

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 638 311**

51 Int. Cl.:

<b>A61K 38/16</b>	(2006.01)
<b>A61K 38/17</b>	(2006.01)
<b>A61K 39/395</b>	(2006.01)
<b>A61P 35/00</b>	(2006.01)
<b>A61P 29/00</b>	(2006.01)
<b>A61P 3/00</b>	(2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **27.04.2011 PCT/US2011/034207**
- 87 Fecha y número de publicación internacional: **10.11.2011 WO11139801**
- 96 Fecha de presentación y número de la solicitud europea: **27.04.2011 E 11777984 (3)**
- 97 Fecha y número de publicación de la concesión europea: **07.06.2017 EP 2563382**

54 Título: **Descubrimiento innovador de composiciones terapéuticas, diagnósticas y de anticuerpos relacionadas con fragmentos de proteína de treonil ARNt sintetasas**

30 Prioridad:

**27.04.2010 US 328373 P**  
**27.04.2010 US 328371 P**  
**27.04.2010 US 328369 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**19.10.2017**

73 Titular/es:

**ATYR PHARMA, INC. (50.0%)**  
**3545 John Hopkins Court, Suite 250**  
**San Diego, CA 92121, US y**  
**PANGU BIOPHARMA LIMITED (50.0%)**

72 Inventor/es:

**GREENE, LESLIE ANN;**  
**CHIANG, KYLE, P.;**  
**HONG, FEI;**  
**VASSEROT, ALAIN, P.;**  
**LO, WING-SZE;**  
**WATKINS, JEFFRY, D.;**  
**MENDLEIN, JOHN, D. y**  
**QUINN, CHERYL, L.**

74 Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

ES 2 638 311 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

Descubrimiento innovador de composiciones terapéuticas, diagnósticas y de anticuerpos relacionadas con fragmentos de proteína de treonil ARNt sintetasas

5

**Campo técnico**

La presente divulgación se refiere en general a composiciones que comprenden fragmentos de proteína recientemente identificados de aminoacil-ARNt sintetasas y otras proteínas, a polinucleótidos que los codifican y a complementos de los mismos, a agentes relacionados y a métodos de uso de los mismos en aplicaciones diagnósticas, de descubrimiento de fármacos, de investigación y terapéuticas.

10

**Antecedentes**

Durante más de cuatro décadas, se consideró que las aminoacil-ARNt sintetasas (AARS) eran proteínas constitutivas esenciales que catalizan la aminoacilación de moléculas de ARNt como parte de la decodificación de la información genética durante el proceso de traducción de proteínas. Las AARS se han estudiado exhaustivamente a este respecto y muchas de sus secuencias de longitud completa se clonaron para analizar la secuencia y para proporcionar una fuente rica de experimentación bioquímica. Sin embargo, algunos fragmentos de AARS y otras proteínas poseen actividades inesperadas no asociadas con la aminoacilación, incluyendo actividades de señalización extracelular que modulan vías aparte de la traducción de proteínas. En general, estas actividades inesperadas no se observan en el contexto de las secuencias de proteína de longitud completa o parentales; en su lugar, se observan después de la eliminación o corte de fragmentos de proteína de AARS de sus secuencias parentales o mediante la expresión y purificación suficiente de fragmentos de secuencias de AARS y después practicando pruebas para nuevas actividades no relacionadas con sintetasas.

15

20

25

Aunque se conocen desde hace algún tiempo secuencias de AARS de longitud completa, no se ha llevado a cabo un análisis experimental sistemático para dilucidar dichos fragmentos de proteína AARS o fragmentos de proteína de proteínas relacionadas o asociadas o para evaluar el papel potencial de las proteínas AARS de longitud completa respecto de nuevas actividades biológicas más allá del contexto de la síntesis de aminoácidos. En partes de la presente memoria descriptiva, dichos fragmentos de proteína AARS, dominios de AAR o variantes de corte y empalme alternativo de AARS se citan en el presente documento como "resectinas". En su contexto más amplio, el término "resectina" se refiere a una porción de una proteína que se ha cortado o restringido (ya sea mediante proteólisis, corte y empalme alternativo, mutagénesis o ingeniería genética recombinante) del contexto de su secuencia de proteína nativa de longitud completa o parental, que de lo contrario normalmente enmascara sus nuevas actividades biológicas. Del mismo modo, no se ha llevado a cabo un análisis experimental sistemático para explorar el uso de dichas resectinas como agentes bioterapéuticos, agentes diagnósticos o dianas farmacológicas en el tratamiento de diversas afecciones médicas o su asociación potencial con enfermedades humanas. Como genes constitutivos esenciales de función conocida en mamíferos que es crucial para la vida, las AARS ni se tomaron en consideración como dianas farmacológicas en mamíferos, ni se analizaron mediante secuenciación genómica, bioinformática o esfuerzos similares convencionales para identificar resectinas que tengan actividades no sintetasa. De manera similar, los esfuerzos de investigación bioquímica convencionales se han apartado de la caracterización de las propiedades biológicas de las resectinas de AARS y su potencial relevancia terapéutica y diagnóstica, debido principalmente al papel previamente asumido de sus AARS parentales de longitud completa correspondientes.

30

35

40

45

La técnica anterior incluye la BASE DE DATOS UniProt [en línea], a 1 de agosto de 1992 (01-08-1992), "RecName: Completo=Treonina-ARNt ligasa, citoplasmática; EC=6.1.1.3; AltName: Completo=Treonil-ARNt sintetasa; Abreviado=ThrRS; ", n.º de referencia EBI UNIPROT:P26639; n.º de referencia de la base de datos P26639, que divulga la treonil-ARNt sintetasa humana.

50

**Breve descripción de los dibujos**

La **figura 1** muestra la estructura de dominios de la treonil aminoacil ARNt sintetasa dispuesta con las posiciones y tamaños relativos de los polipéptidos de AARS N-terminales identificados mostrados de manera esquemática. La figura 1A muestra fragmentos identificados mediante análisis de espectrometría de masas y la figura 1B muestra fragmentos identificados mediante secuenciación profunda de los transcriptomas.

55

La **figura 2** muestra la estructura de dominios de la treonil aminoacil ARNt sintetasa dispuesta con las posiciones y tamaños relativos de los polipéptidos de AARS C-terminales mostrados de manera esquemática. La figura 2A muestra fragmentos identificados mediante análisis de espectrometría de masas, la figura 2B muestra fragmentos identificados mediante secuenciación profunda de los transcriptomas.

60

La **figura 3** muestra la estructura de dominios de la treonil aminoacil ARNt sintetasa dispuesta con las posiciones y tamaños relativos de los polipéptidos de AARS internos identificados mediante análisis de espectrometría de masas mostrados de manera esquemática.

### Breve resumen de la invención

La invención a la que se refiere la presente memoria descriptiva se expone en las reivindicaciones adjuntas a esta descripción.

5 Las realizaciones de la presente divulgación se refieren en general al descubrimiento de fragmentos de proteína de aminoacil-ARNt sintetasas (AARS), que poseen actividades biológicas no canónicas, tales como actividades de señalización extracelular y/u otras características de relevancia terapéutica y diagnóstica. Las AARS son elementos universales y esenciales de la maquinaria de síntesis de proteínas hallados en todos los organismos, pero las AARS  
10 humanas y sus proteínas asociadas tienen variantes cortadas de origen natural, con potentes actividades de señalización celular que contribuyen al funcionamiento normal de los seres humanos. Las actividades de estos fragmentos de proteínas son distintas de las actividades de síntesis de proteínas comúnmente conocidas de las AARS y la presente invención incluye el descubrimiento y desarrollo de estas proteínas cortadas como nuevos agentes bioterapéuticos, nuevos reactivos de investigación y descubrimiento y como nuevos antígenos/dianas para  
15 agentes biológicos y diagnósticos dirigidos que pueden usarse para tratar o diagnosticar potencialmente una gran variedad de enfermedades humanas, tales como enfermedades o trastornos inflamatorios, hematológicos, neurodegenerativos, autoinmunitarios, hematopoyéticos, cardiovasculares y metabólicos.

20 Los fragmentos de proteína AARS de la presente invención pueden denominarse, por lo tanto, "resectinas" o como alternativa, "apendacrininas". Como se ha indicado anteriormente, el término "resectina" se deriva del proceso de cortar o reseccionar un fragmento de proteína AARS dado del contexto de su secuencia de AARS parental de longitud completa, que normalmente enmascara sus actividades no canónicas. En determinados casos, los fragmentos y polinucleótidos de proteína AARS de la presente invención se identificaron mediante la aparición de este proceso de resección, ya sea de origen natural (por ejemplo, proteolítico, variante de corte y empalme),  
25 inducido artificialmente o predicho. El término "apendacrino" se deriva de una combinación de "adjuntar" (del latín *appendere*) y de "separar" o "discernir" (del griego *crines*) y también refleja la separación de uno o más de los dominios adjuntos de los fragmentos de proteína AARS de sus secuencias de AARS de longitud completa o parental correspondientes.

30 Aunque se ha demostrado previamente que unos pocos fragmentos de AARS tienen actividades no de sintetasa, la expresión, aislamiento, purificación y caracterización de dichos fragmentos respecto de su utilidad bioterapéutica, de descubrimiento o diagnóstica es limitada y los expertos en la materia no habrían apreciado fácilmente que dichas actividades se asocien con cada miembro de la familia completa de AARS o con fragmentos alternativos. En el presente documento, se utilizó una estrategia metódica para descubrir y verificar fragmentos de proteína AARS para  
35 las 20 AARS mitocondriales y las 20 citosólicas (y las proteínas asociadas) respecto de su utilidad para descubrimiento bioterapéutico y diagnóstica. Por ejemplo, algunos de los presentes fragmentos de proteína AARS y los polinucleótidos que los codifican se identifican a partir de muestras biológicas usando espectrometría de masas (EM), principalmente para identificar fragmentos proteolíticos y otros se identificaron mediante técnicas de secuenciación profunda, principalmente para identificar variantes de corte y empalme. Otros fragmentos de proteína  
40 AARS se identifican usando predicciones *in silico* de secuencias de aminoácidos, tal como mediante la comparación computarizada de sintetasas de seres humanos y organismos inferiores junto con demarcaciones clave (por ejemplo, sitios de proteasa); esta estrategia utilizó análisis de secuencia de la AARS de longitud completa basándose en criterios específicos para discernir fragmentos proteolíticos y dominios funcionales que poseen actividades biológicas no canónicas.

45 Son inesperadas nuevas resectinas de las AARS y también es inesperada su expresión diferencial. Normalmente, se observan resecciones específicas con diferentes tratamientos (por ejemplo, de células cultivadas en medio con o sin suero), en diferentes estadios de crecimiento (por ejemplo, cerebro adulto frente a cerebro fetal) y para diferentes tipos de tejido (por ejemplo, páncreas frente a hígado). El patrón de expresión no es el mismo para todas las aminoacil ARNt sintetasas a pesar del hecho de ser necesarias las funciones canónicas de todas las aminoacil  
50 ARNt sintetasas en las mismas ubicaciones celulares y en cantidades relativamente proporcionales. No sería de esperar que aumentasen los niveles de actividad de una aminoacil ARNt sintetasa sin un aumento de las cantidades de otras actividades de aminoacil ARNt sintetasa al mismo tiempo. Los datos de espectrometría de masa y de secuenciación profunda indican que las resectinas de aminoacil ARNt sintetasa tienen diversos niveles y se producen en diferentes sitios y en diferentes estadios.

Además, los fragmentos de proteína ARS pueden expresarse y purificarse hasta una pureza lo suficientemente alta para discernir sus propiedades biológicas. Anteriormente, los fragmentos normalmente no tenían una pureza,  
60 plegamiento y estabilidad suficientes para permitir la caracterización biológica adecuada de las actividades no sintetasa. Por ejemplo, los ensayos basados en células se usan conjuntamente con resectinas suficientemente puras, estables y plegadas para revelar sus importantes actividades bioterapéuticas, de descubrimiento o diagnósticas.

En particular, las realizaciones de la presente divulgación se refieren a fragmentos de proteínas de treonil ARNt  
65 sintetasas, agentes relacionados y composiciones con utilidad bioterapéutica, de descubrimiento o diagnóstica y a métodos de uso de los mismos. Las composiciones de la presente invención son útiles en diversas aplicaciones

diagnósticas, de descubrimiento de fármacos y terapéuticas, tal como se describen en el presente documento. Preferentemente, las proteínas y fragmentos de AARS se purifican y almacenan en un estado adecuado hasta el grado necesario para dichos usos bioterapéuticos, de descubrimiento o diagnósticos.

5 Determinadas realizaciones incluyen composiciones, que comprenden un fragmento de proteína aminoacil-ARNt sintetasa (AARS) aislado de al menos aproximadamente 100, 90, 80, 70, 60, 50 o 40 aminoácidos que comprende una secuencia de aminoácidos tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 y tiene una solubilidad de al menos 5 mg/ml y en donde la composición tiene una pureza de al menos aproximadamente el 95 % basándose en la proteína y menos de aproximadamente 10 UE/mg de proteínas endotoxinas. En un aspecto, la  
10 composición es una composición terapéutica. En realizaciones específicas, la composición es sustancialmente aséptica. En algunas realizaciones, el fragmento de proteína AARS comprende una actividad no canónica. En algunas realizaciones, la actividad biológica no canónica se selecciona entre modulación de la señalización extracelular, modulación de la proliferación celular, modulación de la diferenciación celular, modulación de la transcripción génica, modulación de la producción o actividad de citocinas, modulación de la actividad de citocinas receptoras y modulación de la inflamación. En algunas realizaciones, el fragmento de proteína AARS tiene una  $CE_{50}$   
15 de menos de aproximadamente 1 nM, aproximadamente 5 nM, aproximadamente 10 nM, aproximadamente 50 nM, aproximadamente 100 nM o aproximadamente 200 nM para una actividad biológica no canónica basada en células.

20 En determinadas realizaciones, el fragmento de proteína AARS se fusiona a un polipéptido heterólogo. En algunas realizaciones, la proteína de fusión de AARS retiene sustancialmente una actividad no canónica del fragmento de proteína AARS. En algunas realizaciones, la proteína de fusión de AARS suprime una actividad no canónica del fragmento de proteína AARS. En algunas realizaciones, el polipéptido heterólogo se une al extremo N-terminal del fragmento de proteína AARS. En algunas realizaciones, el polipéptido heterólogo se une al C-terminal del fragmento de proteína AARS. En un aspecto de cualquiera de estas realizaciones, el polipéptido heterólogo se selecciona entre  
25 el grupo que consiste en marcadores de purificación, marcadores epitópicos, secuencias de direccionamiento, péptidos de señal, secuencias de translocación de membrana y modificadores PK.

30 En determinadas realizaciones, la composición comprende un fragmento de proteína AARS a una concentración de al menos aproximadamente 10 mg/ml. En determinadas realizaciones, la composición comprende un fragmento de proteína AARS que está al menos un 90 % monodisperso. En determinadas realizaciones, la composición comprende menos de aproximadamente un 3 % de proteínas agregadas de alto peso molecular. En determinadas realizaciones, la composición muestra una agregación menor del 3 % cuando se almacena a una concentración de al menos 10 mg/ml en PBS durante una semana a 4 °C. En determinadas realizaciones, la composición muestra una agregación menor del 3 % cuando se almacena a una concentración de al menos 10 mg/ml en PBS durante una  
35 semana a temperatura ambiente.

40 En el presente documento se describen diversos ensayos para medir dichas características de las resectinas y pueden usarse para definir aspectos de la invención. En determinados aspectos, estas características serán preferibles para la utilidad bioterapéutica de los fragmentos de proteína AARS descritos en el presente documento.

45 Determinadas realizaciones incluyen composiciones, que comprenden un fragmento de proteína aminoacil-ARNt sintetasa (AARS) de al menos 100 aminoácidos que difiere de una secuencia de aminoácidos expuesta en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 por la sustitución, eliminación y/o adición de aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 aminoácidos, en donde el fragmento de proteína alterado retiene sustancialmente una actividad no canónica de la proteína no alterada o tiene un fenotipo negativo dominante en relación con la actividad no canónica, en donde el fragmento de proteína tiene una solubilidad de al menos aproximadamente 5 mg/ml y en donde la composición tiene una pureza de al menos aproximadamente el 95 % basándose en la proteína y menos de aproximadamente 10 UE/mg de proteína endotoxina. En realizaciones  
50 específicas, la composición es sustancialmente aséptica.

Otras realizaciones incluyen composiciones, que comprenden un anticuerpo aislado que se une específicamente a un fragmento aislado de aminoacil-ARNt sintetasa (AARS) tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9, en donde la afinidad del anticuerpo por el fragmento de proteína AARS es aproximadamente 10X más fuerte que su afinidad por un polipéptido AARS correspondiente de longitud completa. Uno de los aspectos sorprendentes de la presente invención incluye determinadas resectinas que poseen "nuevas" superficies accesibles a los anticuerpos u otros agentes biológicos dirigidos, mientras que la AARS de longitud completa "esconde" o cubre estas superficies con otras secuencias o dominios adyacentes. El proceso de resección también puede crear una mayor accesibilidad al agua para revelar actividades biológicas no identificadas previamente. Algunas realizaciones incluyen composiciones, que comprenden un anticuerpo aislado que se une específicamente a un fragmento aislado  
55 de aminoacil-ARNt sintetasa (AARS) tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9, en donde el anticuerpo tiene una afinidad de la menos aproximadamente 10 nM por el fragmento de proteína AARS y una afinidad de al menos aproximadamente 100 nM por un polipéptido AARS de longitud completa correspondiente. En algunas realizaciones, el anticuerpo se une a un epítipo ubicado dentro de una unión de corte y empalme única del polipéptido de AARS tal como se expone en cualquiera de las tablas 1-3 o las tablas 4-6 o las tablas 7-9 o a una secuencia de aminoácidos C-terminal de este sitio de corte y empalme: En determinadas realizaciones, el anticuerpo antagoniza la actividad no canónica del fragmento de proteína AARS. Dichos antagonistas pueden unirse  
60  
65

opcionalmente a l AARS parental o de longitud completa correspondiente.

Otros aspectos se refieren a sistemas de bioensayo, que comprenden un fragmento de proteína aminoacil-ARNt sintetasa (AARS) sustancialmente puro de al menos 100 aminoácidos que comprende una secuencia de aminoácidos tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 y un compañero de unión que se une al fragmento de proteína AARS. En un aspecto, el compañero de unión se selecciona entre el grupo que consiste en una proteína receptora de la superficie celular, ácido nucleico, membrana lipídica, proteína reguladora celular, enzima y factor de transcripción. Opcionalmente, dicho receptor puede formar parte de una célula, preferentemente una célula relevante para la biología revelada dela resectina.

Determinadas realizaciones incluyen composiciones celulares, que comprenden un fragmento de proteína aminoacil-ARNt sintetasa (AARS) aislado de al menos 100 aminoácidos que comprende una secuencia de aminoácidos tal como se expone en las tablas 1-3 o las talas 4-6 o las tablas 7-9 y una población de células modificadas por ingeniería genética en la que al menos una célula comprende un polinucleótido que codifica dicho fragmento de proteína AARS. En un aspecto, las células son capaces de crecer en un medio libre de suero.

También se incluyen sistemas de detección, que comprenden un fragmento de proteína aminoacil-ARNt sintetasa (AARS) sustancialmente puro de al menos 50 o 100 aminoácidos que comprende una secuencia de aminoácidos tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9, una célula que comprende un receptor de la superficie celular o una porción extracelular del mismo que se une al fragmento de proteína y una molécula de menos de aproximadamente 2000 Dalton o un segundo polipéptido, que modula la unión o la interacción entre el fragmento de proteína AARS y el receptor extracelular.

Las realizaciones particulares incluyen sistemas diagnósticos, que comprenden un fragmento de proteína aminoacil-ARNt sintetasa (AARS) sustancialmente puro de al menos 100 aminoácidos que comprende una secuencia de aminoácidos tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 y una célula que comprende un receptor de la superficie celular o una porción extracelular del mismo que se une al fragmento de proteína AARS, en donde el sistema o la célula comprende una molécula indicadora que permite la detección de un cambio en los niveles o la actividad del receptor de la superficie celular o una porción extracelular del mismo.

Determinadas realizaciones incluyen dispositivos para el crecimiento celular, que comprenden un fragmento de proteína aminoacil-ARNt sintetasa (AARS) aislado de la menos 100 aminoácidos que comprende una secuencia de aminoácidos tal como se expone en las talas 1-3 o las tablas 4-6 o las tablas 7-9, una población de células modificadas por ingeniería genética en las que al menos una célula comprende un polinucleótido que codifica dicho fragmento de proteína AARS, al menos aproximadamente 10 litros de medio libre de suero y un recipiente estéril. En realizaciones específicas, las células utilizadas para cualquiera de los métodos o las composiciones descritas en el presente documento pueden crecer en medio libre de suero, opcionalmente con un antibiótico y un inductor.

Algunas realizaciones se refieren a agentes antisentido o de interferencia de ARN (ARNi), que comprenden una secuencia que se dirige contra una unión de corte y empalme única de una variante de corte y empalme de AARS tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9.

También se incluyen composiciones terapéuticas, que comprenden un fragmento de proteína aminoacil-ARNt sintetasa (AARS) aislado de la menos 100 aminoácidos que comprende una secuencia de aminoácidos tal como se expone en las talas 1-3 o las tablas 4-6 o las tablas 7-9, en donde el fragmento de proteína se une específicamente a un compañero de unión y tiene una solubilidad de al menos aproximadamente 5 mg/ml y en donde la composición tiene una pureza de al menos aproximadamente el 95 % basándose en la proteína. En algunos aspectos, la composición puede tener menos de 10 UE de endotoxina/mg de proteína.

También se incluyen composiciones, que comprenden un fragmento de proteína aminoacil-ARNt sintetasa (AARS) de al menos 100 aminoácidos que es al menos un 80 %, 85 %, 90 %, 95 %, 98 %, o 100 % idéntico a una secuencia de aminoácidos expuesta en las tablas 1-3 o las tablas 4-6 o las tablas 7-9, en donde el fragmento de proteína tiene una solubilidad de al menos aproximadamente 5 mg/ml y en donde la composición tiene una pureza de al menos aproximadamente el 95 % basándose en la proteína y menos de aproximadamente 10 UE de endotoxina/mg de proteína. En cualquiera de estas realizaciones, las composiciones pueden comprender un fragmento de proteína AARS que está al menos aproximadamente un 50 %, aproximadamente un 60 %, aproximadamente un 70 %, aproximadamente un 80 %, aproximadamente un 90 % o aproximadamente un 95 % monodisperso con respecto a su masa molecular aparente. En otro aspecto de estas realizaciones, las composiciones comprenden menos de aproximadamente un 10 % (basándose en la proteína) de proteínas agregadas de alto peso molecular o menos de aproximadamente un 5 % de proteínas agregadas de alto peso molecular o menos de aproximadamente un 4 % de proteínas agregadas de alto peso molecular o menos de aproximadamente un 3 % de proteínas agregadas de alto peso molecular o menos de aproximadamente un 2 % de proteínas agregadas de alto peso molecular o menos de aproximadamente un 1 % de proteínas agregadas de alto peso molecular.

En otro aspecto de estas realizaciones, las composiciones muestran una agregación menor de aproximadamente el 10 % cuando se almacenan a una concentración de al menos 10 mg/ml en PBS durante una semana a 4 °C o una

agregación de menos de aproximadamente el 5 % cuando se almacena a una concentración de al menos 10 mg/ml en PBS durante una semana a 4 °C o una agregación de menos de aproximadamente el 3 % cuando se almacenan a una concentración de al menos 10 mg/ml en PBS durante una semana a 4 °C o una agregación menor de aproximadamente el 2 % cuando se almacenan a una concentración de al menos 10 mg/ml en PBS durante una semana a 4 °C o una agregación menor de aproximadamente el 1 % cuando se almacenan a una concentración de al menos 10 mg/ml en PBS durante una semana a 4 °C.

Determinadas realizaciones incluyen composiciones, que comprenden un fragmento de proteína aminoacil-ARNt sintetasa (AARS) sustancialmente puro de al menos 100 aminoácidos que comprende una secuencia de aminoácidos tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 y al menos un resto unido covalentemente o no covalentemente al mismo. En algunas realizaciones, el resto es un marcador detectable. En algunas realizaciones, el resto es un polímero hidrosoluble. En algunas realizaciones, el resto es PEG. En un aspecto de estas realizaciones, el resto está unido al extremo N-terminal del fragmento de proteína. En un aspecto de estas realizaciones, el resto está unido al extremo C-terminal del fragmento de proteína.

Las realizaciones particulares incluyen composiciones, que comprenden un sustrato sólido unido a un fragmento de proteína aminoacil-ARNt sintetasa (AARS) aislado de la menos 100 aminoácidos que comprende una secuencia de aminoácidos tal como se expone en las talas 1-3 o las tablas 4-6 o las tablas 7-9 o un fragmento biológicamente activo o variante del mismo, en donde el fragmento de proteína tiene una solubilidad de al menos aproximadamente 5 mg/ml y la composición tiene una pureza de al menos aproximadamente el 95 % basándose en la proteína.

También se incluyen composiciones, que comprenden un agente de unión que se une específicamente a un fragmento aislado de aminoacil-ARNt sintetasa (AARS) tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9, en donde el agente de unión tiene una afinidad de al menos aproximadamente 1 nM por el fragmento de proteína. En un aspecto, el agente de unión se une a un epítipo ubicado dentro de una unión de corte y empalme única del polipéptido de AARS tal como se expone en cualquiera de las tablas 1-3 o las tablas 4-6 o las tablas 7-9 o a una secuencia de aminoácidos C-terminal de este sitio de corte y empalme. En algunas realizaciones, el agente de unión antagoniza una actividad no canónica del polipéptido AARS.

Determinadas realizaciones incluyen polipéptidos de aminoacil-ARNt sintetasa (AARS), que comprenden una secuencia de aminoácidos de un fragmento de proteína AARS tal como se describe en el presente documento, una secuencia de aminoácidos codificada por un polinucleótido de AARS tal como se describe en el presente documento, o una variante o fragmento para el mismo. Determinados polipéptidos de AARS comprenden una secuencia de aminoácidos que es al menos un 80 %, 85 %, 90 %, 95 %, 98 %, o 100 % idéntico a una secuencia de referencia de AARS tal como se divulga en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 o la tabla E2. Determinados polipéptidos AARS consisten esencialmente en una secuencia de aminoácidos que es al menos un 80 %, 85 %, 90 %, 95 %, 98 %, o 100 % idéntico a una secuencia de referencia de AARS tal como se divulga en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 o la tabla E2. En determinadas realizaciones, el polipéptido comprende una actividad biológica no canónica. En realizaciones específicas, la actividad biológica no canónica se selecciona entre modulación de la señalización celular (por ejemplo, señalización extracelular), modulación de la proliferación celular, modulación de la migración celular, modulación de la diferenciación celular, modulación de la apoptosis o muerte celular, modulación de la angiogénesis, modulación de la unión celular, modulación del metabolismo celular, modulación de la captación celular, modulación de la transcripción génica o la secreción, modulación de la producción o actividad de citocinas, modulación de la actividad de citocinas receptoras y modulación de la inflamación.

Otros aspectos incluyen anticuerpos y otros agentes de unión que muestran especificidad de unión por un polipéptido AARS aislado tal como se describe en el presente documento, un compañero de unión del polipéptido de AARS o el complejo de ambos. En algunas realizaciones, la afinidad del anticuerpo o el agente de unión para el polipéptido de AARS es aproximadamente 10X más fuerte que su afinidad por un polipéptido AARS de longitud completa correspondiente. En realizaciones específicas, el agente de unión se selecciona entre un péptido, peptidomimético, una adnectina, un aptámero y una molécula pequeña. En determinadas realizaciones, el anticuerpo o el agente de unión antagoniza una actividad no canónica del polipéptido AARS. En otras realizaciones, el anticuerpo o el agente de unión agoniza una actividad no canónica del polipéptido AARS.

Determinadas realizaciones incluyen polinucleótidos de aminoacil-ARNt sintetasa (AARS), que comprenden una secuencia de nucleótidos de un polinucleótido de AARS tal como se describe en el presente documento, una secuencia de nucleótidos que codifica un fragmento de proteína AARS tal como se describe en el presente documento o una variante, un fragmento o un complemento del mismo. Determinados polinucleótidos de AARS comprenden una secuencia de nucleótidos que es al menos un 80 %, 85 %, 90 %, 95 %, 98 %, o 100 % idéntica a un polinucleótido de referencia de AARS o un complemento del mismo, tal como se divulga en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 o la tabla E2. En algunas realizaciones, la secuencia de nucleótidos está optimizada por codones para expresión bacteriana. En un aspecto, la secuencia de nucleótidos es al menos un 80 % idéntica a una secuencia de polinucleótidos divulgada en la tabla E2.

Los polinucleótidos de AARS específicos consisten esencialmente en una secuencia de nucleótidos que es al menos

un 80 %, 85 %, 90 %, 95 %, 98 %, o 100 % idéntica a un polinucleótido de referencia de AARS o un complemento del mismo, tal como se divulga en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 o la tabla E2. Otros polinucleótidos de AARS comprenden o consisten esencialmente en una secuencia de nucleótidos que hibrida específicamente con un polinucleótido de referencia de AARS, tal como se divulga en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 o la

5 tabla E2. En determinadas realizaciones, el polinucleótido se selecciona entre un cebador, una sonda y un oligonucleótido antisentido. En realizaciones específicas, el cebador, la sonda o el oligonucleótido antisentido se dirige a una unión de corte y empalme específica o única y/o a una secuencia 3' de este sitio de corte y empalme dentro del polinucleótido de AARS.

