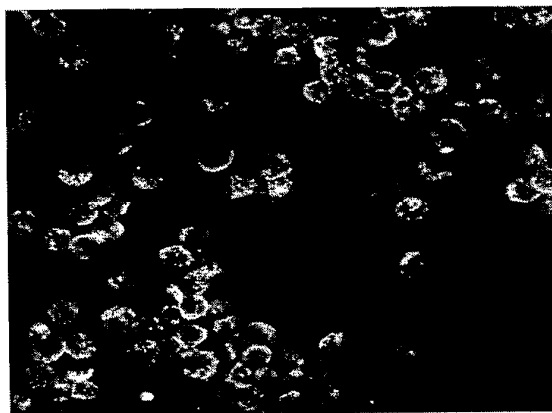
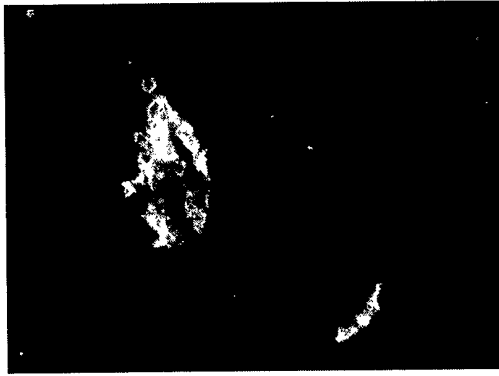
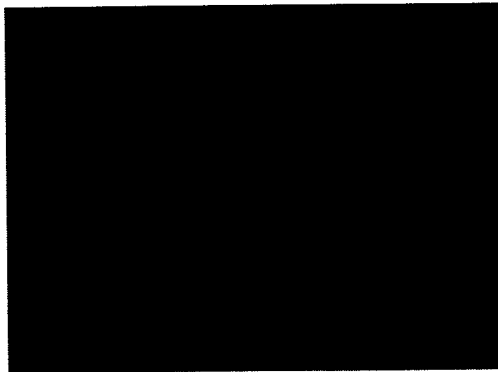


FIGURA 2

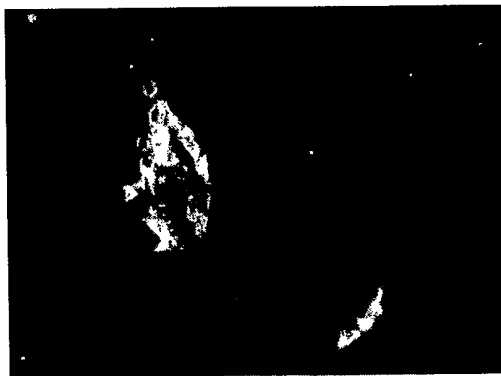
A.



B.



C.



**FIGURA 3**

```

          10          30          50
          .          .          .
atggacttctctgctcctgctcctcctcctggttcctctgctctagcaggcctggcctcg
1  -----+-----+-----+-----+-----+-----+ 60
tacctgaaggacgaggacgaggaggaggaccgaagtagacgagatcgtccggaccggagc
M D F L L L L L L L A S S A L A G L A S

          70          90          110
          .          .          .
tggacggtttccagccccgagaccgtgcagggcatcaagggctcctgcctcatcatcccc
61 -----+-----+-----+-----+-----+ 120
acctgccaaaggtcggggctctggcacgtcccgtagttcccaggacggagtagtagggg
W T V S S P E T V Q G I K G S C L I I P

          130          150          170
          .          .          .
tgcaccttcggcttccccggccaactggaggtgccccatggcatcacagccatctggtac
121 -----+-----+-----+-----+-----+ 180
acgtggaagccgaagggccgggtgcacctccacggggtaccgtagtgctcggtagaccatg
C T F G F P A N V E V P H G I T A I W Y

          190          210          230
          .          .          .
tatgactactcaggcaagcgctggtagtgagccactccaggaacccaaggtggtggag
181 -----+-----+-----+-----+-----+ 240
atactgatgagtcogttcgcggaccatcactcggtgaggtccttgggtttccaccacctc
Y D Y S G K R L V V S H S R N P K V V E

          250          270          290
          .          .          .
aaccacttccaagggccggccctgctgttgGGGCAGGTTGAACAGAGGACGTGCAGCCTG
241 -----+-----+-----+-----+-----+ 300
ttggtgaaggttccggccccgggacgacaacCCCGTCCAACTTGTCTCCTGCACGTCCGGAC
N H F Q G R A L L L G Q V E Q R T C S L

          310          330          350
          .          .          .
CTGCTGAAGGACCTGCAGCCCCAGGACTCGGGCTCCTATAACTTCCGCTTTGAGATCAGC
301 -----+-----+-----+-----+-----+ 360
GACGACTTCCCTGGACGTCCGGGTCCTGAGCCCCGAGGATATTGAAGGCGAAACTCTAGTCG
L L K D L Q P Q D S G S Y N F R F E I S

          370          390          410
          .          .          .
GAGGGCAACCGCTGGTCAGATGTCAAAGGCACAGTTGTACCCGTGACAGAGGTGCCCCAGC
361 -----+-----+-----+-----+-----+ 420
CTCCCGTTGGCGACCAGTCTACAGTTTCCGTGTCAACAGTGGCACTGTCTCCACGGGTCCG
E G N R W S D V K G T V V T V T E V P S
    
```

ES 2 687 029 T3

```

          430              450              470
          .                .                .
GTGCCCACCATTGCCTTGCCAGCCAAGCTGCATGAGGGCATGGAGGTGGACTTCAACTGC
421 -----+-----+-----+-----+-----+-----+-----+ 480
          CACGGGTGGTAACGGAACGGTTCGGTTCGACGTACTCCCGTACCTCCACCTGAAGTTGACG
          V P T I A L P A K L H E G M E V D F N C

          490              510              530
          .                .                .
TCCACTCCCTATGTGTGCCCGACGGAGCCGGTCAACCTACAGTGGCAAGGCCAGGATCCC
481 -----+-----+-----+-----+-----+-----+-----+ 540
          AGGTGAGGGATACACACGGGCTGCCTCGGCCAGTTGGATGTCAACGTTCCGGTCCTAGGG
          S T P Y V C P T E P V N L Q W Q G Q D P

          550              570              590
          .                .                .
ACCCGCTCCGTCACCTCCACCTCCAGAAGCTTGAGCCCTCGGGCACCAGCCACATGGAG
541 -----+-----+-----+-----+-----+-----+-----+ 600
          TGGGCGAGGCAGTGGAGGGTGGAGGTCTTCGAACTCGGGAGCCCGTGGTGGTGTACCTC
          T R S V T S H L Q K L E P S G T S H M E

          610              630              650
          .                .                .
ACCCTGCACATGGCCCTGTCTTGGCAGGACCATGGCCGGATCCTGAGCTGCCAGGTCTCA
601 -----+-----+-----+-----+-----+-----+-----+ 660
          TGGGACGTGTACCGGGACAGGACCGTCTCTGGTACCGCCTAGGACTCGACGGTCCAGAGT
          T L H M A L S W Q D H G R I L S C Q V S

          670              690              710
          .                .                .
GCAGCCGAACGCAGGATGCAGAAGGAGATTCACCTCCAAGTGCAGTATGCCCCAAGGGT
661 -----+-----+-----+-----+-----+-----+-----+ 720
          CGTCGGCTTGGCTCCTACGTCTTCTCTAAGTGGAGGTTACGTCATACGGGGTTCCCA
          A A E R R M Q K E I H L Q V Q Y A P K G

          730              750              770
          .                .                .
GTGGAGATCCTTTTCAGCCACTCCGGACGGAACGTCCTTCCAGGTGATCTGGTCACCCCTC
721 -----+-----+-----+-----+-----+-----+-----+ 780
          CACCTCTAGGAAAAGTCGGTGGAGCCTGCCTTGCAGGAAGGTCCACTAGACCAGTGGGAG
          V E I L F S H S G R N V L P G D L V T L

          790              810              830
          .                .                .
AGCTGCCAGGTGAATAGCAGCAACCCTCAGGTCAGTTCGGTGCAGTGGGTCAAGGATGGG
781 -----+-----+-----+-----+-----+-----+-----+ 840
          TCGACGGTCCACTTATCGTCGTTGGGAGTCCAGTCAAGGCACGTCACCCAGTTCCTACCC
          S C Q V N S S N P Q V S S V Q W V K D G

