



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 364 719**

51 Int. Cl.:
C07K 14/165 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **04727194 .5**

96 Fecha de presentación : **14.04.2004**

97 Número de publicación de la solicitud: **1618126**

97 Fecha de publicación de la solicitud: **25.01.2006**

54 Título: **Proteínas relacionadas con el SRAG.**

30 Prioridad: **15.04.2003 US 462688 P**
22.05.2003 US 472416 P

45 Fecha de publicación de la mención BOPI:
13.09.2011

45 Fecha de la publicación del folleto de la patente:
13.09.2011

73 Titular/es: **Her Majesty the Queen in Right of
Canada as Represented by the Minister of Health
1015 Rue Arlington Piece T2420
Winnipeg, Manitoba R3E 3P6, CA**

72 Inventor/es: **Jones, Steven;**
Feldmann, Heinz;
Bastien, Nathalie;
Li, Yan;
Plummer, Frank;
Ens, Werner;
Standing, Kenneth;
Krokhine, Oleg;
Perreault, Helen y
Tyler, Shaun

74 Agente: **Elzaburu Márquez, Alberto**

ES 2 364 719 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

Proteínas relacionadas con el SRAG.

INFORMACIÓN DE LA SOLICITUD ANTERIOR

5 Esta solicitud reivindica la prioridad de los documentos USSN 60/462.688, presentado el 15 de abril de 2003, y USSN 60/472.416, presentado el 22 de mayo de 2003.

CAMPO DE LA INVENCION

La presente invención se refiere a proteínas novedosas aisladas a partir de un paciente con síndrome respiratorio agudo grave (SRAG). La invención se refiere también a los usos de dichas proteínas en la terapia y el diagnóstico del SRAG.

10 ANTECEDENTES DE LA INVENCION

La reciente identificación clínica de un tipo novedoso de neumonía atípica sin una etiología claramente definida, junto con la evidencia epidemiológica de un alto coeficiente de transmisión, han provocado que la Organización Mundial de la Salud emita una rara alerta de viaje. La nueva entidad se ha denominado síndrome respiratorio agudo grave (SRAG); aparentemente comenzó en la provincia de Guangdong en China en noviembre de 2002, y se ha extendido desde entonces a Hong Kong, Singapur, Vietnam, Canadá, EE.UU., Taiwán y varios países europeos. El brote en Canadá comenzó a finales de febrero de 2003 en un viajero que volvía desde Hong Kong, cuya exposición fue el caso inicial en la epidemia de Hong Kong (un médico que había tratado casos del SRAG en la provincia de Guangdong de la República Popular de China). El caso inicial canadiense murió 9 días después del inicio de la enfermedad y un pariente masculino de 43 años enfermó 2 días después de la exposición y murió de síndrome de dificultad respiratoria aguda 15 días después del inicio de la dolencia (1). Posteriormente, Canadá se ha enfrentado al mayor brote del SRAG fuera de Asia, con al menos 303 casos probables y sospechosos y 24 muertes, la mayoría en la zona de Toronto (2).

Se han remitido muestras de pacientes con sospecha o probabilidad de SRAG en Canadá al Laboratorio nacional de microbiología, Ministerio de sanidad de Canadá "National Microbiology Laboratory (NML), Health Canada", para diagnóstico de laboratorio. Este laboratorio, parte del Centro canadiense de ciencias para la salud humana y animal "Canadian Science Centre for Human and Animal Health", es el centro de referencia nacional de Canadá para enfermedades infecciosas, y alberga las únicas instalaciones de contención de clase 4 del país. El NML ha desempeñado un papel activo en un intenso esfuerzo cooperativo internacional entre 11 laboratorios de todo el mundo que ha sugerido que puede estar etiológicamente implicado un coronavirus preciso. En particular, el laboratorio preparó muestras nucleotídicas para el primer esfuerzo exitoso de determinar la secuencia genómica del coronavirus (3), un resultado pronto confirmado por otros laboratorios; véase, por ejemplo, la ref. (4).

No obstante, la secuencia genómica proporciona meramente un molde para la construcción de las proteínas víricas. Por tanto, una estrategia alternativa es examinar las proteínas mismas, y la espectrometría de masas ha probado ser una herramienta eficaz con este fin (5).

35 SUMARIO DE LA INVENCION

Los presentes inventores han aislado y secuenciado proteínas novedosas a partir de una muestra biológica de un paciente con síndrome respiratorio agudo grave (SRAG).

40 Según una realización de la divulgación, se proporciona una proteína relacionada con el SRAG a partir de un paciente de SRAG. En una realización específica, la proteína relacionada con el SRAG tiene la secuencia aminoacídica mostrada en la Figura 1 (SEC ID N° 1) o 2 (SEC ID N° 2).

Esta divulgación está también dirigida al uso de proteínas relacionadas con el SRAG, o un fragmento de las mismas, para la preparación de una vacuna para tratar el SRAG.

45 Esta divulgación está dirigida al uso de proteínas relacionadas con el SRAG, o un fragmento de las mismas, para la preparación de anticuerpos, incluyendo anticuerpos policlonales, anticuerpos monoclonales o un fragmento funcional de los mismos. Esta divulgación se refiere también a los anticuerpos así producidos.

Esta divulgación está dirigida también a vacunas preparadas a partir de dichos anticuerpos, o proteínas relacionadas con el SRAG o un fragmento de las mismas, para el tratamiento o la prevención de enfermedades mediadas por virus, preferiblemente mediadas por coronavirus.

50 Esta divulgación está dirigida al uso de dichos anticuerpos o proteínas relacionadas con el SRAG, o un fragmento de las mismas, como diagnóstico para cribar la presencia del SRAG.

Esta divulgación está dirigida al uso de una proteína relacionada con el SRAG, o un fragmento de la misma, como herramienta de desarrollo de fármacos.

La divulgación está dirigida adicionalmente a un método de tratamiento de una afección asociada a un virus, preferiblemente un coronavirus, que comprende administrar a una célula o animal necesitado de ello una cantidad eficaz de un agente que module la expresión y/o actividad de una proteína relacionada con el SRAG de la invención. En una realización preferida, la afección a tratar es el SRAG.

5 **BREVE DESCRIPCIÓN DE LOS DIBUJOS**

Estos y otros rasgos de la invención resultarán más evidentes a partir de la siguiente descripción, en la que se hace referencia a los dibujos adjuntos a la misma, en los que:

La Figura 1 muestra la secuencia aminoacídica de la proteína relacionada con el SRAG de 47 kDa (SEC ID N° 1).

10 La Figura 2 muestra la secuencia aminoacídica de la proteína relacionada con el SRAG de 139 kDa (SEC ID N° 2).

La Figura 3 muestra el análisis de proteínas de un nuevo coronavirus asociado al SRAG. A) Se separaron mediante PAGE-SDS las partículas víricas sedimentadas a partir del sobrenadante de células Vero E6 expuestas a muestras derivadas de pacientes con SRAG (carriles 1) o de células infectadas ficticiamente procesadas de forma similar (carriles 2), y se analizaron por transferencia Western con suero de convaleciente de pacientes de SRAG (Tor 2, Tor 3, Tor 4 y BC 1) o un suero de control de un donante no infectado (NML). Los sueros de los pacientes, pero no del control, reaccionaron con una especie de 44-48 kDa presente en los sobrenadantes de cultivos infectados pero no de los tratados ficticiamente (indicados por una flecha). B) Se fraccionó en un gradiente de sacarosa una muestra de virus similar a la descrita en A). Se separó la fracción que contenía material inmunorreactivo en un gel en gradiente de bisacrilamida de 4-12% y se tiñó con azul de Coomassie coloidal. Se observó una banda destacada con un peso molecular aparente de 44-48 kDa junto con una mucho menos intensa a ~180 kDa (indicada por una flecha). Se extrajeron estas bandas y se usaron para los estudios de espectrometría de masas descritos en el texto.

25 Figura 4. A) Espectro de EM individual de la mezcla peptídica obtenida a partir de la digestión trípica de la proteína de 47 kDa antes del fraccionamiento por HPLC. B) Una sección pequeña del espectro de masas MALDI de una mezcla de péptidos trípticos de la proteína de 47 kDa. C) La misma sección del espectro obtenida a partir de la fracción 23 de la separación por HPLC de la mezcla. Las leyendas de pico indican los números de resto correspondientes a la proteína intacta; en un caso, se indica una pérdida de 64 Da. El pico intenso correspondiente a T277-293 en la mezcla está ausente en la fracción 23 (eluye en la fracción 21), pero varios picos más débiles que están presentes, como T389-405, se potencian significativamente por la HPLC. La mejora ayuda a identificarlos, y es esencial para una alta exactitud de la masa y para el posterior análisis de EM/EM. Las masas medidas y predichas para todos los péptidos trípticos pueden encontrarse en la Tabla 2; la Dm es menor de 10 mDa en casi cada caso.

35 Figura 5. Espectro de EM/EM del fragmento trípico de m/z 2297 después de digestión en una mezcla 50/50 de agua normal y marcada con 0-18. Se muestra el espectro completo en A, con la secuencia aminoacídica indicada entre los fragmentos de la serie y. Se muestra en B un ejemplo de un fragmento de la serie b y en C un fragmento de la serie y. Es visible en C el patrón isotópico característico de fragmentos que contienen el extremo C. Las masas medidas y predichas para todos los picos identificados se muestran en la Tabla 1.

La Figura 6 muestra el alineamiento de secuencia de la proteína de nucleocápsida de coronavirus.

40 Figura 7. Detección directa por EM (A, B) y EM/EM (C) de péptidos trípticos glucosilados de la proteína de espícula de 139 kDa. A) Espectro de EM de la fracción n° 21 de HPLC de la digestión trípica de la banda de 180 kDa. Los picos marcados son los iones monoisotópicos $[M+H]^+$ de las formas glucosiladas de T111-126 (véase la Tabla 3). B) Espectro de EM de la fracción n° 25 de HPLC que muestra formas de T222-232 glucosiladas. C) Espectros de EM/EM del pico de 3122,373 Da de B. D) Estructuras de N-glicano sugeridas.

45 Figura 8. Espectro de EM/EM del fragmento trípico de m/z 1144 que representa el extremo N-terminal de la proteína de 47 kDa. La secuencia aminoacídica se indica entre los fragmentos de la serie y. La diferencia entre las masas medida y calculada del resto de Ser N-terminal (42,015 Da) corresponde a la acetilación N-terminal, una modificación post-traduccional (Dm=42,011 Da).

50 Figura 9. Comparación de la digestión (A) trípica y (B) con Lys-C de la misma región de la proteína de 47 kDa. Están presentes dos restos adicionales (QK) en la digestión con Lys-C debido a que no escinde en la arginina. Se efectuaron ambas digestiones en presencia de una mezcla 50/50 de agua normal y 0-18. El patrón isotópico característico mostrado en el inserto (A) es una característica de todos los fragmentos de digestión excepto de aquel que contiene el extremo C de la proteína. Los dos picos a m/z nominal 1410 en A y B corresponden a péptidos que contienen los mismos restos excepto por la diferencia K/Q. La diferencia de masa esperada es de 36 mDa; la medida muestra una diferencia de 35 mDa.

55 Figura 10. Espectro de masas de la fracción n° 12 de la digestión trípica de la proteína de 46 kDa. La digestión

tuvo lugar en presencia de una mezcla 50/50 de agua normal y 0-18. El pico mostrado no exhibe el patrón isotópico característico del marcaje 0-18 C-terminal después de escisión triptica, sugiriendo que es el péptido C-terminal. Esto se confirmó mediante análisis de EM/EM posterior, que revela una secuencia consistente con el extremo C de la proteína de 46 kDa predicha: QLQNSM*SGASADSTQA (M*-resto de metionina oxidado).

5 Figura 11. A) Secuencia deducida de la proteína de nucleocápsida de SRAG de 47 kDa. Las secuencias subrayadas no se identificaron por EM. Los 5 motivos SRXX se indican por letras rojas. B) Análisis de árbol filogenético de diferentes secuencias de proteína N de coronavirus.

Figura 12. Identificación por EM y EM/EM de péptidos desglucosilados a partir de la proteína de espícula de 139 kDa. A) Espectro de EM de la fracción nº 26 de HPLC que muestra T222-232 desglucosilados. B) Espectro de EM/EM de LPLGIDITNFR desglucosilado a 1258,715 Da (véase la Tabla 4).

DESCRIPCIÓN DE LAS REALIZACIONES PREFERIDAS

Los inventores han determinado que los sueros de un paciente con SRAG contenían un antígeno vírico dominante que se identificó por desorción/ionización por láser asistida por matriz-tiempo de vuelo-espectrometría de masas (MALDI-TOF-EM) en forma de una especie de proteína de ~47 kDa. Se encontró que ésta era una proteína de nucleocápsida novedosa. Se ha examinado también por MALDI-TOF-EM (42% de cobertura) (Krokhin *et al.*, 2003, *Mol. Cell. Proteomics* 2: 346-356) una segunda proteína vírica correspondiente a la glucoproteína de espícula de ~139 kDa predicha. Después de digerir con pNGasa F, se confirmaron nueve sitios de glucosilación en esta proteína. Se identificaron también los azúcares unidos a cuatro de los sitios. Estos resultados sugieren que la proteína de nucleocápsida y la glucoproteína de espícula son inmunógenos importantes que pueden ser útiles para el diagnóstico temprano, o de valor profiláctico para combatir el SRAG.

I. Proteínas novedosas

La divulgación se refiere a proteínas aisladas que se han aislado y secuenciado a partir de un paciente con SRAG. Por facilidad de referencia, las proteínas se designan generalmente como proteínas relacionadas con el SRAG. La expresión "proteínas relacionadas con el SRAG" como se usa en la presente memoria incluye las proteínas mostradas en las Figuras 1 (SEC ID N° 1) y 2 (SEC ID N° 2), así como todos los homólogos, análogos, fragmentos o derivados de proteínas relacionadas con el SRAG.

En una realización, la proteína relacionada con el SRAG aislada tiene un peso molecular de aproximadamente 47 kDa y una secuencia aminoacídica como se muestra en la Figura 1. El análisis de secuencia revela que esta proteína comparte un 32% de homología con proteínas de nucleocápsida de coronavirus (Figura 6), aunque hay diferencias significativas en las secuencias. En consecuencia, la presente divulgación proporciona una proteína de nucleocápsida aislada que comprende una secuencia que es al menos un 35% homóloga, preferiblemente al menos un 40% homóloga, de la secuencia mostrada en la Figura 1. Los inventores han caracterizado adicionalmente la proteína de nucleocápsida usando espectroscopia de masas como se describe en el Ejemplo 1 y en las Figuras y Tablas.

En otra realización, la proteína relacionada con el SRAG aislada tiene un peso molecular de 139 kDa y una secuencia aminoacídica mostrada en la Figura 2. El análisis de secuencia revela que esta proteína comparte homología con glucoproteínas de espícula de coronavirus y proteínas estructurales. Los inventores han caracterizado adicionalmente la glucoproteína de espícula usando espectroscopia de masas como se describe en el Ejemplo 1 y en las Figuras y Tablas.

Dentro del contexto de la presente invención, una proteína relacionada con el SRAG puede incluir diversas formas estructurales de la proteína primaria que retienen actividad biológica. Por ejemplo, una proteína relacionada con el SRAG puede estar en forma de sales ácidas o básicas o en forma neutra. Además, los restos aminoacídicos individuales pueden modificarse por oxidación o reducción.

Además de la secuencia aminoacídica completa, las proteínas relacionadas con el SRAG pueden incluir también truncamientos de la proteína, análogos y homólogos de la proteína y truncamientos de los mismos como se describe en la presente memoria. Las proteínas relacionadas con el SRAG truncadas o fragmentos pueden comprender péptidos de al menos 5, preferiblemente 10 y más preferiblemente 15 restos aminoacídicos de la secuencia mostrada en la Figura 1 o 2. En otras realizaciones, las proteínas relacionadas con el SRAG truncadas o fragmentos pueden comprender péptidos de al menos 5, al menos 6, al menos 7, al menos 8, preferiblemente 10 o más preferiblemente 15 restos aminoacídicos consecutivos o contiguos de la secuencia mostrada en la Figura 1 o 2.

La divulgación proporciona adicionalmente polipéptidos que comprenden al menos un dominio funcional o al menos un determinante antigénico de una proteína relacionada con el SRAG.

Los análogos de la proteína y/o truncamientos de la misma como se describen en la presente memoria pueden incluir, pero sin limitación, una secuencia aminoacídica que contiene una o más sustituciones, inserciones y/o deleciones aminoacídicas. Las sustituciones aminoacídicas pueden ser de naturaleza conservativa o no conservativa. Las sustituciones aminoacídicas conservativas implican reemplazar uno o más aminoácidos de las

proteínas por aminoácidos de características de carga, tamaño y/o hidrofobicidad similares. Cuando se realizan solo sustituciones conservativas, el análogo resultante debería ser funcionalmente equivalente. Las sustituciones no conservativas implican reemplazar uno o más aminoácidos de la secuencia aminoacídica por uno o más aminoácidos que posean características de carga, tamaño y/o hidrofobicidad distintas.

5 Pueden introducirse una o más inserciones aminoacídicas en las secuencias aminoacídicas. Las inserciones aminoacídicas pueden consistir en restos aminoacídicos individuales o aminoácidos secuenciales en el intervalo de 2 a 15 aminoácidos de longitud. Por ejemplo, las inserciones aminoacídicas pueden usarse para destruir secuencias diana de modo que la proteína no sea ya activa. Este procedimiento puede usarse *in vivo* para inhibir la actividad de una proteína.

10 Las deleciones pueden consistir en la retirada de uno o más aminoácidos, o porciones discretas de la secuencia aminoacídica, de la proteína relacionada con el SRAG. Los aminoácidos eliminados pueden ser contiguos o no. La longitud límite inferior del análogo resultante con una mutación de deleción es de aproximadamente 10 aminoácidos, preferiblemente de 100 aminoácidos.

15 Según otra realización, se proporcionan moléculas de ácido nucleico deducidas a partir de las secuencias aminoacídicas o fragmentos de las mismas de las proteínas relacionadas con el SRAG. Como se apreciará por un experto en la técnica, las moléculas de ácido nucleico pueden insertarse en cualquier vector de expresión conocido en la técnica para expresión en un hospedador adecuado. Además, ha de observarse que los codones seleccionados para las moléculas de ácido nucleico pueden seleccionarse según tablas de uso de codón para el sistema de expresión hospedador pretendido, maximizando así la expresión.

20 Pueden prepararse análogos de una proteína introduciendo mutaciones en la secuencia nucleotídica que codifica la proteína. Las mutaciones en las secuencias nucleotídicas construidas para la expresión de análogos de una proteína deben conservar el marco de lectura de las secuencias de codificación. Además, las mutaciones preferiblemente no crearán regiones complementarias que puedan hibridar produciendo estructuras de ARNm secundario tales como bucles u horquillas, que podrían afectar adversamente a la traducción del ARNm receptor.

25 Las mutaciones pueden introducirse en loci particulares sintetizando oligonucleótidos que contienen una secuencia mutante flanqueada por sitios de restricción que posibilitan el ligamiento con fragmentos de la secuencia nativa. Después del ligamiento, la secuencia reconstruida resultante codifica un análogo que tiene la inserción, sustitución o deleción aminoacídica deseada.

30 Como alternativa, pueden emplearse procedimientos de mutagénesis específica de sitio dirigida a oligonucleótidos para proporcionar un gen alterado que tenga codones particulares alterados según la sustitución, deleción o inserción necesaria. Puede construirse también una deleción o truncamiento de una proteína utilizando sitios de endonucleasa de restricción convenientes adyacentes a la deleción deseada. Posteriormente a la restricción, los salientes puede rellenarse y religarse el ADN. Se dan a conocer métodos ejemplares para realizar las alteraciones expuestas anteriormente en Sambrook *et al* ("Molecular Cloning: A Laboratory Manual", 2ª Ed., Cold Spring Harbor Laboratory Press, 1989).

35 Las proteínas relacionadas con el SRAG incluyen también homólogos de la secuencia aminoacídica de la proteína relacionada con el SRAG y/o truncamientos de la misma como se describe en la presente memoria. Dichos homólogos son proteínas cuyas secuencias aminoacídicas comprenden secuencias aminoacídicas que hibridan en condiciones de hibridación rigurosas, por ejemplo, dichos homólogos reaccionarán de forma cruzada con anticuerpos generados contra proteínas relacionadas con el SRAG o fragmentos de las mismas. Los homólogos de una proteína relacionada con el SRAG de la invención tendrán las mismas regiones que son características de la proteína.