10 Determinadas realizaciones incluyen métodos para determinar la presencia o los niveles de fragmento de proteína AARS en una muestra, que comprenden poner en contacto la muestra con uno o más agentes de unión que se unen específicamente a un fragmento de proteína AARS tal como se describe en el presente documento, detectar la presencia o ausencia del agente de unión y de este modo determinar la presencia o los niveles del fragmento de proteína AARS. Otras realizaciones incluyen métodos para determinar la presencia o los niveles de fragmento de

15 proteína AARS en una muestra, que comprenden analizar la muestra con un detector que es capaz de identificar de manera específica un fragmento de proteína se describe en el presente documento y de este modo determinar la presencia o los niveles del fragmento de proteína AARS. En realizaciones específicas, el detector es un espectrómetro de masas (EM), un citómetro de flujo, un dispositivo para la obtención de imágenes de proteínas, un ensayo inmunoabsorbente acoplado a enzimas (ELISA), o una micromatriz de proteínas. Determinadas realizaciones

20 comprenden comparar la presencia o los niveles del fragmento de proteína AARS con una muestra de control o con un valor predeterminado. Determinadas realizaciones comprenden caracterizar el estado de la muestra para distinguirla del control. En realizaciones específicas, la muestra y el control comprenden una célula o tejido y el método comprende distinguir entre células o tejidos de diferentes especies, células de diferentes tejidos u órganos, células en diferentes estados de desarrollo celular, células en diferentes estados de diferenciación celular, células en diferentes estados fisiológicos o células sanas y enfermas. Por ejemplo, las resectinas seleccionadas pueden ser

25 más abundantes en condiciones tales como estrés o lesión.

Determinadas realizaciones incluyen métodos de y composiciones relacionadas para, identificar un compuesto que se une específicamente a un polipéptido de aminoacil-ARNt sintetasa (AARS) tal como se describe en el presente

30 documento o uno o más de sus compañeros de unión celulares, que comprenden a) combinar el polipéptido de AARS o su compañero de unión celular o ambos con al menos un compuesto de ensayo en condiciones adecuadas y b) detectar la unión del polipéptido de AARS o su compañero de unión celular o ambos al compuesto de ensayo, identificando de este modo un compuesto que se une específicamente al polipéptido de AARS o su compañero de unión celular o ambos. En determinadas realizaciones, el compuesto de ensayo es un polipéptido o un péptido, un anticuerpo o un fragmento de unión a antígeno del mismo, un peptidomimético o una molécula pequeña. En determinadas realizaciones, el compuesto de ensayo agoniza una actividad biológica no canónica del polipéptido de AARS o su compañero de unión celular. En otras realizaciones, el compuesto de ensayo antagoniza una actividad biológica no canónica del polipéptido de AARS o su compañero de unión celular. Determinadas realizaciones incluyen un compuesto identificado mediante el método anterior, tal como un agonista (por ejemplo, molécula

35 pequeña, péptido).

40

Determinadas realizaciones incluyen métodos para determinar la presencia o los niveles de una secuencia de polinucleótido de una variante de corte y empalme de AARS en una muestra, que comprende poner en contacto la muestra con uno o más oligonucleótidos que hibridan específicamente con un polinucleótido de AARS tal como se describe en el presente documento, detectar la presencia o ausencia de los oligonucleótidos en la muestra y de este modo determinar la presencia o los niveles de la secuencia de polinucleótido de la variante de corte y empalme de AARS. Otras realizaciones incluyen métodos para determinar la presencia o los niveles de una secuencia de polinucleótido de una variante de corte y empalme de AARS en una muestra, que comprenden poner en contacto la muestra con al menos dos oligonucleótidos que amplifican específicamente un polinucleótido de AARS tal como se describe en el presente documento, llevar a cabo una reacción de amplificación, detectar la presencia o ausencia de un producto amplificado y de este modo determinar la presencia o los niveles de la secuencia de polinucleótido de la variante de corte y empalme de AARS. En realizaciones específicas, el o los oligonucleótidos hibridan específicamente con o amplifican específicamente una unión de corte y empalme que es única para la variante de corte y empalme de AARS. Determinadas realizaciones incluyen comparar la presencia o los niveles del fragmento

45 de proteína o variante de corte y empalme de AARS con una muestra de control o con un valor predeterminado. Determinadas realizaciones incluyen caracterizar el estado de la muestra para distinguirla del control. En realizaciones específicas, la muestra y el control comprenden una célula o tejido y el método comprende distinguir entre células o tejidos de diferentes especies, células de diferentes tejidos u órganos, células en diferentes estados de desarrollo celular, células en diferentes estados de diferenciación celular o células sanas y enfermas.

50

55

60 Algunas realizaciones incluyen composiciones farmacéuticas, que comprenden un polinucleótido de AARS descrito en el presente documento, un polipéptido de AARS descrito en el presente documento, un agente de unión como se describe en el presente documento o un compuesto identificado mediante el método anterior o descrito en el presente documento y un excipiente o vehículo farmacéuticamente aceptable.

65

Determinadas realizaciones incluyen métodos para modular una actividad celular de una célula, que comprenden

poner en contacto la célula con un polinucleótido AARS descrito en el presente documento, un polipéptido de AARS descrito en el presente documento, un agente de unión descrito en el presente documento, un compuesto del método anterior o descrito en el presente documento o una composición farmacéutica descrita en el presente documento. En realizaciones específicas, la actividad celular se selecciona entre proliferación celular, migración celular, diferenciación celular, apoptosis o muerte celular, señalización celular, angiogénesis, unión celular, captación celular, secreción celular, metabolismo, producción o actividad de citocinas, actividad de receptor de citocinas, transcripción génica e inflamación. En un aspecto, la célula se selecciona entre el grupo que consiste en preadipocitos, médula ósea, neutrófilos, células sanguíneas, hepatocitos, astrocitos, células madre mesenquimales y células musculares esqueléticas.

En determinadas realizaciones, la célula se encuentra en un sujeto. Determinadas realizaciones comprenden tratar al sujeto, en donde el sujeto tiene una afección asociada con una enfermedad neoplásica, una enfermedad o afección del sistema inmunitario, una enfermedad infecciosa, una enfermedad metabólica, un trastorno inflamatorio, una enfermedad neuronal/neurológica, una enfermedad muscular/cardiovascular, una enfermedad asociada con una hematopoyesis aberrante, una enfermedad asociada con una angiogénesis aberrante o una enfermedad asociada con una supervivencia celular aberrante.

También se incluyen procesos para fabricar un compuesto farmacéutico, que comprenden: a) llevar a cabo una exploración *in vitro* de uno o más compuestos candidatos en presencia de un fragmento de proteína AARS de al menos 100 aminoácidos que comprende una secuencia de aminoácidos tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9, para identificar un compuesto que se une específicamente al fragmento de proteína AARS; b) llevar a cabo un ensayo bioquímico o de receptor basado en células con el compuesto identificado en la etapa a), para identificar un compuesto que modula una o más actividades no canónicas del fragmento de proteína AARS; c) evaluar opcionalmente la relación estructura-actividad (SAR) del compuesto identificado en la etapa b), para correlacionar su estructura con la modulación de la actividad no canónica y opcionalmente derivatizar el compuesto para alterar su capacidad para modular la actividad no canónica; y d) producir cantidades suficientes del compuesto identificado en la etapa b), o el compuesto derivatizado en la etapa c), para su uso en seres humanos, fabricando de este modo el compuesto farmacéutico.

Otras realizaciones incluyen procesos para fabricar un compuesto farmacéutico, que comprenden: a) llevar a cabo una exploración *in vitro* de uno o más compuestos candidatos en presencia de un receptor de la superficie celular o de una porción extracelular del mismo que se une específicamente a un fragmento de proteína AARS de las tablas 1-3 o las tablas 4-6 o las tablas 7-9, para identificar un compuesto que se une específicamente al receptor de la superficie celular o a la porción extracelular del mismo; b) llevar a cabo un ensayo bioquímico o de receptor basado en células con el compuesto identificado en la etapa a), para identificar un compuesto que modula una o más actividades no canónicas del fragmento de proteína AARS; c) evaluar opcionalmente la relación estructura-actividad (SAR) del compuesto identificado en la etapa b), para correlacionar su estructura con la modulación de la actividad no canónica y opcionalmente derivatizar el compuesto para alterar su capacidad para modular la actividad no canónica; y d) producir cantidades suficientes del compuesto identificado en la etapa b), o el compuesto derivatizado en la etapa c), para su uso en seres humanos, fabricando de este modo el compuesto farmacéutico.

Algunas realizaciones incluyen una composición celular, que comprende una población de células modificadas por ingeniería genética en las que al menos una célula comprende un polipéptido que codifica una proteína aminoacil-ARNt sintetasa (AARS) heteróloga de longitud completa, en donde las células son capaces de crecer en un medio libre de suero. En un aspecto, la proteína aminoacil-ARNt sintetasa (AARS) de longitud completa comprende un marcador de purificación o epitópico heterólogo para facilitar la purificación de un fragmento de proteína AARS. En otro aspecto, la proteína aminoacil-ARNt sintetasa (AARS) de longitud completa comprende un sitio de proteólisis heterólogo para permitir la producción del fragmento de proteína AARS tras su escisión.

Algunas realizaciones incluyen un método para producir un polipéptido de AARS tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 o la tabla E2 *in situ* dentro de una célula, que comprende; i) expresar una proteína aminoacil-ARNt sintetasa (AARS) heteróloga de longitud completa dentro de la célula, en donde la célula comprende una proteasa capaz de escindir la proteína aminoacil-ARNt sintetasa (AARS) heteróloga de longitud completa para producir el polipéptido de AARS.

Algunas realizaciones incluyen un método para producir un polipéptido de AARS tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 o la tabla E2 que comprende poner en contacto una proteína aminoacil-ARNt sintetasa (AARS) de longitud completa con una proteasa que es capaz de escindir la proteína aminoacil-ARNt sintetasa (AARS) de longitud completa y producir un polipéptido de AARS.

Algunas realizaciones incluyen una proteína aminoacil-ARNt sintetasa de longitud completa modificada por ingeniería genética que comprende un sitio de proteólisis heterólogo para permitir la generación proteolítica de un fragmento de proteína AARS tal como se expone en cualquiera de las tablas 1-3 o las tablas 4-6 o las tablas 7-9 o la tabla E2.

Algunas realizaciones incluyen una composición, que comprende una proteína aminoacil-ARNt sintetasa aislada de

longitud completa, en donde la composición tiene una pureza de al menos aproximadamente el 95 % basándose en la proteína, menos de aproximadamente 10 UE de endotoxina/mg de proteína y es sustancialmente aséptica. En un aspecto, la proteína aminoacil-ARNt sintetasa de longitud completa está presente a una concentración de al menos 10 mg/ml y está al menos un 90 % monodispersa.

5 Una realización adicional incluye un método para tratar una enfermedad o trastorno mediado por la desregulación de la expresión, la actividad o la ubicación espaciotemporal de una ARNt sintetasa mediante la administración de un fragmento de proteína AARS o un ácido nucleico que codifica el fragmento de proteína AARS, tal como se expone en cualquiera de las tablas 1-3 o las tablas 4-6 o las tablas 7-9 o la tabla E2. En un aspecto de esta realización, la enfermedad se selecciona entre cáncer, neuropatía, diabetes y trastornos inflamatorios.

**Descripción detallada de la invención**

I.	RESUMEN	14
II.	DEFINICIONES	15
III.	FRAGMENTOS Y VARIANTES DE PROTEÍNA AARS PURIFICADOS	27
IV.	POLINUCLEÓTIDOS DE AARS	69
V.	ANTICUERPOS	81
VI.	ALTERNATIVAS A LOS ANTICUERPOS Y OTROS AGENTES DE UNIÓN	86
VII.	BIOENSAYOS Y ENSAYOS ANALÍTICOS	90
VIII.	SISTEMAS DE EXPRESIÓN Y PURIFICACIÓN	93
IX.	MÉTODOS DIAGNÓSTICOS Y COMPOSICIONES	105
X.	AGENTES ANTISENTIDO Y DE IARN	120
	A. AGENTES ANTISENTIDO	121
	B. AGENTES DE INTERFERENCIA DE ARN	129
XI.	DESCUBRIMIENTO DE FÁRMACOS	136
XII.	MÉTODOS DE USO	145
XIII.	FORMULACIONES FARMACÉUTICAS, ADMINISTRACIÓN Y KITS	148
XIV.	EJEMPLOS	157

15 *I. RESUMEN*

La presente invención se refiere, al menos en parte, al descubrimiento de nuevos polipéptidos AARS y a métodos para su preparación y uso, que representan la transformación de proteínas nativas de tipo silvestre en nuevas formas que muestran características notablemente diferentes en comparación con los genes de treonil ARNt sintetasa de origen natural de longitud completa. Dichos polipéptidos de AARS se identificaron basándose en un análisis exhaustivo de la secuencia y de los espectros de masas de la treonil ARNt sintetasa expresada en diferentes tejidos, seguido de la producción y prueba sistemática de cada polipéptido de AARS potencial para identificar secuencias de proteína que representan dominios de proteína estables y solubles que muestran nuevas actividades biológicas y características de fármaco terapéutico favorables.

25 Basándose en este análisis, se han identificado al menos tres nuevas familias de polipéptidos de AARS derivadas de la treonil ARNt sintetasa.

30 En un aspecto, dichos polipéptidos de AARS derivados de treonil ARNt sintetasa comprenden secuencias de polipéptido que comprenden aproximadamente los aminoácidos 1-322 de la thr ARNt sintetasa.

En un segundo aspecto, dichos polipéptidos de AARS derivados de treonil ARNt sintetasa comprenden secuencias de polipéptido que comprenden aproximadamente los aminoácidos 1-146 de la treonil ARNt sintetasa.

35 En un tercer aspecto, dichos polipéptidos de AARS derivados de treonil ARNt sintetasa comprenden secuencias de polipéptido que comprenden aproximadamente los aminoácidos 294-723 de la treonil ARNt sintetasa.

40 Estas nuevas familias de polipéptido de AARS representan nuevos productos proteínicos previamente desconocidos que muestran, entre otras cosas, i) nueva actividad biológica, ii) características de estabilidad y agregación de proteínas favorables e iii) la capacidad de expresarse y producirse a un alto nivel en sistemas de expresión procariotas, que son características materialmente diferentes no encontradas en la proteína de tipo silvestre intacta.

*II. DEFINICIONES*

45 A menos que se defina de otro modo, todos los términos técnicos y científicos utilizados en este documento tienen el

mismo significado que el normalmente entendido por los expertos habituales en la materia a la cual pertenece la invención. Aunque pueden usarse cualesquiera métodos y materiales similares o equivalentes a aquellos descritos en el presente documento en la puesta en práctica o la prueba de la presente invención, se describen métodos y materiales preferidos. Para los fines de la presente invención, a continuación se describen los siguientes términos.

5 Los artículos "un", "una" y "uno" se usan en el presente documento para hacer referencia a uno o a más de uno (es decir, a al menos uno) del objeto gramatical del artículo. A modo de ejemplo, "un elemento" significa un elemento o más de un elemento.

10 Por "aproximadamente" se entiende una cantidad, nivel, valor, número, frecuencia, porcentaje, dimensión, tamaño, cantidad, peso o longitud que varía como mucho en un 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 % respecto de una cantidad, nivel, valor, número, frecuencia, porcentaje, dimensión, tamaño, cantidad, peso o longitud de referencia.

15 Un "agonista" se refiere a una molécula que intensifica o imita una actividad. Por ejemplo, una actividad biológica no canónica de una AARS u otra proteína. Los agonistas pueden incluir proteínas, ácidos nucleicos, hidratos de carbono, moléculas pequeñas o cualquier otro compuesto o composición que modula la actividad de una AARS ya sea interactuando directamente con la AARS o su compañero de unión o actuando sobre componentes de la ruta biológica en la que participa la AARS. Se incluyen agonistas parciales y totales.

20 Tal como se usa en el presente documento, el término "aminoácido" pretende hacer referencia a aminoácidos tanto de origen natural como de origen no natural, así como a análogos y miméticos de aminoácidos. Los aminoácidos de origen natural incluyen los 20 (L)-aminoácidos utilizados durante la biosíntesis de proteínas así como otros, tales como 4-hidroxiprolina, hidroxilisina, desmosina, isodesmosina, homocisteína, citrulina y ornitina, por ejemplo. Los aminoácidos de origen no natural incluyen, por ejemplo, (D)-aminoácidos, norleucina, norvalina, p-fluorofenilalanina, etionina y similares, que son conocidos por un experto en la materia. Los análogos de aminoácidos incluyen formas modificadas de aminoácidos de origen natural y no natural. Dichas modificaciones pueden incluir, por ejemplo, la sustitución o reemplazo de grupos y restos químicos en el aminoácido por derivatización del aminoácido. Los miméticos de aminoácidos incluyen, por ejemplo, estructuras orgánicas que muestran propiedades funcionalmente similares, tales como características de carga y espaciado de la carga del aminoácido de referencia. Por ejemplo, una estructura orgánica que imite a la arginina (Arg o R) podría tener un resto de carga positiva ubicado en un espacio molecular similar y tener el mismo grado de movilidad que el grupo e-amino de la cadena lateral del aminoácido Arg de origen natural. Los miméticos también incluyen estructuras restringidas para mantener interacciones de espaciado y carga óptimas del aminoácido o de los grupos funcionales del aminoácido. Los expertos en la materia saben o pueden determinar qué estructuras constituyen análogos y miméticos de aminoácidos funcionalmente equivalentes.

En determinados aspectos, puede emplearse el uso de aminoácidos no naturales para modificar (por ejemplo, aumentar) una actividad no canónica seleccionada de un fragmento de proteína AARS o para alterar la semivida *in vivo* o *in vitro* de la proteína. Los aminoácidos no naturales también pueden usarse para facilitar modificaciones químicas (selectivas) (por ejemplo, pegilación) de una proteína AARS. Por ejemplo, determinados aminoácidos no naturales permiten la unión selectiva de polímeros, tales como PEG, a una proteína dada y de este modo mejoran sus propiedades farmacocinéticas.

45 Los ejemplos específicos de análogos y miméticos de aminoácidos pueden encontrarse descritos en, por ejemplo, Roberts y Vellaccio, *The Peptides: Analysis, Synthesis, Biology*, Eds. Gross y Meinhofer, Vol. 5, p. 341, Academic Press, Inc., Nueva York, N.Y. (1983), incorporándose el volumen completo al presente documento por referencia. Otros ejemplos incluyen aminoácidos peralquilados, en particular, aminoácidos permetilados. Véase, por ejemplo, *Combinatorial Chemistry*, Eds. Wilson y Czarnik, Cap. 11, p. 235, John Wiley & Sons Inc., Nueva York, N.Y. (1997), incorporándose el libro completo al presente documento por referencia. Otros ejemplos más incluyen aminoácidos cuya porción de amida (y, por lo tanto, la cadena principal de amida del péptido resultante) se ha reemplazado, por ejemplo, por un anillo de azúcar, esteroide, benzodiazepina o carbociclo. Véase, por ejemplo, *Burger's Medicinal Chemistry and Drug Discovery*, Ed. Manfred E. Wolff, Cap. 15, págs. 619-620, John Wiley & Sons Inc., Nueva York, N.Y. (1995), incorporándose el libro completo al presente documento por referencia. Los métodos para sintetizar péptidos, polipéptidos, peptidomiméticos y proteínas se conocen bien en la técnica (véase, por ejemplo, la Patente de los Estados Unidos n.º 5.420.109; M. Bodanzsky, *Principles of Peptide Synthesis* (1ª ed. y 2ª ed. rev.), Springer-Verlag, Nueva York, N.Y. (1984 y 1993), véase el capítulo 7; Stewart y Young, *Solid Phase Peptide Synthesis*, (2ª ed.), Pierce Chemical Co., Rockford, Ill. (1984), incorporándose cada uno de ellos al presente documento por referencia. Por consiguiente, los polipéptidos AARS de la presente invención pueden estar compuestos de aminoácidos de origen natural y no natural así como de análogos y miméticos de aminoácidos.

60 El término "antagonista" se refiere a una molécula que reduce o atenúa una actividad. Por ejemplo, una actividad biológica no canónica de una AARS u otra proteína. Los antagonistas pueden incluir proteínas, tales como anticuerpos, ácidos nucleicos, hidratos de carbono, moléculas pequeñas o cualquier otro compuesto o composición que modula la actividad de un AARS o su compañero de unión, ya sea interactuando directamente con la AARS o su compañero de unión o actuando sobre componentes de la ruta biológica en la que participa la AARS. Se incluyen antagonistas parciales y totales.

El término "aminoacil-ARNt sintetasa" (AARS) se refiere en general a enzimas que en su forma natural o de tipo silvestre son capaces de catalizar la esterificación de un aminoácido específico o de su precursor a uno de todos sus ARNt afines compatibles para formar un aminoacil-ARNt. En esta actividad "canónica", las aminoacil-ARNt sintetetas catalizan una reacción de dos etapas: en primer lugar, activan su aminoácido respectivo formando un aminoacil-adenilato, en el que el carboxilo del aminoácido está unido al alfa-fosfato del ATP mediante el desplazamiento de pirofosfato y después, cuando se ha unido el ARNt correcto, el grupo aminoacilo del aminoacil-adenilato se transfiere al OH 2' o 3' terminal del ARNt.

Las aminoacil-ARNt sintetetas de clase I tienen normalmente dos motivos de secuencia altamente conservados. Estas enzimas aminoacilan en el 2'-OH de un nucleótido de adenosina y normalmente son monoméricas o diméricas. Las aminoacil-ARNt sintetetas de clase II tienen normalmente tres motivos de secuencia altamente conservados. Estas enzimas aminoacilan en el 3'-OH de la misma adenosina y normalmente son diméricas o tetraméricas. Los sitios activos de las enzimas de clase II están formados principalmente de una  $\beta$ -lámina antiparalela de siete hebras flanqueada por  $\alpha$ -hélices. Aunque la fenilalanina-ARNt sintetasa es de clase II, aminoacila en el 2'-OH.

Los polipéptidos de AARS incluyen fuentes de formas mitocondriales y citoplásmicas de tirosil-ARNt sintetasa (TyrRS), una triptofanil-ARNt sintetasa (TrpRS), una glutaminil-ARNt sintetasa (GlnRS), una glicil-ARNt sintetasa (GlyRS), una histidil-ARNt sintetasa (HisRS), una seril-ARNt sintetasa (SerRS), una fenilalanil-ARNt sintetasa (PheRS), una alanil-ARNt sintetasa (AlaRS), una asparaginil-ARNt sintetasa (AsnRS), una aspartil-ARNt sintetasa (AspRS), una cisteinil-ARNt sintetasa (CysRS), una glutamil-ARNt sintetasa (GluRS), una prolil-ARNt sintetasa (ProRS), una arginil-ARNt sintetasa (ArgRS), una isoleucil-ARNt sintetasa (IleRS), una leucil-ARNt sintetasa (LeuRS), una lisil-ARNt sintetasa (LysRS), una treonil-ARNt sintetasa (ThrRS), una metionil-ARNt sintetasa (MetRS) o una valil-ARNt sintetasa (ValRS). Las secuencias de tipo silvestre o parentales de estos polipéptidos de AARS se conocen en la técnica.

Por "secuencia codificante" se entiende cualquier secuencia de ácido nucleico que contribuye al código del producto polipeptídico de un gen. Por el contrario, la expresión "secuencia no codificante" se refiere a cualquier secuencia de ácido nucleico que no contribuye al código del producto polipeptídico de un gen.

A lo largo de esta memoria descriptiva, a menos que el contexto requiera otra cosa, se entenderá que las palabras "comprender", "comprende" y "comprendiendo" implican la inclusión de una etapa o elemento o de un grupo de etapas o elementos indicado, pero no la exclusión de cualquier otra etapa o elemento o de un grupo de etapas o elementos.

Por "que consiste en" se entiende que incluye, y se limita a, cualquier cosa que siga a la expresión "que consiste en". Por lo tanto, la expresión "que consiste en" indica que los elementos listados son necesarios u obligatorios y que no puede haber presentes otros elementos. Por "que consiste esencialmente en" se entiende que incluye cualquier elemento listado después de la expresión y se limita a otros elementos que no interfieren con o contribuyen a la actividad o acción especificada en la divulgación de los elementos listados. Por lo tanto, la expresión "que consiste esencialmente en" indica que los elementos listados son necesarios u obligatorios, pero que otros elementos son opcionales y pueden estar o no presentes, dependiendo de si afectan o no materialmente a la actividad o acción de los elementos listados.

La expresión "libre de endotoxinas" o "sustancialmente libre de endotoxinas" se refiere generalmente a composiciones, disolventes y/o vasos que contienen como máximo cantidades residuales (por ejemplo, cantidades que no tienen efectos fisiológicos clínicamente adversos para un sujeto) de endotoxina y preferentemente, cantidades indetectables de endotoxina. Las endotoxinas son toxinas asociadas con determinadas bacterias, normalmente bacterias gramnegativas, aunque pueden encontrarse endotoxinas en bacterias grampositivas, tales como *Listeria monocytogenes*. Las endotoxinas más prevalentes son los lipopolisacáridos (LPS) o los lipooligosacáridos (LOS) hallados en la membrana externa de diversas bacterias gramnegativas y que representan una característica patógena crucial en la capacidad de estas bacterias para provocar enfermedades. En los seres humanos, pequeñas cantidades de endotoxina pueden producir fiebre, una reducción de la presión sanguínea y la activación de la inflamación y la coagulación, entre otros efectos fisiológicos adversos.

Por lo tanto, en la producción farmacéutica de polipéptidos de AARS, normalmente es deseable retirar la mayor parte de todas las trazas de endotoxina de los productos farmacológicos y/o de los recipientes para fármacos, ya que cantidades incluso pequeñas pueden provocar efectos adversos en seres humanos. Para este fin puede emplearse un horno de despirogenación, ya que se necesitan normalmente temperaturas de más de 300 °C para descomponer la mayor parte de las endotoxinas. Por ejemplo, basándose en el material de embalaje primario, tales como jeringuillas o viales, la combinación de una temperatura del vidrio de 250 °C y un tiempo de retención de 30 minutos es suficiente para lograr una reducción 3 log en los niveles de endotoxinas. Se contemplan otros métodos para eliminar endotoxinas, incluyendo, por ejemplo, métodos de cromatografía y filtración, tal como se describe en el presente documento y se conoce en la técnica. También se incluyen métodos para producir polipéptidos de AARS y aislarlos de células eucariotas, tales como células de mamífero para reducir, si no eliminar, el riesgo de que haya presentes endotoxinas en una composición de la invención. Se prefieren métodos para producir polipéptidos de

AARS y aislados de células asépticas. Dichas composiciones que comprenden polipéptidos de AARS, representan nuevas formulaciones que muestran características terapéuticas nuevas y novedosas no encontradas en las composiciones de polipéptido de AARS contaminadas con suero o endotoxina que tienen el potencial de unirse a y alterar las nuevas propiedades biológicas de los polipéptidos de AARS.

5 Las endotoxinas pueden detectarse usando técnicas rutinarias conocidas en la técnica. Por ejemplo, el ensayo de lisado de amebocitos de *Limulus*, que utiliza sangre del cangrejo herradura, es un ensayo muy sensible para detectar la presencia de endotoxinas y los reactivos, kits e instrumentación para la detección de endotoxinas basándose en este ensayo se encuentran disponibles comercialmente, por ejemplo, del Lonza Group. En esta  
10 prueba, niveles muy bajos de LPS pueden provocar una coagulación detectable del lisado de *Limulus* debido a una fuerte cascada enzimática que amplifica esta reacción. También pueden cuantificarse endotoxinas mediante un ensayo inmunoabsorbente acoplado a enzimas (ELISA). Para encontrarse sustancialmente libre de endotoxinas, los niveles de endotoxinas pueden ser menor de aproximadamente 0,001, 0,005, 0,01, 0,02, 0,03, 0,04, 0,05, 0,06, 0,08, 0,09, 0,1, 0,5, 1,0, 1,5, 2, 2,5, 3, 4, 5, 6, 7, 8, 9 o 10 UE/mg de proteína. normalmente, 1 ng de lipopolisacárido (LPS)  
15 se corresponde a aproximadamente 1-10 UE.

En determinadas realizaciones, la "pureza" de cualquier agente dado (por ejemplo, fragmento de proteína AARS) en una composición puede definirse de manera específica. Por ejemplo, determinadas composiciones pueden comprender un agente que es al menos un 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100 % puro, incluyendo  
20 todos los decimales intermedios, según se mide, por ejemplo y de manera no limitante, mediante cromatografía líquida de alta presión (HPLC), una forma bien conocida de cromatografía en columna frecuentemente usada en bioquímica y en química analítica para separar, identificar y cuantificar compuestos.

Tal como se usa en el presente documento, los términos "función" y "funcional" y similares se refieren a una función biológica, enzimática o terapéutica.  
25

Por "gen" se entiende una unidad de herencia que puede ocupar un locus específico en un cromosoma y que consiste en secuencias reguladoras de la transcripción y/o traducción y/o una región codificante y/o secuencias no traducidas (es decir, intrones, secuencias no traducidas 5' y 3').  
30

"Homología" se refiere al número porcentual de aminoácidos que son idénticos o que constituyen sustituciones conservativas. La homología puede determinarse usando programas de comparación de secuencias, tales como GAP (Deveraux et al., 1984, Nucleic Acids Research 12, 387-395), que se incorpora al presente documento por referencia. De este modo, pueden compararse secuencias con una longitud similar o sustancialmente diferente a aquellas citadas en el presente documento, mediante la inserción de huecos en el alineamiento, determinándose dichos huecos, por ejemplo, por el algoritmo de comparación usado mediante GAP.  
35

La expresión "célula hospedadora" incluye una célula individual o un cultivo celular que puede ser o ha sido receptor de cualquier vector o vectores recombinantes, polinucleótido aislado o polipéptido de la invención. Las células hospedadoras incluyen la progenie de una sola célula hospedadora y la progenie no es necesariamente completamente idéntica (en cuanto a su morfología o complemento de ADN total) a la célula progenitora original debido a mutaciones y/o cambios naturales, accidentales o deliberados. Una célula hospedadora incluye células transfectadas o infectadas *in vivo* o *in vitro* con un vector recombinante o un polinucleótido de la invención. Una célula hospedadora que comprenda un vector recombinante de la invención es una célula hospedadora recombinante.  
40  
45

Por "aislado" se entiende un material que está sustancial o esencialmente libre de los componentes que normalmente lo acompañan en su estado nativo. Por ejemplo, un "polinucleótido aislado", tal como se usa en el presente documento, incluye un polinucleótido que se ha purificado de secuencias que lo flanquean en su estado de origen natural, *por ejemplo*, un fragmento de ADN que se ha retirado de las secuencias que normalmente se encuentran adyacentes al fragmento. Como alternativa, un "péptido aislado" o un "polipéptido aislado" y similares, tal como se usa en el presente documento, incluye el aislamiento y/o la purificación *in vitro* de una molécula de péptido o polipéptido de su ambiente celular natural y de su asociación con otros componentes de la célula; es decir, no se asocia de manera significativa con sustancias *in vivo*.  
50  
55

El término "ARNm" o en ocasiones citado como "transcritos de ARNm", tal como se usa en el presente documento, incluye, pero sin limitación, transcritos de pre-ARNm, intermedios que procesan transcritos, ARNm maduros listos para su traducción y transcritos del gen o genes o ácidos nucleicos derivados de los transcritos de ARNm. El procesamiento del transcrito puede incluir corte y empalme, edición y degradación. Tal como se usa en el presente documento, un ácido nucleico derivado de un transcrito de ARNm se refiere a un ácido nucleico para cuya síntesis ha servido en última instancia como molde el transcrito de ARNm o una subsecuencia del mismo. Un ADNc retrotranscrito a partir de un ARNm, un ARN transcrito a partir de ese ADNc, un ADN amplificado a partir del ADNc, un ARN transcrito a partir del ADN amplificado, etc., se derivan todos del transcrito de ARNm y la detección de dichos productos derivados es indicativa de la presencia y/o abundancia del transcrito original en una muestra. Por lo tanto, las muestras derivadas de ARNm incluyen, pero sin limitación, transcritos de ARNm del gen o los genes, ADNc retrotranscrito a partir del ARNm, ARNc transcrito a partir del ADNc, ADN amplificado a partir de los genes, ARN  
60  
65

transcrito a partir de ADN amplificado y similares.