```

ES 2 687 029 T3

```

      850              870              890
      .               .               .
ACGAAGCTCAAAGACCAGAAACGTGTACTGCAGTTGCGCCGGGCAGCCTGGGCTGATGCT
841 -----+-----+-----+-----+-----+-----+-----+ 900
TGCTTCGAGTTTCTGGTCTTTGCACATGACGTCAACGCGGCCCGTCGGACCCGACTACGA
      T K L K D Q K R V L Q L R R A A W A D A

      910              930              950
      .               .               .
GGCGTCTACACCTGCCAAGCCGGGAATGCCGTGGGCTCTTCAGTCTCACCCCCGGTCAGC
901 -----+-----+-----+-----+-----+-----+-----+ 960
CCGCAGATGTGGACGGTTCGGCCCTTACGGCACCCGAGAAGTCAGAGTGGGGCCAGTCC
      G V Y T C Q A G N A V G S S V S P P V S

      970              990              1010
      .               .               .
CTCCACGTCTTCATGGCTGAGGTCCAGGTAAGCCCTGTGGGCTCCATCCTGGAGAACCAG
961 -----+-----+-----+-----+-----+-----+-----+ 1020
GAGGTGCAGAAGTACCGACTCCAGGTCCATTTCGGACACCCGAGGTAGGACCTTTGGTC
      L H V F M A E V Q V S P V G S I L E N Q

      1030             1050             1070
      .               .               .
ACGGTGACGCTGGCCTGCAATACACCTAAGGAAGCGCCAGCGAGCTGCGCTACAGCTGG
1021 -----+-----+-----+-----+-----+-----+-----+ 1080
TGCCACTGCGACCCGGACGTTATGTGGATTCTTCGCGGGTCGCTCGACGCGATGTCGACC
      T V T L A C N T P K E A P S E L R Y S W

      1090             1110             1130
      .               .               .
TACAAGAACCACGCCCTGCTGGAGGGCTCTCACAGCCGACCCCTCCGGCTGCACTCAGTT
1081 -----+-----+-----+-----+-----+-----+-----+ 1140
ATGTTCTTGGTGCGGGACGACCTCCCAGAGTGTTCGGCGTGGGAGGCCGACGTGAGTCAA
      Y K N H A L L E G S H S R T L R L H S V

      1150             1170             1190
      .               .               .
ACCAGGGCGGATTCGGGCTTCTACTTCTGCGAGGTGCAGAACCCCGGGGCAGAGAGCGC
1141 -----+-----+-----+-----+-----+-----+-----+ 1200
TGGTCCCGCCTAAGCCCGAAGATGAAGACGCTCCACGTCCTGCGGGCCCGTCTCTCGCG
      T R A D S G F Y F C E V Q N A R G R E R

      1210             1230             1250
      .               .               .
TCTCCCCCTGTGAGGTGGTGGTCAGCCACCCACCCCTCACCCGGACCTAACTGCCTTC
1201 -----+-----+-----+-----+-----+-----+-----+ 1260
AGAGGGGGACAGTCCGACCCAGTCCGTGGGTGGGAGTGGGGCCTGGATTGACGGAAG
      S P P V S V V V S H P P L T P D L T A F

```

ES 2 687 029 T3

```

1261      1270      1290      1310
      .           .           .           .
      CTGGAGACACAGGCGGGGCTGGTGGGCATCCTCCAATGCTCTGTGGTCAGCGAGCCCCCA
      -----+-----+-----+-----+-----+-----+
      GACCTCTGTGTCCGCCCCGACCACCCGTAGGAGGTTACGAGACACCAGTCGCTCGGGGGT
      L E T Q A G L V G I L Q C S V V S E P P
1321      1330      1350      1370
      .           .           .           .
      GCTACTCTGGTGTGTGCACACGGGGGCTCATCTTGGCCTCTACCTCCGGGGAGGGTGCAC
      -----+-----+-----+-----+-----+
      CGATGAGACCACAACAGTGTGCCCCGGAGTAGAACCGGAGATGGAGGCCCTCCCACTG
      A T L V L S H G G L I L A S T S G E G D
1381      1390      1410      1430
      .           .           .           .
      CACAGCCCACGCTTCAGTGTGCTCTGCCCCCAACTCCCTGCGCCTGGAGATTCAAGAC
      -----+-----+-----+-----+-----+
      GTGTCCGGTGCGAAGTCACAGCGGAGACGGGGTTGAGGGACCGGACCTCTAAGTTCGT
      H S P R F S V A S A P N S L R L E I Q D
1441      1450      1470      1490
      .           .           .           .
      CTGGGGCCAACAGACAGTGGGAATACATGTGCTCAGCCAGCAGTTCCTTGGGAATGCG
      -----+-----+-----+-----+-----+
      GACCCCGGTTGTCTGTACCCCTTATGTACACGAGTCGGTCGTCAAGAGAACCCTTACGC
      L G P T D S G E Y M C S A S S S L G N A
1501      1510      1530      1550
      .           .           .           .
      TCCTCCACCCTGGACTTCCATGCCAATGCAGCCCGCCTCCTCATCAGCCAGCAGCAGAG
      -----+-----+-----+-----+-----+
      AGGAGGTGGGACCTGAAGGTACGTTACGTGCGGGCGGAGGAGTAGTCGGGTCGTCTC
      S S T L D F H A N A A R L L I S P A A E
1561      1570      1590      1610
      .           .           .           .
      GTGGTGAAGGGCAGGCGGTGACACTGAGCTGCAGGAGCAGCCTGAGCCTGATGCCTGAC
      -----+-----+-----+-----+-----+
      CACCACCTTCCCGTCCGCCACTGTGACTCGAGCTCCTCGTCGGACTCGGACTACGGACTG
      V V E G Q A V T L S C R S S L S L M P D
1621      1630      1650      1670
      .           .           .           .
      ACCCGTTTTTCCTGGTACCTGAACGGGGCCCTGATTCTCGAGGGGCCAGCAGCAGCCTC
      -----+-----+-----+-----+-----+
      TGGGCAAAAAGGACCATGGACTTGCCCCGGGACTAAGAGCTCCCCGGGTCGTCTCGGAG
      T R F S W Y L N G A L I L E G P S S S L

```

ES 2 687 029 T3

```

          1690              1710              1730
          .               .               .
CTGCTCCCAGCAGCCTCCAGCACAGATGCCGGCTCATACCACTGCCGGGCCAGAACAGC
1681 -----+-----+-----+-----+-----+-----+-----+
          GACGAGGGTCGTCGGAGGTCGTGTCTACGGCCGAGTATGGTGACGGCCCGGGTCTTGTGC
          L L P A A S S T D A G S Y H C R A Q N S

          1750              1770              1790
          .               .               .
CACAGCACCAGCGGGCCCTCCTCACCTGCTGTTCTCACCGTGCTCTACGCCCCAGCCAG
1741 -----+-----+-----+-----+-----+-----+-----+
          GTGTGCTGGTCGCCCGGGAGGAGTGGACGACAAGAGTGGCAGAGATGCGGGGTGCGGTC
          H S T S G P S S P A V L T V L Y A P R Q

          1810              1830              1850
          .               .               .
CCCCTGTTCACTGCCAGCTGGACCTGATACTGCAGGAGCTGGGGCCGGACGCCAAGGC
1801 -----+-----+-----+-----+-----+-----+-----+
          GGGCACAAGTGACGGGTCGACCTGGGACTATGACGTCCTCGACCCCGGCTGCGGTTCCG
          P V F T A Q L D P D T A G A G A G R Q G

          1870              1890              1910
          .               .               .
CTCCTCTTGTGCCGTGTGGACAGCGACCCCCAGCCAGCTGCAGCTGCTCCACAGGGGC
1861 -----+-----+-----+-----+-----+-----+-----+
          GAGGAGAACACGGCACACCTGTGCTGGGGGTCGGGTCGACGTCGACGAGGTGTCCCCG
          L L L C R V D S D P P A Q L Q L L H R G

          1930              1950              1970
          .               .               .
CGTGTGTGGCCTCTTCTCTGTGTCATGGGGGGCGGCTGCTGCACCTGCGGAGGCTGTTTC
1921 -----+-----+-----+-----+-----+-----+-----+
          GCACAACACCGGAGAAGAGACAGTACCCCCCGCCGACGACGTCGACGCCTCCGACAAG
          R V V A S S L S W G G G C C T C G G C F

          1990              2010              2030
          .               .               .
CACCGCATGAAGGTCACCAAAGCACCAACCTACTGCGTGTAGAGATCCGAGACCCGGTG
1981 -----+-----+-----+-----+-----+-----+-----+
          GTGGCGTACTTCCAGTGGTTTCGTGGGTTGGATGACGCACATCTCTAGGCTCTGGGCCAC
          H R M K V T K A P N L L R V E I R D P V

          2050              2070              2090
          .               .               .
CTGGAGGATGAGGGTGTGTACCTGTGCGAGGCCAGCAGCGCCCTGGGCAACGCCTCCGCC
2041 -----+-----+-----+-----+-----+-----+-----+
          GACCTCCTACTCCCACACATGGACACGCTCCGGTCGTCGCGGGACCCGTTGCGGAGGCGG
          L E D E G V Y L C E A S S A L G N A S A
    
```