40 Una proteína homóloga incluye una proteína con una secuencia aminoacídica que tiene al menos un 60%, preferiblemente al menos un 70%, más preferiblemente un 80-95% de identidad con la secuencia aminoacídica de la proteína relacionada con el SRAG.

La divulgación contempla también isoformas de las proteínas. Una isoforma contiene el mismo número y clase de aminoácidos que una proteína, pero la isoforma tiene una estructura molecular diferente. Las isoformas son aquellas que tienen las mismas propiedades que una proteína de la invención como se describe en la presente memoria.

45 La presente divulgación incluye también una proteína conjugada con una proteína seleccionada, o una proteína marcadora seleccionable, para producir proteínas de fusión. Por ejemplo, se inserta la secuencia de ADNc de SRAG, por ejemplo como se deduce a partir de la secuencia aminoacídica de una de las proteínas relacionadas con el SRAG como se debate anteriormente, en un vector que contiene una secuencia nucleotídica que codifica otro péptido (por ejemplo, GST-glutation succiniltransferasa). La proteína de fusión se expresa y se recupera de células procaríóticas (por ejemplo, bacterianas o baculovíricas) o eucarióticas. La proteína de fusión puede purificarse entonces mediante cromatografía de afinidad basándose en la secuencia del vector de fusión y la proteína relacionada con el SRAG obtenida mediante escisión enzimática de la proteína de fusión.

Las proteínas (incluyendo truncamientos, análogos, etc.) pueden prepararse usando métodos de ADN recombinante.

En consecuencia, pueden incorporarse moléculas de ácido nucleico que tienen una secuencia que codifica una proteína según procedimientos conocidos en la técnica a un vector de expresión apropiado que asegure una buena expresión de la proteína. Los posibles vectores de expresión incluyen, pero sin limitación, cósmidos, plásmidos o virus modificados (por ejemplo, retrovirus, adenovirus y virus adenoasociados defectivos de replicación), a condición de que el vector sea compatible con la célula hospedadora usada. La expresión “vectores adecuados para la transformación de una célula hospedadora” significa que los vectores de expresión contienen una molécula de ácido nucleico y secuencias reguladoras, seleccionadas basándose en las células hospedadoras que se van a usar para expresión, que están ligadas operativamente a la molécula de ácido nucleico. Por “ligado operativamente” se pretende indicar que el ácido nucleico está ligado a secuencias reguladoras de manera que permita la expresión del ácido nucleico.

La divulgación contempla por lo tanto un vector de expresión recombinante de la invención que contiene una molécula de ácido nucleico, o un fragmento de la misma, y las secuencias reguladoras necesarias para la transcripción y traducción de la secuencia proteica insertada. Las secuencias reguladoras adecuadas pueden derivar de una variedad de fuentes, incluyendo genes bacterianos, fúngicos o víricos (por ejemplo, véanse las secuencias reguladoras descritas en Goeddel, “Gene Expression Technology: Methods in Enzymology” 185, Academic Press, San Diego, CA (1990). La selección de las secuencias reguladoras apropiadas depende de la célula hospedadora elegida, y puede conseguirse fácilmente por un experto en la técnica. Los ejemplos de dichas secuencias reguladoras incluyen: un promotor y potenciador transcripcional o secuencia de unión a ARN polimerasa y una secuencia de unión ribosómica, incluyendo una señal de inicio de la traducción. Adicionalmente, dependiendo de la célula hospedadora elegida y del vector empleado, pueden incorporarse al vector de expresión otras secuencias, tales como un origen de replicación, sitios de restricción de ADN adicionales, potenciadores y secuencias que confieren inducibilidad de transcripción. Se apreciará también que pueden suministrarse las secuencias reguladoras necesarias por la proteína nativa y/o sus regiones flanqueantes.

La divulgación proporciona adicionalmente un vector de expresión recombinante que comprende una molécula de ácido nucleico ADN clonada en el vector de expresión en orientación de sentido contrario. Es decir, la molécula de ADN está ligada operativamente a una secuencia reguladora de manera que permita la expresión, mediante transcripción de la molécula de ADN, de una molécula de ARN que es de sentido contrario a una secuencia nucleotídica. Pueden elegirse secuencias reguladoras ligadas operativamente al ácido nucleico en sentido opuesto que dirijan la expresión continua de la molécula de ARN de sentido contrario.

Los vectores de expresión recombinantes pueden contener también un gen marcador seleccionable que facilite la selección de células hospedadoras transformadas o transfectadas con una molécula recombinante. Los ejemplos de genes marcadores seleccionables son genes que codifican una proteína tal como G418 e higromicina, que confieren resistencia a ciertos fármacos, β -galactosidasa, cloranfenicol acetiltransferasa o luciferasa de luciérnaga. La transcripción del gen marcador seleccionable se monitoriza mediante los cambios de concentración de la proteína marcadora seleccionable tal como β -galactosidasa, cloranfenicol acetiltransferasa o luciferasa de luciérnaga. Si el gen marcador seleccionable codifica una proteína que confiere resistencia a anticuerpos tal como resistencia a neomicina, pueden seleccionarse las células transformantes con G418. Las células que han incorporado el gen marcador seleccionable sobrevivirán, mientras que las otras células mueren. Esto posibilita visualizar y ensayar la expresión de vectores de expresión recombinantes, y en particular de determinar el efecto de una mutación sobre la expresión y el fenotipo. Se apreciará que pueden introducirse marcadores seleccionables en un vector separado desde el ácido nucleico de interés.

Los vectores de expresión recombinantes pueden contener también genes que codifican un resto de fusión que proporciona una expresión aumentada de la proteína recombinante; una solubilidad aumentada de la proteína recombinante y que ayudan en la purificación de una proteína recombinante diana al actuar como ligando en purificación por afinidad. Por ejemplo, puede añadirse un sitio de escisión proteolítica a la proteína recombinante diana para permitir la separación de la proteína recombinante del resto de fusión posteriormente a la purificación de la proteína de fusión.

Pueden introducirse vectores de expresión recombinantes en células hospedadoras para producir una célula hospedadora transformada. En consecuencia, la divulgación incluye una célula hospedadora que comprende un vector de expresión recombinante. La expresión “célula hospedadora transformada” se pretende que incluya células procarióticas y eucarióticas que se han transformado o transfectado con un vector de expresión recombinante. Las expresiones “transformado con”, “transfectado con”, “transformación” y “transfección” se pretende que comprendan la introducción de ácido nucleico (por ejemplo, un vector) en una célula mediante una de las muchas técnicas posibles conocidas en la técnica. Las células procarióticas pueden transformarse con ácido nucleico, por ejemplo, mediante electroporación o transformación mediada por cloruro de calcio. El ácido nucleico puede introducirse en células de mamífero mediante técnicas convencionales tales como coprecipitación con fosfato de calcio o cloruro de calcio, transfección mediada por DEAE-dextrano, lipofectina, electroporación o microinyección. Pueden encontrarse métodos adecuados para transformar y transfectar células hospedadoras en Sambrook *et al.* (“Molecular Cloning: A Laboratory Manual”, 2ª edición, Cold Spring Harbor Laboratory Press (1989)), y otros de dichos libros de texto de laboratorio.

Las células hospedadoras adecuadas incluyen una amplia variedad de células hospedadoras procarióticas y

eucarióticas. Por ejemplo, las proteínas relacionadas con el SRAG pueden expresarse en células bacterianas tales como *E. coli*, *Pseudomonas*, *Bacillus subtilis*, células de insecto (usando baculovirus), células de levadura o células de mamífero. Pueden encontrarse otras células hospedadoras adecuadas en Goeddel, "Gene Expression Technology: Methods in Enzymology" 185, Academic Press, San Diego, CA (1991).

- 5 Las proteínas relacionadas con el SRAG pueden aislarse también a partir de un coronavirus o de una muestra de un paciente con SRAG. La proteína puede purificarse mediante métodos de purificación convencionales conocidos por los expertos en la técnica, tales como métodos de cromatografía, métodos de cromatografía líquida de alta resolución o precipitación.

- 10 Por ejemplo, puede usarse un anticuerpo anti-proteína relacionada con el SRAG (como se describe a continuación) para aislar una proteína relacionada con el SRAG, que se purifica entonces mediante métodos estándar.

Las proteínas pueden prepararse también mediante síntesis química usando técnicas bien conocidas en la química de proteínas tales como síntesis en fase sólida (Merrifield, 1964, *J. Am. Chem. Assoc.* 85: 2149-2154) o síntesis en disolución homogénea (Houbenweyl, 1987, "Methods of Organic Chemistry", ed. E. Wansch, Vol. 15 I y II, Thieme, Stuttgart).

15 **II. Usos**

La presente divulgación incluye todos los usos de las proteínas relacionadas con el SRAG incluyendo, pero sin limitación, la preparación de anticuerpos, el aislamiento de sustancias que modulan la expresión y/o actividad de una proteína relacionada con el SRAG, así como el uso de proteínas y moduladores de la misma en aplicaciones de diagnóstico, biomédicas, terapéuticas y de vacuna. Algunos de los usos se describen detalladamente a continuación.

20 **(a) Anticuerpos**

- El aislamiento de una proteína relacionada con el SRAG posibilita la preparación de anticuerpos específicos del virus, probablemente coronavirus, que causa el SRAG. En consecuencia, la presente divulgación proporciona un anticuerpo que se une a una proteína relacionada con el SRAG. Pueden prepararse anticuerpos que se unen a un epítipo definido en una región no conservada de la proteína relacionada con el SRAG. Es una región no conservada de la proteína relacionada con el SRAG aquella que no tiene homología de secuencia sustancial con otras proteínas, por ejemplo, otras proteínas víricas relacionadas que tienen al menos una región homóloga de una región de dicha proteína relacionada con el SRAG.

- Los anticuerpos de la proteína relacionada con el SRAG pueden prepararse también usando técnicas conocidas en la técnica. Por ejemplo, usando un péptido de la proteína relacionada con el SRAG, pueden prepararse antiseros policlonales o anticuerpos monoclonales usando métodos estándar. Un mamífero (por ejemplo, un ratón, hámster o conejo) puede inmunizarse con una forma inmunogénica del péptido que desencadena una respuesta de anticuerpo en el mamífero. Las técnicas para conferir inmunogenicidad a un péptido incluyen conjugación con portadores u otras técnicas bien conocidas en la técnica. Por ejemplo, la proteína o péptido pueden administrarse en presencia de coadyuvante. La progresión de la inmunización puede monitorizarse mediante la detección de títulos de anticuerpo en plasma o suero. Pueden usarse procedimientos ELISA u otros inmunoensayos estándar con el inmunógeno como antígeno para evaluar los niveles de anticuerpos. Después de inmunizar, pueden obtenerse anticuerpos y, si se desea, anticuerpos policlonales aislados a partir de los sueros.

- Para producir anticuerpos monoclonales, pueden recogerse células productoras de anticuerpo (linfocitos) de un animal inmunizado y fusionarse con células de mieloma mediante procedimientos de fusión celular somática estándar, inmortalizando así estas células y proporcionando células de hibridoma. Dichas técnicas son bien conocidas en la técnica (por ejemplo, la técnica de hibridoma, desarrollada originalmente por Kohler y Milstein (*Nature* 256, 495-497 (1975)), así como otras técnicas tales como la técnica de hibridoma de linfocitos B humanos (Kozbor *et al.*, *Immunol. Today* 4, 72 (1983)), la técnica de hibridoma de EBV para producir anticuerpos monoclonales humanos (Cole *et al.*, "Monoclonal Antibodies in Cancer Therapy" (1985) Allen R. Bliss, Inc., páginas 77-96), y el cribado de colecciones combinatorias de anticuerpos (Huse *et al.*, *Science* 246, 1275 (1989)). Las células de hibridoma pueden cribarse in vitro para la producción de anticuerpos específicamente reactivos con el péptido, y pueden aislarse los anticuerpos monoclonales. Por lo tanto, la divulgación contempla también células de hibridoma que secretan anticuerpos monoclonales con especificidad por una proteína relacionada con el SRAG como se describe en la presente memoria.

- 50 El término "anticuerpo" como se usa en la presente memoria se pretende que incluya fragmentos del mismo que reaccionan específicamente también con una proteína relacionada con el SRAG o fragmentos de la misma. Los anticuerpos pueden fragmentarse usando técnicas convencionales y cribarse en los fragmentos la utilidad de la misma manera que se describe anteriormente. Por ejemplo, los fragmentos F(ab')₂ pueden generarse tratando el anticuerpo con pepsina. El fragmento F(ab')₂ resultante puede tratarse para reducir los puentes disulfuro, produciendo fragmentos Fab'.

Se contemplan también dentro del alcance de la divulgación derivados de anticuerpo quiméricos, concretamente, moléculas de anticuerpo que combinan una región variable de animal no humano y una región constante humana.

Las moléculas de anticuerpo quimérico pueden incluir, por ejemplo, el dominio de unión a antígeno de un anticuerpo de ratón, rata u otra especie, con regiones constantes humanas. Pueden usarse métodos convencionales para preparar anticuerpos quiméricos que contienen la región variable de inmunoglobulina que reconoce el producto génico de antígenos de proteína relacionada con el SRAG (véanse, por ejemplo, Morrison *et al.*, Proc. Natl. Acad. Sci. U. S. A. 81, 6851 (1985); Takeda *et al.*, Nature 314, 452 (1985), Cabilly *et al.*, patente de EE.UU. n° 4.816.567; Boss *et al.*, patente de EE.UU. n° 4.816.397; Tanaguchi *et al.*, publicación de patente europea EP171496; publicación de patente europea 0173494, patente del Reino Unido GB 2177096B). Se espera que los anticuerpos quiméricos sean menos inmunogénicos en un sujeto humano que el correspondiente anticuerpo no quimérico.

Los anticuerpos monoclonales o quiméricos específicamente reactivos con una proteína como se describe en la presente memoria pueden humanizarse adicionalmente produciendo quimeras de región constante humana, en que partes de las regiones variables, particularmente las regiones estructurales conservadas del dominio de unión a antígeno, son de origen humano y solo las regiones hipervariables son de origen no humano. Dichas moléculas de inmunoglobulina pueden prepararse mediante técnicas conocidas en la técnica (por ejemplo, Teng *et al.*, Proc. Natl. Acad. Sci. U.S.A., 80, 7308- 7312 (1983); Kozbor *et al.*, Immunology Today, 4, 7279 (1983); Olsson *et al.*, Meth. Enzymol., 92, 3-16 (1982)) y la publicación PCT WO92/06193 o EP 0239400). Los anticuerpos humanizados pueden producirse también comercialmente (Scotgen Limited, 2 Holly Road, Twickenham, Middlesex, Reino Unido).

Pueden generarse también anticuerpos específicos, o fragmentos de anticuerpo tales como, pero sin limitación, anticuerpos monoclonales Fv monocatenarios reactivos frente a proteínas relacionadas con el SRAG, cribando colecciones de expresión que codifican genes de inmunoglobulina, o porciones de los mismos, expresados en bacterias con péptidos producidos a partir de las moléculas de ácido nucleico de proteínas relacionadas con el SRAG. Por ejemplo, pueden expresarse fragmentos de Fab completos, regiones VH y regiones FV en bacterias usando colecciones de expresión en fago (véanse, por ejemplo, Ward *et al.*, Nature 341, 544-546: (1989); Huse *et al.*, Science 246, 1275-1281 (1989) y McCafferty *et al.* Nature 348, 552-554 (1990)). Como alternativa, puede usarse un ratón SCID-hu, por ejemplo el modelo desarrollado por Genpharm, para producir anticuerpos o fragmentos de los mismos.

(b) Ensayos de diagnóstico

El hallazgo por los presentes inventores de proteínas relacionadas con el SRAG novedosas permite el desarrollo de ensayos de diagnóstico del SRAG.

En una realización, la proteína relacionada con el SRAG, o fragmento funcional de la misma, puede usarse en un método para detectar anticuerpos en una muestra de pacientes sospechosos de tener el SRAG. En consecuencia, la presente divulgación proporciona un método para detectar el SRAG en una muestra que comprende poner en contacto la muestra con una proteína relacionada con el SRAG que puede detectarse después de unirse a un anticuerpo de proteína relacionada con el SRAG en la muestra. Estos métodos incluyen, por ejemplo, ensayo de inmunosorción ligado a enzima (ELISA), radioinmunoensayos (RIA), ensayos inmunoradiométricos (IRMA) y ensayos inmunoenzimáticos (IEMA). Todos estos métodos de detección son bien conocidos en la técnica (por ejemplo, véanse la patente de EE.UU. 4.376.110 o la patente de EE.UU. 4.486.530). En este contexto, un "fragmento funcional" incluye un fragmento de la proteína relacionada con el SRAG que puede unirse a anticuerpos encontrados en una muestra tomada de un paciente infectado por coronavirus.

En una realización específica, el ensayo de diagnóstico usará la proteína de nucleocápsida, o un fragmento de la misma, como se muestra en la Figura 1, ya que la proteína de nucleocápsida parece ser un antígeno inmunogénico importante de los coronavirus.

En otra realización, puede ensayarse en una muestra de un paciente la presencia de proteína relacionada con el SRAG. La proteína relacionada con el SRAG puede detectarse en una muestra usando anticuerpos que se unen a la proteína relacionada con el SRAG que se describen con detalle en la sección II(a). En consecuencia, la presente divulgación proporciona un método para la detección del SRAG en una muestra que comprende poner en contacto la muestra con un anticuerpo que se une a una proteína relacionada con el SRAG que puede detectarse después de unirse a la proteína relacionada con el SRAG en la muestra.

Pueden usarse anticuerpos específicamente reactivos con proteínas relacionadas con el SRAG, o derivados de las mismas tales como conjugados enzimáticos o derivados marcados, para detectar proteínas relacionadas con el SRAG en diversos materiales biológicos; por ejemplo, pueden usarse en cualquier inmunoensayo conocido que se base en la interacción de unión entre un determinante antigénico de proteína relacionada con el SRAG y los anticuerpos. Son ejemplos de dichos ensayos radioinmunoensayos, enzimoimmunoensayos (por ejemplo, ELISA), inmunofluorescencia, inmunoprecipitación, aglutinación con látex, hemaglutinación y pruebas histoquímicas. Por tanto, los anticuerpos pueden usarse para detectar y cuantificar proteínas relacionadas con el SRAG en una muestra para determinar su papel en eventos celulares particulares o estados patológicos, y para diagnosticar y tratar dichos estados patológicos.

Las técnicas citoquímicas conocidas en la técnica para localizar antígenos usando microscopía óptica y electrónica pueden usarse para detectar proteínas relacionadas con el SRAG. En general, un anticuerpo de la invención puede

- 5 marcarse con una sustancia detectable y puede localizarse la proteína relacionada con el SRAG en tejido basándose en la presencia de la sustancia detectable. Los ejemplos de sustancias detectables incluyen diversas enzimas, materiales fluorescentes, materiales luminiscentes y materiales radioactivos. Los ejemplos de enzimas adecuadas incluyen peroxidasa de rábano picante, biotina, fosfatasa alcalina, β -galactosidasa o acetilcolinesterasa; los ejemplos de materiales fluorescentes adecuados incluyen umbeliferona, fluoresceína, isotiocianato de fluoresceína, rodamina, diclorotriazinilaminofluoresceína, cloruro de dansilo o ficoeritrina; un ejemplo de material luminiscente incluye luminol; y los ejemplos de materiales radiactivos adecuados incluyen yodo radiactivo I-125, I-131 o 3-H. Los anticuerpos pueden acoplarse también con sustancias de alta densidad electrónica, tales como ferritina u oro coloidal, que se visualizan fácilmente por microscopia electrónica.
- 10 Pueden emplearse también métodos indirectos en los que la reacción antígeno-anticuerpo primaria se amplifica mediante la introducción de un segundo anticuerpo que tiene especificidad por el anticuerpo reactivo frente a la proteína relacionada con el SRAG. A modo de ejemplo, si el anticuerpo que tiene reactividad frente a la proteína relacionada con el SRAG es un anticuerpo de IgG de conejo, el segundo anticuerpo puede ser gammaglobulina de cabra anti-conejo marcada con una sustancia detectable como se describe en la presente memoria.
- 15 Cuando se usa un marcaje radiactivo como sustancia detectable, la proteína relacionada con el SRAG puede localizarse mediante autorradiografía. Los resultados de la autorradiografía pueden cuantificarse determinando la densidad de partículas en las autorradiografías mediante diversos métodos ópticos, o contando los granos.

(c) Moduladores de proteína relacionada con el SRAG

- 20 Además de los anticuerpos descritos anteriormente, pueden identificarse otras sustancias que modulan la expresión o actividad de una proteína relacionada con el SRAG. En consecuencia, la presente divulgación incluye el uso de proteínas relacionadas con el SRAG para desarrollar o identificar sustancias que modulen la expresión o actividad de una proteína relacionada con el SRAG. Dichas sustancias pueden ser útiles en aplicaciones terapéuticas para tratar el SRAG que se describen a continuación.