Actividad "no canónica", tal como se usa en el presente documento, se refiere generalmente a i) una nueva actividad poseída por un polipéptido de AARS de la invención que no es poseída en cualquier grado significativo por la proteína parental de longitud completa nativa intacta o ii) una actividad que era poseída por la proteína parental de longitud completa nativa intacta, mientras que el polipéptido de AARS muestra una actividad específica significativamente mayor (es decir, al menos un 20 % mayor) en comparación con la proteína parental de longitud completa nativa intacta o muestra la actividad en un nuevo contexto; por ejemplo, aislando la actividad de otras actividades poseídas por la proteína parental de longitud completa nativa intacta. En el caso de polipéptidos de AARS, los ejemplos no limitantes de actividades no canónicas incluyen señalización extracelular, unión a ARN, unión a aminoácidos, modulación de la proliferación celular, modulación de la migración celular, modulación de la diferenciación celular (por ejemplo, hematopoyesis, neurogénesis, miogénesis, osteogénesis y adipogénesis), modulación de la transcripción génica, modulación de la apoptosis u otras formas de muerte celular, modulación de la señalización celular, modulación de la captación o secreción celular, modulación de la angiogénesis, modulación de la unión celular, modulación del metabolismo celular, modulación de la producción o actividad de citocinas, modulación de la actividad de receptor de citocinas, modulación de la inflamación y similares.

La expresión "concentración eficaz semimáxima" o "CE<sub>50</sub>" se refiere a la concentración de un fragmento de proteína AARS, anticuerpo u otro agente descrito en el presente documento a la cual se induce una respuesta a medio camino entre la basal y la máxima después de un determinado tiempo de exposición específico; la CE<sub>50</sub> de una curva de respuesta a la dosis graduada representa por lo tanto la concentración de un compuesto a la cual se observa un 50 % de su efecto máximo. En determinadas realizaciones, la CE<sub>50</sub> de un agente proporcionado en el presente documento se indica en relación con una actividad "no canónica", tal como se ha indicado anteriormente. La CE<sub>50</sub> también representa la concentración plasmática necesaria para obtener un 50 % de un efecto máximo *in vivo*. De manera similar, la "CE<sub>90</sub>" se refiere a la concentración de un agente o composición a la cual se observa un 90 % de su efecto máximo. La "CE<sub>90</sub>" puede calcularse a partir de la "CE<sub>50</sub>" y la pendiente de Hill o puede determinarse directamente a partir de los datos, usando conocimientos rutinarios de la técnica. En algunas realizaciones, la CE<sub>50</sub> de un fragmento de proteína AARS, anticuerpo u otro agente es menor de aproximadamente 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, 60, 70, 80, 90 o 100 nM. Preferentemente, la composición bioterapéutica tendrá un valor de CE<sub>50</sub> de aproximadamente 1 nM o menor.

El término "modular" incluye "aumentar" o "estimular", así como "disminuir" o "reducir", normalmente en una cantidad estadísticamente significativa o fisiológicamente significativa en comparación con un control. Por consiguiente, un "modulador" puede ser un agonista, un antagonista o una mezcla de los mismos dependiendo de las condiciones empleadas. Una cantidad "aumentada" o "potenciada" es normalmente una cantidad "estadísticamente significativa", y puede incluir un aumento es de 1,1, 1,2, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30 o más veces (por ejemplo, 500, 1000 veces) (incluyendo todos los números enteros y puntos decimales entre y por encima de 1, *por ejemplo*, 1,5, 1,6, 1,7, 1,8, etc.) la cantidad producida por ninguna composición (la ausencia de un agente o compuesto) o una composición de control. Una cantidad "disminuida" o reducida es normalmente una cantidad "estadísticamente significativa" y puede incluir una reducción del 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 11 %, 12 %, 13 %, 14 %, 15 %, 16 %, 17 %, 18 %, 19 %, 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, o 100 % en la cantidad producida por ninguna composición (la ausencia de un agente o compuesto) o una composición de control, incluyendo todos los números enteros intermedios. A modo de ejemplo no limitante, un control para comparar actividades canónicas y no canónicas podría incluir el fragmento de proteína AARS de interés comparado con su AARS de longitud completa correspondiente o un fragmento de AARS que tenga una actividad canónica comparable a la de su AARS de longitud completa correspondiente. Otros ejemplos de cantidades "estadísticamente significativas" se describen en el presente documento.

Por "obtenible de" se entiende que una muestra, tal como, por ejemplo, un extracto de polinucleótido o un extracto de polipéptido se aísla de o se deriva de, una fuente particular del sujeto. Por ejemplo, el extracto puede obtenerse de un tejido o un fluido biológico aislado directamente del sujeto. "Derivado" u "obtenido de" también pueden referirse a la fuente de un polipéptido o una secuencia de polinucleótido. Por ejemplo, una secuencia de AARS de la presente invención puede "derivarse" de la información de secuencia de un fragmento proteolítico de AARS o una variante de corte y empalme de AARS o una porción de la misma, ya sea de origen natural o generada artificialmente y por lo tanto puede comprender, consistir esencialmente en o consistir en esa secuencia.

Los términos "polipéptido" y "proteína" se usan de manera indistinta en el presente documento para hacer referencia a un polímero de restos de aminoácidos y a variantes y análogos de origen sintético y natural de los mismos. Por lo tanto, estos términos se aplican a polímeros de aminoácidos en los que uno o más restos de aminoácidos son aminoácidos sintéticos de origen no natural, tales como un análogo químico de un aminoácido de origen natural correspondiente, así como a polímeros de aminoácidos de origen natural y a derivados químicos de origen natural de los mismos. Dichos derivados incluyen, por ejemplo, modificaciones postraduccionales y productos de degradación que incluyen piroglutamilo, isoaspartilo, variantes proteolíticas, fosforiladas, glucosiladas, oxidadas, isomerizadas y desamidadas del fragmento de referencia de AARS.

Las expresiones "identidad de secuencia" o, por ejemplo, que comprende una "secuencia un 50 % idéntica a", tal como se usan en el presente documento, se refieren al grado en que las secuencias son idénticas basándose en la secuencia de nucleótidos o basándose en la secuencia de aminoácidos a lo largo de una ventana de comparación. Por lo tanto, puede calcularse un "porcentaje de identidad de secuencia" comparando dos secuencias alineadas de manera óptima a lo largo de la ventana de comparación, determinando el número de posiciones en las cuales aparece la base de ácido nucleico idéntica (por ejemplo, A, T, C, G, U) o el resto de aminoácidos idéntico (por ejemplo, Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys y Met) en ambas secuencias para proporcionar el número de posiciones emparejadas, dividiendo el número de posiciones emparejadas entre el número total de posiciones en la ventana de comparación (es decir, el tamaño de la ventana), y multiplicando el resultado por 100 para dar el porcentaje de identidad de secuencia.

Los términos usados para describir relaciones de secuencia entre dos o más polinucleótidos o polipéptidos incluyen "secuencia de referencia", "ventana de comparación", "identidad de secuencia", "porcentaje de identidad de secuencia" e "identidad sustancial". Una "secuencia de referencia" tiene una longitud de al menos 12 pero con frecuencia de 15 a 18 y normalmente al menos 25 unidades de monómero, incluyendo nucleótidos y restos de aminoácido. Debido a que dos polinucleótidos pueden comprender cada uno (1) una secuencia (es decir, solo una porción de la secuencia de polinucleótido completa) que sea similar entre los dos polinucleótidos y (2) una secuencia que sea divergente entre los dos polinucleótidos, las comparaciones de secuencia entre dos (o más polinucleótidos) se llevan a cabo normalmente comparando secuencias de los dos polinucleótidos a lo largo de una "ventana de comparación" para identificar y comparar regiones locales de similitud de secuencia. Una "ventana de comparación" se refiere a un segmento conceptual de al menos 6 posiciones contiguas, normalmente de aproximadamente 50 a aproximadamente 100, más normalmente de aproximadamente 100 a aproximadamente 150, en el que se compara una secuencia con una secuencia de referencia con el mismo número de posiciones contiguas después de que se alineen de manera óptima las dos secuencias. La ventana de comparación puede comprender adiciones o eliminaciones (es decir, huecos) de aproximadamente el 20 % o menos en comparación con la secuencia de referencia (que no comprende adiciones o eliminaciones) para un alineamiento óptimo de las dos secuencias. El alineamiento óptimo de secuencias para alinear una ventana de comparación puede llevarse a cabo mediante implementaciones computarizadas de algoritmos (GAP, BESTFIT, FASTA y TFASTA en la versión 7.0 de Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive Madison, WI, EE.UU.) o mediante inspección y el mejor alineamiento (es decir, el que da como resultado el mayor porcentaje de homología a lo largo de la ventana de comparación) generado mediante cualquiera de los diversos métodos seleccionados. También puede hacerse referencia a la familia de programas BLAST, tal como se divulga, por ejemplo, el Altschul et al., 1997, Nucl. Acids Res. 25:3389. Puede obtenerse una descripción detallada del análisis de secuencia en la Unidad 19.3 de Ausubel et al., "Current Protocols in Molecular Biology", John Wiley & Sons Inc, 1994-1998, Capítulo 15.

Los cálculos de similitud de secuencia o de identidad de secuencia entre secuencias (usándose los términos de manera intercambiable en el presente documento) se efectúan del modo siguiente. Para determinar el porcentaje de identidad de dos secuencias de aminoácidos o de dos secuencias de ácido nucleico, las secuencias se alinean con el objetivo de un alineamiento óptimo (por ejemplo, pueden introducirse huecos en una o ambas de una primera y una segunda secuencia de aminoácidos o de ácido nucleico para un alineamiento óptimo y pueden descargarse secuencias no homólogas para fines de comparación). En determinadas realizaciones, la longitud de una secuencia de referencia alineada con fines de comparación tiene al menos un 30 %, preferentemente al menos un 40 %, más preferentemente al menos un 50 %, 60 %, e incluso más preferentemente al menos un 70 %, 80 %, 90 %, 100 % de la longitud de la secuencia de referencia. Después, se comparan los restos de aminoácidos o los nucleótidos en las posiciones de aminoácidos o las posiciones de nucleótidos correspondientes. Cuando una posición en la primera secuencia está ocupada por el mismo resto de aminoácido o nucleótido en la posición correspondiente en la segunda secuencia, se deduce que las moléculas son idénticas en esa posición.

El porcentaje de identidad entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias, teniendo en cuenta el número de huecos y el tamaño de cada hueco, lo que es necesario introducir para un alineamiento óptimo de las dos secuencias.

La comparación de secuencias y la determinación del porcentaje de identidad entre dos secuencias puede llevarse a cabo usando un algoritmo matemático. En una realización preferida, el porcentaje de identidad entre dos secuencias de aminoácido se determina usando el algoritmo de Needleman y Wunsch (1970, J. Mol. Biol. 48: 444-453) que se ha incorporado en el programa GAP en el paquete de programas informáticos GCG (disponible en <http://www.gcg.com>), usando una matriz Blossum 62 o una matriz PAM250 y una ponderación por hueco de 16, 14, 12, 10, 8, 6 o 4 y una ponderación por longitud de 1, 2, 3, 4, 5 o 6. En otra realización preferida más, se determina el porcentaje de identidad entre dos secuencias de nucleótido usando el programa GAP en el paquete de programas informáticos GCG (disponible de <http://www.gcg.com>), usando una NWSgapdna.CMP una ponderación por hueco de 40, 50, 60, 70 o 80 y una ponderación por longitud de 1, 2, 3, 4, 5 o 6. Un conjunto de parámetros preferido (y el que debe emplearse a menos que se especifique de otro modo) es una matriz de puntuación Blossum 62 con una penalización por hueco de 12, una penalización por extensión de hueco de 4 y una penalización por hueco de desplazamiento de fase de 5.

El porcentaje de identidad entre dos secuencias de aminoácidos o de nucleótidos puede determinarse usando el

algoritmo de E. Meyers y W. Miller (1989, *Cabios*, 4: 11-17) que se ha incorporado en el programa ALIGN (versión 2.0), usando una tabla de ponderación de restos PAM120, una penalización por longitud de hueco de 12 y una penalización por hueco de 4.

- 5 Las secuencias de ácido nucleico y de proteína descritas en el presente documento pueden usarse como "secuencia de consulta" para efectuar una búsqueda contra una base de datos pública para, por ejemplo, identificar otros miembros de la familia o secuencias relacionadas. Dichas búsquedas pueden efectuarse usando los programas NBLAST y XBLAST (versión 2.0) de Altschul, et al., (1990, *J. Mol. Biol.*, 215: 403-10). Las búsquedas de nucleótidos BLAST pueden efectuarse con el programa NBLAST, puntuación = 100, longitud de palabra = 12 para obtener
- 10 secuencias de nucleótidos homólogas a moléculas de ácidos nucleicos de la invención. Las búsquedas de proteínas BLAST pueden efectuarse con el programa XBLAST, puntuación = 50, longitud de palabra = 3 para obtener secuencias de aminoácidos homólogas a moléculas de proteína de la invención. Para obtener alineamientos con huecos con fines de comparación, puede utilizarse Gapped Blast, tal como se describe en Altschul et al., (1997, *Nucleic Acids Res.*, 25: 3389-3402). Cuando se utilizan los programas BLAST y Gapped BLAST, pueden usarse los
- 15 parámetros por defecto de los programas respectivos (por ejemplo, XBLAST y NBLAST).

El término "solubilidad" se refiere a la propiedad de un agente proporcionado en el presente documento para disolverse en un disolvente líquido y formar una solución homogénea. La solubilidad se expresa normalmente como una concentración, ya sea en masa o soluto por unidad de volumen de disolvente (g de soluto por kg de disolvente, g por dl (100 ml), mg/ml, etc.), molaridad, molalidad, fracción molar u otras descripciones similares de la concentración. La cantidad máxima en equilibrio de soluto que puede disolverse por cantidad de disolvente es la solubilidad de ese soluto en ese disolvente en las condiciones especificadas, incluyendo la temperatura, presión, pH y la naturaleza del disolvente. En determinadas realizaciones, la solubilidad se mide a pH fisiológico. En determinadas realizaciones, la solubilidad se mide en agua o un tampón fisiológico, tal como PBS. En determinadas realizaciones, la solubilidad se mide en un fluido biológico (disolvente), tal como sangre o suero. En determinadas realizaciones, la temperatura puede ser aproximadamente la temperatura ambiental (por ejemplo, aproximadamente 20, 21, 22, 23, 24, 25 °C) o aproximadamente la temperatura corporal (37 °C). En determinadas realizaciones, un agente, tal como un fragmento de proteína AARS, tiene una solubilidad de al menos aproximadamente 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25 o 30 mg/ml a temperatura ambiente o a 37 °C.

Una "unión de corte y empalme", tal como se usa en el presente documento, incluye la región en un transcrito de ARNm maduro o el polipéptido codificado donde el extremo 3' de un primer exón se une al extremo 5' de un segundo exón. El tamaño de la región puede variar y puede incluir 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 o más (incluyendo todos los números enteros intermedios de) nucleótidos o restos de aminoácidos a ambos lados de los restos exactos donde el extremo 3' de un exón se une al extremo 5' de otro exón. Un "exón" se refiere a una secuencia de ácido nucleico que está representada en la forma madura de una molécula de ARN después de que ambas porciones de un ARN precursor (intrones) se hayan retirado mediante corte y empalme en cis o de que se hayan ligado dos o más moléculas de ARN precursoras mediante corte y empalme en trans. La molécula de ARN madura puede ser un ARN mensajero o una forma funcional de un ARN no codificante, tal como ARNr o ARNt. Dependiendo del contexto, un exón puede referirse a la secuencia en el ADN o su transcrito de ARN. Un "intrón" se refiere a una región de ácido nucleico no codificante dentro de un gen, que no se traduce en una proteína. Las secciones intrónicas no codificantes se transcriben en precursor de ARNm (pre-ARNm) y algunos otros ARN (tales como ARN largos no codificantes), y posteriormente se retiran mediante corte y empalme durante el procesamiento en ARN maduro.

Una "variante de corte y empalme" se refiere a un ARNm maduro y su proteína codificada que se producen mediante corte y empalme alternativo, un proceso mediante el cual los exones del ARN (un transcrito génico primario o pre-ARNm) se reconectan de diversas maneras durante el corte y empalme del ARN. Los diferentes ARNm resultantes pueden traducirse en diferentes isoformas de proteína, permitiendo que un solo gen codifique múltiples proteínas.

Un "sujeto", tal como se usa en el presente documento, incluye cualquier animal que muestre un síntoma o que se encuentre en riesgo de mostrar un síntoma, que pueda tratarse o diagnosticarse con un polinucleótido o polipéptido de AARS de la invención. También se incluyen sujetos para los cuales es deseable elaborar perfiles de los niveles de polipéptidos y/o polinucleótidos de AARS de la invención, para fines diagnósticos u otros. Los sujetos adecuados (pacientes) incluyen animales de laboratorio (tales como un ratón, rata, conejo o cobaya), animales de granja y animales domésticos o mascotas (tales como un perro o un gato). Se incluyen primates no humanos y preferentemente, pacientes humanos.

"Tratamiento" o "tratar", tal como se usa en el presente documento, incluye cualquier efecto deseable sobre los síntomas o la patología de una enfermedad o afección que pueda efectuarse por las actividades no canónicas de un polinucleótido o polipéptido de AARS, tal como se describe en el presente documento y puede incluir cambios o mejoras incluso mínimas en uno o más marcadores medibles de la enfermedad o afección que se esté tratando. También se incluyen tratamientos que se refieren a terapias no con AARS, en las que una secuencia de AARS descrita en el presente documento proporciona un marcador clínico de tratamiento. "Tratamiento" o "tratar" no indica necesariamente una erradicación o cura completa de la enfermedad o afección o un síntoma asociado a la misma. El

sujeto que recibe este tratamiento es cualquier sujeto que lo necesite. Los marcadores a modo de ejemplo de mejora clínica serán evidentes para las personas expertas en la materia.

La práctica de la presente invención empleará, a menos que se indique específicamente lo contrario, métodos convencionales de biología molecular y técnicas de DN recombinante dentro del nivel de capacidad de la técnica, muchas de las cuales se describen más adelante con fines ilustrativos. Dichas técnicas se explican detalladamente en la bibliografía. Véase, por ejemplo, Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (3ª edición, 2000); *DNA Cloning: A Practical Approach*, vol. I y II (D. Glover, ed.); *Oligonucleotide Synthesis* (N. Gait, ed., 1984); *Oligonucleotide Synthesis: Methods and Applications* (P. Herdewijn, ed., 2004); *Nucleic Acid Hybridization* (B. Hames y S. Higgins, eds., 1985); *Nucleic Acid Hybridization: Modern Applications* (Buzdin y Lukyanov, eds., 2009); *Transcription and Translation* (B. Hames y S. Higgins, eds., 1984); *Animal Cell Culture* (R. Freshney, ed., 1986); Freshney, R.I. (2005) *Culture of Animal Cells, a Manual of Basic Technique*, 5ª ed. Hoboken NJ, John Wiley & Sons; B. Perbal, *A Practical Guide to Molecular Cloning* (3ª edición, 2010); Farrell, R., *RNA Methodologies: A Laboratory Guide for Isolation and Characterization* (3ª edición 2005), *Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA Methods in Enzymology*, Academic Press; *Using Antibodies: A Laboratory Manual: Portable Protocol NO. 1* por Edward Harlow, David Lane, Ed Harlow (1999, Cold Spring Harbor Laboratory Press, ISBN 0-87969-544-7); *Antibodies: A Laboratory Manual* de Ed Harlow (Editor), David Lane (Editor) (1988, Cold Spring Harbor Laboratory Press, ISBN 0-87969-3,4-2), 1855. *Handbook of Drug Screening*, editado por Ramakrishna Seethala, Prabhavathi B. Fernandes (2001, Nueva York, N.Y., Marcel Dekker, ISBN 0-8247-0562-9); y *Lab Ref: A Handbook of Recipes, Reagents, and Other Reference Tools for Use at the Bench*, editado por Jane Roskams y Linda Rodgers, (2002, Cold Spring Harbor Laboratory, ISBN 0-87969-630-3).

Todas las publicaciones, patentes y solicitudes de patente citadas en el presente documento se incorporan al presente documento por referencia en su totalidad.

### III. FRAGMENTOS Y VARIANTES DE PROTEÍNA AARS PURIFICADOS PARA APLICACIONES TERAPÉUTICAS Y OTRAS

Sorprendentemente y a diferencia de sus secuencias parentales de longitud completa que se conocen únicamente por sus actividades de aminoacilación, se ha descubierto que los fragmentos de AARS poseen actividades biológicas importantes para aplicaciones bioterapéuticas, de descubrimiento y diagnósticas. Por lo tanto, las realizaciones de la invención incluyen proteínas de longitud completa, isoformas maduras de proteínas y fragmentos de proteína de aminoacil-ARNt sintetasas (AARS), además de variantes biológicamente activas y fragmentos de las mismas. En determinadas realizaciones, las proteínas y los fragmentos pueden surgir mediante proteólisis endógena, proteólisis *in vitro*, variación de corte y empalme o predicción informatizada, entre otros mecanismos.

Los fragmentos de proteína AARS descritos en el presente documento y las variantes de los mismos, pueden poseer al menos una actividad biológica "no canónica". Los fragmentos de proteína AARS de la presente invención también se citan en el presente documento como "polipéptidos de AARS" o "polipéptidos de referencia de AARS". En determinadas realizaciones, los polipéptidos de AARS proporcionados en el presente documento comprenden o consisten esencialmente en la totalidad o una parte de las "secuencias de referencia" del polipéptido de AARS, tal como se exponen en las tablas 1-3 o las tablas 4-6 o las tablas 7-9 más adelante, que representan las secuencias de aminoácidos de diversos fragmentos de treonil-ARNt sintetasas. Las secuencias de proteína AARS de ratón y humanas están altamente relacionadas, difiriendo normalmente en no más de unos pocos aminoácidos dentro de una secuencia completa, un dominio particular o un fragmento de proteína particular.

#### Polipéptidos de AARS N-terminales: (Tablas 1, 2 y 3)

Tabla 1A Polipéptidos de AARS identificados por EM			
Nombre	Tipo / Especie / Restos	Secuencias de aminoácidos y de ácido nucleico	SEQ ID NO
ThrRS1 <sup>N1</sup>	Proteína / Humano / 1-322	MFEEKASSPSGKMGGEKPIGAGEEKQKEG GKKKNKEGSGDGGRAELNPWPEYIYTRLE MYNILKAEHDSILAEKAEKDSKPIKVTLPDG KQVDAESWKTPYQIACGISQGLADNTVIA KVNNVWDLDRPLEEDCTLELLKFEDEEAO	SEQ ID NO: 12

		AVYWHSSAHIMGEAMERVYGGCLCYGPPIE NGFYIDMYLEEGGVSSNDFSSLEALCKKIIK EKQAFERLEVKKETLLAMFKYNKFKCRILN EKVNTPTTTVYRCGPLIDLRCRPHVRHTGKI KALKIHKNSSTYWEGKADMETLQRIYGISFP DPKMLKEWEKFEAKNRD	
ThrRS1 <sup>N1</sup>	ADN / Humano	ATGTTTGAGGAGAAGGCCAGCAGTCCTTC AGGGAAGATGGGAGGCGAGGAGAAGCCG ATTGGTGCTGGTGAAGAGAAGCAAAGG AAGGAGGCAAAAAGAAGAACAAGAAGG ATCTGGAGATGGAGGTCGAGCTGAGTTGA ATCCTTGGCCTGAATATATTTACACACGTC TTGAGATGTATAATAATACTAAAAGCAGAA CATGATTCCATTCTGGCAGAAAAGGCAGA AAAAGATAGCAAGCCAATTAAGTCACTT TGCCTGATGGTAAACAGGTTGATGCGGAA TCTTGAAAACACTACCCATATCAAATTGC CTGTGGAATTAGTCAAGGCCTGGCCGACA ACACCGTTATTGCTAAAGTAAATAATGTT GTGTGGGACCTGGACCGCCCTCTGGAAGA AGATTGTACCTTGGAGCTTCTCAAGTTTGA GGATGAGGAAGCTCAGGCAGTGTATTGGC ACTCTAGTGCTCACATAATGGGTGAAGCC ATGGAAAGAGTCTATGGTGGATGTTTATG CTACGGTCCGCAATAGAAAATGGATTCT ATTATGACATGTACCTCGAAGAAGGGGGT GTGTCTAGCAATGATTTCTTCTCTGGAG GCTTTGTGTAAGAAAATCATTAAAGAAAA ACAAGCTTTTGAAAGACTGGAAGTTAAGA AAGAACTTTACTGGCAATGTTTAAGTAC ACAAGTTCAAATGCCGGATATTGAATGA AAAGGTGAATACTCCAACCTACCACAGTCT ATAGATGTGGCCCTTTGATAGATCTCTGCC GGGGTCCTCATGTTAGACACACGGGCAAA ATTAAGGCTTTAAAAATACACAAAAATTC CTCCACGTACTGGGAAGGCAAAGCAGATA TGGAGACTCTCCAGAGAATTTATGGCATT TCATTCCCAGATCCTAAAATGTTGAAAGA GTGGGAGAAGTTCCAAGAGGAAGCTAAA AACCGAGAT	SEQ ID NO: 13
ThrRS1 <sup>N2</sup>	Proteína / Humano /	MFEKASSPSGKMGGEKPIGAGEEKQKEG GKKKNKEGSGDGGRAELNPWPEYIYTRLE	SEQ ID NO: 14
	1-146	MYNILKAEHDSILAEKAEKDSKPIKVTLPDG KQVDAESWKTPYQIACGISQGLADNTVIA KVNNVVWDLDRPLEEDCTLELLKFED	

ThrRS1 <sup>N2</sup>	ADN / Humano	<p>ATGTTTGAGGAGAAGGCCAGCAGTCCTTC                  AGGGAAGATGGGAGGCGAGGAGAAGCCG                  ATTGGTGCTGGTGAAGAGAAGCAAAGG                  AAGGAGGCAAAAAGAAGAACAAGAAGG                  ATCTGGAGATGGAGGTCGAGCTGAGTTGA                  ATCCTTGGCCTGAATATATTTACACACGTC                  TTGAGATGTATAATACTAAAAGCAGAA                  CATGATTCCATTCTGGCAGAAAAGGCAGA                  AAAAGATAGCAAGCCAATTAAGTCACTT                  TGCCTGATGGTAAACAGGTTGATGCGGAA                  TCTTGGAATACTACCCATATCAAATTGC                  CTGTGGAATTAGTCAAGGCCTGGCCGACA                  ACACCGTTATTGCTAAAGTAAATAATGTT                  GTGTGGGACCTGGACCGCCCTCTGGAAGA                  AGATTGTACCTTGGAGCTTCTCAAGTTTGA                  GGAT</p>	SEQ ID NO: 15
----------------------	--------------	---	------------------

Tabla 1B ThrRS1 <sup>N1</sup> Péptidos detectados por espectrometría de masas y péptidos enlazadores deducidos		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	MSQEKASSPSGKMDGEK	SEQ ID NO: 16
Proteína / ratón	PVDASEEKRKEGGKKKSKDGGGDGGRAELNPWPEYIN TRLDMYNK	SEQ ID NO: 17
Proteína / ratón	LKAEHDSILAEK	SEQ ID NO: 18

Tabla 1C ThrRS1 <sup>N1</sup> Secuencias concatenadas basadas en péptidos detectados por espectrometría de masas		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína /	MSOEKASSPSGKMDGEKPVDASEEKRKEGGKKKSKDG	SEQ ID
ratón	GGDGGRAELNPWPEYINTRLDMYNKLKAEHDSILAEK	NO.19

Tabla 1D ThrRS1 <sup>N2</sup> Péptidos detectados por espectrometría de masas y péptidos enlazadores deducidos		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	LKAEHDSILAEK	SEQ ID NO: 20
Proteína / ratón	AAKDSKPIKVTLPDGKQVDAESWK	SEQ ID NO: 21
Proteína / ratón	TTPYQIACGISQGLADNTVAK	SEQ ID NO: 22

Tabla 1E ThrRS1 <sup>N2</sup> Secuencias concatenadas basadas en péptidos detectados por espectrometría de masas		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	<u>LKAEHDSILA</u> EKA AKDSKPIKVTLPDGKQVDAESWKT <u>T</u> PYQIACGISOGLADNTVVAK	SEQ ID NO: 23