ES 2 687 029 T3

```

                2110                2130                2150
                .                   .                   .
TCTGCAACCTTGGATGCCAGGCCACTGTCCTGGTCATCACACCGTCACACACGCTGCAG
2101 -----+-----+-----+-----+-----+-----+ 2160
AGACGTTGGAACCTACGGGTCCGGTGACAGGACCAGTAGTGTGGCAGTGTGTGCGACGTC
S A T L D A Q A T V L V I T P S H T L Q

                2170                2190                2210
                .                   .                   .
GAAGGCATTGAAGCCAACCTGACTTGCAACGTTGAGCCGTGAAGCCAGCGGCCCTGCCAAC
2161 -----+-----+-----+-----+-----+ 2220
CTTCCGTAAC TTCGGTTGGACTGAACGTTGCACTCGGCAC TTCGGTCGCCGGGACGGTTG
E G I E A N L T C N V S R E A S G P A N

                2230                2250                2270
                .                   .                   .
TTCTCCTGGTTCCGAGATGGGGCGCTATGGGCCAGGGCCCTCTGGACACCGTGACGCTG
2221 -----+-----+-----+-----+-----+ 2280
AAGAGGACCAAGGCTCTACCCCGGATACCCGGTCCCGGGAGACCTGTGGCACTGCGAC
F S W F R D G A L W A Q G P L D T V T L

                2290                2310                2330
                .                   .                   .
CTACCTGTGGCCAGAACTGATGCTGCCCTCTATGCTTGCCGCATCGTCACCGAGGCTGGT
2281 -----+-----+-----+-----+-----+ 2340
GATGGACACCGGCTCTTGACTACGACGGGAGATACGAACGGCGTAGCAGTGGCTCCGACCA
L P V A R T D A A L Y A C R I V T E A G

                2350                2370                2390
                .                   .                   .
GCTGGCCTCTCCACCCCTGTGGCCCTGAATGTGCTCTATCCCCCGATCCTCCAAGTTG
2341 -----+-----+-----+-----+-----+ 2400
CGACCGGAGAGGTGGGGACACCGGACTTACACGAGATAGGGGGCTAGGAGGTTTCAAC
A G L S T P V A L N V L Y P P D P P K L

                2410                2430                2450
                .                   .                   .
TCAGCCCTCCTGGACGTGGACCAGGCCACACGGCTGTGTTCTGCTGTACTGTGGACAGT
2401 -----+-----+-----+-----+-----+ 2460
AGTCGGGAGGACCTGCACCTGGTCCCGTGTGCCGACACAAGCAGACATGACACCTGTCA
S A L L D V D Q G H T A V F V C T V D S

                2470                2490                2510
                .                   .                   .
CGCCCTCTTGCCAGTTGGCCCTGTTCCGTGGGGAACACCTCCTGGCCGCCAGCTCGGCA
2461 -----+-----+-----+-----+-----+ 2520
GCGGGAGAACGGGTCAACCGGGACAAGGCACCCCTTGTGGAGGACCGGCGGTGAGCCGT
R P L A Q L A L F R G E H L L A A S S A

```

ES 2 687 029 T3

```
                2530                2550                2570
      .           .           .           .           .
CTCCGGCTCCCCCTCGTGGCCGCTCCAGGCCAAAGCCTCGGCCAACTCCTTGCAGCTA
2521 -----+-----+-----+-----+-----+-----+ 2580
      GAGGCCGAGGGGGAGCACCGGCGGAGGTCCGGTTTTCGGAGCCGGTTGAGGAACGTCGAT
      L R L P P R G R L Q A K A S A N S L Q L

                2590                2610                2630
      .           .           .           .           .
GAGGTCCGAGACTTGAGCCTTGGGGACTCTGGCAGCTACCACTGTGAGGCCACCAACATC
2581 -----+-----+-----+-----+-----+ 2640
      CTCCAGGCTCTGAACTCGGAACCCCTGAGACCGTCGATGGTGA(CTCCGGTGGTTGTAG
      E V R D L S L G D S G S Y H C E A T N I

                2650                2670                2690
      .           .           .           .           .
CTTGGATCAGCCAACACTTCTCTTACCTTCCAGGTCGAGGAGCCTGGGTCCGGGTGTCA
2641 -----+-----+-----+-----+-----+ 2700
      GAACCTAGTCGGTTGTGAAGAGAATGGAAGGTCCAGGCTCCTCGGACCCAGGCCACAGT
      L G S A N T S L T F Q V R G A W V R V S

                2710                2730                2750
      .           .           .           .           .
CCGTCGCCTGAGCTCCAGGAGGCCAGGCTGTGGTCTGAGCTGCCAGGTACCCATAGGG
2701 -----+-----+-----+-----+-----+ 2760
      GGCAGCGGACTCGAGGTCTCCCGGTCCGACACCAGGACTCGACGGTCCATGGGTATCCC
      P S P E L Q E G Q A V V L S C Q V P I G

                2770                2790                2810
      .           .           .           .           .
GTCTGGAGGGGACCTCATATCGTTGGTATCGGGATGGCCAGCCCCTCCAGGAGTCCACT
2761 -----+-----+-----+-----+-----+ 2820
      CAGGACCTCCCCTGGAGTATAGCAACCATAGCCCTACCGTCCGGGAGGTCTCAGGTGA
      V L E G T S Y R W Y R D G Q P L Q E S T

                2830                2850                2870
      .           .           .           .           .
TCGGCCACGCTCCGTTTTGCAGCCATAACTCTGAGCCAGGCTGGAGCCTACCATGCCAA
2821 -----+-----+-----+-----+-----+ 2880
      AGCCGGTGCAGGGCAAACGTCGGTATGAGACTCGGTCCGACCTCGGATGGTAACGGTT
      S A T L R F A A I T L S Q A G A Y H C Q

                2890                2910                2930
      .           .           .           .           .
GCCCAAGCTCCAGGCTCAGCCACCACGACCTGGCTGCCCTGTGAGCCTCCAGGTGACC
2881 -----+-----+-----+-----+-----+ 2940
      CGGGTTCGAGGTCCGAGTCGGTGGTGCCTGGACCGACGGGACAGTCGGAGGTGCACTGG
      A Q A P G S A T T D L A A P V S L H V T
```

ES 2 687 029 T3

```

                2950                2970                2990
                .                .                .
TACGCACCTCGCCAGGCCACACTCACCACCCTGATGGACTCAGGCCTCGGGCGACTGGGC
2941 -----+-----+-----+-----+-----+-----+ 3000
ATGCGTGGAGCGGTCCGGTGTGAGTGGTGGGACTACCTGAGTCCGGAGCCCGTGACCCG
Y A P R Q A T L T T L M D S G L G R L G

                3010                3030                3050
                .                .                .
CTCCTTCTGTGCGGTGTGAACAGTGACCCTCCTGCCAGCTCCGACTGCTCCATGGGGAGC
3001 -----+-----+-----+-----+-----+ 3060
GAGGAAGACACGGCACACTTGTCACTGGGAGGACGGGTCGAGGCTGACGAGGTACCCCTCG
L L L C R V N S D P P A Q L R L L H G S

                3070                3090                3110
                .                .                .
CGCCTCGTGGCCTCTACTCTACAAGGTGTGGAGGAGCTTGCAGGCAGCTCTCCCCGCCTA
3061 -----+-----+-----+-----+-----+ 3120
GCGGAGCACCGGAGATGAGATGTTCCACACCTCCTCGAACGTCCGTCGAGAGGGGCGGAT
R L V A S T L Q G V E E L A G S S P R L

                3130                3150                3170
                .                .                .
CAGGTGGCCACAGCCCCAACACGCTGCGCCTGGAGATCCACAACGCAGTGCTGGAGGAT
3121 -----+-----+-----+-----+-----+ 3180
GTCCACCGGTGTGCGGGGTGTGCGACGCGGACCTCTAGGTGTTGCGTCACGACCTCCTA
Q V A T A P N T L R L E I H N A V L E D

                3190                3210                3230
                .                .                .
GAAGCGTCTACACCTGCGAGGCCACCAACACCCTGGGTGAGACCTTGGCCTCCGCCGCC
3181 -----+-----+-----+-----+-----+ 3240
CTTCCGCAGATGTGGACGCTCCGGTGGTGTGGGACCCAGTCTGGAACCGGAGGCGGCGG
E G V Y T C E A T N T L G Q T L A S A A

                3250                3270                3290
                .                .                .
TTCGATGCCAGGCTATGAGAGTGCAGGTGTGGCCCAATGCCACCGTGCAAGAGGGGCAG
3241 -----+-----+-----+-----+-----+ 3300
AAGCTACGGGTCCGATACTCTCACGTCCACACCGGTTACGGTGGCACGTTCTCCCCGTC
F D A Q A M R V Q V W P N A T V Q E G Q

                3310                3330                3350
                .                .                .
CTGGTGAACCTGACCTGCCCTTGTATGGACCACGCACCTGGCCAGCTCACCTACACGTGG
3301 -----+-----+-----+-----+-----+ 3360
GACCCTTGGACTGGACCGAACATACCTGGTGCCTGGACCGGTCGAGTGGATGTGCACC
L V N L T C L V W T T H L A Q L T Y T W