(i) Sustancias que se unen a proteínas relacionadas con el SRAG

- 25 Las sustancias que afectan a la actividad de proteínas relacionadas con el SRAG pueden identificarse basándose en su capacidad de unirse a proteínas relacionadas con el SRAG.
- 30 Las sustancias que pueden unirse a proteínas relacionadas con el SRAG pueden identificarse haciendo reaccionar las proteínas relacionadas con el SRAG con una sustancia que se une potencialmente a proteínas relacionadas con el SRAG, y ensayando los complejos, la sustancia libre o las proteínas relacionadas con el SRAG no complejadas o la activación de proteínas relacionadas con el SRAG. En particular, puede usarse un sistema de ensayo de dos híbridos de levadura para identificar proteínas que interactúan con proteínas relacionadas con el SRAG (Fields, S. y Song, O., 1989, *Nature*, 340: 245-247). Los sistemas de análisis que pueden usarse incluyen también ELISA.

En consecuencia, la divulgación proporciona un método de identificación de sustancias que pueden unirse a una proteína relacionada con el SRAG que comprende las etapas de:

- 35 (a) hacer reaccionar una proteína relacionada con el SRAG y una sustancia de ensayo en condiciones que permitan la formación de un complejo entre las proteínas relacionadas con el SRAG y la sustancia de ensayo, y
- (b) ensayar los complejos de una proteína relacionada con el SRAG y la sustancia de ensayo, la sustancia libre o una proteína relacionada con el SRAG no complejada, en el que la presencia de complejos indica que la sustancia de ensayo es capaz de unirse a una proteína relacionada con el SRAG.
- 40 Las proteínas relacionadas con el SRAG usadas en el ensayo pueden tener la secuencia aminoacídica mostrada en la Figura 1 o 2, o pueden ser un fragmento, análogo, derivado, homólogo o mimético de las mismas como se describe en la presente memoria.

Las condiciones que permiten la formación de complejos de sustancia y proteína relacionada con el SRAG pueden seleccionarse con respecto a factores tales como la naturaleza y las cantidades de sustancia y proteína.

- 45 El complejo de sustancia-proteína, sustancia libre o proteínas no complejadas pueden aislarse mediante técnicas de aislamiento convencionales, por ejemplo, precipitación salina, cromatografía, electroforesis, filtración en gel, fraccionamiento, absorción, electroforesis en gel de poliácridamida, aglutinación o combinaciones de las mismas. Para facilitar el ensayo de los componentes, puede utilizarse un anticuerpo contra una proteína relacionada con el SRAG, o la sustancia, o una proteína relacionada con el SRAG marcada, o una sustancia marcada. Los anticuerpos, proteínas o sustancias pueden marcarse con una sustancia detectable como se describe anteriormente.
- 50

La proteína relacionada con el SRAG, o la sustancia usada en el método, puede insolubilizarse. Por ejemplo, la proteína relacionada con el SRAG o sustancia puede unirse a un portador adecuado. Son ejemplos de portadores adecuados agarosa, celulosa, dextrano, Sephadex, Sepharose, carboximetilcelulosa, poliestireno, papel de filtro, resina de intercambio iónico, película de plástico, tubo de plástico, perlas de vidrio, copolímero de poliamina-ácido

metilvinilitermaleico, copolímero aminoacídico, copolímero de etileno-ácido maleico, nailon, seda, etc. El portador puede estar en forma de, por ejemplo, un tubo, placa de ensayo, perlas, disco, esfera, etc.

La proteína o sustancia insolubilizada puede prepararse haciendo reaccionar el material con un portador insoluble adecuado usando métodos químicos o físicos conocidos, por ejemplo, acoplamiento con bromuro de cianógeno.

- 5 Las proteínas o sustancias pueden expresarse también en la superficie de una célula usando los métodos descritos en la presente memoria.

La divulgación contempla también ensayar un antagonista o agonista de la acción de una proteína relacionada con el SRAG.

(ii) Miméticos peptídicos

- 10 La presente divulgación incluye también miméticos peptídicos de proteínas relacionadas con el SRAG. Por ejemplo, un péptido derivado de un dominio de unión de una proteína relacionada con el SRAG interactuará directa o indirectamente con una molécula asociada de tal modo que imite el dominio de unión nativo. Dichos péptidos pueden incluir inhibidores competitivos, potenciadores, miméticos peptídicos y similares. Todos estos péptidos, así como las moléculas sustancialmente homólogas, complementarias o funcional o estructuralmente equivalentes de
15 otro modo a estos péptidos, pueden usarse con los fines de la presente divulgación.

- “Miméticos peptídicos” son estructuras que sirven como sustitutos de péptidos en interacciones entre moléculas (véase Morgan *et al.* (1989), *Ann. Reports Med. Chem.* 24: 243-252 para revisión). Los miméticos peptídicos incluyen estructuras sintéticas que pueden contener o no enlaces aminoacídicos y/o peptídicos pero que retienen los rasgos estructurales y funcionales de un péptido, o potenciador, o inhibidor. Los miméticos peptídicos incluyen también peptoides, oligopeptoides (Simon *et al.* (1972) *Proc. Natl. Acad. Sci USA* 89: 9367) y colecciones de péptidos que contienen péptidos de la longitud deseada que representan todas las secuencias de aminoácidos posibles correspondientes a un péptido.
20

- Los miméticos peptídicos pueden diseñarse basándose en la información obtenida mediante el reemplazo sistemático de L-aminoácidos por D-aminoácidos, el reemplazo de cadenas laterales por grupos que tienen propiedades electrónicas diferentes y mediante el reemplazo sistemático de enlaces peptídicos por reemplazos de enlace amida. Pueden introducirse también limitaciones conformacionales locales para determinar los requisitos conformacionales para la actividad de un mimético peptídico candidato. Los miméticos pueden incluir enlaces amida isoestéricos o D-aminoácidos para estabilizar o promover conformaciones de giro inverso y para ayudar a estabilizar la molécula. Pueden usarse análogos aminoacídicos cíclicos para limitar los restos aminoacídicos a estados conformacionales particulares. Los miméticos pueden incluir también miméticos de estructuras secundarias de péptido inhibidor. Estas estructuras pueden modelizar la orientación tridimensional de los restos aminoacídicos en las conformaciones secundarias conocidas de las proteínas. Pueden usarse también peptoides, que son oligómeros de aminoácidos N-sustituídos, y pueden usarse como motivos para la generación de colecciones químicamente diversas de moléculas novedosas.
25

- Los péptidos pueden usarse también para identificar compuestos de partida para el desarrollo de fármacos. La estructura de los péptidos descritos en la presente memoria puede determinarse fácilmente mediante una serie de métodos tales como RMN y cristalografía de rayos X. Una comparación de las estructuras de péptidos similares en secuencia, pero que difieren en las actividades biológicas que desencadenan en las moléculas diana, puede proporcionar información sobre la relación de estructura-actividad de la diana. La información obtenida a partir del examen de las relaciones de estructura-actividad puede usarse para diseñar péptidos modificados u otras moléculas pequeñas o compuestos de partida en que pueden ensayarse las propiedades predichas con relación a la molécula diana. La actividad de los compuestos de partida puede evaluarse usando ensayos similares a los descritos en la presente memoria.
30

- Puede obtenerse también información sobre las relaciones de estructura-actividad a partir de estudios de cocristalización. En estos estudios, se cristaliza un péptido con la actividad deseada asociado a una molécula diana, y se determina la estructura de rayos X del complejo. La estructura puede compararse entonces con la estructura de la molécula diana en su estado nativo, y puede usarse la información de dicha comparación para diseñar compuestos que se espera que la posean.
35

(iii) Métodos de cribado de fármacos

- Según una realización, la divulgación posibilita un método para el cribado de compuestos candidatos por su capacidad de modular (por ejemplo aumentar o reducir) la actividad de una proteína relacionada con el SRAG. Dichos compuestos pueden ser útiles en el tratamiento del SRAG. El método comprende proporcionar un sistema de ensayo para ensayar la actividad de proteína relacionada con el SRAG, ensayar la actividad en presencia o ausencia del compuesto candidato o de ensayo y determinar si el compuesto tiene una actividad de proteína relacionada con el SRAG aumentada o reducida.
40
55

En consecuencia, la presente divulgación proporciona un método para identificar un compuesto que afecta a la

actividad o expresión de proteína relacionada con el SRAG que comprende:

(a) incubar un compuesto de ensayo con una proteína relacionada con el SRAG o un ácido nucleico que codifica una proteína relacionada con el SRAG; y

5 (b) determinar la cantidad de actividad o expresión de proteína relacionada con el SRAG y compararla con un control (concretamente, en ausencia de la sustancia de ensayo), en el que un cambio en la actividad o expresión de la proteína relacionada con el SRAG en comparación con el control indica que el compuesto de ensayo tiene efecto sobre la actividad o expresión de proteína relacionada con el SRAG.

10 Según una realización adicional, la divulgación posibilita un método para cribar en compuestos candidatos su capacidad de aumentar o reducir la expresión de una proteína relacionada con el SRAG. El método comprende poner una célula en contacto con un compuesto candidato, en el que la célula incluye una región reguladora de un gen de una proteína relacionada con el SRAG unida operativamente a una región de codificación de gen informador, y detectar un cambio en la expresión del gen informador.

15 En una realización, la presente divulgación proporciona sistemas de cultivo en los que se incuban estirpes celulares que expresan el gen de una proteína relacionada con el SRAG, y por tanto productos de proteína relacionada con el SRAG, con compuestos candidatos para ensayar sus efectos sobre la expresión de proteína relacionada con el SRAG. Dichos sistemas de cultivo pueden usarse para identificar compuestos que regulan positivamente o regulan negativamente la expresión de proteína relacionada con el SRAG o su función mediante la interacción con otras proteínas.

20 En otra realización, la presente divulgación proporciona microchips que tienen moléculas de ácido nucleico que codifican una proteína relacionada con el SRAG unidas a los mismos. Dichos microchips son útiles en ensayo de cribado de alto rendimiento para identificar fármacos que interactúan con la proteína relacionada con el SRAG.

25 Dichos compuestos pueden seleccionarse a partir de compuestos proteicos, productos químicos y diversos fármacos que se añaden al medio de cultivo. Después de un periodo de incubación en presencia de uno o unos compuestos de ensayo seleccionados, puede examinarse la expresión de proteína relacionada con el SRAG cuantificando los niveles de ARNm de proteína relacionada con el SRAG usando el procedimiento de transferencia Northern, como se describe en los ejemplos incluidos en la presente memoria, para determinar cualquier cambio en la expresión como resultado del compuesto de ensayo. Las estirpes celulares transfectadas con constructos que expresan proteína relacionada con el SRAG pueden usarse también para ensayar la función de compuestos desarrollados para modificar la expresión proteica. Además, las estirpes celulares transformadas que expresan una proteína relacionada con el SRAG normal podrían mutagenizarse mediante el uso de agentes mutagénicos produciendo un fenotipo alterado en el que puede estudiarse el papel de la proteína relacionada con el SRAG mutada para estudiar las relaciones de estructura/función de los productos proteicos y sus efectos fisiológicos.

35 En la identificación o el diseño de fármacos potenciales, se usa la proteína biológicamente activa relacionada con el SRAG, o fragmento de la misma, para identificar moléculas pequeñas con las que interacciona para adecuarse a fármacos que interferirán con la función del polipéptido *in vivo*. En un enfoque, se determina en primer lugar la estructura tridimensional de la proteína de interés o, por ejemplo, del receptor o complejo con ligando de la proteína relacionada con el SRAG mediante cristalografía de rayos X, mediante modelización informática o, lo más típicamente, mediante una combinación de enfoques. Menos frecuentemente, puede obtenerse información útil respecto a la estructura del polipéptido mediante modelización basada en la estructura de proteínas homólogas. Por tanto, mediante uno de estos métodos, pueden diseñarse fármacos que actúen como agonistas o antagonistas de la actividad polipeptídica de proteína relacionada con el SRAG. Estos inhibidores pueden ser moléculas pequeñas, anticuerpos, anticuerpos monoclonales o fragmentos de anticuerpo. Dichos inhibidores pueden usarse por lo tanto para tratar el SRAG y enfermedades mediadas por coronavirus.

45 Esta divulgación contempla también el uso de ensayos de cribado de fármacos competitivos en que anticuerpos neutralizantes capaces de unión específica a la proteína relacionada con el SRAG compiten con un compuesto de ensayo por la unión a la proteína relacionada con el SRAG o fragmento de la misma. De esta manera, pueden usarse anticuerpos para detectar la presencia de cualquier polipéptido que comparta uno o más determinantes antigénicos de la proteína relacionada con el SRAG. Los anticuerpos pueden prepararse según métodos bien conocidos (Hurn y Chantler, 1980). Se inyecta generalmente un péptido, o un fragmento del mismo, durante un periodo de varios meses en conejos u otro animal adecuado. Se ensaya en los sueros la inmunoreactividad a la proteína relacionada con el SRAG, o fragmento de la misma.

Una vez se identifica un fármaco potencial o modulador de proteína relacionada con el SRAG usando los ensayos de cribado anteriores, puede ensayarse la eficacia terapéutica en una variedad de modelos *in vitro* o *in vivo*.

(e) Usos terapéuticos

55 Como se debate anteriormente, se aislaron las proteínas relacionadas con el SRAG a partir de un paciente con SRAG, y probablemente derivan de un coronavirus o virus similar a coronavirus. En consecuencia, la presente divulgación incluye el uso de proteínas relacionadas con el SRAG y moduladores de las mismas en aplicaciones

terapéuticas, incluyendo el tratamiento o la prevención de afecciones asociadas a coronavirus. Las definiciones usadas en esta sección se aplican a todas las realizaciones a menos que se observe otra cosa.

5 En una realización, la presente divulgación proporciona un método de tratamiento o prevención del SRAG que comprende administrar a una célula o animal necesitado de ello una cantidad eficaz de agente que modula la expresión y/o actividad de una proteína relacionada con el SRAG. La divulgación proporciona también el uso de una cantidad eficaz de un agente que modula la expresión y/o actividad de proteína relacionada con el SRAG para tratar el SRAG o en la fabricación de un medicamento para tratar el SRAG.

10 La expresión “agente que modula la expresión y/o actividad de una proteína relacionada con el SRAG” significa cualquier sustancia que pueda alterar la expresión y/o actividad de una proteína relacionada con el SRAG, e incluye agentes que puedan inhibir la expresión o actividad de una proteína relacionada con el SRAG (antagonistas de proteína relacionada con el SRAG) y agentes que puedan potenciar la expresión o actividad de una proteína relacionada con el SRAG (agonistas de proteína relacionada con el SRAG). El agente puede ser cualquier tipo de molécula incluyendo, pero sin limitación, proteínas, péptidos, ácidos nucleicos (incluyendo ADN, ARN, oligonucleótidos de sentido contrario, ácidos nucleicos peptídicos), carbohidratos, compuestos orgánicos, moléculas pequeñas, productos naturales y extractos de colecciones. Los ejemplos de agentes que pueden usarse para modular una proteína relacionada con el SRAG incluyen moléculas de ácido nucleico que codifican una proteína relacionada con el SRAG; la proteína relacionada con el SRAG, así como fragmentos, análogos, derivados u homólogos de la misma; anticuerpos; ácidos nucleicos de sentido contrario; miméticos peptídicos y sustancias aisladas usando los métodos de cribado descritos en la presente memoria.

20 La expresión “cantidad eficaz” como se usa en la presente memoria significa una cantidad eficaz, a las dosificaciones y durante los periodos de tiempo necesarios, para conseguir los resultados deseados.

El término “animal” como se usa en la presente memoria incluye todos los miembros del reino animal, incluyendo seres humanos.

25 La expresión “tratamiento o tratar” como se usa en la presente memoria significa un enfoque para obtener resultados beneficiosos o deseados, incluyendo resultados clínicos. Los resultados clínicos beneficiosos o deseados pueden incluir, pero sin limitación, el alivio o mejora de uno o más síntomas o afecciones, la reducción de la extensión de la enfermedad, un estado patológico estabilizado (concretamente, que no empeora), la prevención de la difusión de la enfermedad, el retraso o retardo de la progresión de la enfermedad, la mejora o paliación del estado patológico y la remisión (parcial o total) tanto detectable como indetectable. “Tratar” puede significar también prolongar la supervivencia en comparación con la supervivencia esperada si no se recibe tratamiento.

30 En una realización, la divulgación proporciona un método de tratamiento o prevención del SRAG que comprende administrar una cantidad eficaz de un antagonista de una proteína relacionada con el SRAG. La presente divulgación proporciona también el uso de una cantidad eficaz de un antagonista de una proteína relacionada con el SRAG para tratar el SRAG o en la fabricación de un medicamento para tratar el SRAG.

35 La expresión “antagonista de una proteína relacionada con el SRAG” incluye cualquiera sustancia que pueda inhibir la expresión de un gen de una proteína relacionada con el SRAG y/o la actividad de una proteína relacionada con el SRAG. Los ejemplos de antagonistas de una proteína relacionada con el SRAG incluyen anticuerpos como se describen anteriormente.

40 En una realización específica, el antagonista se unirá a o interferirá con la actividad de la glucoproteína de espícula mostrada en la Figura 2.

(f) Composiciones farmacéuticas

45 Las sustancias anteriormente descritas, que incluyen ácidos nucleicos que codifican proteínas relacionadas con el SRAG, proteínas relacionadas con el SRAG, anticuerpos, así como otros agentes que modulan proteínas relacionadas con el SRAG, pueden formularse en composiciones farmacéuticas para administración en sujetos de forma biológicamente compatible adecuada para administración *in vivo*. Por “forma biológicamente compatible adecuada para administración *in vivo*” se pretende una forma de la sustancia a administrar en que cualquier efecto tóxico sea compensado por los efectos terapéuticos. Las sustancias pueden administrarse a organismos vivos, incluyendo seres humanos y animales.

50 La administración de una cantidad terapéuticamente activa de composiciones farmacéuticas de la presente divulgación se define como una cantidad eficaz a las dosificaciones y durante los periodos de tiempo necesarios para conseguir el resultado deseado. Por ejemplo, una cantidad terapéuticamente activa de una sustancia puede variar según factores tales como el estado patológico, edad, sexo y peso del individuo, y la capacidad de la sustancia de desencadenar una respuesta deseada en el individuo. Los regímenes de dosificación pueden ajustarse para proporcionar la respuesta terapéutica óptima. Por ejemplo, pueden administrarse varias dosis divididas diariamente o la dosis puede reducirse proporcionalmente como se indique por las exigencias de la situación terapéutica.

Puede administrarse una sustancia activa de manera conveniente tal como mediante inyección (subcutánea, intravenosa, etc.), administración oral, inhalación, administración transdérmica o administración rectal. Dependiendo de la vía de administración, la sustancia activa puede recubrirse con un material para proteger al compuesto de la acción de enzimas, ácidos y otras condiciones naturales que pueden inactivar el compuesto. Si la sustancia activa es un ácido nucleico que codifica una proteína relacionada con el SRAG, puede suministrarse usando técnicas conocidas en la técnica.

Las composiciones descritas en la presente memoria pueden prepararse mediante métodos conocidos *per se* para la preparación de composiciones farmacéuticamente aceptables que pueden administrarse a sujetos, tal como combinar una cantidad eficaz de la sustancia activa en una mezcla con un vehículo farmacéuticamente aceptable. Los vehículos adecuados se describen, por ejemplo, en "Remington's Pharmaceutical Sciences" (Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., EE.UU. 1985) o "Handbook of Pharmaceutical Additives" (recopilado por Michael e Irene Ash, Gower Publishing Limited, Aldershot, Inglaterra (1995)). Basándose en esto, las composiciones incluyen, aunque no exclusivamente, disoluciones de las sustancias en asociación con uno o más vehículos o diluyentes farmacéuticamente aceptables, y pueden estar contenidas en disoluciones tamponadas con un pH adecuado y/o ser isoosmóticas con fluidos fisiológicos. A este respecto, puede hacerse referencia a la patente de EE.UU. n° 5.843.456. Como se apreciará por los expertos, la administración de sustancias descritas en la presente memoria puede ser mediante un portador vírico inactivo.

(g) Vacunas

La presente divulgación incluye vacunas para tratar o prevenir enfermedades mediadas por coronavirus.