Tabla 2 Polipéptidos de AARS y transcritos alternativos identificados mediante secuenciación profunda			
Nombre	Tipo / Especie / Restos	Secuencias de aminoácidos y de ácido nucleico	SEQ ID NO
ThrRS1 <sup>N3</sup>	Proteína / Humano / 1-328 + 1aa	MFEEKASSPSGKMGGEEKPIGAGEEKQKE GGKKKNKEGSGDGGRAELNPWPEYIYTRL EMYNILKAEHDSILA EKA EKDSKPIKVTLP DGKQVDAESWKTTPYQIACGISOGLADNT VIAKVNNVVWDLDRPLEEDCTLELLKFED EEAQAVYWHSSAHIMGEAMERVYGGCLC YGPIENGFYDMYLEEGGVSSNDFSSLEA LCKKIIKEKQAFERLEVKKETLLAMFKYNK FKCRILNEKVNTPTTTVYRCGPLIDLCRGP VRHTGKIKALKIHKNSSTYWEGKADMETL QRIYGISFPDPKMLKEWEKFQEEAKNRDHR KIGRP	SEQ ID NO: 24

ThrRS1 <sup>N3</sup>	ADN / Humano	<p>ATGTTTGAGGAGAAGGCCAGCAGTCCTT</p> <p>CAGGGAAGATGGGAGGCGAGGAGAAGC  CGATTGGTGCTGGTGAAGAGAAGCAAAA  GGAAGGAGGCCAAAAGAAGAACAAGA  AGGATCTGGAGATGGAGGTCGAGCTGAG  TTGAATCCTTGGCCTGAATATATTTACAC  ACGTCTTGAGATGTATAATACTAAAAG  CAGAACATGATTCCATTCTGGCAGAAAA  GGCAGAAAAAGATAGCAAGCCAATTTAA  GTCACCTTGCCTGATGGTAAACAGGTTGA  TGCGGAATCTTGGAAAACACACCATATC  AAATTGCCTGTGGAATTAGTCAAGGCCTG  GCCGACAACACCGTTATTGCTAAAGTAA  ATAATGTTGTGTGGGACCTGGACCGCCCT  CTGGAAGAAGATTGTACCTTGGAGCTTCT  CAAGTTTGAGGATGAGGAAGCTCAGGCA  GTGTATTGGCACTCTAGTGCTCACATAAT  GGGTGAAGCCATGGAAAGAGTCTATGGT  GGATGTTTATGCTACGGTCCGCCAATAGA  AAATGGATTCTATTATGACATGTACCTCG  AAGAAGGGGGTGTGTCTAGCAATGATTT  CTCTTCTCTGGAGGCTTTGTGTAAGAAAA  TCATTAAGAAAAACAAGCTTTTGAAAG  ACTGGAAGTTAAGAAAGAAACTTTACTG  GCAATGTTTAAGTACAACAAGTTCAAAT  GCCGGATATTGAATGAAAAGGTGAATAC  TCCAACACCACAGTCTATAGATGTGGCC  CTTTGATAGATCTCTGCCGGGGTCTCAT  GTTAGACACACGGGCAAAATTAAGGCTT  TAAAAATACACAAAAATTCCTCCACGTA  CTGGGAAGGCAAAGCAGATATGGAGACT  CTCCAGAGAATTTATGGCATTTCATTCCC  AGATCCTAAAATGTTGAAAGAGTGGGAG  AAGTTCCAAGAGGAAGCTAAAAACCGAG  ATCATAGGAAAATTGGCAGGCCATGA</p>	SEQ ID NO: 25
ThrRS1 <sup>N4</sup>	Proteína / Humano / 1-109 + 9aa	<p>MFEEKASSPSGKMGGEKPIGAGEEKQKE  GGKKKNKEGSDGGRAELNPWEYIYTRL  EMYNILKAEHDSILAEKAEKDSKPIKVTLF  DGKQVDAESWKTPYQIACGIRCIGTLVLT</p>	SEQ ID NO: 26
ThrRS1 <sup>N4</sup>	ADN / Humano	<p>ATGTTTGAGGAGAAGGCCAGCAGTCCTT</p> <p>CAGGGAAGATGGGAGGCGAGGAGAAGC  CGATTGGTGCTGGTGAAGAGAAGCAAAA  GGAAGGAGGCCAAAAGAAGAACAAGA  AGGATCTGGAGATGGAGGTCGAGCTGAG</p>	SEQ ID NO: 27

		<p>TTGAATCCTTGGCCTGAATATATTTACAC  ACGTCTTGAGATGTATAATACTAAAAG  CAGAACATGATTCCATTCTGGCAGAAAA  GGCAGAAAAAGATAGCAAGCCAATTA  GTCACCTTGCCTGATGGTAAACAGGTTGA  TGC GGAATCTTGGAAA ACTACACCATATC  AAATTGCCTGTGGAATTAGGTGTATTGGC  ACTCTAGTGCTCACATAA</p>	
ThrRS1 <sup>N5</sup>	Proteína / Humano / 1-576 + S577R + 613-723	<p>MFE EKASSPSGKMGGE EKPIGAGE EKQKE  GGKKNKEGSGDGGRAELNPWEYIYTRL  EMYNILKAEHDSILA EKA EKDSKPIKVTLP  DGKQVDAESWKTPYQIACGISQGLADNT  VIAKVNNVVWDLDRPLEEDCTLELLKFED  EEAQAVYWHSSAHIMGEAMERVYGGCLC  YGPPIENGFYDMYLEEGGVSSNDFSSLEA  LCKKIIKEKQAFERLEVKKETLLAMFKYNK  FKCRILNEKVNTPTTTVYRCGPLIDLRCRPH  VRHTGKIKALKIHKNSSTYWEGKADMETL  QRIYGISFPDPKMLKEWEKFQEEAKNRDHR  KIGRDQELYFFHELSPGSCFFLPKGAYTYNA  LIEFIRSEYRKRGFQEVVTPNIFNSRLWMTS  GHWQHYSENMFSEFEVEKELFALKPMNCPG  HCLMFDHRPRSWRELPLRLADFGVLHRNE  LSGALTGLTRVRRFQDDAHIFCAMEQIED  EIKGCLDFLRTVYSVFGFSFKLNLSTRPEKF  LGDIEVWDQAEKQLENSLNEFGEKWELNS  GDGAFYGPKIDIQIKDAIGRYHQCATIQ LDF  QLPIRFNLTYVRPFWLSPRQVMVVPVGP TC  DEYAQKVRQQFHDAKFMADIDLDPGCTLN  KKIRNAQLAQYNFILVVGEKEKISGTVNIRT  RDNKVHGERTISETIERLQQLKEFRSKQAE  EF</p>	SEQ ID NO: 28
ThrRS1 <sup>N5</sup>	ADN / Humano	<p>ATGTTTGAGGAGAAGGCCAGCAGTCCTT  CAGGGAAGATGGGAGGCGAGGAGAAGC  CGATTGGTGCTGGTGAAGAGAAGCAAAA  GGAAGGAGGCAAAAAGAAGAACAAGA  AGGATCTGGAGATGGAGGTCGAGCTGAG  TTGAATCCTTGGCCTGAATATATTTACAC  ACGTCTTGAGATGTATAATACTAAAAG  CAGAACATGATTCCATTCTGGCAGAAAA  GGCAGAAAAAGATAGCAAGCCAATTA  GTCACCTTGCCTGATGGTAAACAGGTTGA  TGC GGAATCTTGGAAA ACTACACCATATC  AAATTGCCTGTGGAATTAGTCAAGGCCTG</p>	SEQ ID NO: 126

	<p>GCCGACAACACCGTTATTGCTAAAGTAA  ATAATGTTGTGTGGGACCTGGACCGCCCT  CTGGAAGAAGATTGTACCTTGGAGCTTCT  CAAGTTTGAGGATGAGGAAGCTCAGGCA  GTGTATTGGCACTCTAGTGCTCACATAAT  GGGTGAAGCCATGGAAAGAGTCTATGGT  GGATGTTTATGCTACGGTCCGCCAATAGA  AAATGGATTCTATTATGACATGTACCTCG  AAGAAGGGGGTGTGTCTAGCAATGATTT  CTCTTCTCTGGAGGCTTTGTGTAAGAAAA  TCATTAAAGAAAAACAAGCTTTTGAAAG  ACTGGAAGTTAAGAAAGAACTTTACTG  GCAATGTTTAAGTACAACAAGTTCAAAT  GCCGGATATTGAATGAAAAGGTGAATAC  TCCAACCTACCACAGTCTATAGATGTGGCC  CTTTGATAGATCTCTGCCGGGGTCCTCAT  GTTAGACACACGGGCAAATTAAGGCTT  TAAAAATACACAAAAATTCCTCCACGTA  CTGGGAAGGCAAAGCAGATATGGAGACT  CTCCAGAGAATTTATGGCATTTCATTCCC  AGATCCTAAAATGTTGAAAGAGTGGGAG  AAGTTCCAAGAGGAAGCTAAAAACCGAG  ATCATAGGAAAATTGGCAGGGACCAAGA  ACTATATTTCTTTCATGAACTCAGCCCTG  GAAGTTGCTTTTTTCTGCCAAAAGGAGCC  TACATTTATAATGCACTTATTGAATTCAT  TAGGAGCGAATATAGGAAAAGAGGATTC  CAGGAGGTAGTCACCCCAAACATCTTCA  ACAGCCGACTCTGGATGACCTCGGGCCA  CTGGCAGCACTACAGCGAGAACATGTTT  TCCTTTGAGGTGGAGAAGGAGCTGTTTGC  CCTGAAACCCATGAACTGCCCAGGACAC  TGCCTTATGTTTGATCATCGGCCAAGGTC  CTGGCGAGAAGTGCCTCTGCGGCTAGCTG  ATTTTGGGGTACTTCATAGGAACGAGCTG  TCTGGAGCACTCACAGGACTCACCCGGG  TACGAAGATTCCAACAGGATGATGCTCA  CATATTCTGTGCCATGGAGCAGATTGAAG  ATGAAATAAAAGGTTGTTTGGATTTTCTA  CGTACGGTATATAGCGTATTTGGATTTTC  TTTTAAACTAAACCTTTCTACTCGCCCGG  AAAAATTCCTTGGAGATATCGAAGTATG  GGATCAAGCTGAGAAACAACCTTGAAAAC  AGTCTGAATGAATTTGGTGAAAAGTGGG</p>	
--	---	--

		<p>AGTTAAACTCTGGAGATGGAGCTTTCTAT                  GGCCCAAAGATTGACATACAGATTAAG                  ATGCGATTGGGCGGTACCACCAGTGTGC                  AACCATCCAGCTGGATTTCAGTTGCCCA                  TCAGATTTAATCTTACTTATGTAAGGCC                  TTTTGGCTGTCCCCTCGCCAGGTAATGGT                  AGTTCCAGTGGGACCAACCTGTGATGAA                  TATGCCCAAAGGTACGACAACAATTCC                  ACGATGCCAAATTCATGGCAGACATTGA                  TCTGGATCCAGGCTGTACATTGAATAAAA                  AGATTCGAAATGCACAGTTAGCACAGTA                  TAACTTCATTTTAGTTGTTGGTGAAAAAG                  AGAAAATCAGTGGCACTGTTAATATCCG                  CACAAGAGACAATAAGGTCCACGGGGAA                  CGCACCATTTCTGAAACTATCGAGCGGCT                  ACAGCAGCTCAAAGAGTTCCGCAGCAAA                  CAGGCAGAAGAAGAATTTTAA</p>	
--	--	---	--

<p><b>Tabla 2B</b>                      Uniones de corte y empalme únicas de polipéptidos de AARS</p>			
Nombre	Tipo / Especie	Secuencias de aminoácidos y ácidos nucleicos en la proximidad de la unión de corte y empalme única	SEQ ID NO
T1-AS02	ADN / Humano /	CCGAGATCATAGGAAAATTGGCAGG CCA TGATGGTGTGATAAGAAAAGG	SEQ ID NO: 29
	Proteína / Humano	RDHRKIGRP	SEQ ID NO: 30
T1-AS04	ADN / Humano /	CATATCAAATTGCCTGTGGAATTAG GTG TATTGGCACTCTAGTGCTCAC	SEQ ID NO: 31
	Proteína / Humano	YQIACGIRCIGTLVLT	SEQ ID NO: 32
T1-AS05	ADN / Humano /	TCAGATTTAATCTTACTTATGTAAG GCCC TTTTGGCTGTCCCCTCGCCAG	SEQ ID NO: 33
	Proteína / Humano	RFNLTYVRPFWLSPRQ	SEQ ID NO: 34

<p><b>Tabla 3</b>                      Polipéptidos y ácidos nucleicos de AARS identificados mediante bioinformática</p>			
Nombre	Tipo / Especie / Restos	Secuencias de aminoácidos y de ácido nucleico	SEQ ID NO

Polipéptidos de AARS C-terminales: (Tablas 4, 5 y 6)

5

<p><b>Tabla 4A</b>                      Polipéptidos de AARS identificados por EM</p>			
Nombre	Tipo / Especie / Restos	Secuencias de aminoácidos y de ácido nucleico	SEQ ID NO

ThrRS1 <sup>C1</sup>	Proteína / Humano / 307-723	<p>MLKEWEKFQEEAKNRDHRKIGRDQELYFF          HELSPGSCFFLPKGAYIYNALIEFIRSEYRKR          GFQEVVTPNIFNSRLWMTSGHWQHYSENM          FSFEVEKELFALKPMNCPGHCLMFDHRPRS          WRELPLRLADFGVLHRNELSGALTGLTRVR          RFQQDDAHIFCAMEQIEDEIKGCLDFLRTVY          SVFGFSFKLNLSTRPEKFLGDIEVWDQAEKQ          LENSLENEFGEKWELNSGDGAFYGPKIDIQIK          DAIGRYHQCATIQLDLDFQLPIRFNLTYVSHDG          DDKKRPVIVHRAILGSVERMIAILTENYGGK          WPFWLSRQVMVVPVGPTEDEYAQKVRQQ          FHDAKFMADIDLDPGCTLNKKIRNAQLAQY          NFILVVGEEKISGTVNIRTRDNKVVHGERTI          SETIERLOOLKEFRSKOAEFEF</p>	SEQ ID NO: 50
ThrRS1 <sup>C1</sup>	ADN / Humano /	<p>ATGTTGAAAGAGTGGGAGAAGTTCCAAG          AGGAAGCTAAAAACCGAGATCATAGGAA          AATTGGCAGGGACCAAGAAGTATATTTCT          TTCATGAACTCAGCCCTGGAAGTTGCTTTT          TTCTGCCAAAAGGAGCCTACATTTATAAT          GCACTTATTGAATTCATTAGGAGCGAATA          TAGGAAAAGAGGATTCCAGGAGGTAGTC          ACCCAAACATCTTCAACAGCCGACTCTG          GATGACCTCGGGCCACTGGCAGCACTACA          GCGAGAACATGTTCTCCTTTGAGGTGGAG          AAGGAGCTGTTTGCCCTGAAACCCATGAA          CTGCCCAGGACACTGCCTTATGTTTGATC          ATCGGCCAAGGTCCTGGCGAGAAGTGCCT          CTGCGGCTAGCTGATTTTGGGGTACTTCA          TAGGAACGAGCTGTCTGGAGCACTCACAG          GACTCACCCGGGTACGAAGATTCCAACAG          GATGATGCTCACATATTCTGTGCCATGGA          GCAGATTGAAGATGAAATAAAAGGTTGTT          TGGATTTTCTACGTACGGTATATAGCGTA          TTTGGATTTTCTTTTAAACTAAACCTTTCT          ACTCGCCCGGAAAAATTCCTTGGAGATAT          CGAAGTATGGGATCAAGCTGAGAAACAA          CTTGAAAACAGTCTGAATGAATTTGGTGA          AAAGTGGGAGTTAAACTCTGGAGATGGA          GCTTTCTATGGCCCAAAGATTGACATACA</p>	SEQ ID NO: 51

		<p>GATTAAGATGCGATTGGGCGGTACCACC  AGTGTGCAACCATCCAGCTGGATTTCCAG  TTGCCATCAGATTTAATCTTACTTATGTA  AGCCATGATGGTGATGATAAGAAAAGGC  CAGTGATTGTTTCATCGAGCCATCTTGGGA  TCAGTGGAAGAATGATTGCTATCCTCAC  AGAAAATATGGGGGCAAATGGCCCTTTT  GGCTGTCCCCTCGCCAGGTAATGGTAGTT  CCAGTGGGACCAACCTGTGATGAATATGC  CCAAAAGGTACGACAACAATTCCACGATG  CCAAATTCATGGCAGACATTGATCTGGAT  CCAGGCTGTACATTGAATAAAAAGATTTCG  AAATGCACAGTTAGCACAGTATAACTTCA  TTTTAGTTGTTGGTGAAAAAGAGAAAATC  AGTGGCACTGTTAATATCCGCACAAGAGA  CAATAAGGTCCACGGGGAACGCACCATTT  CTGAAACTATCGAGCGGCTACAGCAGCTC  AAAGAGTTCGCGCAGCAAACAGGCAGAAG  AAGAATTTTAA</p>	
ThrRS1 <sup>cz</sup>	Proteína / Humano / 346-723	<p>LPKGAYIYNALIEFIRSEYRKRGFQEVVTPNI  FNSRLWMTSGHWQHYSENMFSEVEKELF  ALKPMNCPGHCLMFDHRPRSWRELPLRLA  DFGVLHRNELSGALTGLTRVRRFQQDDAHI  FCAMEQIEDEIKGCLDFLRTVYSVFGFSFKL  NLSTRPEKFLGDIEVWDQAEKQLENSLNEF  GEKWELNSGDGAFYGPKIDIQIKDAIGRYH  QCATIQLDFQLPIRFNLTYVSHDGGDKKRPV  IVHRAILGSVERMIAILTENYGGKWPFWLSP  RQVMVVPVGPTEDEYAQKVRQQFHDAKF  MADIDLDPGCTLNKKIRNAQLAQYNFILVV  GEKEKISGTVNIRTRDNKVHGERTISETIERL  OOLKEFRSKOAEFEF</p>	SEQ ID NO: 52
ThrRS1 <sup>cz</sup>	ADN / Humano /	<p>CTGCCAAAAGGAGCCTACATTTATAATGC  ACTTATTGAATTCATTAGGAGCGAATATA  GGAAAAGAGGATTCCAGGAGGTAGTCAC  CCCAAACATCTTCAACAGCCGACTCTGGA  TGACCTCGGGCCACTGGCAGCACTACAGC  GAGAACATGTTCTCCTTTGAGGTGGAGAA  GGAGCTGTTTGCCCTGAAACCCATGAACT  GCCAGGACACTGCCTTATGTTTGATCAT  CGGCCAAGGTCCTGGCGAGAACTGCCTCT  GCGGCTAGCTGATTTTGGGGTACTTCATA  GGAACGAGCTGTCTGGAGCACTCACAGG  ACTCACCCGGGTACGAAGATTCCAACAGG</p>	SEQ ID NO: 53

		<p>ATGATGCTCACATATTCTGTGCCATGGAG  CAGATTGAAGATGAAATAAAAGGTTGTTT  GGATTTTCTACGTACGGTATATAGCGTAT  TTGGATTTTCTTTTAAACTAAACCTTTCTA  CTCGCCCGGAAAAATTCCTTGGAGATATC  GAAGTATGGGATCAAGCTGAGAAACAAC  TTGAAAACAGTCTGAATGAATTTGGTGAA  AAGTGGGAGTTAAACTCTGGAGATGGAG  CTTTCTATGGCCCAAAGATTGACATACAG  ATTAAAGATGCGATTGGGCGGTACCACCA  GTGTGCAACCATCCAGCTGGATTTCAGT  TGCCCATCAGATTTAATCTTACTTATGTAA  GCCATGATGGTGATGATAAGAAAAGGCC  AGTGATTGTTTCATCGAGCCATCTTGGGAT  CAGTGGAAGAATGATTGCTATCCTCACA  GAAAACATGGGGGCAAATGGCCCTTTTG  GCTGTCCCCTCGCCAGGTAATGGTAGTTC  CAGTGGGACCAACCTGTGATGAATATGCC  CAAAGGTACGACAACAATTCCACGATGC  CAAATTCATGGCAGACATTGATCTGGATC  CAGGCTGTACATTGAATAAAAAGATTCGA  AATGCACAGTTAGCACAGTATAACTTCAT  TTAGTTGTTGGTGAAAAAGAGAAAATCA  GTGGCACTGTTAATATCCGCACAAGAGAC  AATAAGGTCCACGGGGAACGCACCATTTC  TGAAACTATCGAGCGGCTACAGCAGCTCA  AAGAGTTCCGCAGCAAACAGGCAGAAGA  AGAATTTTAA</p>	
ThrRS1 <sup>C3</sup>	Proteína / Humano / 432-723	<p>LRLADFGVLHRNELSGALTGLTRVRRFQQD  DAHIFCAMEQIEDEIKGLDFLRTVYSVFGF  SFKLNLSRPEKFLGDIEVWDQAEKQLENSL  NEFGEKWELNSGDGAFYGPKIDIQIKDAIGR  YHQCATIQLDFQLPIRFNLTYVSHDGDGDKK  RPVIVHRAILGSVERMIAILTENYGGKWPFW  LSPRQVMVVPVGPTCDEYAQKVRQQFHDA  KFMADIDLDPGCTLNKKIRNAQLAQYNFIL  VVGEKEKISGTVNIRTRDNKVHGERTISETIE  RLOOLKEFRSKOAEIEEF</p>	SEQ ID NO: 54
ThrRS1 <sup>C3</sup>	ADN / Humano /	<p>CTGCGGCTAGCTGATTTTGGGGTACTTCA  TAGGAACGAGCTGTCTGGAGCACTCACAG  GACTCACCCGGGTACGAAGATTCCAACAG  GATGATGCTCACATATTCTGTGCCATGGA  GCAGATTGAAGATGAAATAAAAGGTTGTT  TGGATTTTCTACGTACGGTATATAGCGTA</p>	SEQ ID NO: 55

		<p>TTTGGATTTTCTTTTAAACTAAACCTTTCT  ACTCGCCCGGAAAAATTCCTTGGAGATAT  CGAAGTATGGGATCAAGCTGAGAAACAA  CTTGAAAACAGTCTGAATGAATTTGGTGA  AAAGTGGGAGTTAAACTCTGGAGATGGA  GCTTTCTATGGCCCAAAGATTGACATACA  GATTAAGATGCGATTGGGCGGTACCACC  AGTGTGCAACCATCCAGCTGGATTTCCAG  TTGCCCATCAGATTTAATCTTACTTATGTA  AGCCATGATGGTGATGATAAGAAAAGGC  CAGTGATTGTTTCATCGAGCCATCTTGGGA  TCAGTGGAAAGAATGATTGCTATCCTCAC  AGAAAATATGGGGGCAAATGGCCCTTTT  GGCTGTCCCCTCGCCAGGTAATGGTAGTT  CCAGTGGGACCAACCTGTGATGAATATGC  CCAAAAGGTACGACAACAATTCCACGATG  CCAAATTCATGGCAGACATTGATCTGGAT  CCAGGCTGTACATTGAATAAAAAGATTCG  AAATGCACAGTTAGCACAGTATAACTTCA  TTTTAGTTGTTGGTGAAAAGAGAAAATC  AGTGGCACTGTTAATATCCGCACAAGAGA  CAATAAGGTCCACGGGGAACGCACCATTT  CTGAAACTATCGAGCGGCTACAGCAGCTC  AAAGAGTTCGCGAGCAAACAGGCAGAAG  AAGAATTTTAA</p>	
ThrRS1 <sup>C4</sup>	Proteína / Humano / 380-723	<p>SRLWMTSGHWQHYSNMFSFEVEKELFAL  KPMNCPGHCLMFDHRPRSWRELPLRLADF  GVLHRNELSGALTGLTRVRRFQDDAHIFC  AMEQIEDEIKGCLDFLRTVYSVFGFSKLN  STRPEKFLGDIEVWDQAEKQLENSLNEFGE  KWELNSGDGAFYGPKIDIQIKDAIGRYHQC  ATIQLDFQLPIRFNLTYVSHDGDGDKKRPVIV  HRAILGSVERMIAILTENYGGKWPFWLSPR  QVMVVPVGPTCDEYAQKVRQQFHDAKFM  ADIDLDPGCTLNKKIRNAQLAQYNFILVVG  EKEKISGTVNIRTRDNKVHGERTISETIERLQ  OLKEFRSKOAEFEF</p>	SEQ ID NO: 56
ThrRS1 <sup>C4</sup>	ADN / Humano /	<p>AGCCGACTCTGGATGACCTCGGGCCACTG  GCAGCACTACAGCGAGAACATGTTCTCCT  TTGAGGTGGAGAAGGAGCTGTTTGCCCTG  AAACCCATGAACTGCCAGGACACTGCCT  TATGTTTGATCATCGGCCAAGGTCCTGGC  GAGAACTGCCTCTGCGGCTAGCTGATTTT  GGGGTACTTCATAGGAACGAGCTGTCTGG</p>	SEQ ID NO: 57

		<p>AGCACTCACAGGACTCACCCGGGTACGAA  GATTCCAACAGGATGATGCTCACATATTC  TGTGCCATGGAGCAGATTGAAGATGAAAT  AAAAGGTTGTTTGGATTTTCTACGTACGG  TATATAGCGTATTTGGATTTTCTTTTAAAC  TAAACCTTTCTACTCGCCCGGAAAAATTC  CTTGGAGATATCGAAGTATGGGATCAAGC  TGAGAAACAACCTTGAAAACAGTCTGAATG  AATTTGGTGAAAAGTGGGAGTTAAACTCT  GGAGATGGAGCTTTCTATGGCCCAAAGAT  TGACATACAGATTAAAGATGCGATTGGGC  GGTACCACCAGTGTGCAACCATCCAGCTG  GATTTCCAGTTGCCCATCAGATTTAATCTT  ACTTATGTAAGCCATGATGGTGATGATAA  GAAAAGGCCAGTGATTGTTTCATCGAGCCA  TCTTGGGATCAGTGGAAAGAATGATTGCT  ATCCTCACAGAAAACATGGGGGCAAATG  GCCCTTTTGGCTGTCCCCTCGCCAGGTAAT  GGTAGTTCAGTGGGACCAACCTGTGATG  AATATGCCCAAAGGTACGACAACAATTC  CACGATGCCAAATTCATGGCAGACATTGA  TCTGGATCCAGGCTGTACATTGAATAAAA  AGATTGCAAATGCACAGTTAGCACAGTAT  AACTTCATTTTAGTTGTTGGTGAAAAAGA  GAAAATCAGTGGCACTGTTAATATCCGCA  CAAGAGACAATAAGGTCCACGGGGAACG  CACCATTTCTGAAACTATCGAGCGGCTAC  AGCAGCTCAAAGAGTTCCGCAGCAAACA  GGCAGAAGAAGAATTTTAA</p>	
<p>ThrRS1<sup>CS</sup></p>	<p>Proteína / Humano /  294-723</p>	<p>LQRIYGISFPDPKMLKEWEKFQEEAKNRDH  RKIGRDQELYFFHELSPGSCFFLPKGAYIYN  ALIEFIRSEYRKRGFQEVVTPNIFNSRLWMT  SGHWQHYSNMFSFEVEKELFALKPMNCP  GHCLMFDHRPRSWRELPLRLADFGVLHRN  ELSGALTGLTRVRRFQDDAHIFCAMEQIE  DEIKGCLDFLRTVYSVFGFSFKLNLSTRPEK  FLGDIEVWDQAEKQLENSLNEFGEKWELNS  GDGAFYGPKIDIQIKDAIGRYHQCATIQLDF  QLPIRFNLTIVSHDGDGDDKKRPVIVHRAILGS  VERMIAILTENYGGKWPFWLSPRQVMVVP  VGPTCDEYAQKVRQQFHDAKFMADIDLDP  GCTLNKKIRNAQLAQYNFILVVGEKEKISGT  VNIRTRDNKVHGERTISETIERLQQLKEFRS  KOAEEEF</p>	<p>SEQ ID  NO: 58</p>

ThrRS1 <sup>CS</sup>	ADN / Humano /	<p>CTCCAGAGAATTTATGGCATTTCATTCCC  AGATCCTAAAATGTTGAAAAGAGTGGGAG  AAGTTCCAAGAGGAAGCTAAAAACCGAG  ATCATAGGAAAATTGGCAGGGACCAAGA  ACTATATTTCTTTCATGAACTCAGCCCTGG  AAGTTGCTTTTTTCTGCCAAAAGGAGCCT  ACATTTATAATGCACTTATTGAATTCATTA  GGAGCGAATATAGGAAAAGAGGATTCCA  GGAGGTAGTCACCCCAAACATCTTCAACA  GCCGACTCTGGATGACCTCGGGCCACTGG  CAGCACTACAGCGAGAACATGTTCTCCTT  TGAGGTGGAGAAGGAGCTGTTTGCCCTGA  AACCCATGAACTGCCCAGGACACTGCCTT  ATGTTTGATCATCGGCCAAGGTCCTGGCG  AGAACTGCCTCTGCGGCTAGCTGATTTTG  GGTACTTCATAGGAACGAGCTGTCTGGA  GCACTCACAGGACTCACCCGGGTACGAAG  ATTCCAACAGGATGATGCTCACATATTCT  GTGCCATGGAGCAGATTGAAGATGAAAT  AAAAGGTTGTTTGGATTTTCTACGTACGG  TATATAGCGTATTTGGATTTTCTTTTAAAC  TAAACCTTTCTACTCGCCCGGAAAAATTC  CTTGGAGATATCGAAGTATGGGATCAAGC  TGAGAAACAACCTTGAAAACAGTCTGAATG  AATTTGGTGAAAAGTGGGAGTTAACTCT  GGAGATGGAGCTTTCTATGGCCCAAAGAT  TGACATACAGATTAAGATGCGATTGGGC  GGTACCACCAGTGTGCAACCATCCAGCTG  GATTTCCAGTTGCCCATCAGATTTAATCTT  ACTTATGTAAGCCATGATGGTGATGATAA  GAAAAGGCCAGTGATTGTTTCATCGAGCCA  TCTTGGGATCAGTGGAAAGAATGATTGCT  ATCCTCACAGAAAATATGGGGGCAAATG  GCCCTTTGGCTGTCCCCTCGCCAGGTAAT  GGTAGTTCAGTGGGACCAACCTGTGATG  AATATGCCCAAAGGTACGACAACAATTC  CACGATGCCAAATTCATGGCAGACATTGA  TCTGGATCCAGGCTGTACATTGAATAAAA  AGATTGCAAATGCACAGTTAGCACAGTAT  AACTTCATTTAGTTGTTGGTGAAAAAGA  GAAAATCAGTGGCACTGTTAATATCCGCA  CAAGAGACAATAAGGTCCACGGGGAACG  CACCATTTCTGAAACTATCGAGCGGCTAC  AGCAGCTCAAAGAGTTCCGCAGCAAACA</p>	SEQ ID NO: 59
		GGCAGAAGAAGAATTTTAA	