```

ES 2 687 029 T3

```

          3370              3390              3410
          .                .                .
TACCGAGACCAGCAGCAGCTCCCAGGTGCTGCCACTCCATCCTCTGCCCAATGTCACT
3361 -----+-----+-----+-----+-----+-----+-----+ 3420
ATGGCTCTGGTCGTCGTCGAGGGTCCACGACCGGTGAGGTAGGAGGACGGGTACAGTGA
Y R D Q Q Q L P G A A H S I L L P N V T

          3430              3450              3470
          .                .                .
GTCACAGATGCCGCCTCCTACCGCTGTGGCATAATTGATCCCTGGCCAGGCACTCCGCCTC
3421 -----+-----+-----+-----+-----+-----+-----+ 3480
CAGTGTCACGGCGGAGGATGGCGACACCGTATAACTAGGGACCGGTCCGTGAGGCGGAG
V T D A A S Y R C G I L I P G Q A L R L

          3490              3510              3530
          .                .                .
TCCAGACCTGTCGCCCTGGATGTCCTCTACGCACCCCGCAGACTGCGCCTGACCCATCTC
3481 -----+-----+-----+-----+-----+-----+-----+ 3540
AGGTCCTGGACAGCGGACCTACAGGAGATGCGTGGGGCGTCTGACGCGGACTGGGTAGAG
S R P V A L D V L Y A P R R L R L T H L

          3550              3570              3590
          .                .                .
TTGGAGAGCCGTGGTGGGCAGCTGGCCGTGGTGCTGTGCACTGTGGACAGTCGCCAGCT
3541 -----+-----+-----+-----+-----+-----+-----+ 3600
AACCTCTCGGCACCACCCGTCGACCGGCACACGACACGTCGACACCTGTCAGCGGGTCTGA
L E S R G G Q L A V V L C T V D S R P A

          3610              3630              3650
          .                .                .
GCCCAGCTGACCCCTCAGCCATGCTGGCCGCCTCCTGGCCCTCCTCAACCGCAGCCTCTGTC
3601 -----+-----+-----+-----+-----+-----+-----+ 3660
CGGGTCGACTGGGAGTCGGTACGACCGCGGAGGACCGGAGGAGTTGGCGTCGGAGACAG
A Q L T L S H A G R L L A S S T A A S V

          3670              3690              3710
          .                .                .
CCCAACACCCCTGCGCCTGGAGCTGTGGGAGCCCCGGCCAGTGATGAGGGTCTCTACAGC
3661 -----+-----+-----+-----+-----+-----+-----+ 3720
GGGTGTGGGACGCGGACCTCGACACCCFCGGGGCCGGGTCACACTCCCAGAGATGTCG
P N T L R L E L W E P R P S D E G L Y S

          3730              3750              3770
          .                .                .
TGCTCGGCCCGCAGTCTCTGGGCCAGGCCAACACATCCCTGGAGCTGCGGCTAGAGGGC
3721 -----+-----+-----+-----+-----+-----+-----+ 3780
ACGAGCCGGGCGTCAGGAGACCCGGTCCGGTTGTGTAGGGACCTCGACGCCGATCTCCCG
C S A R S P L G Q A N T S L E L R L E G

```

ES 2 687 029 T3

```

          3790              3810              3830
          .                .                .
GTGCAGGTGGCACTGGCTCCATCGGCCACTGTGCCGGAGGGGGCCCTGTCACAGTGACC
3781 -----+-----+-----+-----+-----+-----+ 3840
          CACGTCCACCGTGACCGAGGTAGCCGGTGACACGGCCTCCCCCGGGACAGTGTCACTGG
          V Q V A L A P S A T V P E G A P V T V T

          3850              3870              3890
          .                .                .
TGTGAAGACCTGCTGCCCGCCACCCACTCTCTATGTCTGGTACCACAACAGCCGTTGG
3841 -----+-----+-----+-----+-----+ 3900
          ACACTTCTGGGACGACGGGCGGGTGGGTGAGAGATACAGACCAATGGTGTGTCCGCAACC
          C E D P A A R P P T L Y V W Y H N S R W

          3910              3930              3950
          .                .                .
CTGCAGGAGGGGTCGGCTCCCTCCCTCTCGTTTCCAGCGGCTACACGGGCTCACGCGGGC
3901 -----+-----+-----+-----+-----+ 3960
          GACGTCTCTCCCGACCGGAGGGAGAGCAAAGGTGCGCGATGTGCCCGAGTGCGCCCG
          L Q E G S A A S L S F P A A T R A H A G

          3970              3990              4010
          .                .                .
GCCTATACCTGCCAGGTCCAGGATGCCAGGGCACACGCATCTCCAGCCCGCAGCACTG
3961 -----+-----+-----+-----+-----+ 4020
          CGGATATGGACGGTCCAGTCTACGGGTCCCGTGTGCGTAGAGGGTCCGGCGTGTGAC
          A Y T C Q V Q D A Q G T R I S Q P A A L

          4030              4050              4070
          .                .                .
CACATCCTCTATGCCCTCGGGATGctgtcctttcctccttctgggactcaagggccagc
4021 -----+-----+-----+-----+-----+ 4080
          GTGTAGGAGATACGGGAGCCCTACgacaggaaggaggaagaccctgagttcccggtcg
          H I L Y A P R D A V L S S F W D S R A S

          4090              4110              4130
          .                .                .
cctatggccgtggtacagtgactgtggacagcgagccacctgccgagatgaccctgtcc
4081 -----+-----+-----+-----+-----+ 4140
          ggataccggcaccatgtcacgtgacacctgtcgctcggtggacggctctactgggacagg
          P M A V V Q C T V D S E P P A E M T L S

          4150              4170              4190
          .                .                .
catgatggcaaggtgctggccaccagccatggggtccacggcttagcagtgaggacaggc
4141 -----+-----+-----+-----+-----+ 4200
          gtactaccggtccacgaccggtggtcggtaccccaggtgccgaatcgtaaccctgtccg
          H D G K V L A T S H G V H G L A V G T G
    
```

ES 2 687 029 T3

```

          4210              4230              4250
          .                .                .
catgtccaggtagggcccgcaacgccctgcagctgcgggtgcagaatgtgccctcacgtgac
4201 -----+-----+-----+-----+-----+-----+ 4260
          .                .                .
gtacagggtccaccggcgtagcgggacgtcgacgccacgtcttacacgggagtgactg
          .                .                .
H V Q V A R N A L Q L R V Q N V P S R D

          4270              4290              4310
          .                .                .
aaggacacctactgtctgcatggaccgcaactccttgggctcagtcagcaccatggggcag
4261 -----+-----+-----+-----+-----+ 4320
          .                .                .
ttcctgtggatgcagacgtacctggcgttgaggaacccgagtcagtctgtggtaccccgtc
          .                .                .
K D T Y V C M D R N S L G S V S T M G Q

          4330              4350              4370
          .                .                .
ctgcagccagaagggtgacacgtggtagctgagccagggctggatgtgccctgaaggcaca
4321 -----+-----+-----+-----+-----+ 4380
          .                .                .
gacgtcggctctccacacgtgcaccatcgactcgggtcccgcacctacacggacttccgtgt
          .                .                .
L Q P E G V H V V A E P G L D V P E G T

          4390              4410              4430
          .                .                .
gcgctgaacctgagctgtgcctccctagtgccctgggcacataggcaactccaccttt
4381 -----+-----+-----+-----+-----+ 4440
          .                .                .
cgcgacttgactcgacagcggagggatcaccgggacccgtgtatccgttgaggtggaaa
          .                .                .
A L N L S C R L P S G P G H I G N S T F

          4450              4470              4490
          .                .                .
gcttggttccggaacggtagcagctacacacagagtctgtgccacccttacctcacc
4441 -----+-----+-----+-----+-----+ 4500
          .                .                .
cgaaccaaggccttgccagccgtcgatgtgtgtctcagacacgggtgggaatggaagtgg
          .                .                .
A W F R N G R Q L H T E S V P T L T F T

          4510              4530              4550
          .                .                .
catgtggcccgccccaaagctggtgtaccactgccaggctgagctccccgccccggct
4501 -----+-----+-----+-----+-----+ 4560
          .                .                .
gtacaccgggccccgggttcgaccgaacatggtgacgggtccgactcgaggggcccccgga
          .                .                .
H V A R A Q A G L Y H C Q A E L P A G A

          4570              4590              4610
          .                .                .
gccacctctgctccagtctgctcgggtgctctaccctccaagacgccaccatgact
4561 -----+-----+-----+-----+-----+ 4620
          .                .                .
cgggtggagacgaggtcagaacgagccccacgagatgggagggttctgccccgtgactga
          .                .                .
A T S A P V L L R V L Y P P K T P T M T