En una realización, la vacuna comprende una proteína relacionada con el SRAG mezclada con un diluyente o portador adecuado. En otra realización, la vacuna comprende un ácido nucleico que codifica una proteína relacionada con el SRAG mezclado con un diluyente o portador adecuado. La expresión "proteína relacionada con el SRAG" es como se define anteriormente en la presente memoria para incluir todos los homólogos, análogos, fragmentos o derivados de la proteína relacionada con el SRAG de coronavirus, incluyendo las proteínas mostradas en las Figuras 1 y 2. Los fragmentos de la proteína relacionada con el SRAG serán suficientes para desencadenar una respuesta inmunitaria contra coronavirus.

En una realización específica, la vacuna comprenderá la glucoproteína de espícula de la Figura 2 o un fragmento de la misma. La glucoproteína de espícula desempeña un papel importante para las etapas iniciales de la infección al mediar la unión del virus a la membrana celular del hospedador. Por consiguiente, inhibir o neutralizar esta proteína puede ser una estrategia terapéutica eficaz.

En una realización adicional, la vacuna comprende un anticuerpo que se une a una proteína relacionada con el SRAG mezclado con un diluyente o portador adecuado. Dichos anticuerpos pueden prepararse como se describe anteriormente en la presente memoria.

La vacuna puede incluir una proteína (tal como una proteína relacionada con el SRAG o anticuerpo que se une a una proteína o péptido relacionado con el SRAG) o un ácido nucleico que codifica la proteína relacionada con el SRAG. Dichos ácidos nucleicos incluyen ARN o ADN libre o desnudo o en un vector. En una realización preferida, la secuencia de ácido nucleico está contenida en un vector o plásmido. En una realización, el vector puede ser vírico tal como de poxvirus, adenovirus o alfavirus. Preferiblemente, el vector vírico es incapaz de integración en células animales receptoras. Los elementos para la expresión de dicho vector pueden incluir un promotor adecuado para la expresión en células de animales receptores.

Las vacunas pueden estar en mezcla con un portador, diluyente o excipiente adecuado tal como agua estéril, disolución salina fisiológica, glucosa o similar para formar formulaciones de vacuna adecuadas. Las vacunas pueden estar también liofilizadas. Las vacunas pueden contener también sustancias auxiliares tales como agentes humectantes o emulsionantes, agentes tamponadores del pH, aditivos gelificantes o potenciadores de la viscosidad, conservantes, agentes aromatizantes, colorantes y similares, dependiendo de la vía de administración y de la preparación deseada. A este respecto, puede hacerse referencia a la patente de EE.UU. n° 5.843.456. Puede hacerse referencia también al libro de texto "Vaccine Design: the Subunit and Adjuvant Approach", Michael F. Powell y Mark J. Newman, eds. Plenum Press, Nueva York, 1995.

La vacuna puede incluir adicionalmente coadyuvantes para potenciar la respuesta inmunitaria a la proteína relacionada con el SRAG o anticuerpo. Un amplio intervalo de coadyuvantes extrínsecos puede provocar respuestas inmunitarias potentes ante antígenos. Estos incluyen saponinas complejadas con antígenos de proteína de membrana (complejos inmunoestimulantes), polímeros plurónicos con aceite mineral, micobacterias muertas y aceite mineral, coadyuvante completo de Freund, productos bacterianos tales como dipéptido de muramilo (MDP) y lipopolisacárido (LPS), así como lípido A y liposomas.

En consecuencia, la presente divulgación proporciona un método para desencadenar una respuesta inmunitaria frente a coronavirus que comprende administrar una cantidad eficaz de una formulación de vacuna que comprende una proteína relacionada con el SRAG, un ácido nucleico que codifica una proteína relacionada con el SRAG o un anticuerpo de una proteína relacionada con el SRAG a un animal necesitado de ello.

La expresión “desencadenar una respuesta inmunitaria” se define como causar, potenciar o mejorar cualquier respuesta del sistema inmunitario, por ejemplo, de naturaleza humoral o mediada por célula. Que una vacuna o antígeno desencadene una respuesta inmunitaria puede evaluarse usando ensayos conocidos por los expertos en la técnica incluyendo, pero sin limitación, ensayos de anticuerpo (por ejemplo, ensayos ELISA), ensayos de citotoxicidad específica de antígeno y la producción de citocinas (por ejemplo, ensayos ELISPOT).

La expresión “administración” se define como cualquier vía convencional de administración de un antígeno a un animal para uso en el campo de las vacunas como es conocido por un experto en la técnica. Esto puede incluir, por ejemplo, administración por vía tópica, oral y parenteral (concretamente, subcutánea, intradérmica, intramuscular, etc.) e incluye además suministro transdérmico y mucoso, incluyendo suministro mucoso realizado mediante alimentación oral, inhalación y a través de membranas accesibles a través de las porciones terminales del intestino grueso.

Se proporcionan también kits para llevar a cabo los ensayos de detección de SRAG debatidos anteriormente. Por ejemplo, dichos kits pueden incluir uno cualquiera o más de los siguientes: un polipéptido relacionado con el SRAG aislado, uno o más anticuerpos específicos de un polipéptido relacionado con el SRAG, un marcaje de detección e instrucciones para uso del kit. Como se apreciará por un experto en la técnica, pueden incluirse también otros componentes y reactivos adecuados. Ha de observarse que el kit anteriormente descrito puede usarse para detectar la presencia de anticuerpos anti-SRAG en una muestra (usando polipéptido relacionado con el SRAG aislado marcado o fragmentos inmunogénicos del mismo) o para detectar la presencia de polipéptidos relacionados con el SRAG.

Los siguientes ejemplos no limitantes son ilustrativos de la presente invención.

EJEMPLO 1

PROCEDIMIENTOS EXPERIMENTALES

Descubrimiento y aislamiento del coronavirus del SRAG

Se cultivó coronavirus del SRAG en un cultivo celular de Vero E6 usando medio esencial mínimo (designado MEM) suplementado con antibióticos, fungizona, aminoácidos, vitaminas, tampón HEPES y suero fetal bovino al 3%. Se almacenó el medio a 4°C.

Se inocularon en monocapas confluyentes de células Vero E6 cultivadas en tubos de cultivo especímenes de secreción respiratoria de pacientes, se incubaron a 37°C y se examinaron diariamente con microscopio invertido. La observación de un efecto citopático en algunos de los cultivos inoculados (en comparación con las células sanas normales) indicaba la posibilidad de crecimiento vírico.

Toda la bacteriología fue negativa. La identificación del virus como un nuevo coronavirus se realizó mediante pruebas de PCR para la mayoría de familias de virus conocidas. Esto excluyó a muchos agentes como causa de la enfermedad, especialmente virus o bacterias conocidos por estar asociados a la neumonía. Sin embargo, hubo una reacción positiva para 5 de los 8 coronavirus del agrupamiento inicial de SRAG. La secuenciación de los productos de PCR dio una secuencia que confirmaba que un coronavirus era el agente etiológico probable, y que el coronavirus del SRAG es una nueva familia de virus que está relacionada lejanamente con los coronavirus de ser humano y animal conocidos anteriormente. Empleando el antígeno de coronavirus del SRAG cultivado en Vero E6, se desarrollaron pruebas serológicas (ELISA e IFA) para detectar la presencia de anticuerpos en pacientes. Empleando estos métodos, ensayos adicionales confirmaron al coronavirus del SRAG como la causa de la enfermedad. En más de 250 casos sospechosos o probables de SRAG, un 95% se probó positivo de este nuevo coronavirus.

Aislamiento de nucleocápsida, purificación de ARN y extracción de coronavirus del SRAG

Se inocularon células Vero E6 con coronavirus del SRAG a una multiplicidad de entre 1 y 5 unidades infecciosas por célula. Cuando aproximadamente un 80% de las células infectadas con CoV del SRAG mostraron ECP (3 días después de la infección), se recogieron para aislamiento de la nucleocápsida vírica. Se aislaron las nucleocápsidas de CoV del SRAG a partir de células infectadas como sigue. Brevemente, se lisaron 10^7 células infectadas en 6 ml de tampón de lisis (0,5% de desoxicolato, 0,5% de NP-40, Tris 30 mM, pH 7,5, $Mg(C_2H_3O_2) \cdot 2,4H_2O$ 5 mM, KCl 125 mM, EDTA 0,5 mM, $CaCl_2$ 3,6 mM, 2-mercaptoetanol 6 mM) y se sometieron a sonicación (Branson Sonifier 250) en hielo a un 50% de potencia durante 10 ráfagas de 10 segundos. Se confirmó la lisis celular mediante microscopía óptica en un hemocitómetro. Se añadieron 600 unidades de ADNasa I (Invitrogen, Burlington, Ontario) y 6 μ l de ARNasa A 10 mg/ml (Sigma, Oakville, Ontario) y se incubó la mezcla a 30°C durante 90 minutos con 600 unidades de ADNasa I cada 30 minutos. Se añadieron 6 ml de 1,1,2-triclorotrifluoroetano (Sigma, Oakville, Ontario) y se mezcló vigorosamente antes de centrifugar a 800 xg durante 10 minutos. Se retiró la fracción superior y se ultracentrifugó (Beckman Optima XL-100K) a través de un gradiente de glicerol de 5 a 40% en tampón de lisis a 151.000 x g durante 1 hora a 4°C. Se extrajo entonces el ARN de la preparación de nucleocápsida usando el kit RNeasy de Qiagen como sigue:

1. Se resuspende el sedimento en 100 µl de PBS y se añade a 350 µl de RTL (con beta-mercaptoetanol y ARN poliA) para extracción de ARN.
2. Se añaden 450 µl de EtOH al 70% y se agita con vórtex.
3. Se une la columna de purificación a un colector de vacío y se fija el vacío a 4,14 MPa.
- 5 4. Se añade muestra a la columna y se aplica vacío hasta que la muestra entera ha pasado a través.
5. Se añaden 700 µl de tampón RW1 y se aplica vacío.
6. Se añaden 500 µl de tampón RPE y se aplica vacío.
7. Se añaden 500 µl de tampón RPE y se aplica vacío (concretamente, 2 lavados).
- 10 8. Se transfiere la columna a un tubo de recogida y se centrifuga durante 2 min a 10 krpm para retirar las trazas de RPE.
9. Se transfiere la columna a un tubo Eppendorf de 1,5 ml y se añaden 50 µl de agua DEPC a la membrana de columna. Se deja reposar 1 min.
10. Se centrifuga durante 1 min a 10 krpm y se almacena el ARN a -70°C para uso posterior en secuenciación. Se confirmó la pureza y rendimiento del ARN vírico mediante espectrofotometría

15 Medios de crecimiento de células Vero E6

Reactivos:

Suero fetal bovino, nº de cat. Gibco 16000-044,

Medio esencial mínimo (MEM) con sales de Earle sin L-glutamina, nº de cat. Gibco BRL 11090-081,

20 Penicilina-estreptomicina-glutamina 100x (10.000 U/ml de penicilina G; 10.000 µg/ml de sulfato de estreptomicina y 29,2 mg de L-glutamina) nº de cat. Gibco BRL 10378-057,

Tampón HEPES, disolución madre 1 M, nº de cat. Gibco BRL 15630-080,

Disolución de piruvato de sodio MEM 100 mM (100x), líquido, nº de cat. Gibco BRL 11360-070,

Disolución de aminoácidos no esenciales MEM 10 mM (100x), líquido, nº de cat. Gibco BRL 11140-050

Para preparar 500 ml de medio de crecimiento celular:

Reactivo	Volumen	Concentración final
Suero fetal bovino	50 ml	10%
Penicilina-estreptomicina- glutamina	5,0 ml	
Penicilina		100 U/ml
Estreptomicina		100 µg/ml
Glutamina		0,292 mg/ml
Tampón HEPES	12,5 ml	25 mM
AA no esenciales MEM	5 ml	0,1 mM
Piruvato de sodio MEM	5 ml	1,0 mM

25 *Preparación del material primario en el NML -*

Como se debate anteriormente, se investigó extensamente en los especímenes clínicos obtenidos a partir del grupo de casos original la presencia de patógenos bacterianos y víricos (1). Se encontró que la torunda nasofaríngea y los fluidos de lavado broncoalveolar de varios de estos pacientes eran positivos por reacción en cadena de la polimerasa-transcripción inversa (PCR-TI) de metapneumovirus humano y el coronavirus novedoso (1). La inoculación del fluido de lavado broncoalveolar de un paciente masculino de 43 años en células Vero E6 produjo un fuerte efecto citopático cuatro días después de la infección. El segundo pase de este aislamiento vírico se usó adicionalmente para producir grandes cantidades del virus. Inicialmente, se usó este material vírico para evaluar su antigenicidad con muestras de suero de convaleciente de pacientes de SRAG. Los sueros de convaleciente que se

encontró anteriormente que eran positivos de anticuerpos del virus mediante ensayo de inmunofluorescencia indirecta (IFA) (11) reaccionaron fuertemente en transferencia Western con una proteína de ~46 kDa (Fig. 3A) similar en tamaño a la proteína de nucleocápsida de coronavirus (12).

- 5 Para preparar esto (y quizás otras proteínas relacionadas con el SRAG) para digestión proteolítica, se purificó el virus mediante gradiente lineal de sacarosa de 20-60%. La transferencia Western de las fracciones de gradiente mostraron que la fracción nº 4 (densidad 1,18 g/cm³) reaccionaba fuertemente con suero de convaleciente de un paciente de SRAG. Se procesó esta fracción en PAGE-SDS y se tiñó con azul de Coomassie (Fig. 3B), se extrajo entonces y se transfirió a los laboratorios de la universidad para digestión con diversas enzimas proteolíticas.

Digestiones proteolíticas

- 10 Se digirieron las bandas de proteína extraídas en gel con una de tres enzimas diferentes [tripsina modificada (Promega), Lys-C, o Asp-N (ambas de Roche Biochemical) con pureza de secuenciación]. Se efectuaron las digestiones según el procedimiento descrito por Shevchenko *et al.* (13) en H₂¹⁶O, o en una mezcla 1:1 de H₂¹⁶O:H₂¹⁸O (14-16), preparada a partir de H₂¹⁸O al 98% (Isotec). A menos que se indique otra cosa, todos los demás productos químicos se adquirieron en Sigma.
- 15 Se liofilizaron los extractos que contenían la mezcla peptídica y se resuspendieron en 5,5 ml de TFA al 0,5% en agua. Se mezclaron 0,5 ml de la muestra resultante 1:1 con una disolución matriz de ácido 2,5-dihidroxibenzoico (DHB, 150 mg/ml en agua:acetonitrilo 1:1) y se depositaron en la superficie de oro de una diana MALDI. Se usaron los 5 ml restantes para análisis mHPLC-MALDI-QqTOF.

Cromatografía

- 20 Se efectuaron las separaciones cromatográficas usando un sistema Agilent 1100 Series. Se usaron agua desionizada (18 MΩ) y acetonitrilo de pureza HPLC para la preparación de los eluyentes. Se inyectaron muestras (5 µl) en una columna de 150 mm x 150 mm (Vydac 218 TP C18, 5m) y se eluyeron con un gradiente lineal de 1-80% de acetonitrilo (0,1% de TFA) en 60 min. Se mezcló el efluente de columna (4 ml/min) en línea con una disolución matriz de DHB (0,5 µl/min) y se depositó mediante un pequeño robot controlado electrónicamente sobre una diana de oro móvil a intervalos de 1 min (17). La amplia mayoría de los fragmentos trípticos eluyó al cabo de 40 min en las condiciones de HPLC usadas, así que se recogieron normalmente 40 fracciones.
- 25

Análisis de glucoproteína

- La intención original era posponer cualquier análisis detallado de la proteína de mayor masa a una investigación posterior. Cuando los inventores decidieron incluir este esfuerzo en las presentes medidas, los únicos materiales que tenían disponibles eran dos muestras liofilizadas de digestiones de la proteína mayor (banda de ~180 kDa), una de una digestión tríptica y otra de una digestión con Lys-C. Se separó la muestra de la digestión tríptica mediante HPLC y se usó para análisis de péptidos glucosilados. Se digirió la muestra de la digestión con Lys-C dos veces más, primero con pNGasa F (Roche Biochemical) para retirar los azúcares y entonces con tripsina para producir fragmentos menores (ambas digestiones en H₂¹⁶O).
- 30

Espectrometría de masas-tiempo de vuelo

- Se analizaron las manchas en las dianas de oro individualmente, tanto por espectrometría de masas individual (EM) como por espectrometría de masas en tándem (EM/EM) en el espectrómetro de masas de cuadrupolo/tiempo de vuelo prototípico de Manitoba/Sciex (QqTOF) (18). Se produjeron iones mediante irradiación de la diana con pulsos de fotones de un láser de nitrógeno de 20 Hz (Laser Science) con 300 mJ de energía por pulso. La inyección ortogonal de iones desde el cuadrupolo a la sección de TOF producía normalmente una potencia de resolución de masa de 10.000 FWHM y una exactitud de unos pocos mDa en el instrumento de TOF tanto para espectros de EM como de EM/EM, a condición de que el pico iónico fuera razonablemente intenso.
- 35
- 40

RESULTADOS

Espectros de masas de digestiones proteolíticas de la proteína de ~47 kDa

- 45 La Figura 4A muestra el espectro de m/z de la mezcla de péptidos resultante de la digestión tríptica de la proteína de ~47 kDa en H₂¹⁶O, antes de fraccionamiento por HPLC. Se muestra una pequeña región de este espectro en la Fig. 4B, y se muestra una fracción de HPLC que muestra algunos de los mismos iones en la Fig. 4C. Aquí, el ión más intenso en la Fig. 5B se ha trasladado a una fracción diferente, pero algunos de los iones más débiles son mucho más destacados, de modo que resulta claro que los picos peptídicos individuales son considerablemente más fáciles de distinguir después de la separación por HPLC. Los espectros de las fracciones son drásticamente más sencillos y tienen una relación de señal/ruido mejorada por un factor ~10 o más.
- 50

Los esfuerzos iniciales para identificar la proteína (basados en una búsqueda en base de datos frente a la huella peptídica) no consiguieron proporcionar ninguna coincidencia significativa, sugiriendo que era una proteína novedosa. Se emprendió por lo tanto una secuenciación peptídica *de novo* para caracterizarla. Con este fin, se

5 digirieron las muestras en presencia de una mezcla 50/50 de $H_2^{16}O$ y $H_2^{18}O$, como se describe anteriormente, puesto que la adición de ^{18}O o ^{16}O durante la escisión enzimática identifica los fragmentos que contienen el extremo C por sus patrones isotópicos distintivos (14-16). Para determinar la secuencia aminoacídica de los fragmentos proteolíticos, se seleccionó a su vez cada ión peptídico claramente observado como original por el cuadrupolo selector de masa del instrumento de QqTOF, y se sometió a disociación inducida por colisión en la celda de colisión. Por ejemplo, se muestra en la Fig. 5 el espectro del ión derivado resultante a partir del original de $m/z = 2297$, en que resultan claramente evidentes las ventajas de la adición de $^{16}O/^{18}O$ para distinguir los iones C- y N-terminales. Los iones y, que contienen el extremo C, muestran toda la estructura de doblete superpuesta con el patrón isotópico habitual, mientras que los iones b, que contienen el extremo N, tienen un patrón normal. Se da en la Tabla 1 una comparación entre los valores de m/z medidos y las masas calculadas a partir de la secuencia deducida.

10 Se proporcionan ejemplos adicionales en el material suplementario. La Fig. 8 muestra los iones derivados de la disociación del péptido N-terminal de 1144 Da, que indican delección de la metionina N-terminal y acetilación de la serina N-terminal resultante. La Fig. 9 muestra una comparación entre los derivados de originales separados por HPLC de masas 1411 y 1795 Da, de digestiones tríplica y por Lys-C, respectivamente. La Fig. 10 muestra un espectro del ión original que contiene el extremo C, el único péptido C-terminal que no muestra estructura de doblete.

15 Se da en la Tabla 2 una comparación de los valores de m/z experimentales y las masas calculadas para las secuencias deducidas de todos los péptidos observados en digestiones tríplicas. En ambas tablas 1 y 2, los valores de m/z más observados y las masas calculadas a partir de las secuencias aminoacídicas deducidas concuerdan en ~10 mDa, dando crédito a las asignaciones; los valores anómalamente altos observados para unos pocos iones de la Tabla 1 corresponden a picos de muy baja intensidad.