<b>Tabla 4B</b> <b>ThrRS1<sup>C1</sup></b> <b>Péptidos detectados por espectrometría de masas y péptidos enlazadores deducidos</b>		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	DQELYFFHELSPGSCFFLPK	SEQ ID NO: 60
Proteína / ratón	GAYIYNTLMEFIR	SEQ ID NO: 61
Proteína / ratón	SEYRKRGFQEVVTPNIFNSRLWMTSGHWQHYSENMFSEVEKEQFALKPMNCPGHCLMFDHRPRSWRELPLRLADFGVLHR	SEQ ID NO: 62
Proteína / ratón	NELSGALTGLTR	SEQ ID NO: 63
Proteína / ratón	VRR	SEQ ID NO: 64
Proteína / ratón	FQQDDAHIFCAMEQIEDEIK	SEQ ID NO: 65

<b>Tabla 4C</b> <b>ThrRS1<sup>C1</sup></b> <b>Secuencias concatenadas basadas en péptidos detectados por espectrometría de masas</b>		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	<u>DQELYFFHELSPGSCFFLPK</u> GAYIYNTLMEFIRSEYRKRGFQEVVTPNIFNSRLWMTSGHWQHYSENMFSEVEKEQFALKPMNCPGHCLMFDHRPRSWRELPLRLADFGVLHRNELSGALTGLTRVRRFOODDAHIFCAMEOIEDEIK	SEQ ID NO: 66

<b>Tabla 4D</b> <b>ThrRS1<sup>C2</sup></b> <b>Péptidos detectados por espectrometría de masas y péptidos enlazadores deducidos</b>		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	GAYIYNTLMEFIR	SEQ ID NO: 67
Proteína / ratón	SEYRKRGFQEVVTPNIFNSRLWMTSGHWQHYSENMFSEVEKEQFALKPMNCPGHCLMFDHRPRSWRELPLRLADFGVLHRNELSGALTGLTRVRRFOODDAHIFCAMEQIEDEIKGCLDFLRTVYSVFGFSFKLNLSTRPEKFLGDIEIWNQAEKQLENSLNEFGEKWE LNPGDGAFYGP KIDIQIKDAIGRYHQCATIQ LDFQLPIRFNLTYVSHDGD DKKR	SEQ ID NO: 68
Proteína / ratón	PVIVHRAILGSVERMIAILTENYGG	SEQ ID NO: 69

<b>Tabla 4E</b> <b>ThrRS1<sup>C2</sup></b> <b>Secuencias concatenadas basadas en péptidos detectados por espectrometría de masas</b>		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	<u>GAYIYN</u> <u>TLMEFIRSEYRKR</u> <u>GFQEVVTPNIFNSRLWMTSG</u> HWQHYSNMFSFEVEKEQFALKPMNCPGHCLMFDHRP RSWRELPLRLADFGVLHRNELSGALTGLTRVRRFQDD AHIFCAMEQIEDEIKGCLDFLRTVYSVFGFSFKLNLSTRP EKFLGDIEIWNQAEKQLENSLNEFGEKWELNPGDGAFY GPKIDIQIKDAIGRYHQCATIQDFQLPIRFNLTYVSHDG DDKKRPVIVHRAILGSVERMIAILTENYGG	SEQ ID NO: 70

<b>Tabla 4F</b> <b>ThrRS1<sup>C3</sup></b> <b>Péptidos detectados por espectrometría de masas y péptidos enlazadores deducidos</b>		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	NELSGALTGLTR	SEQ ID NO: 71
Proteína / ratón	VRRFQDDAHIFCAMEQIEDEIKGCLDFLRTVYSVFGFS FKLNLSTRPEKFLGDIEIWNQAEKQLENSLNEFGEKWEL NPGDGAFYGPKIDIQIKDAIGRYHQCATIQDFQLPIRFN LTYVSHDGDDKKRPVIVHRAILGSVERMIAILTENYGGK WPFWLSR	SEQ ID NO: 72
Proteína / ratón	QVMVVPVGPTCDEYAQK	SEQ ID NO: 73

<b>Tabla 4G</b> <b>ThrRS1<sup>C3</sup></b> <b>Secuencias concatenadas basadas en péptidos detectados por espectrometría de masas</b>		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	<u>NELSGALTGLTRVRRFQDDAHIFCAMEQIEDEIKGCLD</u> <u>FLRTVYSVFGFSFKLNLSTRPEKFLGDIEIWNQAEKQLE</u> <u>NSLNEFGEKWELNPGDGAFYGPKIDIQIKDAIGRYHQCA</u> <u>TIQLDFQLPIRFNLTYVSHDGDDKKRPVIVHRAILGSVER</u> <u>MIAILTENYGGKWPFWLSRQVMVVPVGPTCDEYAQK</u>	SEQ ID NO: 74

<b>Tabla 4H</b> <b>ThrRS1<sup>C4</sup></b> <b>Péptidos detectados por espectrometría de masas y péptidos enlazadores deducidos</b>		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	NELSGALTGLTR	SEQ ID NO: 75

Proteína / ratón	VRRFQQDDAHIFCAMEQIEDEIKGCLDFLRTVYVSVFGFS FKLNLSTRPEKFLGDIEIWNQAEKQLENSLNEFGEKWEL NPGDGAFYGPKIDIQIKDAIGRYHQCATIQLDFQLPIRFN LTYVSHDGDDKKRPVIVHRAILGSVER	SEQ ID NO: 76
Proteína / ratón	MIAILTENYGGK	SEQ ID NO: 77

<b>Tabla 4I</b> <b>ThrRS1<sup>C4</sup></b> <b>Secuencias concatenadas basadas en péptidos detectados por espectrometría de masas</b>		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	<u>NELSGALTGLTR</u> VRRFQQDDAHIFCAMEQIEDEIKGCLD FLRTVYVSVFGFSFKLNLSTRPEKFLGDIEIWNQAEKQLE NSLNEFGEKWELNPGDGAFYGPKIDIQIKDAIGRYHQCA TIQLDFQLPIRFNLTYVSHDGDDKKRPVIVHRAILGSVER MIAILTENYGGK	SEQ ID NO: 78

<b>Tabla 4J</b> <b>ThrRS1<sup>C5</sup></b> <b>Péptidos detectados por espectrometría de masas y péptidos enlazadores deducidos</b>		
Tipo / Especie	Secuencia	SEQ ID NO
Proteína / ratón	IYGISFPDPK	SEQ ID NO: 79
Proteína / ratón	LLKEWEKFQEEAKNRDHRKIGRDQELYFFHELSPGSCFF LPKGAYIYNTLMEFIRSEYRKR	SEQ ID NO: 80
Proteína / ratón	GFQEVVTPNIFNSR	SEQ ID NO: 81
Proteína / ratón	LWMTSGHWQHYSNMFSFEVEKEQFALKPMNCPGHCL MFDHRPRSWRELPLRLADFGVLHRNELSGALTGLTRVR RFQDDAHIFCAMEQIEDEIKGCLDFLRTVYVSVFGFSFK LNLSTRPEKFLGDIEIWNQAEKQLENSLNEFGEKWELN PGDGAFYGPKIDIQIKDAIGRYHQCATIQLDFQLPIRFNLT YVSHDGDDKKRPVIVHRAILGSVER	SEQ ID NO: 82
Proteína / ratón	MIAILTENYGGK	SEQ ID NO: 83

<b>Tabla 4K</b> <b>ThrRS1<sup>C5</sup></b> <b>Secuencias concatenadas basadas en péptidos detectados por espectrometría de masas</b>		
Tipo / Especie	Secuencia	SEQ ID NO

Proteína / ratón	<p><u>IYGISFPDPKLLKEWEKFQEEAKNRDHRKIGRDQELYFF</u>  <u>HELSPGSCFFLPKGAYIYNLMEFIRSEYRKRGFQEVVT</u>  <u>PNIFNSRLWMTSGHWQHYSNMFSFEVEKEQFALKPM</u>  <u>NCPGHCLMFDHRPRSWRELPLRLADFGVLHRNELSGAL</u>  <u>TGLTRVRRFQQDDAHIFCAMEQIEDEIKGCLDFLRTVYS</u>  <u>VFGFSFKLNLSTRPEKFLGDIEIWNQAEKQLENSLNEFG</u>  <u>EKWELNPGDGAIFYGPKIDIQIKDAIGRYHQCATIQLDFQ</u>  <u>LPIRFNLTYVSHDGGDKKRPVIVHRAILGSVERMIAILTE</u>  <u>NYGGK</u></p>	SEQ ID NO: 84
------------------	--	------------------

<p align="center"><b>Tabla 5</b>  <b>Polipéptidos de AARS y transcritos alternativos identificados mediante secuenciación profunda</b></p>			
Nombre	Tipo / Especie / Restos	Secuencias de aminoácidos y de ácido nucleico	SEQ ID NO
ThrRS1 <sup>C6</sup>	Proteína / Humano / 600-723	<p>MIAILTENYGGKWPFWLSRQVMVVPVGPT  CDEYAQKVRQQFHDAKFMADIDLDPGCTL  NKKIRNAQLAQYNFILVVGEKEKISGTVNIR  TRDNKVHGERTISETIERLQQLKEFRSKQAE  EEF</p>	SEQ ID NO: 85
ThrRS1 <sup>C6</sup>	ADN / Humano	<p>ATGATTGCTATCCTCACAGAAAACATGG  GGGCAAATGGCCCTTTTGGCTGTCCCCTC  GCCAGGTAATGGTAGTTCAGTGGGACCA  ACCTGTGATGAATATGCCAAAAGGTACG  ACAACAATTCCACGATGCCAAATTCATGG  CAGACATTGATCTGGATCCAGGCTGTACA</p>	SEQ ID NO: 86
		<p>TTGAATAAAAAGATTTCGAAATGCACAGTT  AGCACAGTATAACTTCATTTTAGTTGTTGG  TGAAAAAGAGAAAATCAGTGGCACTGTTA  ATATCCGCACAAGAGACAATAAGGTCCAC  GGGGAACGCACCATTTCTGAAACTATCGA  GCGGCTACAGCAGCTCAAAGAGTTCCGCA  GCAAACAGGCAGAAGAAGAATTTTAA</p>	

<p>ThrRS1<sup>cr</sup></p>	<p>Proteína / Humano / 1-19 + 47-723</p>	<p>MFE EKASSPSGKMG GEEKPLNPWPEYIYTR          LEMYNILKAEHDSILAEKAEKDSKPIKVTLP          DGKQVDAESWKTPPYQIACGISQGLADNTV          IAKVNNVVDLDRPLEEDCTLELLKFEDEE          AQAVYWHSSAHIMGEAMERVYGGCLCYGP          PIENGFYYDMYLEEGGVSSNDFSSLEALCKK          IIKEKQAFERLEVKKETLLAMFKYNKFKCRI          LNEKVNTPTTTVYRCGPLIDLRCRGPVHRHT          GKIKALKIHKNSSTYWEGKADMETLQRIYGI          SFPDPKMLKEWEKFQEEAKNRDHRKIGRDQ          ELYFFHELSPGSCFFLPKGAYTYNALIEFIRSE          YRKRGFQEVVTPNIFNSRLWMTSGHWQHY          SENMFSFEVEKELFALKPMNCPGHCLMFDH          RPRSWRELPLRLADFGVLHRNELSGALTGL          TRVRRFQQDDAHIFCAMEQIEDEIKGCLDFL          RTVYSVFGFSFKLNLSTRPEKFLGDIEVWDQ          AEKQLENSLNEFGEKWELNSGDGAFYGPKI          DIQIKDAIGRYHQCATIQLDLDFQLPIRFNLTYV          SHDGGDDKKRPVIVHRAILGSVERMIAILTEN          YGGKWPFWLSRQVMVVPVGPTCDEYAQK          VRQQFHDAKFMADIDLDPGCTLNKKIRNAQ          LAQYNFILVVGEKEKISGTVNIRTRDNKVHG          ERTISETIERLOOLKEFRSKOAE EEF</p>	<p>SEQ ID NO: 87</p>
<p>ThrRS1<sup>cr</sup></p>	<p>ADN / Humano</p>	<p>ATGTTTGAGGAGAAGGCCAGCAGTCCTTC          AGGGAAGATGGGAGGCGAGGAGAAGCCG          TTGAATCCTTGGCCTGAATATATTTACACA          CGTCTTGAGATGTATAATACTAAAAGC          AGAACATGATTCCATTCTGGCAGAAAAGG          CAGAAAAAGATAGCAAGCCAATTAAGT          CACTTTGCCTGATGGTAAACAGGTTGATG          CGGAATCTTGGA AACTACACCATATCAA          ATTGCCTGTGGAATTAGTCAAGGCCTGGC          CGACAACACCGTTATTGCTAAAGTAAATA          ATGTTGTGTGGGACCTGGACCGCCCTCTG          GAAGAAGATTGTACCTTGGAGCTTCTCAA          GTTTGAGGATGAGGAAGCTCAGGCAGTGT          ATTGGCACTCTAGTGCTCACATAATGGGT</p>	<p>SEQ ID NO: 88</p>

		<p>GAAGCCATGGAAAGAGTCTATGGTGGATG          TTTATGCTACGGTCCGCCAATAGAAAATG          GATTCTATTATGACATGTACCTCGAAGAA          GGGGGTGTGTCTAGCAATGATTTCTCTTCT          CTGGAGGCTTTGTGTAAGAAAATCATTAA          AGAAAAACAAGCTTTTGAAAGACTGGAA          GTTAAGAAAGAACTTTACTGGCAATGTT          TAAGTACAACAAGTTCAAATGCCGGATAT          TGAATGAAAAGGTGAATACTCCAACCTACC          ACAGTCTATAGATGTGGCCCTTTGATAGA          TCTCTGCCGGGGTCCCTCATGTTAGACACA          CGGGCAAATTAAGGCTTTAAAAATACAC          AAAAATTCCTCCACGTACTGGGAAGGCAA          AGCAGATATGGAGACTCTCCAGAGAATTT          ATGGCATTTCATTCCCAGATCCTAAAATGT          TGAAAGAGTGGGAGAAGTTCCAAGAGGA          AGCTAAAACCGAGATCATAGGAAAATTG          GCAGGGACCAAGAACTATATTTCTTTTCAT          GAACTCAGCCCTGGAAGTTGCTTTTTTTCTG          CCAAAGGAGCCTACATTTATAATGCACT          TATTGAATTCATTAGGAGCGAATATAGGA          AAAGAGGATTCCAGGAGGTAGTCACCCCA          AACATCTTCAACAGCCGACTCTGGATGAC          CTCGGGCCACTGGCAGCACTACAGCGAGA          ACATGTTCTCCTTTGAGGTGGAGAAGGAG          CTGTTTGCCCTGAAACCCATGAACTGCC          AGGACTGCCTTATGTTTGATCATCGGC          CAAGGTCCTGGCGGAGAACTGCCTCTGCGG          CTAGCTGATTTTGGGGTACTTCATAGGAA          CGAGCTGTCTGGAGCACTCACAGGACTCA          CCCGGGTACGAAGATTCCAACAGGATGAT          GCTCACATATTCTGTGCCATGGAGCAGAT          TGAAGATGAAATAAAAGGTTGTTTGGATT          TTCTACGTACGGTATATAGCGTATTTGGAT          TTTCTTTTAAACTAAACCTTTCTACTCGCC          CGGAAAATTCCTTGGAGATATCGAAGTA          TGGGATCAAGCTGAGAAACAACCTTGAAAA          CAGTCTGAATGAATTTGGTGAAAAGTGGG          AGTTAAACTCTGGAGATGGAGCTTTCTAT          GGCCCAAAGATTGACATACAGATTAAAGA          TGCGATTGGGCGGTACCACCAGTGTGCAA          CCATCCAGCTGGATTTCCAGTTGCCCATCA          GATTTAATCTTACTTATGTAAGCCATGATG          GTGATGATAAGAAAAGGCCAGTGATTGTT</p>	
--	--	--	--

		<p>CATCGAGCCATCTTGGGATCAGTGGAAAG  AATGATTGCTATCCTCACAGAAAACATG  GGGGCAAATGGCCCTTTTGGCTGTCCCCT  CGCCAGGTAATGGTAGTTCCAGTGGGACC  AACCTGTGATGAATATGCCCAAAGGTAC  GACAACAATTCCACGATGCCAAATTCATG  GCAGACATTGATCTGGATCCAGGCTGTAC  ATTGAATAAAAAGATTTCGAAATGCACAGT  TAGCACAGTATAACTTCATTTTAGTTGTTG  GTGAAAAAGAGAAAATCAGTGGCACTGTT  AATATCCGCACAAGAGACAATAAGGTCCA  CGGGGAACGCACCATTCTGAAACTATCG  AGCGGCTACAGCAGCTCAAAGAGTCCCGC  AGCAAACAGGCAGAAGAAGAATTTTAA</p>	
ThrRS1 <sup>C8</sup>	Proteína / Humano / 161-723	<p>MGEAMERVYGGCLCYGPPIENGFYYDMYL  EEGGVSSNDFSSLEALCKKIIKEKQAFERLE  VKKETLLAMFKYNKFKCRILNEKVNTPTTT  VYRCGPLIDLRCRPHVRHTGKIKALKIHKNS  STYWEGKADMETLQRIYGISFPDPKMLKEW  EKFQEEAKNRDHRKIGRDQELYFFHELSPGS  CFFLPKGAYIYNALIEFIRSEYRKRGFQEVVT  PNIFNSRLWMTSGHWQHYSNMFSFEVEKE  LFALKPMNCPGHCLMFDHRPRS WRELPLRL  ADFGVLHRNELSGALTGLTRVRRFQDDAH  IFCAMEQIEDEIKGCLDFLRTVYSVFGFSFKL  NLSTRPEKFLGDIEVWDQAEKQLENSLNEF  GEKWELNSGDGAFYGPKIDIQIKDAIGRYHQ  CATIQDFQLPIRFNLTYVSHDGDGDKRPVI  VHRAILGSVERMIAILTENYGGKWPFWLSPR  QVMVVPVGPTCDEYAQKVRQQFHDAKFM  ADIDLDPGCTLNKKIRNAQLAQYNFILVVGE  KEKISGTVNIRTRDNKVHGERTISETIERLQQ  LKEFRSKOAEFEF</p>	SEQ ID NO: 89
ThrRS1 <sup>C8</sup>	ADN / Humano	<p>ATGGGTGAAGCCATGGAAAGAGTCTATGG  TGGATGTTTATGCTACGGTCCGCCAATAG  AAAATGGATTCTATTATGACATGTACCTC  GAAGAAGGGGGTGTGTCTAGCAATGATT  CTCTTCTCTGGAGGCTTTGTGTAAGAAAAT  CATTAAAGAAAAACAAGCTTTTGAAAGAC  TGGAAGTTAAGAAAGAACTTTACTGGCA  ATGTTTAAGTACAACAAGTTCAAATGCCG  GATATTGAATGAAAAGGTGAATACTCCAA  CTACCACAGTCTATAGATGTGGCCCTTTG  ATAGATCTCTGCCGGGGTCCTCATGTTAG</p>	SEQ ID NO: 90

		<p>ACACACGGGCAAATTAAGGCTTTAAAAA  TACACAAAAATTCCTCCACGTACTGGGAA  GGCAAAGCAGATATGGAGACTCTCCAGAG  AATTTATGGCATTTCATTCCCAGATCCTAA  AATGTTGAAAGAGTGGGAGAAGTTCCAAG  AGGAAGCTAAAAACCGAGATCATAGGAA  AATTGGCAGGGACCAAGAACTATATTTCT  TTCATGAACTCAGCCCTGGAAGTTGCTTTT  TTCTGCCAAAAGGAGCCTACATTTATAAT  GCACTTATTGAATTCATTAGGAGCGAATA  TAGGAAAAGAGGATTCCAGGAGGTAGTC  ACCCAAACATCTTCAACAGCCGACTCTG  GATGACCTCGGGCCACTGGCAGCACTACA  GCGAGAACATGTTCTCCTTTGAGGTGGAG  AAGGAGCTGTTTGCCCTGAAACCCATGAA  CTGCCCAGGACACTGCCTTATGTTTGATCA  TCGGCCAAGGTCCTGGCGAGAACTGCCTC  TGCGGCTAGCTGATTTTGGGGTACTTCATA  GGAACGAGCTGTCTGGAGCACTCACAGGA  CTCACCCGGGTACGAAGATTCCAACAGGA  TGATGCTCACATATTCTGTGCCATGGAGC  AGATTGAAGATGAAATAAAAAGGTTGTTG  GATTTTCTACGTACGGTATATAGCGTATTT  GGATTTTCTTTTAAACTAAACCTTTCTACT  CGCCCGGAAAAATTCCTTGGAGATATCGA  AGTATGGGATCAAGCTGAGAAACAACCTTG  AAAACAGTCTGAATGAATTTGGTGAAAAG  TGGGAGTTAAACTCTGGAGATGGAGCTTT  CTATGGCCCAAAGATTGACATACAGATTA  AAGATGCGATTGGGCGGTACCACCAGTGT  GCAACCATCCAGCTGGATTTCCAGTTGCC  CATCAGATTTAATCTTACTTATGTAAGCCA  TGATGGTGATGATAAGAAAAGGCCAGTGA  TTGTTTCATCGAGCCATCTTGGGATCAGTG  GAAAGAATGATTGCTATCCTCACAGAAA  CTATGGGGGCAAATGGCCCTTTTGGCTGT  CCCCTCGCCAGGTAATGGTAGTTCCAGTG  GGACCAACCTGTGATGAATATGCCCAAAA  GGTACGACAACAATTCCACGATGCCAAAT  TCATGGCAGACATTGATCTGGATCCAGGC  TGTACATTGAATAAAAAGATTTCGAAATGC  ACAGTTAGCACAGTATAACTTCATTTTAGT  TGTTGGTGAAAAAGAGAAAATCAGTGCCA  CTGTTAATATCCGCACAAGAGACAATAAG</p>	
		<p>GTCCACGGGGAACGCACCATTTCTGAAAC  TATCGAGCGGCTACAGCAGCTCAAAGAGT  TCCGCAGCAAACAGGCAGAAGAAGAATTT  TAA</p>	

Tabla 5B Uniones de corte y empalme únicas de polipéptidos de AARS			
Nombre	Tipo / Especie	Secuencias de aminoácidos y ácidos nucleicos en la proximidad de la unión de corte y empalme única	SEQ ID NO
T1-AS02	ADN / Humano /	CCGAGATCATAGGAAAATTGGCAGG CCA TGATGGTGTATGATAAGAAAAGG	SEQ ID NO: 91
	Proteína / Humano /	N/A	SEQ ID NO: 92
T1-AS03	ADN / Humano /	GAAGATGGGAGGCGAGGAGAAGCCG TTG AATCCTTGGCCTGAATATATT	SEQ ID NO: 93
	Proteína / Humano /	KMGGEEKPLNPWEYI	SEQ ID NO: 94
T1-AS04	ADN / Humano /	CATATCAAATTGCCTGTGGAATTAG GTGT ATTGGCACTCTAGTGCTCACA	SEQ ID NO: 95
	Proteína / Humano /	N/A	SEQ ID NO: 96

Tabla 6 Polipéptidos y ácidos nucleicos de AARS identificados mediante bioinformática			
Nombre	Tipo / Especie / Restos	Secuencias de aminoácidos y de ácido nucleico	SEQ ID NO:

Polipéptidos de AARS internos: (Tablas 7, 8 y 9)

Tabla 7A Polipéptidos de AARS identificados por EM			
Nombre	Tipo / Especie / Restos	Secuencias de aminoácidos y de ácido nucleico	SEQ ID NO
ThrRS <sup>11</sup>	Proteína / Humano / 147-620	EAAQAVYWHSSAHIMGEAMERVYGGCLCY GPPIENGFYDYDMYLEEGGVSSNDFSSLEALC KKIIEKQAFERLEVKKETLLAMFKYNKFK CRILNEKVNTPTTTVYRCGPLIDLRCRPHVR HTGKIKALKIHKNSSTYWEGKADMETLQRI YGISFPDPKMLKEWEKFQEEAKNRDHRKIG	SEQ ID NO: 115
		RDQELYFFHELSPGSCFFLPKGAYIYNALIEF IRSEYRKRGFQEVVTPNIFNSRLWMTSGHW QHYSNMFSFEVEKELFALKPMNCPGHCL MFDHRPRSWRELPLRLADFGVLHRNELSGA LTGLTRVRRFQQDDAHIFCAMEQIEDEIKGC LDFLRTVYSVFGFSFKLNLSTRPEKFLGDIE VWDQAEKQLENSLNEFGEKWELNSGDGAF YGPKIDIQIKDAIGRYHQCATIQLDFQLPIRF NLTYVSHDGDGDKKRPVIVHRAILGSVERMI AILTENYGGKWPFWLSPRO	

ThrRS1 <sup>11</sup>	ADN / Humano /	<p>GAGGAAGCTCAGGCAGTGTATTGGCACTC  TAGTGCTCACATAATGGGTGAAGCCATGG  AAAGAGTCTATGGTGGATGTTTATGCTAC  GGTCCGCCAATAGAAAATGGATTCTATTA  TGACATGTACCTCGAAGAAGGGGGTGTGT  CTAGCAATGATTTCTTCTCTGGAGGCTT  TGTGTAAGAAAATCATTAAAGAAAAACA  AGCTTTTGAAAGACTGGAAGTTAAGAAAG  AACTTTACTGGCAATGTTTAAGTACAAC  AAGTTCAAATGCCGGATATTGAATGAAAA  GGTGAATACTCCAACACTACCACAGTCTATA  GATGTGGCCCTTTGATAGATCTCTGCCGG  GGTCTCATGTTAGACACACGGGCAAAT  TAAGGCTTTAAAAATACACAAAATTCCT  CCACGTACTGGGAAGGCAAAGCAGATAT  GGAGACTCTCCAGAGAATTTATGGCATT  CATTCCCAGATCCTAAAATGTTGAAAGAG  TGGGAGAAGTTC AAGAGGAAGCTAAAA  ACCGAGATCATAGGAAAATTGGCAGGGA  CCAAGAACTATATTTCTTTCATGAACTCA  GCCCTGGAAGTTGCTTTTTTCTGCCAAAA  GGAGCCTACATTTATAATGCACTTATTGA  ATTCATTAGGAGCGAATATAGGAAAAGA  GGATTCCAGGAGGTAGTCACCCCAAACAT  CTTCAACAGCCGACTCTGGATGACCTCGG  GCCACTGGCAGCACTACAGCGAGAACAT  GTTCTCCTTTGAGGTGGAGAAGGAGCTGT  TTGCCCTGAAACCCATGAACTGCCCAGGA  CACTGCCTTATGTTTGATCATCGGCCAAG  GTCCTGGCGAGAACTGCCTCTGCGGCTAG  CTGATTTTGGGGTACTTCATAGGAACGAG  CTGTCTGGAGCACTCACAGGACTCACCCG  GGTACGAAGATTCCAACAGGATGATGCTC  ACATATTCTGTGCCATGGAGCAGATTGAA</p>	SEQ ID NO: 116
----------------------	----------------	---	-------------------

		<p>GATGAAATAAAAAGGTTGTTTGGATTTTCT  ACGTACGGTATATAGCGTATTTGGATTTT  CTTTTAAACTAAACCTTTCTACTCGCCCGG  AAAAATTCCTTGGAGATATCGAAGTATGG  GATCAAGCTGAGAAACAACCTGAAAACA  GTCTGAATGAATTTGGTGAAAAGTGGGAG  TTAAACTCTGGAGATGGAGCTTTCTATGG  CCCAAAGATTGACATACAGATTAAGATG  CGATTGGGCGGTACCACCAGTGTGCAACC  ATCCAGCTGGATTTCCAGTTGCCCATCAG  ATTTAATCTTACTTATGTAAGCCATGATG  GTGATGATAAGAAAAGGCCAGTGATTGTT  CATCGAGCCATCTTGGGATCAGTGGAAAG  AATGATTGCTATCCTCACAGAAAACATG  GGGGCAAATGGCCCTTTTGGCTGTCCCCT  CGCCAG</p>	
--	--	---	--

<p align="center"><b>Tabla 7B</b>  <b>ThrRS1<sup>11</sup></b>  <b>Péptidos detectados por espectrometría de masas y péptidos enlazadores deducidos</b></p>		
<b>Tipo / Especie</b>	<b>Secuencia</b>	<b>SEQ ID NO</b>
Proteína / ratón	EFEDAFPADFIAEGIDQTR	SEQ ID NO: 117
Proteína / ratón	GWFYLLVLATALFGQPPFK	SEQ ID NO: 118
Proteína / ratón	NVIVNGLILASDGQK	SEQ ID NO: 119
Proteína / ratón	MSKRKKNYPDPVSIIDKYGADALRLYLINSPVVRAENL RFKEEGVRDVLKDVLLPWYNA YRFFIQNVFRLHKEEEV KFLYNEHTVRESPNITDRWVLSFMQSLLGFFETEMAA Y RLYTVVPRLVKFVDILT NWYVRMNRRLKGESGVEDC VMALETLSVLLSLCRLMAPYTPFLTELMYQNLKLLIDP ASLRDKDTLSIHYLMLPRVREELIDKKTENAVSRMQSVI ELGRVIRDRKTIPIKYPLKEIVVIHQDPEALEDIRSLEKYII EELNVRKVTLSTDKNKYGIRLRAEPDHMVLGKRLKGA FKAVMMAIKRLSNEELERFQKSGSIVVEGHHELHEEDIRL MYTFDQATGGTAQFEAHSDAQALVLLDVTPDQSMVDE GMAREVINRIQKLRKKNL VPTDEITVYYNAKSEGRYL NNVIESHTDFVFATIK	SEQ ID NO: 120
Proteína / ratón	APLKPYVPVPTSDNILIQEQTLK	SEQ ID NO: 121

<b>Tabla 7C</b> <b>ThrRS1<sup>11</sup></b> <b>Secuencias concatenadas basadas en péptidos detectados por espectrometría de masas</b>		
<b>Tipo / Especie</b>	<b>Secuencia</b>	<b>SEQ ID NO</b>
Proteína / ratón	<u>EFEDAFPADFIAEGIDQTRGWFYTLVLATALFGQPPFK</u> <u>NVIVNGLILASDGQKMSKRKKNYPDPVSIIDKYGADAL</u> RLYLINSPPVRAENLRFKEEGVRDVLKDVLLPWYNAYR FFIQNVFRLHKEEEVKFLYNEHTVRESPNITDRWVLSFM QSLGFFETEMAA YRLYTVVPRLVKFVDILTWNWYVRM NRRRLKGESGVEDCVMALETLFSVLLSLCRLMAPYTPF LTELMYQNLKLLIDPASLRDKDTLSIHYLMLPRVREELI DKKTENA VSRMQSVIELGRVIRDRKTIPIKYPLKEIVVIH QDPEALEDIRSLEKYIIEELNVRKVTLSTDKNKYGIRLRA EPDHMVLGKRLKGAFKAVMMAIKRLSNEELERFQKSG SIVVEGHELHEEDIRLMYTFDQATGGTAQFEAHSDAQA LVLLDVTDPQSMVDEGMAREVINRIQKLRKCCNLVPTD EITVYNAKSEGRYLNNVIESHTDFVFATIK <u>APLKPYPV</u> PTSNLIQIEOTLK	SEQ ID NO: 122

<b>Tabla 8</b> <b>Polipéptidos de AARS y transcritos alternativos identificados mediante secuenciación profunda</b>			
<b>Nombre</b>	<b>Tipo / Especie / Restos</b>	<b>Secuencias de aminoácidos y de ácido nucleico</b>	<b>SEQ ID NO</b>

<b>Tabla 8B</b> <b>Uniones de corte y empalme únicas de polipéptidos de AARS</b>			
<b>Nombre</b>	<b>Tipo / Especie</b>	<b>Secuencias de aminoácidos y ácidos nucleicos en la proximidad de la unión de corte y empalme única</b>	<b>SEQ ID NO</b>

<b>Tabla 9</b> <b>Polipéptidos y ácidos nucleicos de AARS identificados mediante bioinformática</b>			
<b>Nombre</b>	<b>Tipo / Especie / Restos</b>	<b>Secuencias de aminoácidos y de ácido nucleico</b>	<b>SEQ ID NO</b>

5

Los "fragmentos de proteína" o la secuencia de aminoácidos de los fragmentos de proteína, tales como los fragmentos proteolíticos o los fragmentos variantes de corte y empalme, pueden caracterizarse, identificarse o derivarse de acuerdo con diversas técnicas. Por ejemplo, las variantes de corte y empalme pueden identificarse mediante técnicas, tales como secuenciación profunda (véase, *por ejemplo*, Xing et al., RNA. 14:1470-1479, 2008; y Zhang et al., Genome Research. 17:503-509, 2007). A modo de ejemplo adicional, pueden identificarse fragmentos de proteína, tales como fragmentos proteolíticos *in vitro*, tal como incubando polipéptidos de AARS de longitud completa u otros con proteasas seleccionadas o pueden identificarse de manera endógena (por ejemplo, *in vivo*). En determinadas realizaciones, pueden generarse o identificarse fragmentos de proteína, tales como fragmentos proteolíticos endógenos, por ejemplo, expresando de manera recombinante polipéptidos de AARS de longitud completa u otros en un microorganismo o célula eucariota seleccionada que se ha modificado para que contenga una o más proteasas seleccionadas o que contiene naturalmente una o más proteasas que son capaces de actuar en un polipéptido AARS seleccionado y aislar y caracterizar los fragmentos de proteína producidos por las mismas.