```

ES 2 687 029 T3

```

          4630              4650              4670
          .                .                .
gtttttgtggagcccgaggggtggcatccagggcattctggactgccgagtgacagtgag
4621 -----+-----+-----+-----+-----+-----+ 4680
caaaaaacacctcgggctcccaccgtaggtcccgttaagacctgacggctcacctgtcactc
V F V E P E G G I Q G I L D C R V D S E

          4690              4710              4730
          .                .                .
cccctagccagcctgaccctccacctgggcagtcggctggcctccagccagcctcag
4681 -----+-----+-----+-----+-----+ 4740
ggggatcggctcggactgggaggtggaccgtcagccgaccaccggaggtcggctcggagtc
P L A S L T L H L G S R L V A S S Q P Q

          4750              4770              4790
          .                .                .
gctgccctgccaaagccgcacatccgcgtctcagccagtcaccaatgcCTTGCGAGTGGAC
4741 -----+-----+-----+-----+-----+ 4800
cgacggggacggctcggcgtgtaggcgcagagtcggtcagggttacgGAACGCTCACCTG
A A P A K P H I R V S A S P N A L R V D

          4810              4830              4850
          .                .                .
ATGGAGGAGCTGAAGCCCAGTGACCAGGGGAGTATGTGTGCTCGGCCTCCAATGCCCTG
4801 -----+-----+-----+-----+-----+ 4860
TACCTCCTCGACTTCGGGTCACTGGTCCCCCTCATAACACGAGCCGGAGGTTACGGGAC
M E E L K P S D Q G E Y V C S A S N A L

          4870              4890              4910
          .                .                .
GGCTCTGCCTCTGCTGCCACCTACTTCGGAACCAGAGCCCTGCATCGCCTGCATCTGTTC
4861 -----+-----+-----+-----+-----+ 4920
CCGAGACGGAGACGACGGTGGATGAAGCCTTGGTCTCGGGACGTAGCGGACGTAGACAAG
G S A S A A T Y F G T R A L H R L H L F

          4930              4950              4970
          .                .                .
CAGCACCTTCTCTGGTTCCTGGGGCTGCTGGCGAGCCTCCTTCTTCTACTGTGGGCCTG
4921 -----+-----+-----+-----+-----+ 4980
GTCGTGGAAGAGACCAAGGACCCCGACACCCTCGGAGGAGAAGGATGACAACCCGGAC
Q H L L W F L G L L A S L L F L L L G L

          4990              5010              5030
          .                .                .
GGGGTCTGGTACGCCTGGAGACGGGAAATTTTACAAGCTGAGAATGGGCGAATATTCA
4981 -----+-----+-----+-----+-----+ 5040
CCCCAGACCATGCGGACCTCTGCCCTTTAAAAATGTTTCGACTCTTACCCGCTTATAAGT
G V W Y A W R R G N F Y K L R M G E Y S

```

ES 2 687 029 T3

```

                    5050                5070                5090
                    .                   .                   .
gtagagatggatctcgggaaggaaaccacgcagatgtccactgaccaggaagaagtact
5041 -----+-----+-----+-----+-----+-----+ 5100
                    .                   .                   .
catctctaccatagagccttcctttggtgctctacaggtgactggccttcttcaatga
V E M V S R K E T T Q M S T D Q E E V T

                    5110                5130                5150
                    .                   .                   .
ggaatcggatgatgacgggctctgtgaaccaggcggcatttgatcctgccacctctgt
5101 -----+-----+-----+-----+-----+ 5160
                    .                   .                   .
ccttagccactactacgcccgagacacttggccgccaataggacgggtggagaca
G I G D D A G S V N Q A A F D P A H L C

                    5170                5190
                    .                   .
gaaaacacacagtctgtGaaaagcacagtctga
5161 -----+-----+-----+-----+ 5193
                    .                   .
cttttgtgtgcagacaCttttcgtgtcagact
E N T Q S V K S T V *
```

FIGURA 4

```

          10          30          50
          .           .           .
atggacttctctgctcctgctcctcctcctggcttcatccgctctagcaggcctggcctcg
1  -----+-----+-----+-----+-----+-----+ 60
tacctgaaggacgaggacgaggaggaggaccgaagtaggcgagatcgtccggaccggagc
M D F L L L L L L L A S S A L A G L A S

          70          90          110
          .           .           .
tggacggtttccaaccccgagaccgtgcagggcatcaagggctcctgcctcatcatcccc
61  -----+-----+-----+-----+-----+ 120
acctgccaaaggttggggctctggcacgtcccgtagttcccgaggacggagtagtagggg
W T V S N P E T V Q G I K G S C L I I P

          130          150          170
          .           .           .
tgcaccttcggcttcccggccaacgtggaggtgccccatggcatcacagccatctggtac
121  -----+-----+-----+-----+-----+ 180
acgtggaagccgaagggccggttgcacctccacggggtaccgtagtgctcggtagaccatg
C T F G F P A N V E V P H G I T A I W Y

          190          210          230
          .           .           .
tatgactactcaggcaagcgctggtagtgagccactccaggaacccaaaggtggtggag
181  -----+-----+-----+-----+-----+ 240
atactgatgagtcggttcgcgaccatcactcggtgaggtccttggggttccaccacctc
Y D Y S G K R L V V S H S R N P K V V E

          250          270          290
          .           .           .
aaccacttccaagaccgggcccctgctgttggggcaggttgagcagaggacgtgcagcctg
241  -----+-----+-----+-----+-----+ 300
ttggtgaaggttctggccccgggacgacaacCCCGTCCAACCTCGTCTCCTGCACGTCCGAC
N H F Q D R A L L L G Q V E Q R T C S L

          310          330          350
          .           .           .
CTGCTGAAGGACCTGCAGCCCCAGGACTCGGGCTCCTATAACTCCGCTTTGAGATCAGC
301  -----+-----+-----+-----+-----+ 360
GACGACTTCTTGACGTCCGGGTCCTGAGCCCCGAGGATATTGAAGGCGAAACTCTAGTCG
L L K D L Q P Q D S G S Y N F R F E I S

          370          390          410
          .           .           .
GAGGGCAACCGCTGGTCAGATGTCAAAGGCACAGTTGTACCCGTGACAGAGGTGCCACG
361  -----+-----+-----+-----+-----+ 420
CTCCCGTTGGCGACCAGTCTACAGTTTCCGTGTCAACAGTGGCACTGTCTCCACGGGTCG
E G N R W S D V K G T V V T V T E V P S
    
```



ES 2 687 029 T3

```

      850              870              890
      .              .              .
ACGAAGCTCAAAGACCAGAAACGTGTACTGCAGTTGCGCCGGGCAGCCTGGGCTGATGCT
841 -----+-----+-----+-----+-----+-----+-----+ 900
TGCTTCGAGTTTCTGGTCTTTGCACATGACGTCAACGCGGCCCGTCGGACCCGACTACGA
T K L K D Q K R V L Q L R R A A W A D A

      910              930              950
      .              .              .
GGCGTCTACACCTGCCAAGCCGGGAATGCCGTGGGCTCTTCAGTCTCACCCCCGGTCAGC
901 -----+-----+-----+-----+-----+-----+-----+ 960
CCGCAGATGTGGACGGTTCGGCCCTTACCGCACCCGAGAAGTCAGAGTGGGGCCAGTGC
G V Y T C Q A G N A V G S S V S P P V S

      970              990              1010
      .              .              .
CTCCACGTCTTCATGGCTGAGGTCCAGGTAAGCCCTGTGGGCTCCATCCTGGAGAACCAG
961 -----+-----+-----+-----+-----+-----+-----+ 1020
GAGGTGCAGAAGTACCGACTCCAGGTCCATTCCGGACACCCGAGGTAGGACCTCTTGGTC
L H V F M A E V Q V S P V G S I L E N Q

      1030              1050              1070
      .              .              .
ACGGTGACGCTGGCCTGCAATACACCTAAGGAAGCGCCAGCGAGCTGCGCTACAGCTGG
1021 -----+-----+-----+-----+-----+-----+-----+ 1080
TGCCACTGCGACCCGACGTTATGTGGATTCTTCGCGGGTCGCTCGACGCGATGTCGACC
T V T L A C N T P K E A P S E L R Y S W

      1090              1110              1130
      .              .              .
TACAAGAACCACGCCCTCCTGGAGGGCTCTCACAGCCGACCCCTCCGGCTGCACTCAGTC
1081 -----+-----+-----+-----+-----+-----+-----+ 1140
ATGTTCTTGGTGCGGGAGGACCTCCCAGAGTGTGCGCGTGGGAGGCCGACGTGAGTCAG
Y K N H A L L E G S H S R T L R L H S V

      1150              1170              1190
      .              .              .
ACCAGGGCGGATTCGGGCTTCTACTTCTGCGAGGTGCAGAACCCCGGGGCAGAGAGCGC
1141 -----+-----+-----+-----+-----+-----+-----+ 1200
TGGTCCCAGCTAAGCCCGAAGATGAAGACGCTCCACGCTCTGCGGGCCCGTCTCTCGCG
T R A D S G F Y F C E V Q N A R G R E R

      1210              1230              1250
      .              .              .
TCTCCCCCTGTGAGCGTGGTGGTTCAGCCACCCACCCCTCACCCCGACCTAACTGCCTTC
1201 -----+-----+-----+-----+-----+-----+-----+ 1260
AGAGGGGGACAGTCGACACCACAGTCGGTGGGTGGGAGTGGGCCTGGATTGACGGAAG
S P P V S V V V S H P P L T P D L T A F