20 Las medidas de EM y EM/EM recién descritas se aplicaron por primera vez a los péptidos resultantes de las digestiones tríplicas de la banda de gel, enumerados en la Tabla 1, y entonces a los productos de una digestión con Lys-C. Se emprendió entonces la búsqueda BLAST (20, 21) de la base de datos de proteína de Genbank total con estos péptidos para buscar homología. El ejemplo más definitivo se proporcionó por el péptido tríplico de 2297 Da. En ese caso, los resultados con mejor puntuación de la búsqueda BLAST se muestran en la Fig. 6; todos son proteínas de nucleocápsida de coronavirus, y todos proporcionan puntuaciones BLAST de 40 a 41, con valores de E de 0,003. Además, la mayor puntuación conseguida en la búsqueda BLAST que no es una proteína de coronavirus tenía una puntuación de solo 29 y un alto de valor de E de 9,4. Por tanto, la proteína de ~47 kDa es claramente una proteína de nucleocápsida de coronavirus; hay una concordancia completa entre los primeros 10 restos y los encontrados por BLAST en una región de los coronavirus que está altamente conservada. Por otro lado, solo 3 de los siguientes 9 restos concuerdan con cualquiera de los otros virus, de modo que el virus del SRAG es significativamente diferente de cualquiera de los otros coronavirus. Las búsquedas BLAST con otros péptidos reforzaron la evidencia de diferencias significativas entre el coronavirus del SRAG y cualquier otro coronavirus en la base de datos, pero condujeron a conclusiones similares.

25 Para el 12 de abril, se habían llevado a cabo estas medidas y analizado la mayoría de ellas, proporcionando una información de secuencia casi completa sobre los péptidos individuales, como se resume en la Tabla 2. La tarea de ajustar los péptidos individuales entre sí no se realizó, sin embargo, puesto que seguía habiendo una serie de ambigüedades en su orden. Para resolver este problema, se había llevado a cabo también una digestión con Asp-N (pero todavía no separado por HPLC) y se planearon digestiones con Glu-C y quizás Arg-C en cuanto estuviera disponible suficiente material. Se planearon también búsquedas BLAST más sofisticadas (20). Sin embargo, estas medidas y análisis resultaron ser innecesarias porque en esa etapa se obtuvo una secuencia nucleotídica de material infeccioso (también preparada por el NML), por un grupo del Centro de genoma Michael Smith "Michael Smith Genome Centre" en Vancouver (acceso a la base de datos de Genbank AY274119), pronto seguidas por resultados similares de varios otros laboratorios.

30 Pronto resultó evidente que el marco de lectura abierto identificado por el grupo de Vancouver como específico de la proteína de nucleocápsida de coronavirus predecía de hecho la secuencia aa de la proteína de ~47 kDa que se estaba analizando, de modo que fue posible eliminar las ambigüedades restantes para ordenar los fragmentos proteolíticos enumerados en la Tabla 2. Se muestra en la Fig. 11A una comparación de estos resultados con la secuencia predicha; los datos de espectros de masas cubren más de un 96% de la secuencia predicha, e incluyen ambos extremos C y N. Los espectros de masas indican también la retirada de la metionina N-terminal y la oxidación de todas las demás metioninas, así como la acetilación de la serina N-terminal resultante, como se muestra en la Fig. 8. La delección y acetilación N-terminales ocurren presuntamente como resultado de modificaciones post-traduccionales que, por supuesto, no pudieron predecirse por los datos nucleotídicos. Por lo demás, los resultados están en completo acuerdo con la secuencia predicha (acceso a la base de datos de Genbank AY274119), un resultado consistente con que las muestras derivan de la misma fuente infecciosa en el NML.

Espectros de masas de digestiones proteolíticas de la proteína de espícula

35 Además de la proteína de ~47 kDa casi completamente definida, se ha caracterizado parcialmente una proteína que aparecía como una banda muy débil a una masa aparente de ~180 kDa en la separación en gel (Fig. 3B). A pesar de la baja intensidad, se encontró que 39 péptidos de la digestión tríplica inicial pertenecían a la "proteína de

espícula" de ~139 kDa predicha por la secuencia nucleotídica (acceso a la base de datos de Genbank AY274119), y 36 de ellos eran suficientemente intensos para medidas de EM/EM, que confirmaron la identificación (30% de cobertura). Se da en la Tabla 5 un resumen de los datos y de la cobertura para esta proteína.

5 Esta proteína es homóloga de las proteínas de espícula de otros coronavirus, que se considera que están extensamente glucosiladas y que actúan como proteínas de unión. Es más, la secuencia predicha de la proteína de espícula del coronavirus del SRAG contiene 23 sitios de N-glucosilación potenciales (NXT o NXS), de los cuales 17 se identifican como sitios probables por Netglyc 1.0 server (21). (La O-glucosilación puede ser también posible, pero no se ha investigado aquí).

10 Por lo tanto, se trató una digestión triptica de la proteína de espícula con PNGasa F para retirar los glucanos, como se describe anteriormente. Esta etapa convierte los restos de asparagina en ácido aspártico, especificando por tanto los correspondientes péptidos desglucosilados mediante la observación de diferencias de masas de 0,984 Da por sitio desglucosilado de los valores calculados a partir de la secuencia aa predicha. Este procedimiento identificó 9 glucopéptidos y sus productos desglucosilados (Tabla 3) y elevó la cobertura de secuencia a un 42%. Las medidas de EM/EM en todos los péptidos desglucosilados confirmaron los sitios de N-glucosilación predichos, y se encontró que T111-126, T316-333 y T1140-1163 tenían dos sitios de glucosilación cada uno (Tabla 3). Por ejemplo, la Fig. 5 muestra los espectros de EM y EM/EM del péptido desglucosilado T222-232. No se observaron péptidos que contuvieran sitios posibles que no estuvieran glucosilados, sugiriendo que algunos de los otros sitios pueden estar modificados también por glucosilación.

20 La digestión triptica de la proteína de espícula sin desglucosilación con PNGasa F proporcionó espectros de cuatro de estos glucopéptidos que eran suficientemente intensos para análisis detallado, como se resumen en la Tabla 4. Se observó microheterogeneidad en los cuatro glucopéptidos, y a cada glucoforma se le asignó una composición de glucano ligado por N. Se encontró que dos péptidos, T222-232 y T778-796, tenían una sustitución M_xN_y rica en manosa, mientras que los otros dos (T111-126, T226-287) mostraron estructuras de glucano complejas $M_xN_yG_zF$; (M= manosa, N= N-acetilglucosamina, G= galactosa, F= fucosa), similares al patrón observado en la proteína hemaglutinina de coronavirus bovino (22). Estas composiciones comprenden cada una más de una estructura isomérica; se dan algunos ejemplos en la Figura 7D. Es importante observar que algunas glucoformas observadas pueden ser el resultado de fragmentación en origen, aunque los resultados sugieren que esto es improbable (véase a continuación).

30 La Figura 7A muestra el espectro de masas MALDI de la fracción de HPLC que contenía T111-126 glucosilado. Este espectro se interpreta en la Tabla 3 que contiene uno o dos sitios de O-glucosilación; mientras que los picos entre m/z 3000 y 3800 muestran un sitio posible de glucosilación, aquellos mayores de m/z 3800 corresponden posiblemente a T111-126 diglucosilado (véase la Fig. 5D). En la Fig. 5B, la EM de la fracción de HPLC T222-232 glucosilada destaca claramente la presencia de estructuras ricas en manosa. Aquí, el predominio de M_2N_9 , la forma mayor posible de oligosacárido rico en manosa ligado por N, sugiere que hay poca fragmentación en origen. La Figura 7C es el espectro de masas en tándem de T222-232 con una unión M_9N_2 . Este espectro muestra pérdidas de uno a cinco restos de manosa, pérdida del resto oligosacárido completo (m/z 1257,733), pérdida del resto -1N completo (m/z 1460,808) y pérdida del resto de carbohidrato por escisión intraanular (m/z 1340,768). El análisis de EM/EM de los iones de m/z 3404,572, 3607,682 y 4093,818 confirmó la presencia de estructuras de glucano complejas a partir de las pérdidas observadas de restos [G-N] (no mostrados). Se efectuó también EM en tándem en las glucoformas de T226-287 y T778-796, con resultados consistentes con las composiciones de glucano complejas sugeridas en la Tabla 4 (espectros no mostrados). Todos los espectros de EM/EM registrados en este estudio mostraron que el modo de fragmentación preferido era la pérdida del azúcar entero en lugar de la pérdida de un resto cada vez, lo que de nuevo va en contra de una fragmentación extensa en origen.

45 Se buscaron en otros péptidos candidatos sus posibles formas glucosiladas, pero no se detectaron, quizás debido a la baja cantidad de muestra. Como alternativa, pueden no ser detectables como iones positivos debido a la presencia de ácidos siálicos cargados negativamente. Es más, las varias composiciones oligosacáridas complejas galactosiladas encontradas en este estudio sugieren la presencia no detectada de ácido siálico, puesto que este último compuesto se une a la galactosa terminal en dichas estructuras.

DISCUSIÓN

50 *Proteína N de nucleocápsida*

La comparación de las secuencias aminoacídicas deducidas de diferentes proteínas N de coronavirus reveló un 32% de identidad entre el coronavirus relacionado con el SRAG y virus conocidos de los tres grupos de coronavirus. El árbol filogenético (23) de las secuencias aminoacídicas deducidas de la proteína N indicaba que el virus relacionado con el SRAG está relacionado solo lejanamente con cualquiera de los otros grupos (Fig. 11B). La distancia evolutiva, basada en este análisis de árbol filogenético, hace difícil especular sobre el origen del virus.

A pesar de la notable heterogenicidad de la proteína N de coronavirus del SRAG cuando se compara con otras nucleoproteínas de coronavirus, ciertos dominios parecen funcionalmente conservados (24). La región rica en SR de la proteína N del SRAG se parece a la de los coronavirus de murino y bovino; en un corto intervalo de 36 restos

(aminoácidos 176-212), contiene 14 serinas y 7 argininas. La secuencia aminoacídica en esta región es altamente variable entre coronavirus, excepto por un motivo central SRXX para el que las repeticiones dobles o triples son un rasgo distinguible entre todas las proteínas N de coronavirus. Esta región se ha cartografiado como el dominio de unión a ARN de la proteína N. Es un rasgo intrigante de la proteína N del SRAG que contiene cinco motivos SRXX (Fig. 11A, véase la secuencia aa destacada); que se traduzcan en una actividad de unión a ARN mucho mayor está por ver. Sin embargo, este hallazgo apoya el concepto de función conservada en el dominio rico en SR.

Se ha observado la aparición de una forma más corta de la proteína N tardíamente en infecciones por coronavirus TGEV, MHV, FIPV, BCoV, IBV y TCoV aviar (~2 a 5 kDa menor) en el cultivo celular. Se ha demostrado que las caspasas de la célula hospedadora, que se activan durante la infección por coronavirus, son responsables de esta escisión (25). Está presente un motivo de escisión de caspasa común en todas las proteínas N de coronavirus mencionadas. Además, la acumulación de forma acortada de la proteína N se correlacionaba con una reducción logarítmica doble de la producción de virus. Esta observación sugiere que la escisión de la proteína de nucleocápsida vírica por caspasas de la célula hospedadora podría ser un mecanismo general mediante el que las células infectadas eliminan coronavirus. De forma interesante, no está presente ningún motivo de escisión de caspasa en la proteína N de coronavirus relacionado con el SRAG.

Proteína S de espícula

La proteína de espícula es una diana importante de la respuesta inmunitaria celular a los coronavirus, y desempeña un papel importante en las etapas iniciales de infección. Media la unión del virus a receptores de la superficie celular e induce la fusión de membranas víricas y celulares. En varios tipos de virus, se ha destacado a menudo la importancia de la N-glucosilación de las proteínas de unión en interacciones de virus-receptor.

En los virus de la gripe C y A, que son mixovirus, Rosenthal *et al.* mostraron que la N-glucosilación de proteínas de fusión de hemaglutinina-esterasa puede tener efectos drásticos sobre el escape inmunitario, la virulencia y las interacciones con receptores celulares (26). Se ha mostrado que las hemaglutininas interaccionan con restos de ácido siálico en los receptores, y es conocido que los inhibidores de neuraminidasa inhiben la replicación de los virus de la gripe A y B (27).

Se ha mostrado que los virus de la hepatitis geman en el RE y dependen de la N-glucosilación de proteínas de cubierta para formar partículas víricas infecciosas (28,29). Dwek y sus colaboradores investigaron el efecto de la desoxinójirimicina, un inhibidor de α -glucosidasa, y encontraron que bloqueaba el procesamiento de oligosacáridos en la etapa de monoglucosilación de los sitios de Asn. Se mostró que las proteínas glucosiladas se plegaban erróneamente. Incluso a una concentración muy baja de inhibidor, los títulos víricos cayeron casi 100 veces (30). Los estudios en animales mostraron que la desoxinójirimicina tenía un efecto despreciable sobre la glucosilación del hospedador (31) y por tanto los fármacos tales como este inhibidor de α -glucosidasa se ven como buenos candidatos para el tratamiento de la hepatitis B. El virus de la hepatitis C también puede responder a estos inhibidores (32). Estudios similares usaron inhibidores de azúcar en VIH, que tiene muchos sitios ligados por N, pero mostraron menos sensibilidad a la glucosilación errónea (33).

Rossen y colaboradores modificaron las características de N-glucosilación de proteínas de espícula de coronavirus en células epiteliales cultivadas, y encontraron que los N-glucanos tenían un impacto importante sobre la formación y el comportamiento de los virus. Por ejemplo, la inhibición de la N-glucosilación de espícula por tunamicina, que inhibe la síntesis de N-glucanos, dio como resultado la síntesis de viriones sin espícula, que sin embargo seguían teniendo la capacidad de liberar viriones (34). Los mismos autores debatieron también las implicaciones de la N-glucosilación de proteínas de hemaglutinina de coronavirus de células epiteliales (34).

La secuencia genómica del coronavirus asociado al SRAG muestra que no contiene un gen que codifique hemaglutinina ni genes grandes derivados de otro virus o célula hospedadora (35). Se cree que el intervalo de hospedadores, el tropismo de tejido y la virulencia de los coronavirus de animales pueden cambiarse mutando el gen S, modificando por tanto las proteínas S (35).

Aunque no está actualmente aprobado ningún fármaco con eficacia probada contra coronavirus, existen dianas potenciales para nuevos fármacos. Por ejemplo, los inhibidores de proteasa podrían prevenir el procesamiento de la ARN polimerasa o la escisión de la glucoproteína S vírica. Encontrar anticuerpos contra la glucoproteína S o contra el receptor de coronavirus del SRAG no identificado son caminos posibles que tomar (35). También el uso de inhibidores de la glucosilación que tengan efectos mínimos sobre las células hospedadoras sería una ruta interesante (30). Muy recientemente, una contribución importante de Hilgenfeld *et al.* (36) esbozó un plan para el diseño de fármacos basado en la inhibición de la proteinasa vírica principal, denominada M pro o 3CL pro, que controla la actividad del complejo de replicación de coronavirus (36).

Las ideas para el desarrollo de vacunas contra el SRAG incluyen el uso de vacunas inactivadas o de subunidades que contienen la glucoproteína de espícula con otras proteínas víricas.

Resumen

Los presentes estudios proporcionan las primeras descripciones de proteínas derivadas del coronavirus novedoso

que se cree que es el agente etiológico del SRAG. De forma similar al patrón observado con coronavirus de animales (21), la nucleoproteína de 46 kDa parece ser el antígeno inmunogénico principal, ya que era la única proteína vírica reconocida por sueros de convalciente agudo y temprano de varios pacientes que se recuperaban del SRAG. Aunque la respuesta inmunitaria a NP podría servir como marcador de diagnóstico temprano para la infección, es improbable que una respuesta inmunitaria a esta proteína ofrezca protección, puesto que es una proteína interna y los anticuerpos neutralizantes es más probable que se orienten a proteínas de superficie (6). Sin embargo, se ha mostrado para otros coronavirus que algunos péptidos antigénicos de la proteína N pueden reconocerse sobre la superficie de células infectadas por linfocitos T (6).

La aplicación de secuenciación *de novo* mediante espectrometría de masas proporciona una alternativa al enfoque genómico habitual para la identificación de proteínas. Por tanto, era posible identificar la proteína de 46 kDa como una proteína de nucleocápsida de coronavirus novedosa incluso en ausencia de la secuencia genómica inicial. Se proporcionó también la caracterización inequívoca de las proteínas producidas, incluyendo la definición de modificaciones post-traduccionales. Sin embargo, quizás su contribución más importante es distinguir las proteínas reales expresadas de las que simplemente se teorizan o predicen por la secuencia nucleotídica.

MENCIONES COMPLETAS DE LAS REFERENCIAS DESIGNADAS EN LA MEMORIA DESCRIPTIVA

1. Poutanen, S. M. *et al.* (2003) "Identification of severe acute respiratory syndrome in Canada". N. Engl. J. Med. 348: 1995-2005

2. Booth, C. M. *et al.*, (2003) "Clinical features and short-term outcomes of 144 patients with SRAG in the greater Toronto area". JAMA 6 de mayo, [publicación electrónica antes de impresión]

3. Marra, M. A. *et al.*, (2003) "The genome sequence of the SRAG- associated coronavirus". Science 1 de mayo, [publicación electrónica antes de impresión]

4. Rota, P. A. *et al.*, (2003) "Characterization of a novel coronavirus associated with severe acute respiratory syndrome". Science 1 de mayo, [publicación electrónica antes de impresión]

5. Thomas, J. J., R. Bakhtiar, R., G. Siuzdak, G. (2001) "Mass spectrometry in viral proteomics". Acc. Chem. Res. 33,179-187

6. Seifers, D. L., Harvey, T. L., Haber, S., She, Y-M., Chernushevich, I., Ens, W. y Standing, K. G. (1999) "Natural infection of sorghum by foxtail mosaic disease in Kansas". Plant Disease 83, 905-912

7. Seifers, D. L., Salomon, R., Marie-Jeanne, V., Alliot, B., Signoret, P., Haber, S., Loboda, A., Ens, W., She, Y-M. y Standing, K. G. (2000) "Characterization of a novel potyvirus isolated from maize in Israel", Phytopathology 90, 505-513.

8. She, Y-M, Haber, S., Seifers, D. L., Loboda, A., Chernushevich, I., Perreault, H., Ens, W. y Standing, K. G. (2001) "Determination of the complete amino acid sequence for the coat protein of brome mosaic virus by time-of-flight mass spectrometry", J. Biol. Chem. 276, 20039-20047.

9. Swanson, M. I., She, Y-M, Ens, W., Brown, E. G. y Coombs, K. M. (2002) "Mammalian reovirus core protein micro 2 initiates at the first start codon and is acetylated". Rapid Commun. Mass Spectrom. 16, 2317-2324.

10. Méndez, I. I., She, Y. M., Ens, W. y Coombs, K. M. (2003) "Digestion pattern of reovirus outer capsid protein s3 determined by mass spectrometry". Virology, en prensa.

11. Li, Y., no publicado.

12. Lai, M. M. C., Holmes, K. V. (2001) en "Fields Virology" (4ª ed), pág. 1163-1185.

13. Shevchenko, A., Wilm, M., Vorm, O. y Mann, M. (1996) "Mass spectrometric sequencing of proteins from polyacrylamide gels". Anal. Chem. 68, 850-858.

14. Shevchenko, A., Chernushevich, I., Ens, W., Standing, K. G., Thomson, B., Wilm, M. y Mann, M. (1997) "Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer". Rapid Commun. Mass Spectrom. 11, 1015-1024.

15. Stewart, II, Thomson, T. y Fi geys, D. (2001) "¹⁸O labeling: a tool for proteomics". Rapid Commun. Mass Spectrom. 15, 2456-2465.

16. Yao, X., Alfonso, C. y Fenselau C. (2003) "Dissection of proteolytic ¹⁸O labeling: endoprotease catalyzed 160 to 180 exchange of truncated peptide substrates". J. Proteome Res. 2, 147-152.

17. Krokhin, O., Qian, Y., McNabb, J. R., Spicer, V., Standing, K. G. y Ens, W. (2002) "An off-line interface between HPLC and orthogonal MALDI TOF". Am. Soc. for Mass Spectrometry, 50th ASMS Conference on Mass

Spectrometry and Allied Topics, Orlando FL, EE.UU., 2-6 de junio, artículo WPP 233.