20

En determinadas realizaciones, pueden generarse o identificarse fragmentos de proteína, tales como fragmentos proteolíticos endógenos (por ejemplo, de origen natural), por ejemplo, a partir de varias fracciones celulares (por ejemplo, citosólicas, membrana, nucleares) y/o medio de crecimiento de diversos tipos celulares, incluyendo, por ejemplo, células inmunitarias, tales como monocitos, células dendríticas, macrófagos (por ejemplo, macrófagos RAW 264.7), neutrófilos, eosinófilos, basófilos y linfocitos, tales como células B y células T (por ejemplo, células CD4+ colaboradoras y CD8+ citolíticas), incluyendo células T primarias y líneas de células T, tal como células T Jurkat, así

como linfocitos citolíticos naturales (NK).

En determinadas realizaciones, pueden identificarse fragmentos de proteína, tales como fragmentos proteolíticos endógenos, generados de cualquier modo, mediante técnicas, tales como espectrometría de masas o técnicas equivalentes. Una vez que se ha generado o identificado un fragmento de proteína identificado *in vitro* o de manera endógena, puede mapearse o secuenciarse y, por ejemplo, clonarse en un vector de expresión para su producción recombinante o producirse de manera sintética.

Pueden emplearse una gran variedad de proteasas para producir, identificar, derivar o caracterizar la secuencia de fragmentos de proteína AARS, tales como fragmentos proteolíticos. En general, las proteasas se clasifican normalmente según tres criterios principales: (i) la reacción catalizada, (ii) la naturaleza química del sitio catalítico e (iii) la relación evolutiva, revelada por la estructura. Los ejemplos generales de proteasas o proteinasas, clasificadas por el mecanismo de catálisis, incluyen aspártico proteasas, serina proteasas, cisteína proteasas y metaloproteasas.

La mayor parte de las aspártico proteasas pertenecen a la familia de pepsina. Esta familia incluye las enzimas digestivas, tales como pepsina y quimiosina, así como las catepsinas D lisosómicas y enzimas de procesamiento, tales como renina y determinadas proteasas fúngicas (por ejemplo, penicilopepsina, rhizopuspepsina, endotiapepsina). Una segunda familia de aspártico proteasas incluye proteinasas víricas, tales como la proteasa del virus del SIDA (VIH), también denominada retropepsina.

Las serina proteasas incluyen dos familias distintas. En primer lugar, la familia de quimiotripsina, que incluye las enzimas de mamífero tales como quimiotripsina, tripsina, elastasa y calicreína y en segundo lugar, la familia de subtilisina, que incluye las enzimas bacterianas, tales como subtilisina. La estructura en 3D general entre estas dos familias es diferente, pero tienen la misma geometría de sitio activo y la catálisis se produce mediante el mismo mecanismo. Las serina proteasas muestran diferentes especificidades de sustrato, diferencias que están relacionadas principalmente con sustituciones de aminoácidos en los diversos subsitios enzimáticos (sitios que interactúan con resto de sustrato). Algunas serina proteasas tienen un sitio de interacción extendido con el sustrato mientras que otros tienen una especificidad que está restringida al resto de sustrato P1.

Esta familia de cisteína proteasa incluye las proteasas vegetales, tales como papaína, actinidina y bromelaína, varias catepsinas lisosómicas de mamífero, las calpains citosólicas (activadas por calcio), así como varias proteasas parasitarias (por ejemplo, *Trypanosoma*, *Schistosoma*). La papaína es el arquetipo y el miembro mejor estudiado de la familia. La reciente dilucidación de la estructura de rayos X de la enzima convertidora de interleucina-1-beta ha revelado un nuevo tipo de plegamiento para las cisteína proteinasas.

Las metaloproteasas son una de las clases de proteasas más antiguas, encontradas en bacterias, hongos y organismos superiores. Difieren ampliamente en cuanto a sus secuencias y estructuras 3D, pero la gran mayoría de las enzimas contienen un átomo de cinc que es catalíticamente activo. En algunos casos, el cinc puede reemplazarse por otro metal, tal como cobalto o níquel sin pérdida de actividad proteolítica. La termolisina bacteriana se ha caracterizado bien y su estructura cristalográfica indica que el cinc está unido mediante dos histidinas y un ácido glutámico. Muchas metaloproteasas contiene el motivo de secuencia HEXXH, que proporciona dos ligandos de histidina para el cinc. El tercer ligando es un ácido glutámico (termolisina, neprilisina, alanil aminopeptidasa) o una histidina (astacina, serralisina).

Las proteasas ilustrativas incluyen, por ejemplo, acromopeptidasa, aminopeptidasa, ancrod, enzima convertidora de angiotensina, bromelaína, calpaína, calpaína I, calpaína II, carboxipeptidasa A, carboxipeptidasa B, carboxipeptidasa G, carboxipeptidasa P, carboxipeptidasa W, carboxipeptidasa Y, caspasa 1, caspasa 2, caspasa 3, caspasa 4, caspasa 5, caspasa 6, caspasa 7, caspasa 8, caspasa 9, caspasa 10, caspasa 11, caspasa 12, caspasa 13, catepsina B, catepsina C, catepsina D, catepsina E, catepsina G, catepsina H, catepsina L, quimiopapaína, quimasa, quimiotripsina, clostripaína, colagenasa, C1r de complemento, C1s de complemento, factor D de complemento, factor I de complemento, cucumisina, dipeptidil peptidasa IV, elastasa (leucocitos), elastasa (pancreática), endoproteinasa Arg-C, endoproteinasa Asp-N, endoproteinasa Glu-C, endoproteinasa Lys-C, enterocinasa, factor Xa, ficina, furina, granzima A, granzima B, proteasa del VIH, IGasa, calicreína tisular, leucina aminopeptidasa (general), leucina aminopeptidasa (citosol), leucina aminopeptidasa (microsómica), metaloproteasa, metionina aminopeptidasa, neutrasa, papaína, pepsina, plasmina, prolidasa, pronasa E, antígeno específico de próstata, proteasa alcalófila de *Streptomyces griseus*, proteasa de *Aspergillus*, proteasa de *Aspergillus saitoi*, proteasa de *Aspergillus sojae*, proteasa (*B. licheniformis*) (alcalina o alcalasa), proteasa de *Bacillus polymyxa*, proteasa de *Bacillus* sp., proteasa de *Rhizopus* sp., proteasa S, proteasomas, proteinasas de *Aspergillus oryzae*, proteinasa 3, proteinasa A, proteinasa K, proteína C, piroglutamato aminopeptidasa, renina, renina, estreptocinasa, subtilisina, termolisina, trombina, activador del plasminógeno tisular, tripsina, triptasa y urocinasa.

Determinadas realizaciones se refieren a polipéptidos AARS aislados, que comprenden, consisten esencialmente en o consisten en secuencias de aminoácidos que se han derivado de fragmentos de polipéptido de AARS endógenos de origen natural y a composiciones farmacéuticas que comprenden dichos fragmentos y métodos de uso de los mismos. Pueden generarse o identificarse estas realizaciones y otras relacionadas *in vivo*, *ex vivo*, y/o *in vitro*. En determinadas realizaciones *in vitro* preferidas, se generan o identifican fragmentos proteolíticos de AARS incubando

- un polipéptido de AARS, tal como un polipéptido de AARS de longitud completa, con una o más proteasas humanas aisladas, principalmente aquellas proteasas que son endógenas o naturales para seres humanos, tales como elastasa y otras descritas en el presente documento y conocidas en la técnica. Otras realizaciones se refieren a polipéptidos AARS aislados, que comprenden, consisten esencialmente en o consisten en secuencias de aminoácidos que se han derivado de variantes de corte y empalme de AARS endógenas de origen natural y a composiciones farmacéuticas que comprenden dichos fragmentos y métodos de uso de los mismos. Esencialmente, el fragmento de proteína AARS puede aislarse de muestras que se han expuesto a proteasas, ya sea *in vivo* o *in vitro*.
- En determinadas realizaciones, pueden identificarse fragmentos de proteína AARS mediante técnicas tales como espectrometría de masas o técnicas equivalentes. Solo a título ilustrativo y sin limitación, en determinadas realizaciones pueden separarse los proteomas de diversos tipos celulares, tejidos o fluidos corporales de una serie de estados fisiológicos (por ejemplo, hipoxia, dieta, edad, enfermedad) o fracciones de los mismos mediante SDS-PAGE en 1D y cortarse las bandas de gel a intervalos fijos; después de esto pueden digerirse opcionalmente las bandas con una proteasa adecuada, tal como tripsina, para liberar los péptidos, que después pueden analizarse mediante CL-EM/EM en fase reversa en 1D. Los datos proteómicos resultantes pueden integrarse en los denominados peptogramas, que representan, en el panel izquierdo, la cobertura de secuencia para una proteína dada en la dimensión horizontal (del extremo N al C, de izquierda a derecha) frente a la migración por SDS-PAGE en la dimensión vertical (de elevado a bajo peso molecular, de arriba a abajo). Los fragmentos peptídicos específicos pueden secuenciarse o mapearse. En determinadas realizaciones, el fragmento de referencia de AARS puede caracterizarse por su peso molecular único, en comparación, por ejemplo, con el peso molecular del AARS de longitud completa correspondiente.
- Como se ha indicado anteriormente, las realizaciones de la presente divulgación incluyen los polipéptidos de AARS expuestos en las tablas 1-3 o las tablas 4-6 o las tablas 7-9. También se incluyen "variantes" de los polipéptidos de referencia de AARS. La expresión "variante" de polipéptido se refiere a polipéptidos que se distinguen respecto de un polipéptido de AARS de referencia por la adición, eliminación y/o sustitución de al menos un resto de aminoácido y que normalmente retienen (por ejemplo, imitan) o modulan (por ejemplo, antagonizan) una o más actividades no canónicas de un polipéptido de AARS de referencia.
- Además, las treonil ARNt sintetetas humanas incluyen varios cientos de formas polimórficas altamente relacionadas y se conoce en la técnica que estas son funcionalmente intercambiables, al menos parcialmente. Por lo tanto, sería un hecho rutinario la selección de una variante de origen natural de treonil ARNt sinteteta, incluyendo, por ejemplo, las formas polimórficas de un solo nucleótido enunciadas en la tabla A para crear un polipéptido de AARS que contiene uno o más cambios de aminoácidos, basándose en la secuencia de cualquiera de los homólogos, ortólogos e isoformas de origen natural de los seres humanos, así como otras especies de treonil ARNt sinteteta.

<b>Tabla A</b>			
<b>SNP de treonil ARNt sinteteta humana</b>			
<b>N.º de referencia de Gen Bank</b>	<b>Cambio de nucleótido</b>	<b>N.º de referencia de Gen Bank</b>	<b>Cambio de nucleótido</b>
rs117847077	A/G	rs3849683	G/T
rs117577672	A/G	rs3849682	C/T
rs116809288	A/G	rs3830415	-/GAA
rs116793555	A/G	rs3777083	C/T
rs116686295	A/G	rs3777082	C/T
rs116583264	C/T	rs3777081	C/T
rs116548419	A/C	rs3777080	A/C
rs116483994	A/T	rs3777079	C/T
rs116483298	A/T	rs3777078	C/T
rs116477391	A/G	rs3777077	C/G
rs116390906	C/T	rs3777076	C/G
rs116389220	C/T	rs3777075	C/G

ES 2 638 311 T3

rs116202547	C/T	rs3777074	A/C
rs116175137	C/T	rs3777073	C/T
rs116106015	A/G	rs3765147	A/G
rs116034587	A/T	rs3765146	C/T
rs115949173	A/G	rs3765145	A/G
rs115916716	A/G	rs3765144	A/G
rs115854398	A/G	rs3756664	A/G
rs115792337	A/G	rs3756663	G/T
rs115791161	A/G	rs3756662	A/G
rs115757567	C/T	rs3756661	A/C
rs115722196	G/T	rs3736393	A/G
rs115716728	G/T	rs3215116	-/TT
rs115631256	A/C	rs2642669	A/C
rs115615283	C/T	rs2642668	A/C
rs115606888	C/T	rs2453288	A/C
rs115521824	C/T	rs2304621	C/G
rs115457182	C/T	rs2292015	A/T
rs115335493	C/T	rs2292014	A/G
rs115269494	A/C	rs2270905	A/G
rs115142319	A/G	rs2270904	A/G
rs115094360	C/T	rs2270903	A/T
rs115022219	C/G	rs2270902	C/T
rs114861363	A/G	rs2164991	C/T
rs114781111	C/T	rs1804513	G/T
rs114771001	C/G	rs1687511	A/C
rs114460446	C/T	rs11960553	A/G
rs114430658	A/G	rs11955289	A/G
rs114345473	C/G	rs11955272	C/G
rs114325362	A/G	rs11952399	A/G
rs114216008	C/T	rsll747025	C/T
rs114076009	C/T	rs11742862	A/C
rs114050082	A/T	rs11541416	A/G
rs113811412	A/G	rs11457738	-/G

ES 2 638 311 T3

rs113783190	C/T	rs11422595	-/T
rs113761016	A/G	rs11414360	-/T
rs113591967	A/G	rs11383852	-/C
rs113541140	C/T	rs11356243	-/T
rs113423654	C/T	rs10941055	C/T
rs113336276	G/T	rs9885230	A/C
rs113272480	C/T	rs9686398	C/T
rs113224643	A/G	rs7734078	C/T
rs113183080	A/G	rs7705177	C/T
rs113177776	C/T	rs6893613	A/T
rs113130518	-/A	rs6889780	A/G
rs112838270	A/G	rs6887261	A/C
rs112825149	A/G	rs6885094	A/C
rs112725208	C/T	rs6884566	C/G
rs112715860	C/G	rs6880623	A/G
rs112666038	G/T	rs6880616	A/C
rs112663402	C/T	rs6870521	C/T
rs112327269	C/G	rs6866808	A/C
rs112295695	A/G	rs6450991	A/G
rs112111323	A/G	rs6450990	A/C
rs111991770	A/G	rs5867191	-/CGCG
rs111989999	A/G	rs5867190	-/T
rs111861274	A/C	rs5867189	-/G
rs111820262	A/G	rs5022546	G/T
rs111784154	G/T	rs4867513	C/G
rs111754118	C/T	rs4867512	C/T
rs111751492	C/T	rs4421109	A/G
rs111627406	C/T	rs4398642	C/T
rs111585777	A/T	rs4267874	C/T
rs111524308	A/G	rs1051211	C/T
rs111503761	A/G	rs1051210	C/T
rs111244859	A/G	rs35817175	-/G
rs80171437	A/C	rs35558849	-/G

ES 2 638 311 T3

rs80050750	G/T	rs35375523	-/A
rs79877895	A/T	rs35331115	-/G
rs79826178	A/G	rs35121278	-/T
rs79824382	A/G	rs35079782	-/TT
rs79651403	-/T	rs34910729	A/C
rs79630452	-/AA	rs34875186	-/TTG
rs79607007	C/T	rs34725061	-/T
rs79542443	G/T	rs34715628	C/T
rs79519321	A/G	rs34673293	-/T
rs79384125	A/G	rs34630295	-/A
rs79338682	G/T	rs34596299	-/A
rs79329976	C/G	rs34580645	-/C
rs79122509	A/G	rs34374932	-/G
rs79053234	C/G	rs34334786	A/G
rs79032562	C/T	rs33932086	-/T
rs78636333	C/T	rs28475744	C/T
rs78609486	A/G	rs28362597	C/T
rs78571879	C/T	rs17565038	C/T
rs78521314	C/G	rs17455135	A/G
rs78454230	-/AAG	rs16891192	C/T
rs78347783	A/G	rs16891189	A/G
rs78297441	C/T	rs16891180	A/G
rs78207732	C/T	rs16891143	C/T
rs78020132	A/G	rs16891141	C/G
rs77972413	A/G	rs16891129	A/G
rs77949355	-/T	rs16891125	C/G
rs77756346	A/C	rs13184927	A/T
rs77571586	G/T	rs13164529	A/G
rs77568569	A/G	rs13154531	A/G
rs77381964	-/TTT	rs12659798	C/T
rs77276541	A/G	rs12522185	C/T
rs77257774	A/G	rs12520517	A/G
rs77082161	G/T	rs12517884	A/G

ES 2 638 311 T3

rs77078848	G/T	rs12188515	G/T
rs77025977	A/T	rs12187755	C/T
rs77018010	C/T	rs12187727	C/G
rs76879463	C/T	rs12187715	A/C
rs76852333	G/T	rs72737344	G/T
rs76795002	A/T	rs72737343	C/T
rs76751197	A/C	rs72532932	-/GAA
rs76642289	A/G	rs72532931	-/T
rs76371310	C/T	rs72438851	-/TA
rs76365125	A/G	rs72422014	-/TATA
rs76360820	C/T	rs72267455	-/AG
rs76116449	A/G	rs72205520	-/TA
rs75998770	A/G	rs72136033	-/TG
rs75982140	G/T	rs72087182	-/TA
rs75786537	G/T	rs72062875	-/TT
rs75716245	G/T	rs71996427	-/TT
rs75690412	A/T	rs71824161	-/T
rs75560674	A/T	rs71801512	-/A
rs75334927	C/T	rs71749653	-/T
rs75319890	C/T	rs71678725	-/GA
rs75309293	A/C	rs71600909	-/GAA
rs75248757	A/G	rs71600908	-/T
rs75238388	A/C	rs67503660	-/T
rs75075633	A/G	rs67450756	A/C
rs74889236	-/TT	rs66916211	-/TT
rs74728320	C/T	rs66544947	-/TT
rs74618872	A/C	rs62351066	A/G
rs74417746	A/C	rs62351065	A/G
rs74316832	C/T	rs61734318	C/G
rs73758555	A/G	rs60881601	-/TA
rs73758552	C/T	rs60484211	G/T
rs73076308	C/G	rs60163332	C/T
rs73076306	C/T	rs59747499	A/G

ES 2 638 311 T3

rs73074392	C/G	rs59128488	A/T
rs73074385	A/G	rs59078367	A/G
rs73074377	A/G	rs58771215	A/G
rs73074376	C/G	rs58510147	A/T
rs73074374	C/T	rs58463740	A/T
rs73074360	C/T	rs56938839	A/G
rs73074357	C/T	rs56718466	C/T
rs72737350	A/G	rs55759090	A/C
rs72737349	A/G	rs55749081	-/T
rs72737345	A/C	rs35931457	-/T
rs1003000	C/T		
rs893551	C/T		
rs7261	A/G		
rs3842066	-/AAC/CAA		
rs4365842	-/C/CACGTG/T		
rs10583985	-/CTCTCTCT		
rs36102914	-/TA/TATATA		
rs57691465	-/TTTTTTTT		
rs59141272	-/ATATATATAG		
rs66723019	-/ATATA/ATATATATAG		
rs71676557	(ELIMINACIÓN GRANDE)/-		
rs67402626	-/TTTTTTTT		
rs71692029	(ELIMINACIÓN GRANDE)/-		
rs78884891	-/CGCGTATA		

En determinadas realizaciones, una variante de polipéptido se distingue de un polipéptido de referencia por una o más sustituciones, que pueden ser conservativas o no conservativas, tal como se describe en el presente documento y se conoce en la técnica. En determinadas realizaciones, la variante de polipéptido comprende sustituciones conservativas y, a este respecto, se entiende en la técnica que pueden cambiarse algunos aminoácidos por otros con propiedades ampliamente similares sin cambiar la naturaleza de la actividad del polipéptido.

5

En determinadas realizaciones, una variante de polipéptido incluye una secuencia de aminoácidos que tiene al menos aproximadamente un 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o más de identidad o similitud de secuencia con una secuencia correspondiente de un polipéptido de AARS de referencia, tal como se describe en el presente documento y retiene sustancialmente la actividad no canónica de ese polipéptido de referencia. También se incluyen secuencias que difieren de las secuencias de AARS de referencia por la adición, eliminación o sustitución de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 o más aminoácidos pero que mantienen las propiedades del polipéptido de AARS de referencia. En determinadas realizaciones, las adiciones o eliminaciones de aminoácidos se producen en el extremo C-terminal y/o el extremo N-terminal del polipéptido de AARS de referencia. En determinadas realizaciones, las adiciones de aminoácidos incluyen 1, 2, 3, 4, 5, 6, 7, 8, 9,

10

15

10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50 o más restos de tipo silvestre (es decir, del polipéptido de AARS de longitud completa correspondiente) que están próximos al extremo C-terminal y/o al extremo N-terminal del polipéptido de referencia de AARS.

- 5 En determinadas realizaciones, los polipéptidos variantes difieren respecto de las secuencias de referencia de AARS correspondientes en al menos un 1 % pero en menos de un 20 %, 15 %, 10 % o 5 % de los restos. (En caso de que esta comparación requiera de alineamiento, las secuencias deberían alinearse para máxima similitud. Se consideran diferencias las secuencias sacadas del bucle a causa de eliminaciones, inserciones o desemparejamientos). De manera adecuada, las diferencias son diferencias o cambios en un resto no esencial o una sustitución conservativa.
- 10 En determinadas realizaciones, el peso molecular de un polipéptido de AARS variante difiere respecto del polipéptido de AARS de referencia en aproximadamente un 1 %, 2 %, 3 %, 4 %, 5 %, 6 %, 7 %, 8 %, 9 %, 10 %, 11 %, 12 %, 13 %, 14 %, 15 %, 16 %, 17 %, 18 %, 19 %, 20 %, o más.

- 15 También se incluyen "fragmentos" biológicamente activos de los polipéptidos de AARS de referencia, es decir, fragmentos biológicamente activos de los fragmentos de proteína AARS. Los fragmentos activos biológicamente representativos participan generalmente en una interacción, *por ejemplo*, una interacción intramolecular o intermolecular. Una interacción intermolecular puede ser una interacción de unión específica o una interacción enzimática. Una interacción intermolecular puede ser entre un polipéptido de AARS y un compañero de unión celular, tal como un receptor celular o la molécula del hospedador que participa en la actividad no canónica del polipéptido de AARS. En algunas realizaciones, las proteínas AARS, variantes y fragmentos biológicamente activos de las mismas, se unen a uno o más compañeros de unión celulares con una afinidad de al menos aproximadamente 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40 o 50 nM. La afinidad de unión de un fragmento de proteína AARS por un compañero de unión celular seleccionado, en particular un compañero de unión que participa en una actividad no canónica, es normalmente más fuerte que la del polipéptido de AARS de longitud completa correspondiente del fragmento de proteína AARS, en al menos aproximadamente 1,5x, 2x, 2,5x, 3x, 3,5x, 4x, 4,5x, 5x, 6x, 7x, 8x, 9x, 10x, 15x, 20x, 25x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x, 300x, 400x, 500x, 600x, 700x, 800x, 900x, 1000x o más (incluyendo todos los números enteros intermedios). La afinidad de unión de un fragmento de proteína AARS por un compañero de unión que participa en al menos una actividad canónica de una AARS es normalmente más débil que la del polipéptido de AARS de longitud completa correspondiente del fragmento de proteína AARS, en al menos aproximadamente 1,5x, 2x, 2,5x, 3x, 3,5x, 4x, 4,5x, 5x, 6x, 7x, 8x, 9x, 10x, 15x, 20x, 25x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x, 300x, 400x, 500x, 600x, 700x, 800x, 900x, 1000x o más.

- 35 Normalmente, los fragmentos biológicamente activos comprenden un dominio o motivo con al menos una actividad de un polipéptido de AARS de referencia y pueden incluir uno o más (y en algunos casos todos) los diversos dominios activos e incluyen fragmentos que tienen una actividad no canónica. En algunos casos, los fragmentos biológicamente activos de un polipéptido AARS tienen una actividad biológica que es única para el fragmento truncado particular, de tal forma que el polipéptido de AARS de longitud completa puede no tener esa actividad. En determinados casos, puede revelarse la actividad biológica separando el fragmento de polipéptido de AARS biológicamente activo de las demás secuencias de polipéptido de AARS de longitud completa o alterando determinados restos de la secuencia de polipéptido de tipo silvestre de AARS de longitud completa para desenmascarar los dominios biológicamente activos.

- 45 Un fragmento biológicamente activo de un polipéptido de AARS de referencia puede ser un fragmento de polipéptido que tiene, por ejemplo, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 450, 500, 550, 600, 650, 700, 750 o más aminoácidos contiguos o no contiguos, incluyendo todos los números enteros (por ejemplo, 101, 102, 103) e intervalos (por ejemplo, 50-100, 50-150, 50-200) intermedios, de las secuencias de aminoácidos expuestas en uno cualquiera de los polipéptidos de AARS de referencia descritos en el presente documento, pero normalmente excluyen la AARS de longitud completa. En determinadas realizaciones, un fragmento biológicamente activo comprende una secuencia, dominio o motivo relacionado con una actividad no canónica. En determinadas realizaciones, puede truncarse la región C-terminal o N-terminal de cualquier polipéptido de AARS de referencia en aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650 o 700 o más aminoácidos o en aproximadamente 10-50, 20-50, 50-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, 450-500, 500-550, 550-600, 600-650, 650-700 o más aminoácidos, incluyendo todos los números enteros e intervalos intermedios (por ejemplo, 101, 102, 103, 104, 105), en tanto que el polipéptido de AARS truncado conserve la actividad no canónica del polipéptido de referencia. Normalmente, el fragmento biológicamente activo tiene no menos de aproximadamente un 1 %, aproximadamente un 5 %, aproximadamente un 10 %, aproximadamente un 25 %, o aproximadamente un 50 % de una actividad (es decir, una actividad no canónica) del polipéptido de AARS de referencia biológicamente activo del que procede. Los métodos a modo de ejemplo para medir dichas actividades no canónicas se describen en los ejemplos.