```

ES 2 687 029 T3

```

1261      1270      1290      1310
      .           .           .           .           .
CTGGAGACACAGGCGGGGCTGGTGGGCATCCTCCAATGCTCTGTGGTCAGCGAGCCCCCA
-----+-----+-----+-----+-----+-----+
1320      GACCTCTGTGTCCGCCCCGACCACCCGTAGGAGGTTACGAGACACCAGTCGCTCGGGGGT
      L E T Q A G L V G I L Q C S V V S E P P

      1330      1350      1370
      .           .           .           .           .
GCTACTCTGGTGTGTGCACACGGGGGCTCATCTTGACCTCTACCTCCGAGGAGGGTGAC
1321      -----+-----+-----+-----+-----+
1380      CGATGAGACCACAACAGTGTGCCCCGGAGTAGAACTGGAGATGGAGGCTCCTCCCACTG
      A T L V L S H G G L I L T S T S E E G D

      1390      1410      1430
      .           .           .           .           .
CACAGCCCACGCTTCAGTGTACCTCTGCCCCCAACTCCCTGCGCCTGGAGATTCAAGAC
1381      -----+-----+-----+-----+-----+
1440      GTGTCGGGTGCGAAGTCACAGTGGAGACGGGGTTGAGGGACGCGGACCTCTAAGTTCG
      H S P R F S V T S A P N S L R L E I Q D

      1450      1470      1490
      .           .           .           .           .
CTGGGGCCAACAGACAGTGGGAATACATGTGCTCAGCCAGCAGTTCCTTGGGAATGCG
1441      -----+-----+-----+-----+-----+
1500      GACCCCGGTTGTCTGTACCCCTTATGTACACGAGTCGGTCGTCAGAGAACCCTTACGC
      L G P T D S G E Y M C S A S S S L G N A

      1510      1530      1550
      .           .           .           .           .
TCCTCCACCCTGGACTTCCATGCCAATGCAGCCCGCCTCCTCATCAGCCAGCAGCAGAG
1501      -----+-----+-----+-----+-----+
1560      AGGAGTGGGACCTGAAGGTACGGTTACGTCGGGCGGAGGAGTAGTCGGGTCGTCGTC
      S S T L D F H A N A A R L L I S P A A E

      1570      1590      1610
      .           .           .           .           .
GTGGTGAAGGGCAGGCGGTGACACTGAGCTGCAGGAGCAGCCTGAGCCTGATGCCTGAC
1561      -----+-----+-----+-----+-----+
1620      CACCACCTTCCCGTCCGCCACTGTGACTCGACGTCCFCGTCGGACTCGGACTACGGACTG
      V V E G Q A V T L S C R S S L S L M P D

      1630      1650      1670
      .           .           .           .           .
ACCCGTTTTTCTGGTACCTGAACGGGGCCCTGATTCTCGAGGGGCCAGCAGCAGCCTC
1621      -----+-----+-----+-----+-----+
1680      TGGGCAAAAAGGACCATGGACTTGCCCCGGGACTAAGAGCTCCCCGGGTCGTCGCGGAG
      T R F S W Y L N G A L I L E G P S S S L

```

ES 2 687 029 T3

```

                1690                1710                1730
                .                .                .
CTGCTCCCAGCAGCCTCCAGCACAGATGCCGGCTCATACCACTGCCGGGCCAGAACAGC
1681 -----+-----+-----+-----+-----+-----+-----+-----+-----+
GACGAGGGTTCGTCGGAGGTCGTGTCTACGCCGAGTATGGTGACGGCCCGGGTCTTGTGC
L L P A A S S T D A G S Y H C R A Q N S

                1750                1770                1790
                .                .                .
CACAGCACCAGCGGGCCCTCCTCACCTGCTGTTCTCACCGTCTCTACGCCCCAGCCAG
1741 -----+-----+-----+-----+-----+-----+-----+-----+
GTGTCTGGTTCGCCCGGGAGGAGTGGACGACAAGAGTGGCACGAGATGCGGGGTGCGGTC
H S T S G P S S P A V L T V L Y A P R Q

                1810                1830                1850
                .                .                .
CCCGTGTTCCTACTGCCAGCTGGACCCCTGATACTGCAGGAGCTGGGGCCGGACGCCAAGGC
1801 -----+-----+-----+-----+-----+-----+-----+-----+
GGGCACAAGTGACGGGTCGACCTGGGACTATGACGTCTCGACCCCGGCTGCGGTTCCG
P V F T A Q L D P D T A G A G A G R Q G

                1870                1890                1910
                .                .                .
CTCCTCTTGTGCCGTGTGGACAGCGACCCCCAGCCAGCTGCAGCTGCTCCACAGGGGC
1861 -----+-----+-----+-----+-----+-----+-----+-----+
GAGGAGAACACGGCACACCTGTGCTGGGGGGTTCGGGTCGACGTCGACGAGGTGTCCCCG
L L L C R V D S D P P A Q L Q L L H R G

                1930                1950                1970
                .                .                .
CGTGTGTGGCCTCTTCTCTGTGATGGGGGGCGGCTGCTGCACCTGCGGAGGCTGTTC
1921 -----+-----+-----+-----+-----+-----+-----+-----+
GCACAACACCGGAGAAGAGACAGTACCCCCCGCGACGCTGGACGCCTCCGACAAG
R V V A S S L S W G G G C C T C G G C F

                1990                2010                2030
                .                .                .
CACCGCATGAAGGTCACCAAGCACCCAACCTACTGCGGTGTAGAGATCCGAGACCCGGTG
1981 -----+-----+-----+-----+-----+-----+-----+-----+
GTGGCGTACTTCCAGTGGTTTCGTGGGTGGATGACGCACATCTCTAGGCTCTGGGCCAC
H R M K V T K A P N L L R V E I R D P V

                2050                2070                2090
                .                .                .
CTGGAGGATGAGGGTGTGTACCTGTGCGAGGCCAGCAGCACCCCTGGGCAACGCCTCCGCC
2041 -----+-----+-----+-----+-----+-----+-----+-----+
GACCTCCTACTCCACACATGGACACGCTCCGGTCTGTCGTTGGGACCCGTTGCGGAGGCGG
L E D E G V Y L C E A S S T L G N A S A

```

ES 2 687 029 T3

```

                2110                2130                2150
                .                    .                    .
TCTGCAACCTTGGATGCCAGGCCACTGTCCTGGTCATCACACCGTCACACACGCTGCAG
2101 -----+-----+-----+-----+-----+-----+ 2160
AGACGTTGGAACCTACGGGTCCGGTGACAGGACCAGTAGTGTGGCAGTGTGTGCGACGTC
S A T L D A Q A T V L V I T P S H T L Q

                2170                2190                2210
                .                    .                    .
GAAGGCATTGAAGCCAACCTGATTTGCAACGTGAGCCGTGAAGCCAGCGGCCCTGCCAAC
2161 -----+-----+-----+-----+-----+ 2220
CTTCCGTAACTTCGGTTGGACTAAACGTTGCACTCGGCACCTTGGTTCGCCGGGACGGTTG
E G I E A N L I C N V S R E A S G P A N

                2230                2250                2270
                .                    .                    .
TTCTCCTGGTTCCGAGATGGGGCGCTATGGGCCAGGGCCCTCTGGACACCGTGACACTG
2221 -----+-----+-----+-----+-----+ 2280
AAGAGGACCAAGGCTCTACCCCGGATACCCGGGTCCCGGAGACCTGTGGCACTGTGAC
F S W F R D G A L W A Q G P L D T V T L

                2290                2310                2330
                .                    .                    .
CTACCTGTGGCCAGAACTGATGCTGCCCTCTATGCTTGCCGCATCGTCACCGAGGCTGGT
2281 -----+-----+-----+-----+-----+ 2340
GATGGACACCGGTCTTGACTACGACGGGAGATACGAACGGCGTAGCAGTGGCTCCGACCA
L P V A R T D A A L Y A C R I V T E A G

                2350                2370                2390
                .                    .                    .
GCTGGCCTCTCCACCCTGTGGCCCTGAATGTGCTCTATCCCCCGATCCTCCAAAGTTG
2341 -----+-----+-----+-----+-----+ 2400
CGACCGGAGAGGTGGGGACACCGGGACTTACACGAGATAGGGGGCTAGGAGGTTTCAAC
A G L S T P V A L N V L Y P P D P P K L

                2410                2430                2450
                .                    .                    .
TCAGCCCTCCTGGACGTGGACCAGGCCACACGGCTGTGTTCTGCTGTACTGTGGACAGT
2401 -----+-----+-----+-----+-----+ 2460
AGTCGGGAGGACCTGCACCTGGTCCCCTGTGCGGACACAAGCAGACATGACACCTGTCA
S A L L D V D Q G H T A V F V C T V D S

                2470                2490                2510
                .                    .                    .
CGCCCTCTTGCCAGTTGGCCCTGTTCCGTGGGGAACACCTCCTGGCCGCCAGCTCGGCA
2461 -----+-----+-----+-----+-----+ 2520
GCGGGAGAACGGGTCAACCGGGACAAGGCACCCCTTGTGGAGGACCGGCGGTTCGAGCCGT
R P L A Q L A L F R G E H L L A A S S A