18. Loboda, A. V., Krutchinsky, A. N., Bromirski, M., Ens, W. y Standing, K. G. (2000). "A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: design and performance". Rapid Commun. Mass Spectrom. 14, 1047-1057.
- 5 19. Altschul, S. F., Stephen, F., Madden, T. L., Schaeffer, Zhang, J., Miller, W. y Lipman, D. J. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". Nucleic Acids Res. 25, 3389-3402.
- 10 20. A. Shevchenko, A., Loboda, A., Sunyaev, S., Shevchenko, A., Bork, P., Ens, W. y Standing, K. G. (2001). "Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching". Anal. Chem. 73, 1917- 1926.
21. NetNglyc 1.0 server: <http://www.cbs.dtu.dk/services/NetNGlyc/>
22. Hogue, B. G. y Brian, D. A. (1987). "Glycosylation of the bovine coronavirus hemagglutinin protein". Adv. Exp. Med. Biol. 218,131-136
- 15 23. Thomson, J. D., Higgins, D. G., Gibson, T. J. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice". Nucleic Acids Res. 22, 4673-4680.
24. Nelson, G. W., Stohlman, S. A., Tahara, S. M. (2000). "High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA". J. Gen. Virol. 81, 181-188.
- 20 25. Eleouet, J. F., Slee, E. A., Saurini, F., Castagne, N., Poncet, D., Garrido, C., Solary, E y Martin S. J. (2000). "The viral nucleocapsid protein of transmissible gastroenteritis coronavirus (TGEV) is cleaved by caspase-6 and-7 during TGEV-induced apoptosis". J. Virol. 74, 3975-3983.
26. Rosenthal, P. B., Zhang, X., Formanowski, F., Fitz, W., Wong, C. H., Meier-Ewert, H., Skehel, J. J. y Wiley, D. C. (1998) "Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus". Nature 396, 92-96.
- 25 27. Oxford, J. S., Bossuyt, S., Balasingam, S., Mann, A., Novelli, P., Lambkin, R. (2003). Clin Microbiol Infect. 9, 1-14.
28. Lu, X., Mehta, A., Dwek, R., Butters, T., Block, T. (1995). "Evidence that N-linked glycosylation is necessary for hepatitis B virus secretion". Virology 213, 660-665.
- 30 29. Mehta, A., Block, T. M., Dwek, R. A. (1998). "The role of N- glycosylation in the secretion of hepatitis B virus". Adv. Exp. Med. Biol. 435, 195-205.
30. Mehta, A., Rudd, P. M., Block, T. M., Dwek, R. A. (1997). "A strategy for anti-viral intervention: the use of alpha-glycosidase inhibitors to prevent chaperone-mediated folding of viral envelope glycoproteins". Biochem. Soc. Trans. 25, 1188-1193.
- 35 31. Freeze, H. H., Westphal, V. C. (2001). "Balancing N-linked glycosylation to avoid disease". Biochimie 83, 791-799.
32. Zitzmann, N., Mehta, A. S., Carrouee, S., Butters, T. D., Platt, F. M., McCauley, J., Blumberg, B. S., Dwek, R. A., Block, T. M. (1999). "Imino sugars inhibit the formation and secretion of bovine diarrhea virus, a pestivirus model of hepatitis C virus: implications for the development of broad spectrum anti-hepatitis virus agents". Proc. Natl. Acad. Sci. USA 96, 11878-11882.
- 40 33. Mehta, A., Zitzmann, N., Rudd, P. M., Block, T. M., Dwek, R. A. (1998). "Alpha-glucosidase inhibitors as potential broad-based anti-viral agents". FEBS Lett. 430, 17-22.
34. Rossen, J. W., de Beer, R., Godeke, G. J., Raamsman, M. J., Horzinek, M. C., Vennema, H. y Rottier, P. J. (1998) "The viral spike protein is not involved in the polarized sorting of coronaviruses in epithelial cells". J Virol. 72, 497-503.
- 45 35. Peiris, J. S. M., Lai, S. T., Poon, L. L. M., Guan, Y., Yam, L. Y. C., Lim, W., Nicholls, J., Yee, W. K. S., Yan, W. W., Cheung, M. T., Cheng, V. C. C., Chan, K. H., Tsang, D. N. C., Yung, R. W. H., Ng, T. N., Yuen, K. Y., "SRAG study group (2003). Coronavirus as a possible cause of severe acute respiratory syndrome". Lancet 361, 1319-1325.
- 50 36. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., Hilgenfeld, R. (2003). "Coronavirus Main Proteinase (3CL pro) Structure: Basis for Design of Anti-SRAG Drugs". Science, publicado en línea el 13 de mayo,

10.1126/science. 1085658.

Tabla 1. Masas calculada y medida para iones b e y a partir de medidas de EM/EM del fragmento triptico de 2297,092 Da

ión y	m/z enc.	m/z calc.	Δm (mDa)	Resto	Δm (mDa)	m/z enc.	m/z calc.	ión b
y ¹	-	147,113	-	K	2	2279,083	2279,081	b20
y ²	261,138	261,156	-18	N	-	-	2150,986	b19
y ³	332,193	332,193		A			2036,943	b18
y ⁴	389,197	389,215	-18	G	58		1965,906	b17
y ⁵	552,283	552,278	5	Y	0	1908,885	1908,885	b16
y ⁶	649,328	649,331	-3	P	-13	1745,809	1745,822	b15
y ⁷	762,413	762,415	-2	L	21	1648,790	1648,769	b14
y ⁸	849,438	849,447	-9	S	8	1535,695	1535,685	b13
y ⁹	920,478	920,484	-6	A	-15	1448,638	1448,653	b12
y ¹⁰	1049,522	1049,526	-4	E	19	1377,637	1377,616	b11
y ¹¹	1146,571	1146,579	-8	P	-	-	1248,573	b10
y ¹²	1203,592	1203,601	-9	G	-4	1151,516	1151,520	b9
y ¹³	1304,636	1304,648	-8	T	-5	1094,494	1094,499	b8
y ¹⁴	1361,654	1361,670	-16	G	1	993,452	993,451	b7
y ¹⁵	1474,739	1474,754	-15	L	-21	936,409	936,430	b6
y ¹⁶	1637,818	1637,817	1	Y	11	823,357	823,346	b5
y ¹⁷	1800,884	1800,881	3	Y	9	660,291	660,282	b4
y ¹⁸	1947,970	1947,949	21	F	1	497,220	497,219	b3
y ¹⁹	2111,022	2111,012	10	Y	-5	350,145	350,150	b2
y ²⁰	2297,086	2297,092	-6	W	-	-	187,087	b1

5 **Tabla 2.** Valores de m/z medidos y calculados de péptidos encontrados en la digestión triptica de la proteína de 46 kDa. M* representa restos de metionina oxidados, Ac-SDN es el extremo N acetilado de la proteína, Q** representa los restos Gln N-terminales convertidos en piro-Glu, N*** representa un resto Asn convertido en Asp debido a la desamidación

m/z medida	MH calculada (Da)	Δm (mDa)	Restos de inicio a fin	Secuencia peptídica
601,303	601,305	-3	204-209	GNSPAR
601,322	601,331	1	103-107	ELSPR
698,357	698,358	-1	144-149	DHIGTR
708,330	708,335	-5	96-102	GGDGKM*K
711,333	711,331	2	294-299	QGTDYK
746,387	746,383	4	356-361	HIDAYK
749,354	749,354	0	178-185	GGSQASSR
805,378	805,380	-2	196-203	NSTPGSSR

ES 2 364 719 T3

831,462	831,457	5	227-233	LNQLESK
876,452	876,461	-9	101-107	M*KELSPR
886,401	886,406	-5	170-177	GFYAEGSR
916,478	916,478	0	362-369	TFPPTPEPK
928,543	928,546	-3	348-355	DNVILLNK
946,513	946,511	2	62-68	EELRFPR
1105,548	1105,553	-5	339-347	LDDKDPQFK
1144,493	1144,499	-6	1-10	Ac-SDNGPQSNQR
1154,579	1154,580	-1	376-387	TDEAQPLPQR
1166,557	1166,559	-2	267-276	Q**YNVTQAGFR
1183,589	1183,586	3	267-276	QYNVTQAFGR
1202,610	1202,613	-3	238-248	GQQQQGQTVTK
1282,678	1282,675	3	375-385	KTDEAQPLPQR
1330,698	1330,708	-10	238-249	GQQQQGQTVTKK
1410,774	1410,771	3	376-387	KKTDEAQPLPQR
1611,698	1611,692	6	406-421	QLQNSM*SGASADSTQA
1684,895	1684,891	4	128-293	EGIVVWATEGALNTPK
1687,898	1687,905	-7	210-226	MASGGGETALALLLLDR
1703,897	1703,900	-3	210-226	M*ASGGGETALALLLLDR
1774,838	1774,836	2	278-293	GPEQTQGNFGDQDLIR
1850,833	1850,827	6	15-32	ITFGGPTDSTDNNQNGGR
1851,814	1851,811	3	15-32	ITFGGPTDSTDNNQN***GGR
1875,879	1875,879	0	389-405	Q**PTVLLPAADM*DDFSR
1930,944	1930,937	7	277-293	RGPEQTQGNFGDQDLIR
2005,008	2005,006	2	388-405	KQPTVLLPAADMDDFSR
2015,081	2015,081	0	339-355	LDDKDPQFKDNVILLNK
2021,005	2021,001	4	388-405	KQPTVLLPAADM*DDFSR
2077,048	2077,043	5	320-338	IGM*EVTPSGTWLTYHGAIK
2091,126	2091,120	-2	150-169	NPNNNAATVLQLPQGTTLPK
2151,002	2151,010	6	69-88	GQGVPIINTNSGPDDQIGYYR
2252,062	2252,071	-9	300-319	HWPQIAQFAPSASAFFGMSR
2297,078	2297,092	-14	108-127	WYFYLLGTGPEASLPYGANK
2307,128	2307,111	17	69-89	GQGVPIINTNSGPDDQIGYYRR
2324,187	2324,190	-3	41-61	RPQGLPNNTASWFTALTQH GK
2516,325	2516,339	-14	210-233	M*ASGGGETALALLLLDRLNQLESK

Tabla 3. Péptidos desglucosilados encontrados en la digestión trípica tratada con PNGasa F de la proteína de espícula de 139 kDa. Los símbolos N en negrita representaban sitios desglucosilados

Péptido	Secuencia*	[M+H] ⁺ sin glucosil.	calc.	[M+H] ⁺ después PNGasa	medido de	Dif. de m/z	Nº de sitios	EM/EM de péptido desgl.
T111-126	SQSVIIINNSTNVVIR	1756,992		1757,973		0,981	1	X
	SQSVIIINNSTNVVIR			1758,963		1,971	2	X
T222-232	LPLGINITNFR	1257,732		1257,715		0,983	1	X
T226-287	YDENGITIDAV	2453,114		2454,101		0,987	1	X
	DCSQNPLAELK							
T316-333	FPNITNLCPF	2069,017		2071,001		1,984	2	X
	GEVFNATK							
T778-796	YFGGFNFSQ	2169,138		2170,121		0,983	1	X
	ILPDPLKPTK							
T1056-1068	NFITAPAICHEGK	1445,685		1446,672		0,987	1	X
T1074-1089	EGVVFVNG	1887,939		1888,931		0,922	1	X
	TSWFITQR							
T1140-1163	NHTSPDVDLGGI	2493,259		2495,23		1,971	2	X
	SGINASVVINQK							
T1174-1187	NLNESLIDLQELGK	1585,844		1586,823		0,979	1	X

Tabla 4. Péptidos glucosilados encontrados en la digestión trípica de la proteína de espícula de 139 kDa sin tratamiento con PNGasa F

Péptido	Secuencia	[M+H] ⁺ calc. m/z	[M+H] ⁺ glucop. medido	Masa carb. resid.	Compos. carboh.	Masa carboh. calc.	Masa (med- calc.)	EM/EM glucop.
T111-126	SQSVIIINNSTNVVIR	1756,992	3201,514	1444,522	M3N4F	1444,534	-0,012	
			3404,572	1647,580	M3N5F	1647,613	-0,033	X
			3525,669	1768,677	M3N4G2F	1768,640	0,037	
			3566,670	1809,678	M3N5GF	1809,666	0,012	
			3607,682	1850,690	M3N6F	1850,693	-0,003	X
			3728,687	1971,695	M3N5G2F	1971,719	-0,024	
			3769,678	2012,686	M3N6GF	2012,745	-0,059	
			3890,761	2133,769	M3N5G3F	2133,772	-0,003	
			3931,793	2174,801	M3N6G2F	2174,798	0,030	
			4093,818	2336,826	M3N6G3F	2336,859	-0,025	X
			4255,844	2498,852	M3N6G4F	2498,904	0,052	
			4296,905	2539,913	M3N7G3F	2539,930	-0,017	

T222-232	LPLGINIT	1257,732	2636,215	1378,483	M6N2	1378,746	0,007	
	NFR		2798,26	1540,528	M7N2	1540,528	0,000	
			2960,306	1702,574	M8N2	1702,574	-0,007	
			3122,373	1864,641	M9N2	1864,641	0,007	X
T226-287	YDENGIT	2453,114	4100,666	1647,552	M3N5F	1647,613	-0,061	
	DAVDCSQ		4303,735	1850,621	M3N6F	1850,693	0,072	X
	NPLAELK		4427,1*	1974,0	M3N5FG2	1971,719		
			4468,9*	2015,8	M3N6FG	2012,745		
			4589,1*	2136,0	M3N5G3F	2133,772		
			4792,2*	2399,1	M3N6G3F	2336,859		
			4953,9*	2500,8	M3N6G4F	2498,904		
T778-796	YFGGFN	2169,138	3385,59	1216,452	M5N2	1216,423	-0,029	
	FSQILPD		3547,644	1378,506	M6N2	1378,476	-0,030	
	PLKPTK		3709,738	1540,600	M7N2	1540,529	-0,071	X
			3871,702	1702,564	M8N2	1702,581	0,017	

Símbolos peptídicos:

N: primer sitio de glucosilación

N: segundo sitio de glucosilación

Símbolos de composición de carbohidrato:

M= manosa

F= fucosa

G= galactosa

N= N-acetilglucosamina

*los valores de m/z no pudieron determinarse exactamente (véase el texto)

Tabla 5. Valores de m/z medidos y calculados de péptidos encontrados en digestiones tripticas de la proteína de 139 kDa. Los valores de masa se toman directamente del programa de hallazgo de picos automático, y son por lo tanto algo menos fiables que cuando se comprueban manualmente. A) Sin desglucosilación de los 39 péptidos observados (34% de cobertura). Se confirmaron 36 por EM/EM (30% de cobertura). B) Después de desglucosilación con PNGasa F (véase el texto), se encontraron 9 péptidos desglucosilados y se confirmaron mediante EM/EM, dando una cobertura adicional de un 12%. M*-restos de Met oxidados; **-identificaciones peptídicas no confirmadas por EM/EM; D: restos de Asn desglucosilados.

m/z medido	MH (Da)	calculado	Δm (mDa)	Restos de inicio a fin	Secuencia peptídica
A					
653,344	653,341		3	1069-1073	AYFPR
669,361	669,361		0	184-188	EFVFK
673,389	673,388		1	1168-1173	LNEVAK**
725,332	725,329		3	1022-1027	VDFCGK
738,370	738,367		3	292-297	SFEIDK

ES 2 364 719 T3

817,478	817,468	10	448-453	LRPFER
820,470	820,457	13	762-768	EVFAQVK
846,476	846,468	8	1002-1010	ASANLAATK
847,453	847,434	19	830-836	DLICAQK
881,446	881,430	16	1021-1027	RVDFCGK
885,420	885,414	6	366-373	CYGVSATK
927,543	927,526	17	307-315	VVPSGDVVR
962,523	962,506	17	904-911	QIANQFNK
1046,556	1046,563	-7	199-207	GYQPIDVVR
1068,533	1068,540	-7	808-817	VTLADAGFM*K**
1085,553	1085,538	15	298-306	GIYQTSNFR
1114,534	1114,542	-8	85-94	DGIYFAATEK
1130,592	1130,581	11	748-758	ALSQIAAEQDR
1154,555	1154,563	-8	334-342	FPSVYAWER
1154,564	1154,574	-10	545-553	FOFPQQFGR
1225,641	1125,647	-6	798-807	SFIEDLLFNK
1246,646	1246,647	-1	189-198	NKDGFLYVYK
1258,602	1258,610	-8	39-48	GVYYFDEIFR
1310,663	1310,675	-12	544-553	RFOPFQQFQR
1395,622	1395,632	-10	818-829	QYGECLGDINAR
1414,754	1414,754	0	966-977	LDKVEAEVQIDR
1460,690	1460,666	24	427-439	NIDATSTGNVNYK
1480,687	1480,692	-5	554-566	DVSDFTDSVRDPK
1575,866	1576,874	-8	208-221	DLPSQFNTLKIPIK
1690,951	1690,949	2	983-996	LQSLQTYVTQQLIR
1737,879	1737,881	-2	396-411	QIAPGQTGVIADYNYK
1804,885	1804,887	-2	292-306	SFEIDKGIYQTSNFR
1810,834	1810,825	-1	412-436	LPDDFM*QCVLAWNTR
1848,979	1848,992	-13	912-929	AISQIQESLTTTSTALGK
1868,012	1868,024	-12	930-946	LQDVVNQNAQALNLTLVK
1990,954	1990,958	-4	374-390	LNDLCFSNVYADSFVVK
2014,091	2014,104	-13	496-514	VVLSFELLNAPATVCGPK
2355,139	2355,140	-1	522-543	NQCVNFNFNGLTGTGVLTPSSK
3183,578	3183,584	-6	1028-1055	GYHLM*SFPQAAPHGVVFLHVTVVPSOER**

B. Péptidos desglucosilados encontrados en la digestión triptica tratada con PNGasa

1258,715	1258,716	-1	222-232	LPLGIDITNFR
----------	----------	----	---------	-------------

1446,672	1446,669	3	1056-1068	DFTTAPAICHEGK
1586,803	1586,808	-5	1174-1187	NLDESLIDLQELGK
1757,973	1757,976	-3	111-126	SQSVIIINDSTNVVIR
1758,963	1758,960	3	111-126	SQSVIIIDDSTNVVIR
1888,931	1888,923	8	1074-1089	EGVFVFDGTSWFITQR
2071,001	2070,985	16	316-333	FPDITNLCPFGEVFDATK
2170,121	2170122	-1	778-796	YEGGFDFSILPDPLKPTK
2454,101	2454,098	3	226-287	YDEDGTITDAVDCSONFLAELK
2495,230	2495,227	3	1140-1163	DHTSPDVLGDIGIDASVVNIQK

LISTADO DE SECUENCIAS

<110> Universidad de Manitoba

Su Majestad la Reina de derecho de Canadá representada por el Ministro de Sanidad

<120> PROTEÍNAS RELACIONADAS CON EL SRAG

5 <130> 85084-1003

<150> US 60/462688

<151> 15-04-2003

<150> US 60/472416

<151> 22-05-2003

10 <160> 2

<170> PatentIn versión 3.2

<210> 1

<211> 411

<212> PRT

15 <213> Coronavirus

<400> 1

ES 2 364 719 T3

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

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

<210> 2
 <211> 1254
 <212> PRT
 <213> Coronavirus

5 <300>
 <301> Krokhin *et al.*
 <302> Mass spectrometric characterization of proteins from the SARS virus: a preliminary report
 <303> Mol. Cell. Proteomics
 <304> 2
 10 <305> 5
 <306> 346-356
 <307> 29-05-2003
 <400> 2

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

Thr Gly Val Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe Met
 405 410 415
 Gly Cys Val Leu Ala Trp Asn Thr Arg Asn Ile Asp Ala Thr Ser Thr
 420 425 430
 Gly Asn Tyr Asn Tyr Lys Tyr Arg Tyr Leu Arg His Gly Lys Leu Arg
 435 440 445
 Pro Phe Glu Arg Asp Ile Ser Asn Val Pro Phe Ser Pro Asp Gly Lys
 450 455 460
 Pro Cys Thr Pro Pro Ala Leu Asn Cys Tyr Trp Pro Leu Asn Asp Tyr
 465 470 475 480
 Gly Phe Tyr Thr Thr Thr Gly Ile Gly Tyr Gln Pro Tyr Arg Val Val
 485 490 495
 Val Leu Ser Phe Glu Leu Leu Asn Ala Pro Ala Thr Val Cys Gly Pro
 500 505 510
 Lys Leu Ser Thr Asp Leu Ile Lys Asn Gln Cys Val Asn Phe Asn Phe
 515 520 525
 Asn Gly Leu Thr Gly Thr Gly Val Leu Thr Pro Ser Ser Lys Arg Phe
 530 535
 Gln Pro Phe Gln Gln Phe Gly Arg Asp Val Ser Asp Phe Thr Asp Ser
 545 550 555 560
 Val Arg Asp Pro Lys Thr Ser Glu Ile Leu Asp Ile Ser Pro Cys Ala
 565 570 575
 Phe Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Ala Ser Ser Glu
 580 585 590
 Val Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Asp Val Ser Thr Ala
 595 600 605
 Ile His Ala Asp Gln Leu Thr Pro Ala Trp Arg Ile Tyr Ser Thr Gly
 610 615 620
 Asn Asn Val Phe Gln Thr Gln Ala Gly Cys Leu Ile Gly Ala Glu His
 625 630 635 640
 Val Asp Thr Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile Cys
 645 650 655