- 65 Como se ha indicado anteriormente, puede alterarse un polipéptido de AARS de diversas maneras, incluyendo sustituciones de aminoácidos, eliminaciones, truncamientos e inserciones. Los métodos para dichas manipulaciones se conocen de manera general en la técnica. Por ejemplo, pueden prepararse variantes de secuencia de

- aminoácidos de un polipéptido de AARS de referencia mediante mutaciones en el ADN. Los métodos para la mutagénesis y las alteraciones de secuencias de nucleótidos se conocen bien en la técnica. Véase, por ejemplo, Kunkel (1985, Proc. Natl. Acad. Sci. USA. 82: 488-492), Kunkel et al., (1987, Methods in Enzymol, 154: 367-382), Patente de los Estados Unidos n.º 4.873.192, Watson, J. D. et al., ("Molecular Biology of the Gene", cuarta edición, Benjamin/Cummings, Menlo Park, Calif., 1987) y las referencias citadas en el mismo. Puede encontrarse orientación acerca de las sustituciones de aminoácidos adecuadas que no afectan a la actividad biológica de la proteína de interés en el modelo de Dayhoff et al., (1978) Atlas of Protein Sequence and Structure (Natl. Biomed. Res. Found., Washington, D.C.).
- 5
- 10 De manera similar, se encuentra dentro de las capacidades de la técnica abordar y/o mitigar los problemas de inmunogenicidad en caso de que surjan usando un polipéptido de AARS, *por ejemplo*, mediante el uso de programas de reconocimiento informático automatizado para identificar potenciales epítopos de células T y estrategias de evolución dirigida para identificar formas menos inmunogénicas.
- 15 Se conocen en la técnica métodos para explorar productos génicos de biblioteca combinatorias elaboradas mediante mutaciones puntuales o truncamiento y para explorar bibliotecas de ADNc respecto de productos génicos que tienen una propiedad seleccionada. Dichos métodos son adaptables para una exploración rápida de las genotecas generadas mediante mutagénesis combinatoria de polipéptidos de AARS. La mutagénesis de conjunto recursiva (REM), una técnica que potencia la frecuencia de mutantes funcionales en las bibliotecas, puede usarse en combinación con los ensayos de exploración para identificar variantes de polipéptido de AARS (Arkin y Yourvan (1992) Proc. Natl. Acad. Sci. USA 89: 7811-7815; Delgrave et al., (1993) Protein Engineering, 6: 327-331). Pueden ser deseables las sustituciones conservativas, tales como el intercambio de un aminoácido por otro que tenga propiedades similares, tal como se describe con mayor detalle más adelante.
- 20
- 25 Los polipéptidos de AARS truncados y/o variantes biológicamente activos pueden contener sustituciones conservativas de aminoácidos en varias ubicaciones a lo largo de su secuencia, en comparación con un resto de aminoácido de AARS de referencia. Además, se han secuenciado variantes de origen natural de proteínas AARS y se conoce en la técnica que son funcionalmente intercambiables, al menos parcialmente. Por lo tanto, será una cuestión rutinaria seleccionar una posición de aminoácido para introducir una mutación conservativa o no conservativa en un polipéptido de AARS basándose en la variación de secuencia de origen natural entre los homólogos, ortólogos e isoformas de origen natural de proteína AARS conocidos de seres humanos, así como de otras especies de una proteína AARS.
- 30
- 35 Una "sustitución de aminoácidos conservativa" es una en la que se reemplaza el resto de aminoácido por un resto de aminoácido que tenga una cadena lateral similar. Se han definido en la técnica familias de aminoácidos que tienen cadenas laterales similares, que generalmente pueden subclasificarse del siguiente modo:
- Ácidas: El resto tiene una carga negativa debida a la pérdida de ion H a pH fisiológico y el resto es atraído por una solución acuosa para buscar las posiciones superficiales en la conformación de un péptido en el que está contenido cuando el péptido se encuentra en un medio acuoso a pH fisiológico. Los aminoácidos que tienen una cadena lateral ácida incluyen ácido glutámico y ácido aspártico.
- 40
- Básicas: El resto tiene una carga positiva debida a la asociación con ion H a pH fisiológico o a una o dos unidades de pH del mismo (por ejemplo, histidina) y el resto es atraído por una solución acuosa para buscar las posiciones superficiales en la conformación de un péptido en el que está contenido cuando el péptido se encuentra en un medio acuoso a pH fisiológico. Los aminoácidos que tienen una cadena lateral básica incluyen arginina, lisina e histidina.
- 45
- Cargadas: Los restos están cargados a pH fisiológico y, por lo tanto, incluyen aminoácidos que tienen cadenas laterales ácidas o básicas (es decir, ácido glutámico, ácido aspártico, arginina, lisina e histidina).
- 50
- Hidrófobas: Los restos no están cargados a pH fisiológico y el resto es repelido por una solución acuosa para buscar las posiciones internas en la conformación de un péptido en el que está contenido cuando el péptido se encuentra en medio acuoso. Los aminoácidos que tienen una cadena lateral hidrófoba incluyen tirosina, valina, isoleucina, leucina, metionina, fenilalanina y triptófano.
- 55
- Neutras/polares: Los restos no están cargados a pH fisiológico, pero el resto no es suficientemente repelido por soluciones acuosas de tal forma que podría buscar posiciones internas en la conformación de un péptido en el que está contenido cuando el péptido se encuentra en medio acuoso. Los aminoácidos que tienen una cadena lateral Neutras/polares incluyen asparagina, glutamina, cisteína, histidina, serina y treonina.
- 60
- 65 Esta descripción también caracteriza determinados aminoácidos como "pequeños", ya que sus cadenas laterales no son lo suficientemente grandes, incluso si hay ausencia de grupos polares, para conferir hidrofobia. Con la excepción de prolina, los aminoácidos "pequeños" son aquellos de cuatro carbonos o menos cuando al menos un grupo polar está presente en la cadena lateral y de tres carbonos o menos cuando no. Los aminoácidos que tienen una cadena lateral pequeña incluyen glicina, serina, alanina y treonina. El aminoácido secundario, prolina, codificado génicamente es un caso especial debido a sus efectos conocidos en la conformación secundaria de cadenas

- peptídicas. La estructura de la prolina difiere de la de todos los demás aminoácidos de origen natural en que su cadena lateral está unida al nitrógeno del grupo  $\alpha$ -amino, así como al carbono  $\alpha$ . Se conocen en la técnica varias matrices de similitud de aminoácidos (véase, por ejemplo, la matriz PAM120 y la matriz PAM250 divulgadas, por ejemplo, por Dayhoff et al., 1978, A model of evolutionary change in proteins). Las matrices para determinar relaciones distantes en M. O. Dayhoff, (ed.), Atlas of protein sequence and structure, Vol. 5, págs. 345-358, National Biomedical Research Foundation, Washington DC; y de Gonnet et al., (Science, 256: 14430-1445, 1992), sin embargo, incluyen a la prolina en el mismo grupo que la glicina, serina, alanina y treonina. Por consiguiente, para los fines de la presente invención, la prolina se clasifica con un aminoácido "pequeño".
- 10 El grado de atracción o repulsión necesario para la clasificación como polar o no polar es arbitrario y, por lo tanto, los aminoácidos contemplados específicamente por la invención se han clasificado como de uno u otro modo. La mayor parte de los aminoácidos no nombrados específicamente pueden clasificarse basándose en un comportamiento conocido.
- 15 Los restos aniónicos pueden subclasificarse adicionalmente como cíclicos o no cíclicos y aromáticos o no aromáticos, clasificaciones que se explican por sí mismas respecto de los grupos sustituyentes de la cadena lateral de los restos y como pequeños o grandes. El resto se considera pequeño en caso de que contenga un total de cuatro átomos de carbono o menos, incluyendo el carbono del carboxilo, siempre que esté presente un sustituyente polar adicional; de tres o menos en caso contrario. Evidentemente, los restos pequeños siempre son no aromáticos.
- 20 Dependiendo de sus propiedades estructurales, los restos de aminoácidos pueden encontrarse en dos o más clases. Para los aminoácidos de proteínas de origen natural, en la tabla B se presenta una subclasificación de acuerdo con este esquema.

**Tabla B: Subclasificación de aminoácidos**

Subclases	Aminoácidos
Ácidos	Ácido aspártico, Ácido glutámico
Básicos	No cíclicos: Arginina, Lisina; Cíclicos: Histidina
Cargados	Ácido aspártico, Ácido glutámico, Arginina, Lisina, Histidina
Pequeños	Glicina, Serina, Alanina, Treonina, Prolina
Polares/neutros	Asparagina, Histidina, Glutamina, Cisteína, Serina, Treonina
Polares/grandes	Asparagina, Glutamina
Hidrófobos	Tirosina, Valina, Isoleucina, Leucina, Metionina, Fenilalanina, Triptófano
Aromáticos	Triptófano, Tirosina, Fenilalanina
Restos que influyen en la orientación de la cadena	Glicina y Prolina

- 25 La sustitución conservativa de aminoácidos también incluye agrupamientos basados en las cadenas laterales. Por ejemplo, un grupo de aminoácidos que tienen cadenas laterales alifáticas está formado por glicina, alanina, valina, leucina e isoleucina; un grupo de aminoácidos que tienen cadenas laterales alifáticas-de hidroxilo está formado por serina y treonina; un grupo de aminoácidos que tienen cadenas laterales que contienen amida está formado por asparagina y glutamina; un grupo de aminoácidos que tienen cadenas laterales aromáticas está formado por fenilalanina, tirosina y triptófano; un grupo de aminoácidos que tienen cadenas laterales básicas está formado por lisina, arginina e histidina; un grupo de aminoácidos que tienen cadenas laterales que contienen azufre está formado por cisteína y metionina. Por ejemplo, sería razonable esperar que el reemplazo de una leucina por una isoleucina o valina, de un aspartato por un glutamato, de una treonina por una serina o un reemplazo similar de un aminoácido por un aminoácido relacionado estructuralmente no tenga un efecto importante en las propiedades del polipéptido variante resultante. Puede determinarse fácilmente si un cambio de aminoácido da como resultado un polipéptido de AARS truncado y/o variante funcional ensayando su actividad no canónica, tal como se describen en el presente documento. En la tabla C se muestran sustituciones conservativas bajo el encabezado de sustituciones a modo de ejemplo. Las sustituciones de aminoácidos que se encuentran dentro del alcance de la invención se llevan a cabo en general seleccionando sustituciones que no difieren significativamente en su efecto en el mantenimiento de (a) la estructura de la cadena principal del péptido en el área de la sustitución, (b) la carga o hidrofobicidad de la molécula en el sitio diana, (c) el volumen de la cadena lateral o (d) la función biológica. Después de haber introducido las sustituciones, se explora la actividad biológica de las variantes.

45

Tabla C: Sustituciones de aminoácidos a modo de ejemplo

Resto original	Sustituciones a modo de ejemplo	Sustituciones preferidas
Ala	Val, Leu, He	Val
Arg	Lys, Gln, Asn	Lys
Asn	Gln, His, Lys, Arg	Gln
Asp	Glu	Glu
Cys	Ser	Ser
Gln	Asn, His, Lys,	Asn
Glu	Asp, Lys	Asp
Gly	Pro	Pro
His	Asn, Gln; Lys, Arg	Arg
Ile	Leu, Val, Met, Ala, Phe, Norleu	Leu
Leu	Norleu, He, Val, Met, Ala, Phe	Ile
Lys	Arg, Gln, Asn	Arg
Met	Leu, Ile, Phe	Leu
Phe	Leu, Val, He, Ala	Leu
Pro	Gly	Gly
Ser	Thr	Thr
Thr	Ser	Ser
Trp	Tyr	Tyr
Tyr	Trp, Phe, Thr, Ser	Phe
Val	He, Leu, Met, Phe, Ala, Norleu	Leu

5 Como alternativa, pueden agruparse en tres categorías aminoácidos similares para producir sustituciones conservativas basándose en la identidad de las cadenas laterales. El primer grupo incluye ácido glutámico, ácido aspártico, arginina, lisina, histidina, que tienen todas las cadenas laterales cargadas; el segundo grupo incluye glicina, serina, treonina, cisteína, tirosina, glutamina, asparagina; y el tercer grupo incluye leucina, isoleucina, valina, alanina, prolina, fenilalanina, triptófano, metionina, tal como se describe en Zubay, G., Biochemistry, tercera edición, Wm.C. Brown Publishers (1993).

10 Por lo tanto, un resto de aminoácido no esencial predicho en un polipéptido de AARS truncado y/o variante se reemplaza normalmente por otro resto de aminoácido de la misma familia de cadenas laterales. Como alternativa, pueden introducirse mutaciones aleatoriamente a lo largo de la totalidad o parte de una secuencia codificante de AARS, tal como mediante mutagénesis de saturación y pueden explorarse los mutantes resultantes respecto de una actividad del polipéptido progenitor para identificar mutantes que conservan esa actividad.

15 Después de la mutagénesis de las secuencias codificantes, el péptido codificado puede expresarse de manera recombinante y determinarse la actividad del péptido. Un resto de aminoácido "no esencial" es un resto que puede alterarse respecto de la secuencia de referencia de una realización de polipéptido sin suprimir o alterar sustancialmente una o más de sus actividades. De manera adecuada, la alteración no suprime sustancialmente una de estas actividades, por ejemplo, la actividad es al menos un 20 %, 40 %, 60 %, 70 % u 80 % o un 100 %, 500 %, 1000 % o más de la

20 secuencia de AARS de referencia. Un resto de aminoácido "esencial" es un resto que, cuando se altera respecto de la secuencia de referencia de un polipéptido de AARS, da como resultado la supresión de una actividad de la molécula parental, de tal forma que está presente menos de un 20 % de la actividad de referencia. Por ejemplo, dichos restos de aminoácidos esenciales incluyen aquellos que están conservados en polipéptidos de AARS entre diferentes especies, incluyendo aquellas secuencias que se encuentran conservadas en los sitios o motivos de unión

25 activos de polipéptidos de AARS de diversos orígenes.

En general, los polipéptidos y los polipéptidos de fusión (así como sus polinucleótidos codificantes) se encuentran aislados. Un polipéptido o polinucleótido "aislado" es uno que se extrae de su ambiente natural. Por ejemplo, una proteína de origen natural se encuentra aislada en caso de que esté separada de algunos o de todos los materiales coexistentes en el sistema natural. Preferentemente, dichos polipéptidos son al menos un 90 % puros, más

30 preferentemente, al menos aproximadamente un 95 % puros y lo más preferentemente, al menos aproximadamente un 99 % puros. Se considera que un polinucleótido está aislado si, por ejemplo, se clona en un vector que no forma parte del ambiente natural.

Determinadas realizaciones también abarcan dímeros de polipéptidos de AARS. Los dímeros pueden incluir, por ejemplo, homodímeros entre dos polipéptidos de AARS idénticos, heterodímeros entre dos polipéptidos de AARS diferentes (por ejemplo, un polipéptido de YRS de longitud completa y un polipéptido de YRS truncado; un polipéptido de YRS truncado y un polipéptido de YRS truncado), y/o heterodímeros entre un polipéptido de AARS y un polipéptido heterólogo. Determinados heterodímeros, tales como aquellos entre un polipéptido de AARS y un polipéptido heterólogo, pueden ser bifuncionales, tal como se describen en el presente documento.

También se incluyen monómeros de polipéptidos de AARS, incluyendo monómeros de polipéptidos de AARS que no dimerizan sustancialmente con un segundo polipéptido de AARS, ya sea debido a una o más sustituciones, truncamientos, eliminaciones, adiciones, modificaciones químicas o una combinación de estas alteraciones. En determinadas realizaciones, los polipéptidos monoméricos de AARS poseen actividades biológicas, incluyendo actividades no canónicas, que no poseen los complejos multiméricos de polipéptido de AARS diméricos.

Determinadas realizaciones de la presente divulgación también contemplan el uso de polipéptidos de AARS modificados, incluyendo modificaciones que mejoran las características deseadas del polipéptido de AARS, tal como se describen en el presente documento. Las modificaciones de los polipéptidos de AARS de la invención incluyen derivatizaciones químicas y/o enzimáticas en uno o más aminoácidos constituyentes, incluyendo modificaciones de la cadena lateral; modificaciones de la cadena principal y modificaciones N y C-terminales, incluyendo acetilación, hidroxilación, metilación, amidación y la unión de restos de carbohidrato o lípido, cofactores y similares. Las modificaciones a modo de ejemplo también incluyen pegilación de un polipéptido de AARS (véase, *por ejemplo*, Veronese y Harris, *Advanced Drug Delivery Reviews* 54: 453-456, 2002; y Pasut et al., *Expert Opinion. Ther. Patents* 14(6) 859-894 2004, ambas incorporadas al presente documento por referencia).

El PEG es un polímero de sobra conocido que tiene las propiedades de solubilidad en agua y en muchos disolventes orgánicos, ausencia de toxicidad y ausencia de inmunogenicidad. También es transparente, incoloro, inodoro y químicamente estable. Por estos y otros motivos, se ha seleccionado el PEG como el polímero preferido para la unión, pero se ha empleado únicamente con fines ilustrativos y no limitantes. Pueden obtenerse productos similares con otros polímeros hidrosolubles, incluyendo, sin limitación: alcohol polivinílico, otros poli(óxidos de alqueno), tales como poli(propilenglicol) y similares, poli(poliolios oxietilados), tales como poli(glicerol oxietilado) y similares, carboximetilcelulosa, dextrano, alcohol polivinílico, polivinilpirrolidona, poli-1,3-dioxolano, poli-1,3,6-trioxano, etileno/anhídrido maleico y poliaminoácidos. Un experto en la materia podrá seleccionar el polímero deseado basándose en la dosis deseada, el tiempo de circulación, la resistencia a la proteólisis y otras consideraciones.

En particular, se encuentra disponible una gran variedad de derivados de PEG y son adecuados para su uso en la preparación de conjugados de PEG. Por ejemplo, Los reactivos de PEG de NOF Corp., comercializados bajo la marca SUNBRIGHT® Series, proporciona numerosos derivados de PEG, incluyendo metoxipolietilenglicoles y derivados de PEG activados, tales como metoxi-PEG amins, maleimidias, ésteres de N-hidroxisusccinimida y ácidos carboxílicos, para su acoplamiento mediante diversos métodos al aminoácido N-terminal, C-terminal o cualquiera interno del polipéptido de AARS. La tecnología de Pegilación avanzada de Nektar Therapeutics también ofrece diversas tecnologías de acoplamiento de PEG para mejorar potencialmente la seguridad y eficacia de un agente terapéutico a base de polipéptido de AARS.

Una búsqueda de patentes, solicitudes de patente publicadas y publicaciones relacionadas también dota al experto en la materia que lea la presente divulgación de tecnologías significativas posibles de acoplamiento de PEG y de derivados de PEG. Por ejemplo, las Patentes de los Estados Unidos n.º 6.436.386; 5.932.462; 5.900.461; 5.824.784; y 4.904.584; cuyos contenidos se incorporan al presente documento por referencia en su totalidad, describen dichas tecnologías y derivados y métodos para su fabricación.

En determinados aspectos, puede utilizarse tecnología de ligadura quimioselectiva para modificar los polipéptidos de AARS de la invención, tal como uniendo polímeros de un modo específico de sitio y controlado. Dicha tecnología se basa normalmente en la incorporación de anclas quimioselectivas en la cadena principal de proteína por medios tanto químicos como recombinantes y la posterior modificación con un polímero que porta un enlazador complementario. Como resultado, pueden controlarse el proceso de ensamblaje y la estructura covalente del conjugado de proteína-polímero resultante, permitiendo una optimización racional de las propiedades farmacológicas, tales como la eficacia y las propiedades farmacocinéticas (véase, *por ejemplo*, Kochendoerfer, *Current Opinion in Chemical Biology* 9:555-560, 2005).

En otras realizaciones, también se incluyen proteínas de fusión del polipéptido de AARS a otras proteínas y estas proteínas de fusión pueden aumentar la actividad biológica, la secreción, el direccionamiento, la vida biológica, la capacidad para penetrar membranas celulares o la barrera hematoencefálica o las propiedades farmacocinéticas del polipéptido de AARS. Los ejemplos de proteínas de fusión que mejoran las propiedades farmacocinéticas ("modificadores PK") incluyen, sin limitación, fusiones a albúmina humana (Osborn et al.: *Eur. J. Pharmacol.* 456(1-3): 149-158, (2002)), dominios Fc de anticuerpo, secuencias de poli Glu o poli Asp y transferrina. Además, la fusión con secuencias de polipéptido conformacionalmente desordenadas compuestas por los aminoácidos Pro, Ala y Ser ("PASilación") o hidroxietilalmidón (comercializado bajo la marca HESYLATION®) proporciona una manera sencilla de aumentar el volumen hidrodinámico del polipéptido de AARS. Esta extensión adicional adopta una estructura

aleatoria voluminosa, que aumenta significativamente el tamaño de la proteína de fusión resultante. Por este medio, se retrasa en varios órdenes de magnitud la normalmente rápida eliminación de polipéptidos de AARS más pequeños mediante filtración renal. Además, se ha demostrado que el uso de proteínas de fusión a IgG permite que ciertas proteínas de fusión penetren la barrera hematoencefálica (Fu et al., (2010) Brain Res. 1352:208-13).

5 Los ejemplos de proteínas de fusión que mejoran la penetración a través de membranas celulares incluyen fusiones a secuencias de translocación de membrana. En este contexto, la expresión "secuencias de translocación de membrana" se refiere a secuencias de aminoácidos de origen natural y sintéticas que son capaces de translocarse en la membrana a través de una membrana celular. Las secuencias de translocación de membrana representativas  
10 incluyen aquellas basadas en las secuencias de translocación de membrana de origen natural derivadas de la proteína Tat y la proteína de transcripción homeótica Antennapedia, así como secuencias de translocación de membrana sintéticas basadas en su totalidad o en parte en restos de poli arginina y lisina. Las secuencias de translocación de membrana representativas incluyen, por ejemplo, aquellas divulgadas en las siguientes patentes,  
15 US5.652.122; US 5.670.617; US5.674.980; US5.747.641; US5.804.604; US6.316.003; US7.585.834; US7.312.244; US7.279.502; US7.229.961; US7.169.814; US7.453.011; US7.235.695; US6.982.351; US6.605.115; US7.306.784; US7.306.783; US6.589.503; US6.348.185; US6.881.825; US7.431.915; WO0074701A2; WO2007111993A2; WO2007106554A2; WO02069930A1; WO03049772A2; WO03106491A y WO2008063113A1.

20 Se apreciará que puede interponerse, y unirse covalentemente, un enlazador (o espaciador) molecular flexible entre el polipéptido de AARS y cualquiera de las proteínas de fusión divulgadas en el presente documento.

Además, en algunas realizaciones, el polipéptido de AARS puede incluir secuencias de señal de secreción sintéticas o de origen natural, derivadas de otras proteínas secretadas bien caracterizadas. En algunas realizaciones, dichas proteínas pueden procesarse mediante escisión proteolítica para formar el polipéptido de AARS *in situ*. Dichas  
25 proteínas de fusión incluyen, por ejemplo, fusiones del polipéptido de AARS a ubiquitina para proporcionar un nuevo aminoácido N-terminal o el uso de una señal de secreción para mediar la secreción a alto nivel del polipéptido de AARS al medio extracelular o marcadores epitópicos N o C-terminales para mejorar la purificación o la detección.

Los polipéptidos de AARS descritos en el presente documento pueden prepararse mediante cualquier procedimiento adecuado conocido por los expertos en la materia, tal como mediante técnicas recombinantes. Además de los métodos de producción recombinantes, los polipéptidos de la invención pueden producirse mediante síntesis peptídica directa usando técnicas de fase sólida (Merrifield, J. Am. Chem. Soc. 85:2149-2154 (1963)). La síntesis de proteínas puede efectuarse usando técnicas manuales o mediante automatización. Puede lograrse la síntesis automatizada, por ejemplo, usando el sintetizador peptídico 431A de Applied Biosystems (Perkin Elmer). Como  
30 alternativa, pueden sintetizarse químicamente por separado varios fragmentos y combinarse usando métodos químicos para producir la molécula deseada.  
35

#### IV. POLINUCLEÓTIDOS DE AARS

40 Las realizaciones de la presente divulgación incluyen polinucleótidos que codifican uno o más fragmentos de proteína recientemente identificados de una aminoacil-ARNt sintetasa (AARS), además de complementos, variantes y fragmentos de la misma. En determinadas realizaciones, un polinucleótido de AARS codifica la totalidad o una parte de las secuencias de referencia del polipéptido de AARS tal como se expone en las tablas 1-3 o las tablas 4-6 o las tablas 7-9, que representan variantes de corte y empalme, fragmentos proteolíticos u otro tipo de fragmentos  
45 de treonil ARNt sintetasa. Determinadas realizaciones incluyen polinucleótidos, polipéptidos codificantes o proteínas que comprenden la secuencia de una o más uniones de corte y empalme de aquellas variantes de corte y empalme, además de complementos, variantes y fragmentos de los mismos. En determinadas realizaciones, normalmente debido a la naturaleza singular de una variante de corte y empalme seleccionada, que combina exones de un nuevo modo excepcional, las secuencias de referencia del polinucleótido de AARS comprenden una unión de corte y empalme única o excepcional. Determinadas realizaciones excluyen un polinucleótido de AARS de longitud completa correspondiente.  
50

También se incluyen entre los polinucleótidos de AARS de la presente divulgación cebadores, sondas, oligonucleótidos antisentido y agentes de interferencia de ARN que comprenden la totalidad o parte de estos  
55 polinucleótidos de referencia, que son complementarios a la totalidad o una parte de estos polinucleótidos de referencia o que hibridan específicamente con estos polinucleótidos de referencia, tal como se describen en el presente documento.

El término "polinucleótido" o "ácido nucleico", tal como se usa en el presente documento, indica ARNm, ARN, ARNc, ADNc o ADN. El término se refiere normalmente a la forma polimérica de nucleótido de al menos 10 bases de longitud, ya sean ribonucleótidos o desoxinucleótidos o una forma modificada de cualquier tipo de nucleótido. El término incluye formas de ADN mono y bicatenarias. Los términos "ADN" y "polinucleótido" y "ácido nucleico" se refieren a una molécula de ADN que se ha aislado del ADN genómico total de una especie particular. Por lo tanto, un segmento de ADN aislado que codifica un polipéptido se refiere a un segmento de ADN que contiene una o más  
60 secuencias codificantes, aunque esta sustancialmente aislado de o purificado respecto de, el ADN genómico total de la especie de la que se obtiene el ADN genómico. También se incluyen polinucleótidos no codificantes (por ejemplo,  
65

cebadores, sondas, oligonucleótidos), que no codifican un polipéptido de AARS. En los términos "segmento de ADN" y "polinucleótido" se incluyen segmentos de ADN y fragmentos más pequeños de dichos segmentos y también vectores recombinantes, incluyendo, por ejemplo, plásmidos, cósmidos, fagémidos, fagos, virus y similares.

- 5 Pueden estar presentes, aunque no necesariamente, secuencias codificantes o no codificantes adicionales dentro de un polinucleótido de la presente divulgación y un polinucleótido puede, pero no necesariamente, estar unido a otras moléculas y/o materiales de soporte. Por tanto, los polinucleótidos de la presente invención, independientemente de la longitud de la propia secuencia codificante, pueden combinarse con otras secuencias de ADN, tales como promotores, señales de poliadenilación, sitios de enzimas de restricción adicionales, sitios de clonación múltiple, otros segmentos codificantes y similares, de tal forma que su longitud general puede variar considerablemente.

15 Por lo tanto, se contempla que pueda emplearse un fragmento de polinucleótido de prácticamente cualquier longitud; estando preferentemente limitada la longitud general por la facilidad de preparación y uso en el protocolo de ADN recombinante previsto. Se incluyen polinucleótidos de aproximadamente 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 41, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 270, 280, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000 o más (incluyendo todos los números enteros intermedios) bases de longitud, incluyendo cualquier porción o fragmento (por ejemplo, más de aproximadamente 6, 7, 8, 9 o 10 nucleótidos de longitud) de un polinucleótido de AARS de referencia (por ejemplo, un número de bases X-Y, en el que X es aproximadamente 1-3000 o más e Y es aproximadamente 10-3000 o más), o su complemento.

25 Las realizaciones de la presente divulgación también incluyen "variantes" de las secuencias de polinucleótido de AARS de referencia. Las "variantes" de polinucleótido pueden contener una o más sustituciones, adiciones, eliminaciones y/o inserciones en relación a un polinucleótido de referencia. En general, las variantes de una secuencia de polinucleótido de AARS de referencia pueden tener al menos aproximadamente un 30 %, 40 %, 50 %, 55 %, 60 %, 65 %, 70 %, generalmente al menos un 75 %, 80 %, 85 %, deseablemente de aproximadamente un 90 % a un 95 % o más y de manera más adecuada aproximadamente un 98 % o más de identidad de secuencia con dicha secuencia de nucleótidos particular, según se determina mediante programas de alineamiento de secuencias descritos en otras partes del presente documento usando parámetros por defecto. En determinadas realizaciones, las variantes pueden diferir respecto de una secuencia de referencia en aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 41, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100 (incluyendo todos los números enteros intermedios) o más bases. En determinadas realizaciones, tal como cuando la variante de polinucleótido codifica un polipéptido de AARS que tiene una actividad no canónica, la actividad deseada del polipéptido de AARS codificado no se reduce sustancialmente en relación al polipéptido no modificado. El efecto sobre la actividad del polipéptido codificado puede evaluarse generalmente tal como se describe en el presente documento.

40 Determinadas realizaciones incluyen polinucleótidos que hibridan con una secuencia de polinucleótido de AARS de referencia o con sus complementos, en condiciones de rigurosidad descritas más adelante. Tal como se usa en el presente documento, la expresión "hibrida en condiciones de baja rigurosidad, media rigurosidad, alta rigurosidad o muy alta rigurosidad" describe condiciones para la hibridación y lavado. Puede obtenerse orientación para llevar a cabo reacciones de hibridación en Ausubel *et al.*, (1998, anteriormente citado), Secciones 6.3.1-6.3.6. En esa referencia se describen métodos acuosos y no acuosos y pueden usarse cualquiera de los dos.

50 La referencia en el presente documento a condiciones de baja rigurosidad abarca de al menos aproximadamente un 1 % v/v a al menos aproximadamente un 15 % v/v de formamida y de al menos aproximadamente 1 M a al menos aproximadamente 2 M de sal para la hibridación a 42 °C y de al menos aproximadamente 1 M a al menos aproximadamente 2 M de sal para el lavado a 42 °C. Las condiciones de baja rigurosidad también pueden incluir un 1 % de seroalbúmina bovina (BSA), EDTA 1 mM, NaHPO<sub>4</sub> 0,5 M (pH 7,2), un 7 % de SDS para hibridación a 65 °C e (i) 2 x SSC, SDS al 0,1 %; o (ii) BSA al 0,5 %, EDTA 1 mM, NaHPO<sub>4</sub> 40 mM (pH 7,2), SDS al 5 % para el lavado a temperatura ambiente. Una realización de condiciones de baja rigurosidad incluye hibridación en cloruro de sodio/citrato de sodio (SSC) 6 x a aproximadamente 45 °C, seguido de dos lavados en 0,2 x SSC, SDS al 0,1 % a al menos 50 °C (puede aumentarse la temperatura de los lavados a 55 °C para las condiciones de baja rigurosidad).