```

ES 2 687 029 T3

	2530	2550	2570	
2521	CTCCGGCTCCCCCTCGTGGCCGCTCCAGGCCAAAGCCTCGGCCAACTCCTTGACAGCTA -----+-----+-----+-----+-----+-----+			2580
	GAGGCCGAGGGGGAGCACCGCGGAGGTCCGGTTTCGGAGCCGGTTGAGGAACGTCGAT L R L P P R G R L Q A K A S A N S L Q L			
	2590	2610	2630	
2581	GAGGTCCGAGACTTGAGCCTTGGGACTCTGGCAGCTACCACTGTGAGGCCACCAACATC -----+-----+-----+-----+-----+-----+			2640
	CTCCAGGCTCTGAACCTCGGAACCCCTGAGACCGTCGATGGTGA <sup>^</sup> CACTCCGGTGGTTGTAG E V R D L S L G D S G S Y H C E A T N I			
	2650	2670	2690	
2641	CTTGGATCAGCCAACACTTCTTACCTTCCAGGTCCGAGGAGCCTGGGTCCGGGTGTCA -----+-----+-----+-----+-----+-----+			2700
	GAACCTAGTCCGTTGTGAAGAGAATGGAAGGTCCAGGCTCCTCGGACCCAGGCCACAGT L G S A N T S L T F Q V R G A W V R V S			
	2710	2730	2750	
2701	CCGTCGCCTGAGCTCCAGGAGGGCCAGGCTGTGGTCTGAGCTGCCAGGTACCCATAGGG -----+-----+-----+-----+-----+-----+			2760
	GGCAGCGACTCGAGTCTCCCGTCCGACACCAGGACTCGACGGTCCATGGGTATCCC P S P E L Q E G Q A V V L S C Q V P I G			
	2770	2790	2810	
2761	GTCCTGGAGGGACCTCATATCGTTGGTATCGGGATGGCCAGCCCCTCCAGGAGTCCACT -----+-----+-----+-----+-----+-----+			2820
	CAGGACCTCCCCTGGAGTATAGCAACCATAGCCCTACCGGTCCGGGAGGTCTCAGGTGA V L E G T S Y R W Y R D G Q P L Q E S T			
	2830	2850	2870	
2821	TCGGCCACGCTCCGTTTTCAGCCATAACTCTGAGCCAGGCTGGAGCCTACCATTGCCAA -----+-----+-----+-----+-----+-----+			2880
	AGCCGGTCCGAGGCAAACGTCGGTATTGAGACTCGGTCCGACCTCGGATGGTAACGGTT S A T L R F A A I T L S Q A G A Y H C Q			
	2890	2910	2930	
2881	GCCCAAGCTCCAGGCTCAGCCACCACGGACCTGGCTGCCCTGTGAGCCTCCACGTGACC -----+-----+-----+-----+-----+-----+			2940
	CGGGTTCGAGGTCCGAGTCCGGTGGTGCCTGGACCGACGGGACAGTCCGAGGTGCACTGG A Q A P G S A T T D L A A P V S L H V T			

ES 2 687 029 T3

```

                2950                2970                2990
                .                .                .
TACGCACCTCGCCAGGCCACACTCACCACCCTGATGGACTCAGGCCTCGGGCGACTGGGC
2941 -----+-----+-----+-----+-----+-----+-----+ 3000
ATGCGTGGAGCGGTCCGGTGTGAGTGGTGGGACTACCTGAGTCCGGAGCCCGCTGACCCG
Y A P R Q A T L T T L M D S G L G R L G

                3010                3030                3050
                .                .                .
CTCCTTCTGTGCGGTGTGAACAGTGACCCTCCTGCCAGCTCCGACTGCTCCATGGGAGC
3001 -----+-----+-----+-----+-----+-----+ 3060
GAGGAAGACACGGCACACTTGTCACTGGGAGGACGGGTCGAGGCTGACGAGGTACCCTCG
L L L C R V N S D P P A Q L R L L H G S

                3070                3090                3110
                .                .                .
CGCCTCGTGGCCTCTACTCTACAAGGTGTGGAGGAGCTTGCAAGCAGCTCTCCCCGCTA
3061 -----+-----+-----+-----+-----+-----+ 3120
GCGGAGCACCGGAGATGAGATGTTCCACACCTCCTCGAACGTCCGTCGAGAGGGCGGAT
R L V A S T L Q G V E E L A G S S P R L

                3130                3150                3170
                .                .                .
CAGGTGGCCACAGCCCCAACACGCTGCGCCTGGAGATCCACAACGCAGTGTGGAGGAT
3121 -----+-----+-----+-----+-----+-----+ 3180
GTCCACCGGTGTGCGGGGTTGTGCGACGCGGACCTCTAGGTGTTGCGTCACGACCTCCTA
Q V A T A P N T L R L E I H N A V L E D

                3190                3210                3230
                .                .                .
GAAGGCGTCTACACCTGCGAGGCCACCAACACCCCTGGGTGAGACCTTGGCCTCCGCCGCC
3181 -----+-----+-----+-----+-----+-----+ 3240
CTTCCGCAGATGTGGACGCTCCGGTGGTTGTGGGACCCAGTCTGGAACCGGAGGCGGCGG
E G V Y T C E A T N T L G Q T L A S A A

                3250                3270                3290
                .                .                .
TTCGATGCCCAGGCTATGAGAGTGCAGGTGTGGCCCAATGCCACCGTGCAAGAGGGGCAG
3241 -----+-----+-----+-----+-----+-----+ 3300
AAGCTACGGGTCCGATACTCTCACGTCCACACCGGTTACGGTGGCACGTTCTCCCCGTC
F D A Q A M R V Q V W P N A T V Q E G Q

                3310                3330                3350
                .                .                .
CTGGTGAACCTGACCTGCCTTGTATGGACCACGCACCTGGCCAGCTCACCTACACATGG
3301 -----+-----+-----+-----+-----+-----+ 3360
GACCACTTGGACTGGACGGAACATACCTGGTGCCTGGACCGGGTCGAGTGGATGTGTACC
L V N L T C L V W T T H L A Q L T Y T W

```

ES 2 687 029 T3

```

          3370              3390              3410
          .                .                .
TACCGAGACCAGCAGCAGCTCCAGGTGCTGCCACTCCATCCTCCTGCCCAATGTCAC
3361 -----+-----+-----+-----+-----+-----+ 3420
ATGGCTCTGGTCGTCGTCGAGGGTCCACGACGGGTGAGGTAGGAGGACGGGTTACAGTGA
Y R D Q Q Q L P G A A H S I L L P N V T

          3430              3450              3470
          .                .                .
GTCACAGATGCCGCCTCCTACCGCTGTGGCATATTGATCCCTGGCCAGGCACTCCGCCTC
3421 -----+-----+-----+-----+-----+ 3480
CAGTGTCTACGGCGGAGGATGGCGACACCGTATAACTAGGGACCGGTCCGTGAGGCGGAG
V T D A A S Y R C G I L I P G Q A L R L

          3490              3510              3530
          .                .                .
TCCAGACCTGTGCGCCTGGATGTCCTCTACGCACCCCGCAGACTGCGCCTGACCCATCTC
3481 -----+-----+-----+-----+-----+ 3540
AGGTCTGGACAGCGGGACCTACAGGAGATGCGTGGGGCGTCTGACGCGGACTGGGTAGAG
S R P V A L D V L Y A P R R L R L T H L

          3550              3570              3590
          .                .                .
TTGGAGAGCCGTGGTGGGCAGCTGGCCGTGGTGTGCTGTGCACTGTGGACAGTCGCCAGCT
3541 -----+-----+-----+-----+-----+ 3600
AACCTCTCGGCACCCCGTCGACCGGCACCCGACACCGTGACACCTGTCAGCGGGTCGA
L E S R G G Q L A V V L C T V D S R P A

          3610              3630              3650
          .                .                .
GCCAGCTGACCCCTCAGCCATGCTGGCCGCCTCCTGGCCTCCTCAACCGCAGCCTCTGTGC
3601 -----+-----+-----+-----+-----+ 3660
CGGGTCGACTGGGAGTCGGTACGACCGGCGGAGGACCGGAGGAGTTGGCGTCGGAGACAG
A Q L T L S H A G R L L A S S T A A S V

          3670              3690              3710
          .                .                .
CCCAACACCCTGCGCCTGGAGCTGTGGGAGCCCCGGCCAGTGATGAGGGTCTCTACAGC
3661 -----+-----+-----+-----+-----+ 3720
GGGTTGTGGGACGCGGACCTCGACACCCCTCGGGGCCGGGTCACTACTCCAGAGATGTTCG
P N T L R L E L W E P R P S D E G L Y S

          3730              3750              3770
          .                .                .
TGCTCGGCCCGCAGTCCTCTGGGCCAGGCCAACACATCCCTGGAGCTGCGGCTAGAGGGC
3721 -----+-----+-----+-----+-----+ 3780
ACGAGCCGGCGTCCAGGAGACCCGGTCCGGTTGTGTAGGGACCTCGACGCGGATCTCCCG
C S A R S P L G Q A N T S L E L R L E G