Ala Ser Tyr His Thr Val Ser Leu Leu Arg Ser Thr Ser Gln Lys Ser
660 665 670

Ile Val Ala Tyr Thr Met Ser Leu Gly Ala Asp Ser Ser Ile Ala Tyr
675 680 685

Ser Asn Asn Thr Ile Ala Ile Pro Thr Asn Phe Ser Ile Ser Ile Thr
690 695 700

Thr Glu Val Met Pro Val Ser Met Ala Lys Thr Ser Val Asp Cys Asn
705 710 715 720

Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ala Asn Leu Leu Leu Gln
725 730 735

Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Ser Gly Ile Ala
740 745 750

Ala Glu Gln Asp Arg Asn Thr Arg Glu Val Phe Ala Gln Val Lys Gln
755 760 765

Met Tyr Lys Thr Pro Thr Leu Lys Tyr Phe Gly Gly Phe Asn Phe Ser
770 775 780

Gln Ile Leu Pro Asp Pro Leu Lys Pro Thr Lys Arg Ser Phe Ile Glu
785 790 800

Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Met Lys
805 810 815

Gln Tyr Gly Glu Cys Leu Gly Asp Ile Asn Ala Arg Asp Leu Ile Cys
820 825 830

Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr Asp
835 840 845

Asp Met Ile Ala Ala Tyr Thr Ala Ala Leu Val Ser Gly Thr Ala Thr
850 855 860

Ala Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe Ala
865 870 875 880

Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn Val
885 890 895

Leu Tyr Glu Asn Gln Lys Gln Ile Ala Asn Gln Phe Asn Lys Ala Ile
900 905 910

Ser Gln Ile Gln Glu Ser Leu Thr Thr Thr Ser Thr Ala Leu Gly Lys
 915 920 925

Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu Val
 930 935 940

Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn Asp
 945 950 955 960

Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile Asp Arg
 965 970 975

Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln Glr
 980 985 990

Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala T
 995 1000 1005

Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp Phe
 1010 1015 1020

Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ala Ala Pro
 1025 1030 1035

His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ser Gln Glu
 1040 1045 1050

Arg Asn Phe Thr Thr Ala Pro Ala Ile Cys His Glu Gly Lys Ala
 1055 1060 1065

Tyr Phe Pro Arg Glu Gly Val Phe Val Phe Asn Gly Thr Ser Trp
 1070 1075 1080

Phe Ile Thr Gln Arg Asn Phe Phe Ser Pro Gln Ile Ile Thr Thr
 1085 1090 1095

Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly Ile
 1100 1105 1110

Ile Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp Ser
 1115 1120 1125

Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser Pro
 1130 1135 1140

Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val Val

1145						1150						1155			
Asn	Ile	Gln	Lys	Glu	Ile	Asp	Arg	Leu	Asn	Glu	Val	Ala	Lys	Asn	
1160						1165					1170				
Leu	Asn	Glu	Ser	Leu	Ile	Asp	Leu	Gln	Glu	Leu	Gly	Lys	Tyr	Glu	
1175						1180					1185				
Gln	Tyr	Ile	Lys	Trp	Pro	Trp	Tyr	Val	Trp	Leu	Gly	Phe	Ile	Ala	
1190						1195					1200				
Gly	Leu	Ile	Ala	Ile	Val	Met	Val	Thr	Ile	Leu	Leu	Cys	Cys	Met	
1205						1210					1215				
Thr	Ser	Cys	Cys	Ser	Cys	Leu	Lys	Gly	Ala	Cys	Ser	Cys	Gly	Ser	
1220						1225					1230				
Cys	Cys	Lys	Phe	Asp	Glu	Asp	Asp	Ser	Glu	Pro	Val	Leu	Lys	Gly	
1235						1240					1245				
Val	Lys	Leu	His	Tyr	Thr										
1250															

REIVINDICACIONES

1. Una proteína relacionada con el SRAG aislada que tiene una secuencia como se expone en la SEC ID N° 1.
2. Un anticuerpo reactivo contra una proteína según la reivindicación 1.
- 5 3. Un método de detección de un coronavirus o de una afección asociada a un coronavirus que comprende ensayar en una muestra (a) la proteína relacionada con el SRAG según la reivindicación 1, o un fragmento inmunogénico de la misma, o (b) un anticuerpo que se une a la proteína relacionada con el SRAG según la reivindicación 1.
- 10 4. Un método según la reivindicación 3, que comprende poner en contacto una proteína relacionada con el SRAG de la reivindicación 1, o un fragmento inmunogénico de la misma, con la muestra y determinar la presencia de anticuerpos de la proteína relacionada con el SRAG en la muestra.
5. Un método según la reivindicación 3, que comprende poner en contacto el anticuerpo de la reivindicación 2 con la muestra y determinar la presencia de proteína relacionada con el SRAG en la muestra.
6. Un método según una cualquiera de las reivindicaciones 3 a 5, en el que la afección es SRAG.
- 15 7. Un reactivo que comprende una cantidad de (a) una proteína relacionada con el SRAG según la reivindicación 1, o un fragmento inmunogénico de la misma, o (b) un anticuerpo que se une a una proteína relacionada con el SRAG según la reivindicación 1 mezclado con un diluyente o portador adecuado.
- 20 8. Un método de cribado de un compuesto capaz de unirse a una proteína según la reivindicación 1, o un fragmento inmunogénico de la misma, que comprende poner en contacto la proteína, o fragmento de la misma, con un compuesto de ensayo y determinar la capacidad de dicho compuesto de ensayo de unirse a dicha proteína o fragmento de la misma.

FIGURA 1

Secuencia de proteína de 47 kDa, mediodía del 15 de abril de 2003

PM 45936,94 de media, 421 aa, 93% de cobertura de secuencia (restos confirmados indicados en rojo). @S indica acetilación del resto de serina N-terminal (proporcionando un cambio de masa de +42,011 Da).

```

1 @SDNGPQSNQR SAPRITFGGP TDSTDNNQNG GRNGARPKQR RPQGLPNNNTA SWFTALTQHG
61 KEELRFPRGQ GVPINTNSGP DDQIGYYRRA TRRVGGDGK MKELSPRWYF YYLGTGPEAS
121 LPYGANKEGI VVATEGALN TPKDHIGTRN PNNNAATVLQ LPQGTLLPKG FYAEGSRGGS
181 QASSRSSRS RGNSRNSTPG SSRGNSPARM ASGGGETALA LLLDRLNQL ESKVSGKGQQ
241 QQGQTVTKKS AAEASKKPRQ KRTATKQYNV TQAFRRGPE QTOGNFGDQD LIRQGTDYKH
301 WPQIAQFAPS ASAFFGMSRI GMEVTPSGTW LTYHGAIKLD DKDPQFKDNV ILLNKHIDAY
361 KTFPPTEPKK DKKKKTDEAQ PLPQRQKKQP TVTLLPAADM DDFSRQLQNS MSGASADSTQ
421 A

```

FIGURA 2

Proteína de SRAG de 139 kDa

```

FIFLLFL TLTSGSDLDR CTFDDVQAP NYTQHTSSMR GVYYPDEIFR SDTLYLTQDL
FLPFYSNVTG FHTINHTEFGN PVIPFKDGIY FAATEKSNV V RGWVFGSTMN NKSQSVIIIN
NSTNVVIRAC NFEICDNPFF AVSKPMGTQT HTMIFDNAFN CTFEYISDAF SLDVSEKSGN
FKHLREFVFK NKDGFVLYYK GYQPIDVVRD LPSGFNTLKP IFKLPLGINI TNFRAILTAF
SPAQDIWGTS AAAYFVGYLK PTFMFLKYDE NGTITDAVDC SQNPPLAELKC SVKSFEIDKG
IYQTSNFRV V PSGDVVRFPN ITNLCPFGEV FNATKFP SVY AWERKKISNC VADYSVLYNS
TFFSTFKCYG VSATKLN D LC FSNVYADSFV VKGDDVRQIA PGQTGVIADY NYKLPDDFMG
CVLAWNTRNI DATSTGN YNY KYRYLRHGKL RPFERDISNV PFSPDGK PCT PPALNCYWPL
NDYGFYTTTG IGYQPYRVV V LSFELNAPA TVCGPKLSTD LIKNQCVNFN ENGLTGTGVL
TPSSKRFQPF QQFGRDV SDF TDSVRDPKTS EILDISPCAF GGVSVITPGT NASSEVAVLY
QDVNCTDVST AIHADQLTPA WRIYSTGNNV FQTQAGCLIG AEHVDTSYEC DIPIGAGICA
SYHTVSL LRS TSQKSIVAYT MSLGADSSIA YSNNTIAIPT NFSISITTEV MPVSMAKTSV
DCNMYICGDS TECANLL LQY GSFCTQLNRA LSGIAAEQDR NTREVFAQVK QMYKTPTLKY
FGGFNFSQIL PDPLKPTKRS FIEDLLFNKV TLADAGFMKQ YGECLGDINA RDLICAQKFN
GLTVLPP LLLT DDMIAAYTAA LVSGTATAGW TFGAGAALQI PFAMQMAYRF NGIGVTQNVL
YENQKQIANQ FNKAISQIQE SLTTTSTALG KLQDVVNQNA QALNTLVKQL SSNFGAISSV
LNDILSR LDK VEAQVQIDRL ITGRLQSLQT YVTQQLIRAA EIRASANLAA TKMSECVLGQ
SKRVDFCGKG YHLMSFPQAA PHGVVFLHVT YVPSQERNFT TAPAICHEGK AYFPREGV FV
FNGTSWFITQ RNFFSPQ IIT TDNTFVSGNC DVVIGIINNT VYDPLQPELD SFKEELDKYF
KNHTSPD VDL GDISGINASV VNIQKEIDRL NEVAKNLNES LIDLQELGKY EQYIKWPWYV
WLGFIAGLIA IVMVTIL LCC MTSCCSCLKG ACSCGSCCKF DEDDSEPVLK GVKLHYT

```

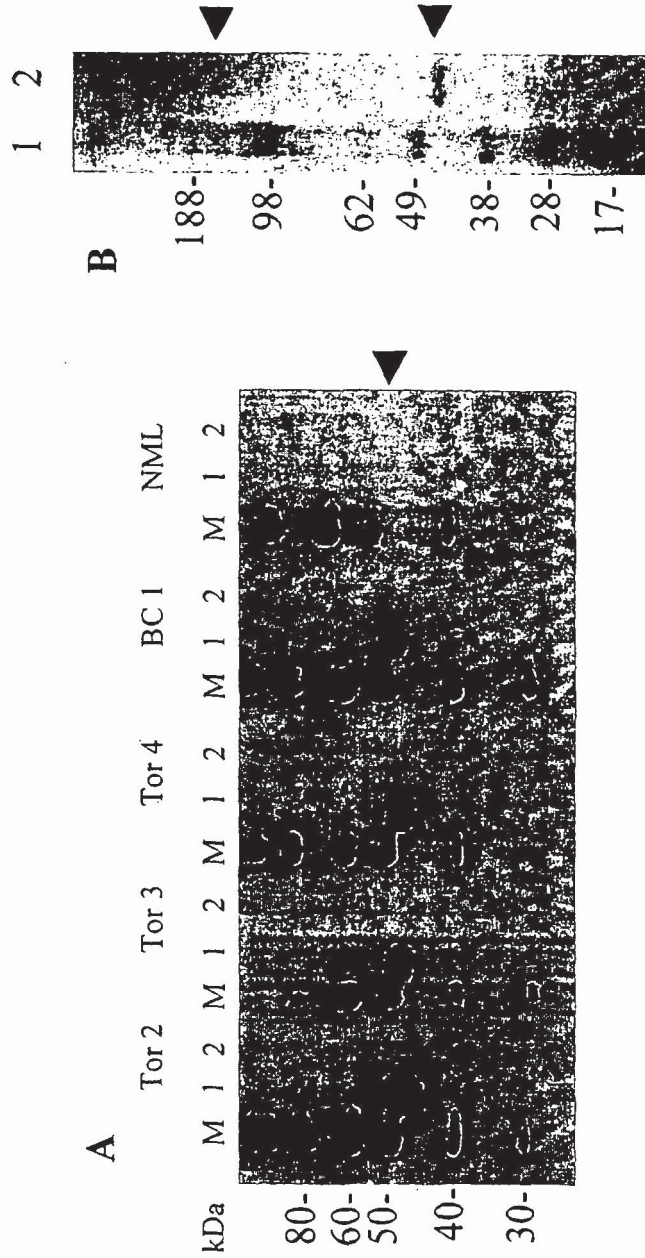



FIGURA 3

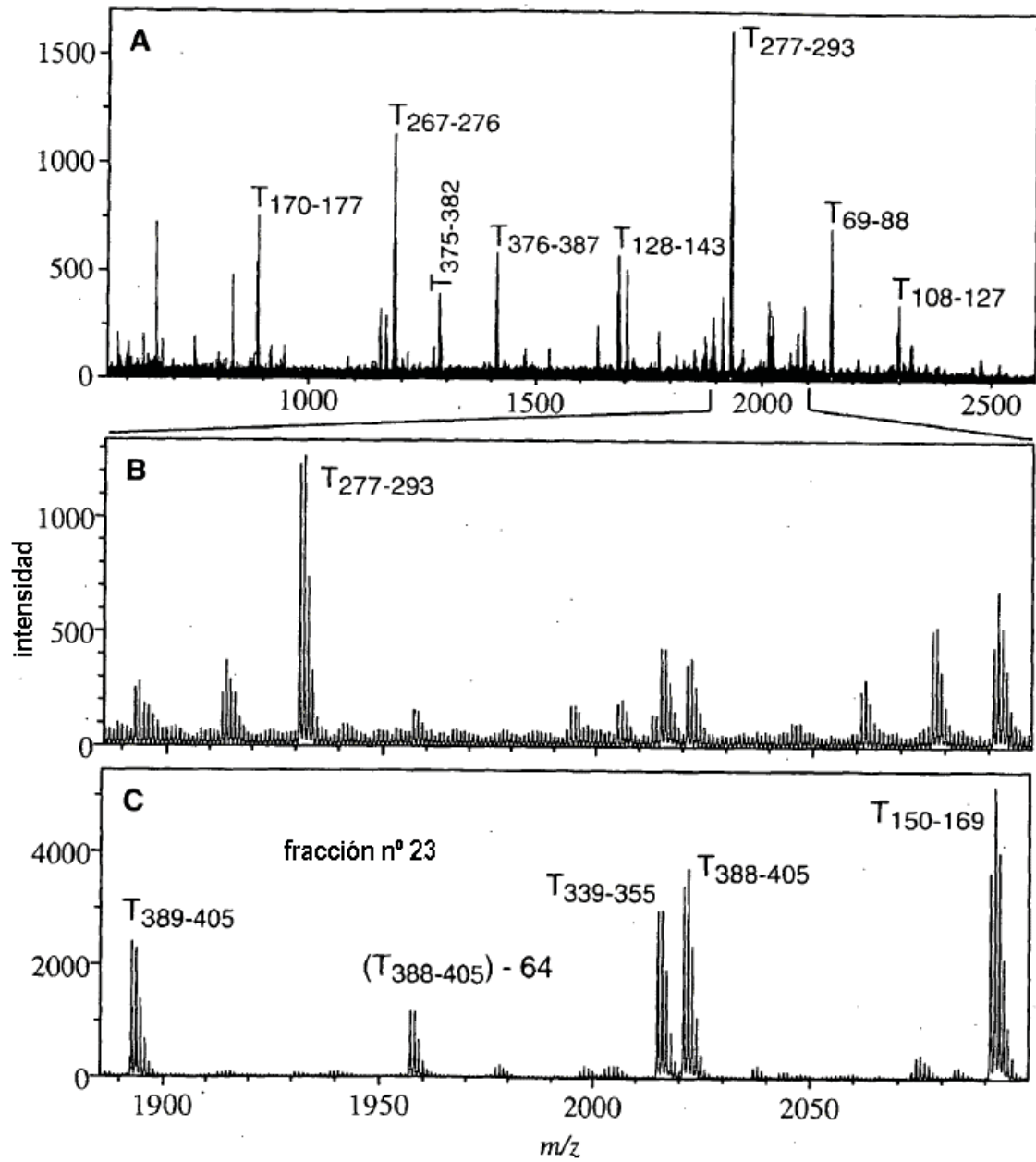


FIGURA 4

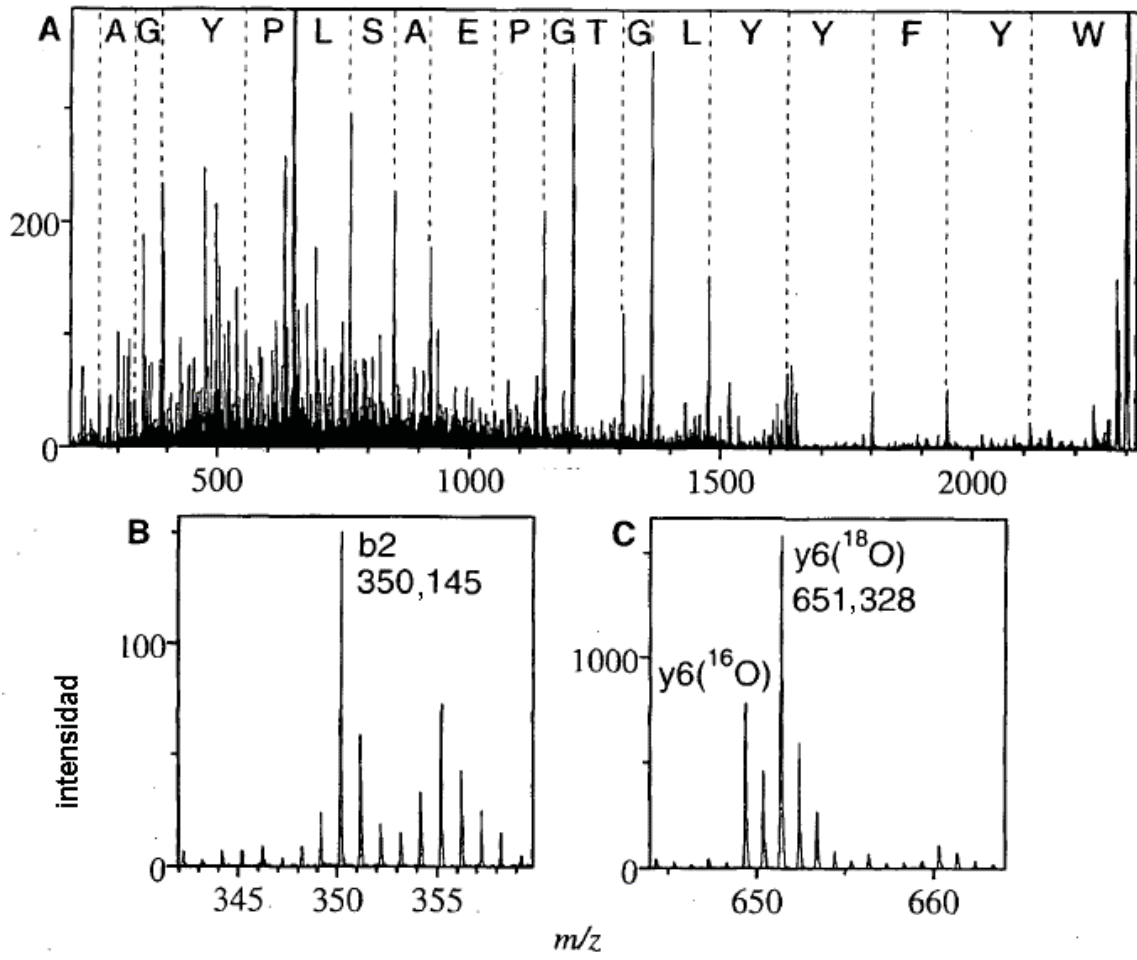


FIGURA 5

Homología de proteína de nucleocápsida de coronavirus

...	WYFYLLGTGPEASLPYGAN	...	SRAG
...	WYFYLLGTGPHAKDQYGTD	...	Humano
...	WYFYLLGTGPHAKDQYGTD	...	Bovino
...	WYFYLLGTGPHAKDQYGTD	...	de pavo
...	WYFYLLGTGPHAKHQYGTD	...	Porcino
...	WYFYLLGTGPHAKAQYGTN	...	Equino
...	WYFYLLGTGPHAGASYGDD	...	Múrido
...	WYFYLLGTGPHAGASFGDS	...	de rata

FIGURA 6

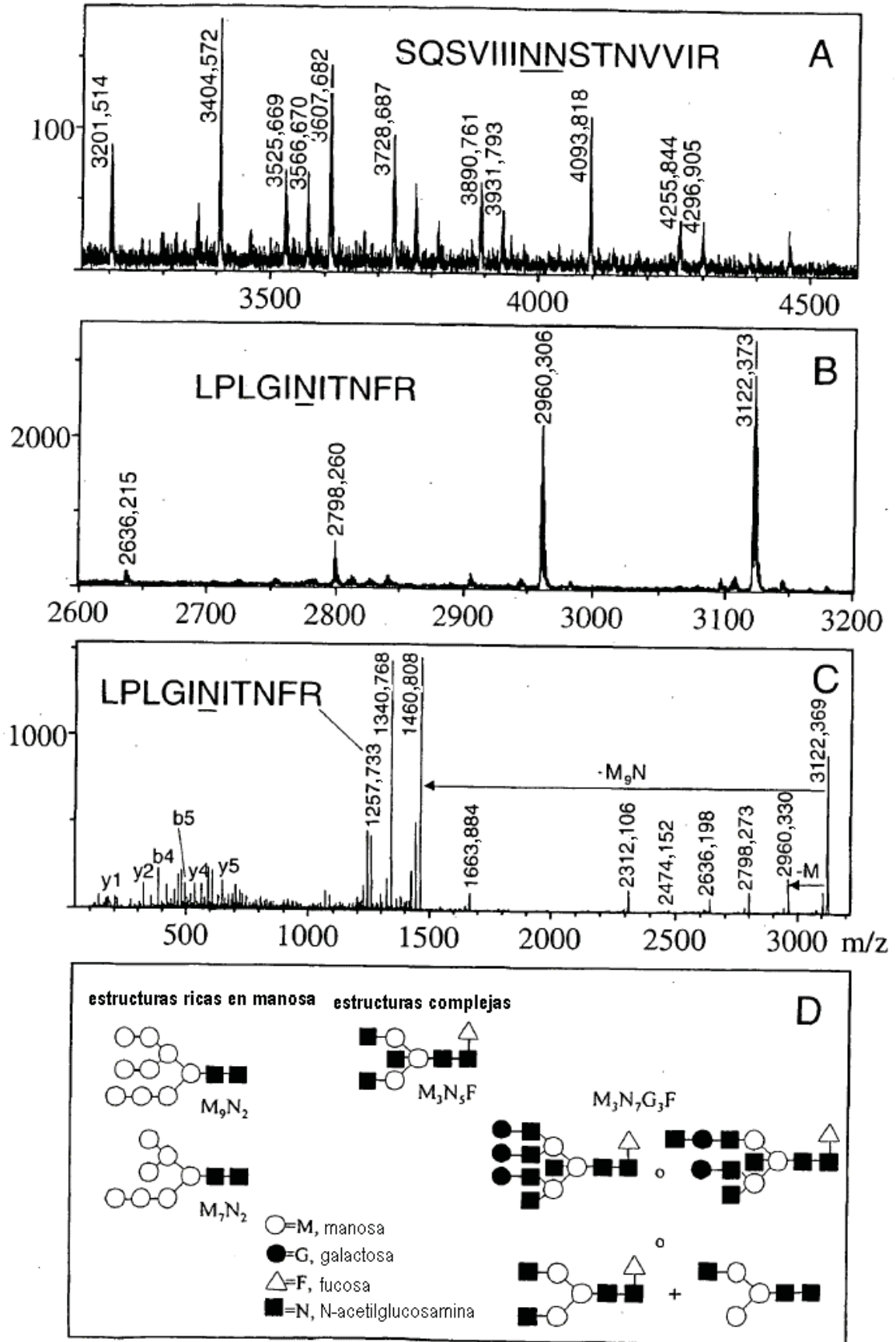


FIGURA 7

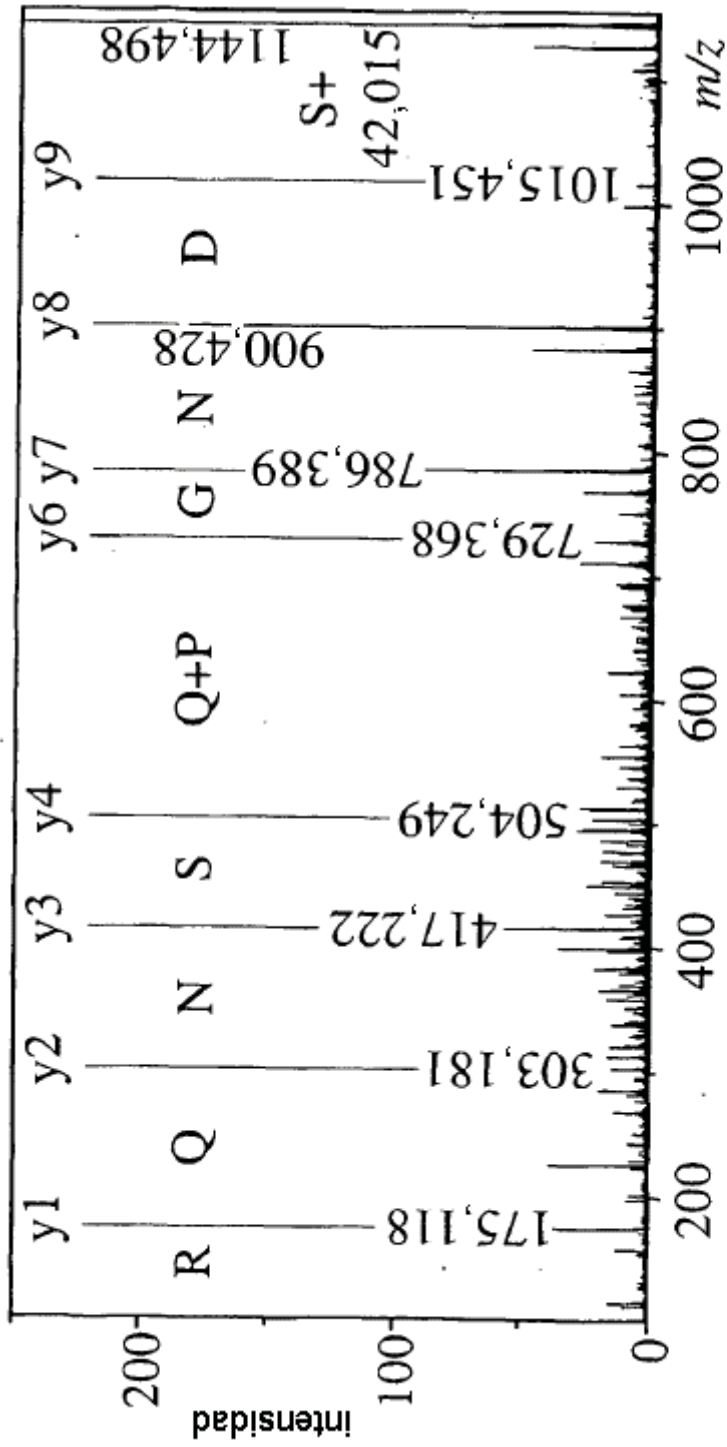


FIGURA 8

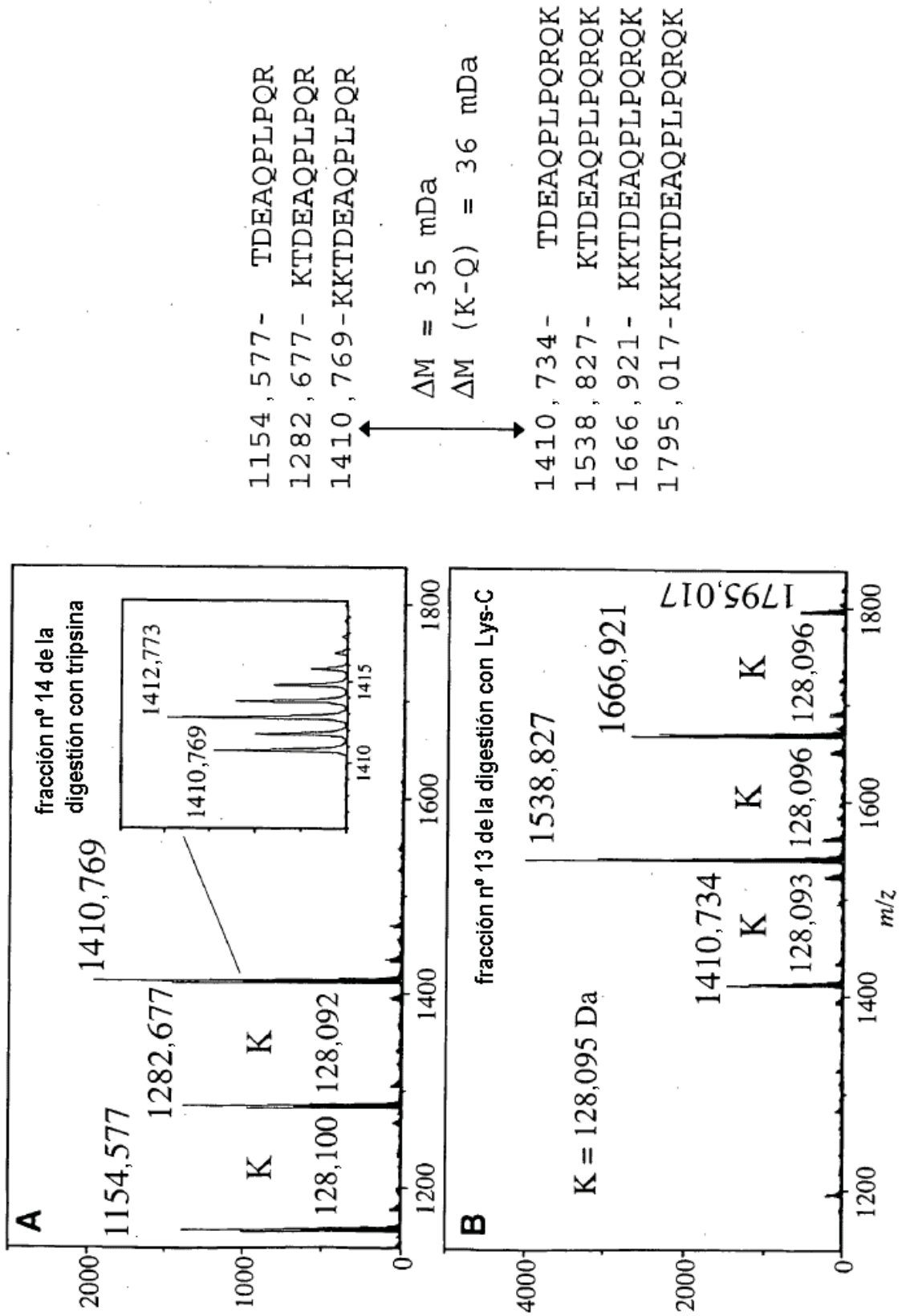


FIGURA 9

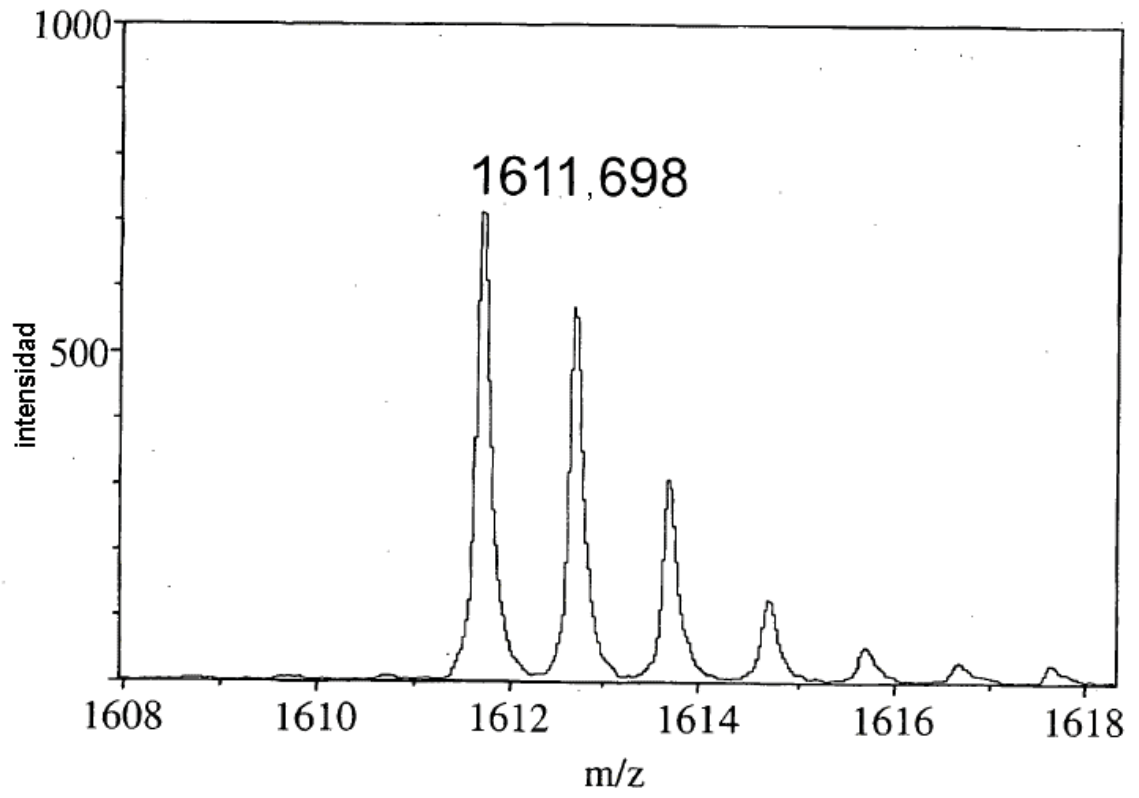


FIGURA 10

A

1 SDNGPQSNQR SAPRITFGGP TDSTDNNQNG GRNGARPKQR RPQGLPNNTA SWFTALTQHG
 61 KEELRFPRGQ GVPINTNSGP DDQIGYYRRA TRRVGGDGK MKELSPRWYF YYLGTGPEAS
 121 LPYGANKEGI VVWATEGALN TPKDHIGTRN PNNNAATVLQ LPQGTTLPKG FYAEGSRGGS
 181 QASRRSSRS RGNSRNSTPG SRRGNSPARM ASGGGETALA LLLDLRLNQL ESKVSGKGQQ
 241 QQGQTVTKKS AAEASKKPRQ KRTATKQYNV TQAFGRRGPE QTQGNFGDQD LIRQGTDYKH
 301 WPQIAQFAPS ASAFFGMSRI GMEVTPSGTW LTYHGAIKLD DKDPQFKDNV ILLNKHIDAY
 361 KTFPPTEPKK DKKKKTDEAQ PLPQRQKKQP TVTLLPAADM DDFSRLQNS MSGASADSTQ
 421 A

B

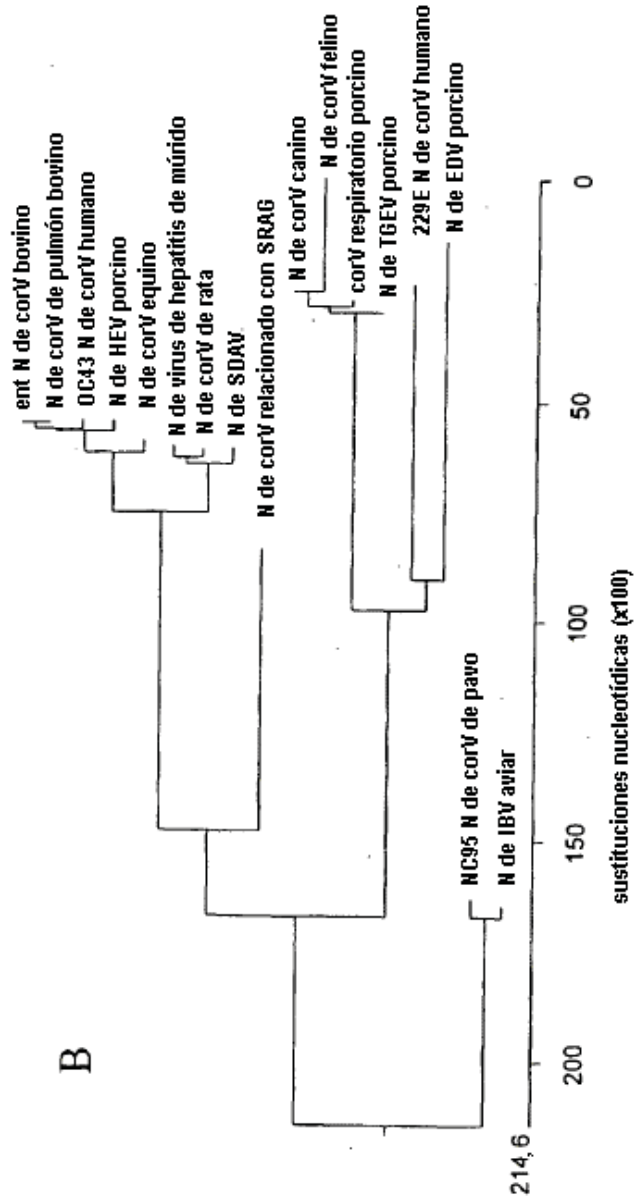


FIGURA 11

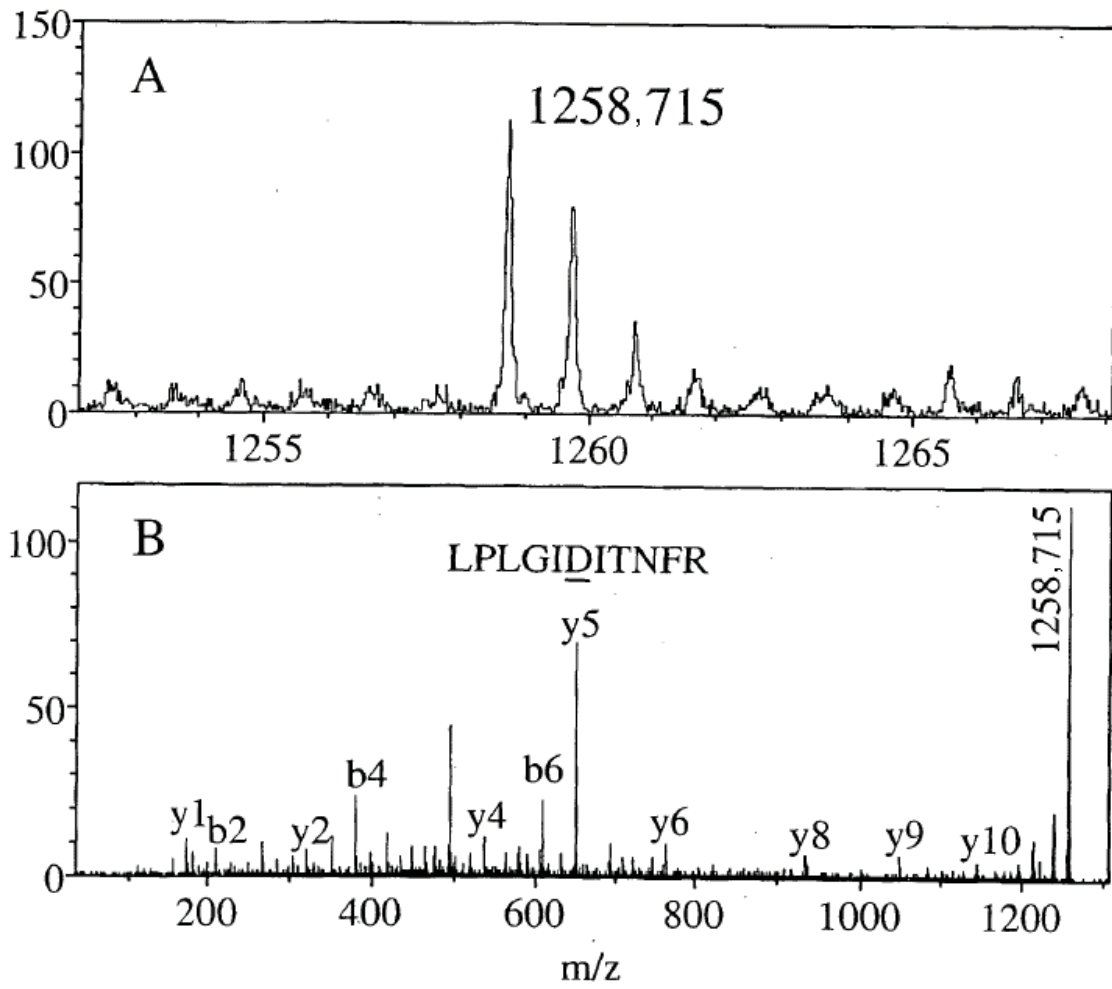


FIGURA 12