```

ES 2 687 029 T3

```

          3790              3810              3830
          .                .                .
GTGCAGGTGACTGGCTCCATCGACCACTGTGCCGGAGGGGGCCCTGTACAGTGACC
3781 -----+-----+-----+-----+-----+-----+ 3840
          CACGTCCACTGTGACCGAGGTAGCTGGTGACACGGCCTCCCCGGGACAGTGTCACTGG
          V Q V T L A P S T T V P E G A P V T V T

          3850              3870              3890
          .                .                .
TGTGAAGACCCCTGCTGCCCGCCACCCACCCCTCTATGTCTGGTACCACAACAGCCGTGG
3841 -----+-----+-----+-----+-----+-----+ 3900
          AACTTCTGGGACGACGGGCGGGTGGGTGGGAGATACAGACCATGGTGTGTCGGCAACC
          C E D P A A R P P T L Y V W Y H N S R W

          3910              3930              3950
          .                .                .
CTGCAGGAGGGGTCGGCTGCCCTCCCTCTCGTTTCCAGCGGCTACACGGGCTCACGGGGC
3901 -----+-----+-----+-----+-----+-----+ 3960
          GACGTCTCCCCAGCCGACGGAGGGAGAGCAAAGGTCCGCGATGTGCCCGAGTGCGCCG
          L Q E G S A A S L S F P A A T R A H A G

          3970              3990              4010
          .                .                .
GCCTATACCTGCCAGGTCCAGGATGCCAGGGCACACGCATCTCCAGCCCGCAGCACTG
3961 -----+-----+-----+-----+-----+-----+ 4020
          CGGATATGGACGGTCCAGGTCTACGGGTCCCGTGTGCGTAGAGGGTCGGGCGTCGTGAC
          A Y T C Q V Q D A Q G T R I S Q P A A L

          4030              4050              4070
          .                .                .
CACATCCTCTATGCCCTCGGGATGctgtcctttcctccttctgggactcaagggccagc
4021 -----+-----+-----+-----+-----+-----+ 4080
          GTGTAGGAGATACGGGGAGCCCTACgacaggaaggaggaagaccctgagttcccggtcg
          H I L Y A P R D A V L S S F W D S R A S

          4090              4110              4130
          .                .                .
cctatggccgtggtacagtgcactgtggacagcgagccacctgccgagatgacctgtcc
4081 -----+-----+-----+-----+-----+-----+ 4140
          ggataccggcaccatgtcacgtgacacctgtcgctcggtggacggctctactgggacagg
          P M A V V Q C T V D S E P P A E M T L S

          4150              4170              4190
          .                .                .
cgtgatggcaaggtgctggccaccagccatggggcccacggcttagcagtggggacaggc
4141 -----+-----+-----+-----+-----+-----+ 4200
          gcactaccggttccacgaccggtggtcggtaccccggtgccgaatcgtcaccctgtccg
          R D G K V L A T S H G A H G L A V G T G

```

ES 2 687 029 T3

```

          4210              4230              4250
          .                .                .
catgtccagggtggccccgaacgccctgcagctgcgggtgcagaatgtgccctcacgtgac
4201 -----+-----+-----+-----+-----+-----+ 4260
          .                .                .
gtacagggtccaccggcggtgcgggacgtgcagcccacgtcttacacgggagtgactg
          .                .                .
H V Q V A R N A L Q L R V Q N V P S R D

          4270              4290              4310
          .                .                .
aaggacacctacgtctgcatggcccgaactccttgggctcagtcagcaccatggggcag
4261 -----+-----+-----+-----+-----+ 4320
          .                .                .
ttcctgtggatgcagacgtaccggcggttgaggaacccgagtcâgtcgtggtaccccgtc
          .                .                .
K D T Y V C M A R N S L G S V S T M G Q

          4330              4350              4370
          .                .                .
ctgcagccagaagggtgtgcacgtggtagccgagccagggtggatgtgcccaaggcaca
4321 -----+-----+-----+-----+-----+ 4380
          .                .                .
gacgtcggctctccacacgtgcaccatcggctcgggtcccgacctacacgggcttccgtgt
          .                .                .
L Q P E G V H V V A E P G L D V P E G T

          4390              4410              4430
          .                .                .
gcgctgaacctgagctgtgcctccctagtgccctgggcacatgggcaactccaccttt
4381 -----+-----+-----+-----+-----+ 4440
          .                .                .
cgcgacttgactcgacagcggaggatcaccgggacccggtgtacccttgaggtggaaa
          .                .                .
A L N L S C R L P S G P G H M G N S T F

          4450              4470              4490
          .                .                .
gcttgggtccggaacgggtcggcagctacacacagagtctgtgccacccttaccttacc
4441 -----+-----+-----+-----+-----+ 4500
          .                .                .
cgaaccaaggccttgccagccgtcgatgtgtgtctcagacacgggtgggaatggaagtgg
          .                .                .
A W F R N G R Q L H T E S V P T L T F T

          4510              4530              4550
          .                .                .
catgtggcccgcgcccgaagctggcttgtaccactgccaggctgagctcccgcggggct
4501 -----+-----+-----+-----+-----+ 4560
          .                .                .
gtacaccgggcgcgggttcgaccgaacatggtgacgggtccgactcgaggggcgccccga
          .                .                .
H V A R A Q A G L Y H C Q A E L P A G A

          4570              4590              4610
          .                .                .
gccacctctgctccagtctgctcgggtgctctaccctccaagacgccaccatgact
4561 -----+-----+-----+-----+-----+ 4620
          .                .                .
cggtgagacgaggtcagaacgagcccacgagatgggagggttctgcgggtgtactga
          .                .                .
A T S A P V L L R V L Y P P K T P T M T

```

ES 2 687 029 T3

```

          4630              4650              4670
          .                .                .
gtttttgtggagcccgaggggtggcatccagggcattctggactgccgagtgacagtgag
4621 -----+-----+-----+-----+-----+-----+ 4680
caaaaaacacctcgggctcccaccgtaggtcccgtaagacctgacggctcacctgtcactc
V F V E P E G G I Q G I L D C R V D S E

          4690              4710              4730
          .                .                .
cccctagccagcctgacctccacctgggcagtcggctggcctccagccagccccag
4681 -----+-----+-----+-----+-----+ 4740
gggatcggctcggactgggaggtggaccctcagccgaccaccggaggtcggctcggggtc
P L A S L T L H L G S R L V A S S Q P Q

          4750              4770              4790
          .                .                .
gctgccctgccaaagccgcacatccgcgtctcagccagtcccaatgcCTTGCGAGTGGAC
4741 -----+-----+-----+-----+-----+ 4800
cgacggggacggttcggcgtgtaggcgcagagtcggtcagggttacgGAACGCTCACCTG
A A P A K P H I R V S A S P N A L R V D

          4810              4830              4850
          .                .                .
ATGGAGGAGCTGAAGCCCAGTGACCAGGGGAGTATGTGTGCTCGGCCTCCAATGCCCTG
4801 -----+-----+-----+-----+-----+ 4860
TACCTCCTCGACTTCGGGTCACCTGGTCCCCCTCATAACACGAGCCGGAGGTTACGGGAC
M E E L K P S D Q G E Y V C S A S N A L

          4870              4890              4910
          .                .                .
GGCTCTGCCTCTGCTGCCACCTACTTCGGAACCAGAGCCCTGCATCGCCTGCATCTGTTC
4861 -----+-----+-----+-----+-----+ 4920
CCGAGACGGAGACGACGGTGGATGAAGCCTTGGTCTCGGGACGTAGCGGACGTAGACAAG
G S A S A A T Y F G T R A L H R L H L F

          4930              4950              4970
          .                .                .
CGGCACCTTCTCTGGTTCCTGGGGCTGCTGGCGAGCCTCCTCTTCTACTGTTGGGCCTG
4921 -----+-----+-----+-----+-----+ 4980
GCCGTGGAAGAGACCAAGGACCCCGACGACCGCTCGGAGGAGAAGGATGACAACCCGGAC
R H L L W F L G L L A S L L F L L L G L

          4990              5010              5030
          .                .                .
GGGGTCTGGTACGCCTGGAGACGGGGAAATTTTCACAAGCTGAGAATGGGCGAATATTCA
4981 -----+-----+-----+-----+-----+ 5040
CCCCAGACCATGCGGACCTCTGCCCTTTAAAAGTGTTCGACTCTTACCCGCTTATAAGT
G V W Y A W R R G N F H K L R M G E Y S

```

ES 2 687 029 T3

```

                    5050                5070                5090
                    .                  .                  .
gtagagatggatatctcggaaggaaccacgcagatgtccactgaccaggaagaagtact
5041 -----+-----+-----+-----+-----+-----+ 5100
catctctaccatagagccttcctttggtgctctacaggtgactggtccttcttcaatga
V E M V S R K E T T Q M S T D Q E E V T

                    5110                5130                5150
                    .                  .                  .
ggaatcggtgatgatgcgggctctgtgaaccaggcggcatttgatcctgcccacctctgt
5101 -----+-----+-----+-----+-----+ 5160
ccttagccactactacgcccgagacacttggcccgtaaaactaggacgggtggagaca
G I G D D A G S V N Q A A F D P A H L C

                    5170                5190
                    .                  .
gaaaacacacagtctgtGaaaagcacagtctga
5161 -----+-----+-----+----- 5193
cttttggtgtcagacaCttttcgtgtcagact
E N T Q S V K S T V *
```