60 Las condiciones de rigurosidad media incluyen y abarcan de al menos aproximadamente un 16 % v/v a al menos aproximadamente un 30 % v/v de formamida y de al menos aproximadamente 0,5 M a al menos aproximadamente 0,9 M de sal para la hibridación a 42 °C; y de al menos aproximadamente 0,1 M a al menos aproximadamente 0,2 M de sal para el lavado a 55 °C. Las condiciones de rigurosidad media también pueden incluir seroalbúmina bovina (BSA) al 1 %, EDTA 1 mM, NaHPO<sub>4</sub> 0,5 M (pH 7,2), un 7 % de SDS para hibridación a 65 °C e (i) 2 x SSC, SDS al 0,1 %; o (ii) BSA al 0,5 %, EDTA 1 mM, NaHPO<sub>4</sub> 40 mM (pH 7,2), SDS al 5 % para el lavado a 60-65 °C. Una realización de condiciones de rigurosidad media incluye hibridar en 6 x SSC a aproximadamente 45 °C, seguido de uno o más lavados en 0,2 x SSC, SDS al 0,1 % a 60 °C. Las condiciones de rigurosidad elevada incluyen y abarcan de al menos aproximadamente un 31 % v/v a al menos aproximadamente un 50 % v/v de formamida y de aproximadamente 0,01 M a aproximadamente 0,15 M de sal para la hibridación a 42 °C y de aproximadamente 0,01

M a aproximadamente 0,02 M de sal para el lavado a 55 °C.

Las condiciones de rigurosidad elevada también pueden incluir BSA al 1 %, EDTA 1 mM, NaHPO<sub>4</sub> 0,5 M (pH 7,2), un 7 % de SDS para hibridación a 65 °C e (i) 0,2 x SSC, SDS al 0,1 %; o (ii) BSA al 0,5 %, EDTA 1 mM, NaHPO<sub>4</sub> 40 mM (pH 7,2), SDS al 1 % para el lavado a una temperatura superior a 65 °C. Una realización de condiciones de rigurosidad elevada incluye hibridar en 6 x SSC a aproximadamente 45 °C, seguido de uno o más lavados en 0,2 x SSC, SDS al 0,1 % a 65 °C. Una realización de condiciones de muy alta rigurosidad incluye hibridación en fosfato de sodio 0,5 M, SDS al 7 % a 65 °C, seguido de uno o más lavados en 0,2 x SSC, SDS al 1 % a 65 °C.

Se conocen en la técnica otras condiciones de rigurosidad y un experto en la materia reconocerá que pueden manipularse diversos factores para optimizar la especificidad de la hibridación. La optimización de la rigurosidad de los lavados finales puede servir para asegurar un alto grado de hibridación. Para ejemplos detallados, véase Ausubel *et al.*, anteriormente citado, en las páginas 2.10.1 a 2.10.16 y Sambrook *et al.* (1989, anteriormente citado) en las secciones 1.101 a 1.104.

Aunque los lavados rigurosos se suelen llevar a cabo a temperaturas de aproximadamente 42 °C a 68 °C, un experto en la materia apreciará que pueden ser adecuadas otras temperaturas para las condiciones rigurosas. La velocidad máxima de hibridación se produce normalmente de aproximadamente 20 °C a 25 °C por debajo de la T<sub>m</sub> para la formación de un híbrido de ADN-ADN. Es de sobra conocido en la técnica que la T<sub>m</sub> es la temperatura de fusión o la temperatura a la cual se disocian dos secuencias de polinucleótido complementarias. Los métodos para estimar la T<sub>m</sub> se conocen bien en la técnica (véase Ausubel *et al.*, anteriormente citado, en la página 2.10.8).

En general, la T<sub>m</sub> de un dúplex de ADN perfectamente emparejado puede predecirse de manera aproximada mediante la fórmula:  $T_m = 81,5 + 16,6 (\log_{10} M) + 0,41 (\%G+C) - 0,63 (\% \text{ formamida}) - (600/\text{longitud})$  en donde: M es la concentración de Na<sup>+</sup>, preferentemente en el intervalo de 0,01 molar a 0,4 molar; %G+C es la suma de bases de guanosina y citosina como porcentaje del número total de bases, dentro del intervalo entre un 30 % y un 75 % de G+C; % formamida es el porcentaje de concentración de formamida en volumen; longitud es el número de pares de bases en el dúplex de ADN. La T<sub>m</sub> de un dúplex de ADN se reduce en aproximadamente 1 °C con cada aumento del 1 % en el número de pares de bases desemparejadas al azar. El lavado se lleva a cabo generalmente a T<sub>m</sub> - 15 °C para alta rigurosidad o T<sub>m</sub> - 30 °C para rigurosidad moderada.

En un ejemplo de un procedimiento de hibridación, se hibrida una membrana (por ejemplo, una membrana de nitrocelulosa o una membrana de nylon) que contiene ADN inmovilizado durante una noche a 42 °C en un tampón de hibridación (formamida desionizada al 50 %, 5 x SSC, 5 x solución de Denhardt (ficoll al 0,1 %, polivinilpirrolidona al 0,1 % y seroalbúmina bovina al 0,1 %), SDS al 0,1 % y 200 mg/ml de ADN de esperma de salmón desnaturalizado) que contiene una sonda marcada. Después, se somete a la membrana a dos lavados de rigurosidad media secuenciales (es decir, 2 x SSC, SDS al 0,1 % durante 15 min a 45 °C, seguido de 2 x SSC, SDS al 0,1 % durante 15 min a 50 °C), seguido de dos lavados secuenciales de mayor rigurosidad (es decir, 0,2 x SSC, SDS al 0,1 % durante 12 min a 55 °C seguido de 0,2 x SSC y solución de SDS al 0,1 % durante 12 min a 65-68 °C).

Como se ha indicado anteriormente, determinadas realizaciones se refieren a polinucleótidos de AARS que codifican un polipéptido de AARS. Entre otros usos, estas realizaciones pueden utilizarse para producir de manera recombinante un polipéptido de AARS deseado o una variante del mismo o para expresar el polipéptido de AARS en una célula o sujeto seleccionado. Los expertos en la materia apreciarán que, como consecuencia de la degeneración del código genético, hay muchas secuencias de nucleótidos que codifican un polipéptido tal como se describe en el presente documento. Algunos de estos polinucleótidos pueden portar una homología mínima para la secuencia de nucleótidos de cualquier gen nativo. Sin embargo, los polinucleótidos que varían debido a diferencias en el uso de codones se contemplan específicamente por la presente invención, por ejemplo, polinucleótidos que están optimizados para la selección de codones humanos y/o de primate.

Por lo tanto, los polinucleótidos de la invención pueden codificar los polipéptidos de AARS de la invención. Asimismo, puede manipularse la secuencia de polinucleótido por diversos motivos. Los ejemplos incluyen, pero sin limitación, la incorporación de codones preferidos para potenciar la expresión del polinucleótido en diversos organismos (véase, de manera general Nakamura *et al.*, Nuc. Acid. Res. (2000) 28 (1): 292). Además, pueden incorporarse mutaciones silentes para introducir o eliminar sitios de restricción, reducir la densidad de motivos de dinucleótido CpG (véase, por ejemplo, Kameda *et al.*, Biochem. Biophys. Res. Commun. (2006) 349(4): 1269-1277) o reducir la capacidad de las secuencias monocatenarias para formar estructuras de tallo-bucle: (véase, *por ejemplo*, Zuker M., Nucl. Acid Res. (2003); 31(13): 3406-3415). Además, puede optimizarse adicionalmente la expresión en mamíferos incluyendo una secuencia consenso Kozak [es decir, (a/g)cc(a/g)ccATGg] en el codón de inicio. Las secuencias consenso Kozak útiles para este fin se conocen en la técnica (Mantyh *et al.* PNAS 92: 2662-2666 (1995); Mantyh *et al.* Prot. Exp. & Purif. 6.124 (1995)).

Los polinucleótidos de la presente invención, independientemente de la longitud de la propia secuencia codificante, pueden combinarse con otras secuencias de ADN, tales como promotores, señales de poliadenilación, sitios de enzimas de restricción adicionales, sitios de clonación múltiple, otros segmentos codificantes y similares, de tal forma que su longitud general puede variar considerablemente. Por lo tanto, se contempla que pueda emplearse un

fragmento de polinucleótido de prácticamente cualquier longitud; estando preferentemente limitada la longitud general por la facilidad de preparación y uso en el protocolo de ADN recombinante previsto.

5 Pueden prepararse, manipularse y/o expresarse polinucleótidos y fusiones de los mismos usando cualquiera de diversas técnicas bien establecidas conocidas y disponibles en la técnica. Por ejemplo, pueden usarse secuencias de polinucleótido que codifican polipéptidos de la invención o proteínas de fusión o equivalentes funcionales de las mismas en moléculas de ADN recombinante para dirigir la expresión de un polipéptido de AARS en células hospedadoras adecuadas. Debido a la degeneración inherente del código genético, pueden producirse otras secuencias de ADN que codifican sustancialmente la misma secuencia de aminoácidos o una funcionalmente equivalente y pueden usarse estas secuencias para clonar y expresar un polipéptido dado.

15 Como apreciarán los expertos en la materia, puede ser ventajoso, en algunos casos, producir secuencias de nucleótido codificantes de polipéptidos que posean codones de origen no natural. Por ejemplo, los codones preferidos por un hospedador procarionta o eucariota particular pueden seleccionarse para aumentar la tasa de expresión de proteína o para producir un transcrito de ARN recombinante que tiene propiedades deseables, tales como una semivida que es más larga que la de un transcrito generado a partir de la secuencia de origen natural. Dichos polinucleótidos se citan comúnmente como "optimizado por codones". Pueden utilizarse cualquiera de los polinucleótidos descritos en el presente documento de un modo optimizado por codones. En determinadas realizaciones, puede optimizarse por codones un polinucleótido para su uso en bacterias específicas, tales como *E. coli* o levaduras, tales como *S. cerevisiae* (véase, *por ejemplo*, Burgess-Brown et al., Protein Expr Purif. 59:94-102, 2008; Ermolaeva MD (2001) Curr. Iss. Mol. Biol. 3 (4) 91-7; Welch et al., PLoS ONE 4(9): e7007 doi:10.1371/journal.pone.0007002).

25 Asimismo, las secuencias de polinucleótido de la presente invención pueden modificarse por ingeniería genética usando métodos conocidos de manera general en la técnica para alterar secuencias codificantes de polipéptido por diversas razones, incluyendo, pero sin limitación, alteraciones que modifican la clonación, procesamiento, expresión y/o actividad del producto génico.

30 De acuerdo con otro aspecto de la divulgación, pueden suministrarse a un sujeto *in vivo* polinucleótidos que codifican polipéptidos de la invención, *por ejemplo*, usando técnicas de terapia génica. La terapia génica se refiere generalmente a la transferencia de ácidos nucleicos heterólogos a determinadas células, células diana, de un mamífero, en particular un ser humano, con un trastorno o afección para la que está pensada dicha terapia. El ácido nucleico se introduce en las células diana seleccionadas de un modo tal que se expresa el ADN heterólogo y se produce un producto terapéutico codificado por el mismo.

35 Los diversos vectores víricos para la terapia génica tal como se enseña en el presente documento que pueden usarse incluyen adenovirus, herpes virus, vaccinia, virus adenoasociados (AAV), o preferentemente, un virus de ARN, tal como un retrovirus. Preferentemente, el vector retroviral es un derivado de un retrovirus murino o aviar o es un vector lentiviral. El vector retroviral preferido es un vector lentiviral. Los ejemplos de vectores retrovirales en los que puede insertarse un solo gen exógeno incluyen, pero sin limitación: el virus de la leucemia murina Moloney (MoMuLV), el virus de sarcoma murino Harvey (HaMuSV), el virus del tumor mamario murino (MuMTV), VIS, VIB, VIH y el virus del sarcoma de Rous (VSR). Una serie de vectores retrovirales adicionales pueden incorporar múltiples genes. Todos estos vectores pueden transferir o incorporar un gen para un marcador de selección, de tal forma que puedan identificarse y generarse las células transducidas. Al insertar una secuencia de polipéptido de unión a ADN derivada de dedo de cinc de interés en el vector vírico, junto con otro gen que codifica el ligando para un receptor en una célula diana específica, por ejemplo, el vector puede hacerse específico para la diana. Los vectores retrovirales pueden hacerse específicos para la diana insertando, por ejemplo, un polinucleótido que codifica una proteína (dímero). Puede lograrse un direccionamiento a modo de ejemplo usando un anticuerpo para dirigir el vector retroviral. Los expertos en la materia serán conocedores de o podrán determinar fácilmente sin experimentación innecesaria, secuencias de polinucleótido específicas que pueden insertarse en el genoma retroviral para permitir un direccionamiento específico del vector retroviral que contiene el polinucleótido de proteína de unión a nucleótidos de dedo de cinc.

55 Ya que los retrovirus recombinantes son defectuosos, necesitan asistencia para producir partículas víricas infecciosas. Esta asistencia puede proporcionarse, por ejemplo, usando líneas celulares colaboradoras que contienen plásmidos que codifican la totalidad de los genes estructurales del retrovirus bajo el control de secuencias reguladoras dentro del LTR. Estos plásmidos carecen de una secuencia de nucleótido que posibilite el mecanismo de empaquetado para reconocer un transcrito de ARN para su encapsulación. Las líneas celulares colaboradoras que tienen eliminaciones de la señal de empaquetado incluyen, pero sin limitación, PSI.2, PA317 y PA12, por ejemplo. Estas líneas celulares producen viriones vacíos, ya que no se empaqueta genoma alguno. En caso de que se introduzca un vector retroviral en dichas células en las que la señal de empaquetado está intacta, pero los genes estructurales están reemplazados por otros genes de interés, puede empaquetarse el vector y producirse el virión del vector. Los viriones de vector producidos mediante este método pueden usarse entonces para infectar una línea celular de tejido, tal como células NIH 3T3, para producir grandes cantidades de viriones retrovirales quiméricos.

65 También pueden emplearse técnicas de suministro "no vírico" para terapia génica que incluyen, por ejemplo,

complejos de ADN-ligando, complejos de adenovirus-ligando-ADN, inyección directa de ADN, precipitación con  $\text{CaPO}_4$ , técnicas de cañón de genes, electroporación, liposomas, lipofección y similares. Se encuentra ampliamente disponible cualquiera de estos métodos para un experto en la materia y serían adecuados para su uso en la presente invención. Se encuentran disponibles para los expertos en la materia otros métodos adecuados y se entenderá que la presente invención puede llevarse a cabo usando cualquiera de los métodos de transfección disponibles. Puede lograrse la lipofección encapsulando una molécula de ADN aislado dentro de una partícula liposómica y poniendo en contacto la partícula liposómica con la membrana celular de la célula diana. Los liposomas son partículas coloidales autoensamblantes en las que una bicapa lipídica, compuesta de moléculas anfífilas, tales como fosfatidil serina o fosfatidil colina, encapsulan una porción del medio circundante, de tal forma que la bicapa lipídica rodea un interior hidrófilo. Pueden construirse liposomas unilamelares o multilamelares, de tal forma que el interior contiene un agente químico deseado, un fármaco o, como en la presente invención, una molécula de ADN aislada.

En otro aspecto, los polinucleótidos que codifican polipéptidos de la invención pueden usarse para expresar y suministrar un polipéptido de AARS mediante terapia celular. De acuerdo con otro aspecto, la presente divulgación incluye una terapia celular para tratar una enfermedad o trastorno, que comprende administrar una célula hospedadora que expresa o que es capaz de expresar, un polipéptido de AARS.

La terapia celular implica la administración de células que se han seleccionado, multiplicado y tratado o alterado farmacológicamente (es decir, modificado por ingeniería genética) fuera del organismo (Bordignon, C. et al, Cell Therapy: Achievements and Perspectives (1999), Haematologica, 84, págs. 1110-1149). Dichas células hospedadoras incluyen, por ejemplo, células primarias, incluyendo macrófagos y células madre que se han modificado genéticamente para expresar un polipéptido de AARS. El objetivo de la terapia celular es reemplazar, reparar o potenciar la función biológica de tejidos u órganos dañados.

Se ha investigado el uso de células trasplantadas para el tratamiento de numerosos trastornos endocrinos, tales como la anemia y el enanismo, trastornos hematológicos, insuficiencia renal y hepática, deficiencias pituitarias y del SNC y diabetes mellitus (Uludag et al., Technology of Mammalian Cell Encapsulation (2000), Advanced Drug Delivery Reviews, 42, págs. 29-64). Las células trasplantadas pueden funcionar liberando compuestos bioactivos, tales como un polipéptido de AARS de la invención, para reemplazar polipéptidos de AARS endógenos que están ausentes o se producen en cantidades insuficientes en un sistema afectado.

Las realizaciones de la presente divulgación también incluyen oligonucleótidos, ya sea para la detección, amplificación, terapias antisentido u otros fines. Para estos fines y otros relacionados, el término "oligonucleótido" u "oligo" u "oligómero" pretende abarcar un oligonucleótido individual, así como múltiples "oligonucleótidos" y se refiere a cualquier polímero de dos o más nucleótidos, nucleósidos, nucleobases o compuestos relacionados usados como reactivo en los métodos de amplificación de la presente invención, así como métodos de detección posteriores. El oligonucleótido puede ser ADN y/o ARN y/o análogos de los mismos.

El término oligonucleótido no indica necesariamente cualquier función particular para el reactivo, en cambio, se usa de manera genérica para abarcar todos los reactivos similares descritos en el presente documento. Un oligonucleótido puede cumplir diversas funciones diferentes, *por ejemplo*, puede funcionar como cebador si es capaz de hibridar con una hebra complementaria y puede además extenderse en presencia de una polimerasa de ácido nucleico, puede proporcionar un promotor en caso de que contenga una secuencia reconocida por una ARN polimerasa y permita la transcripción y puede funcionar previniendo la hibridación o impedir la extensión de cebador en caso de que se sitúe y/o modifique de manera adecuada. Un oligonucleótido también puede funcionar como sonda o como agente antisentido. Un oligonucleótido puede ser prácticamente de cualquier longitud, limitada únicamente por su función específica, *por ejemplo*, en una reacción de amplificación, detectando un producto de amplificación de la reacción de amplificación o en una aplicación antisentido o de interferencia de ARN. Puede usarse cualquiera de los oligonucleótidos descritos en el presente documento como un cebador, una sonda, un oligómero antisentido o un agente de interferencia de ARN.

El término "cebador", tal como se usa en el presente documento, se refiere aun oligonucleótido monocatenario capaz de actuar como punto de inicio para la síntesis de ADN dirigida por molde en condiciones adecuadas definidas, *por ejemplo*, por el tampón y la temperatura, en presencia de cuatro nucleósidos trifosfato diferentes y un agente para la polimerización, tal como una ADN o ARN polimerasa o retrotranscriptasa. La longitud del cebador, en cualquier caso concreto, depende, *por ejemplo*, del uso previsto del cebador y generalmente varía de aproximadamente 15 a 30 nucleótidos, aunque pueden usarse cebadores más cortos y más largos. Las moléculas de cebador cortas requieren generalmente temperaturas más frías para formar complejos híbridos con el molde lo suficientemente estables. Un cebador no necesita reflejar la secuencia exacta del molde, sino que debe ser lo suficientemente complementario para hibridar con dicho molde. El sitio del cebador es la zona del molde con la que hibrida un cebador. El par de cebador es un conjunto de cebadores que incluye un cebador cadena arriba 5' que hibrida con el extremo 5' de la secuencia que se va a amplificar y un cebador cadena abajo 3' que hibrida con el complemento del extremo 3' de la secuencia que se va a amplificar.

El término "sonda", tal como se usa en el presente documento, incluye una molécula inmovilizada sobre una

superficie o soluble pero que puede inmovilizarse que puede ser reconocida por una diana particular. Véase, *por ejemplo*, la Patente de los Estados Unidos n.º 6.582.908 para un ejemplo de matrices que tienen todas las combinaciones de sondas posibles con 10, 12 y más bases. Las sondas y los cebadores usados en el presente documento comprenden normalmente al menos 10-15 nucleótidos contiguos de una secuencia conocida. Para

5 mejorar la especificidad, también pueden emplearse cebadores y sondas más largos, tales como sondas y cebadores que comprenden al menos 20, 25, 30, 40, 50, 60, 70, 80, 90, 100 o al menos 150 nucleótidos de una secuencia de referencia de AARS o su complemento. Las sondas y los cebadores pueden ser considerablemente más largos que estos ejemplos y se entenderá que puede emplearse cualquier longitud soportada por el conocimiento de la técnica y la memoria descriptiva, incluyendo las tablas, las figuras y el listado de secuencias.

10 Los métodos para preparar y usar sondas y cebadores se describen en las referencias, por ejemplo, Sambrook, J. et al. (1989) *Molecular Cloning: A Laboratory Manual*, 2ª ed., vol. 1-3, Cold Spring Harbor Press, Plainview N.Y.; Ausubel, F. M. et al. (1987) *Current Protocols in Molecular Biology*, Greene Publ. Assoc. & Wiley-Intersciences, Nueva York N.Y.; Innis, M. et al. (1990) *PCR Protocols. A Guide to Methods and Applications*, Academic Press, San

15 Diego Calif. Los pares de cebadores de PCR pueden obtenerse de una secuencia conocida, por ejemplo, usando programas informáticos previstos para ese fin, tales como Primer (versión 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge Mass.).

20 Los oligonucleótidos para su uso como cebadores o sondas pueden seleccionarse usando programas informáticos conocidos en la técnica. Por ejemplo, el programa informático OLIGO 4.06 es útil para la selección de pares de cebadores de PCR de hasta 100 nucleótidos cada uno y para el análisis de oligonucleótido y polinucleótidos más largos de hasta 5000 nucleótidos a partir de una secuencia de polinucleótido de entrada de hasta 32 kilobases. Otros programas similares para la selección de cebadores han incorporado características adicionales para

25 capacidades expandidas. Por ejemplo, el programa para la selección de cebadores PrimOU (disponible para el público a través del Genome Center at University of Texas South West Medical Center, Dallas Tex.) tiene la capacidad de seleccionar cebadores específicos procedentes de secuencias de megabases y por lo tanto es útil para diseñar cebadores con una perspectiva de todo el genoma.

30 El programa para la selección de cebadores Primer3 (disponible para el público a través del Whitehead Institute/MIT Center for Genome Research, Cambridge Mass.) permite al usuario introducir una "biblioteca de cebado incorrecto", en la que el usuario especifica qué secuencias se han de evitar como sitios de unión a cebador. Primer3 es útil, en particular, para la selección de oligonucleótidos para micromatrices. (El código fuente para estos dos últimos programas de selección de cebadores puede obtenerse también de sus fuentes respectivas y modificarse para

35 satisfacer las necesidades específicas del usuario). El programa PrimeGen (disponible para el público a través del UK Human Genome Mapping Project Resource Centre, Cambridge, R.U.) diseña cebadores basándose en múltiples alineamientos de secuencia, permitiendo de este modo la selección de cebadores que hibridan con las regiones más conservadas o las regiones menos conservadas de secuencias de ácido nucleico alineadas. Por tanto, este programa es útil para la identificación de oligonucleótidos y fragmentos de polinucleótido conservados. Los oligonucleótidos y fragmentos de polinucleótido identificados mediante cualquiera de los métodos de selección

40 anteriores son útiles en las tecnologías de hibridación, por ejemplo, como cebadores de secuenciación PCR, elementos de micromatriz o sondas específicas para identificar polinucleótidos total o parcialmente complementarios en una muestra de ácido nucleico. Los métodos para la selección de oligonucleótidos no se limitan a los descritos en el presente documento.

45 En determinadas realizaciones, pueden prepararse oligonucleótidos mediante síntesis escalonada en fase sólida, empleando métodos detallados en las referencias citadas anteriormente y más adelante respecto de la síntesis de oligonucleótidos que tienen una mezcla de enlaces de cadena principal no cargados y catiónicos. En algunos casos, puede ser deseable añadir restos químicos adicionales al oligonucleótido, *por ejemplo*, para mejorar la farmacocinética o para facilitar la captura o la detección del compuesto. Dicho resto puede unirse covalentemente,

50 normalmente a un extremo del oligómero, de acuerdo con método sintéticos convencionales. Por ejemplo, la adición de un resto de polietilenglicol u otro polímero hidrófilo, *por ejemplo*, uno que tenga 10-100 subunidades monoméricas, puede ser útil para mejorar la solubilidad. Uno o más grupos cargados, *por ejemplo*, grupos cargados aniónicos, tales como un ácido orgánico, pueden mejorar la captación celular.

55 Pueden emplearse diversas moléculas detectables para hacer que un oligonucleótido o una proteína sea detectable, tales como radioisótopos, fluorocromos, tintes, enzimas, nanopartículas, marcadores quimioluminiscentes, biotina u otro monómero conocido en la técnica que pueda detectarse directamente (por ejemplo, por emisión de luz) o indirectamente (por ejemplo, por unión de un anticuerpo marcado fluorescentemente).

60 Los radioisótopos proporcionan ejemplos de moléculas detectables que pueden utilizarse en determinados aspectos de la presente invención. Pueden usarse varios radioisótopos como moléculas detectables para marcar nucleótidos o proteínas, incluyendo, por ejemplo, <sup>32</sup>P, <sup>33</sup>P, <sup>35</sup>S, <sup>3</sup>H y <sup>125</sup>I. Estos radioisótopos tienen semividas, tipos de desintegración y niveles de energía diferentes que pueden ajustarse a medida para satisfacerlas necesidades de un protocolo particular. Por ejemplo, el <sup>3</sup>H es un emisor de baja energía que da como resultado bajos niveles de fondo,

65 aunque esta baja energía también da como resultado largos periodos de tiempo de autorradiografía. Se encuentran disponibles comercialmente ribonucleótidos, desoxirribonucleótidos y aminoácidos marcados radiactivamente. Hay

disponibles nucleótidos que están marcados radiactivamente en el primer grupo fosfato o  $\alpha$ , o en el tercer grupo fosfato o  $\gamma$ . Por ejemplo, se encuentran disponibles comercialmente tanto  $[\alpha - ^{32}\text{P}] \text{ dATP}$  como  $[\gamma - ^{32}\text{P}] \text{ dATP}$ . Además, también se encuentran disponibles comercialmente diferentes actividades específicas para nucleótidos marcados radiactivamente y pueden diseñarse a medida para diferentes protocolos.

5 Otros ejemplos de moléculas detectables que pueden utilizarse para detectar un oligonucleótido incluyen fluoróforos. Pueden emplearse varios fluoróforos para el marcaje de nucleótidos incluyendo, por ejemplo, fluoresceína, tetrametilrodamina, rojo Texas y varios otros (por ejemplo, Haugland, Handbook of Fluorescent Probes - 9ª Ed., 2002, Molec. Probes, Inc., Eugene OR; Haugland, The Handbook: A Guide to Fluorescent Probes and Labeling  
10 Technologies-10ª Ed., 2005, Invitrogen, Carlsbad, CA).

Como ejemplo, los oligonucleótidos pueden marcarse fluorescentemente durante la síntesis química, ya que la incorporación de aminas o tioles durante la síntesis de nucleótidos permite la adición de fluoróforos. Se encuentran disponibles comercialmente nucleótidos marcados fluorescentemente. Por ejemplo, hay disponibles uridina y desoxiuridina trifosfatos que están conjugados a diez diferentes fluoróforos que abarcan el espectro. Los colorantes fluorescentes que pueden unirse directamente a los nucleótidos también pueden usarse como moléculas detectables. Por ejemplo, FAM, JOE, TAMRA y ROX son colorantes fluorescentes de amina reactiva que se han unido a nucleótidos y se usan en la secuenciación de ADN automatizada. Estos nucleótidos marcados fluorescentemente, por ejemplo, ROX-ddATP, ROX-ddCTP, ROX-ddGTP y ROX-ddUTP, se encuentran disponibles  
15 comercialmente.

También hay disponibles moléculas detectables no radiactivas y no fluorescentes. Como se ha indicado anteriormente, la biotina puede unirse directamente a los nucleótidos y detectarse mediante la unión específica y de alta afinidad a avidina o estreptavidina que se ha acoplado químicamente a una enzima que cataliza una reacción colorimétrica (tal como fosfatasa, luciferasa o peroxidasa). De manera similar, pueden usarse nucleótidos marcados con digoxigenina para la detección no isotópica de ácidos nucleicos. Se encuentran disponibles comercialmente nucleótidos biotinilados y marcados con digoxigenina.  
25

También pueden usarse partículas muy pequeñas, denominadas nanopartículas, para marcar sondas de oligonucleótido. Estas partículas tienen un tamaño en el intervalo de 1-1000 nm e incluyen diversas estructuras químicas, tales como partículas de oro y plata y puntos cuánticos. Cuando se irradian con una luz blanca incidente con un ángulo determinado, las nanopartículas de plata u oro en el intervalo de 40-120 nm dispersarán luz monocromática con una alta intensidad. La longitud de onda de la luz dispersada depende del tamaño de la partícula. De cuatro a cinco partículas diferentes muy próximas entre sí dispersarán cada una luz monocromática, que cuando se superponga proporcionará un color único específico. Las partículas se fabrican por compañías tales como Genicon Sciences (Carlsbad, CA). Las partículas de plata u oro derivatizadas pueden unirse a una gran variedad de moléculas que incluyen proteínas, anticuerpos, moléculas pequeñas, ligandos de receptor y ácidos nucleicos. Por ejemplo, puede derivatizarse químicamente la superficie de la partícula para permitir la unión a un nucleótido.  
30

Otros tipos de nanopartículas que pueden usarse para la detección de una molécula detectable incluyen puntos cuánticos. Los puntos cuánticos son cristales fluorescentes de 1-5 nm de diámetro que son excitables mediante luz dentro de un amplio intervalo de longitudes de onda. Tras su excitación mediante luz que tenga una longitud de onda adecuada, estos cristales emiten luz, tal como luz monocromática, con una longitud de onda que depende de su composición química y tamaño. Los puntos cuánticos tales como CdSe, ZnSe, InP o InAs poseen propiedades ópticas únicas; estos puntos cuánticos y otros similares se encuentran disponibles de una serie de fuentes comerciales (por ejemplo, NN-Labs, Fayetteville, AR; Ocean Nanotech, Fayetteville, AR; Nanoco Technologies, Manchester, R.U.; Sigma-Aldrich, St. Louis, MO).  
35

Pueden crearse decenas de clases de partículas según el número de clases de tamaño de los cristales de puntos cuánticos. Las clases de tamaño de los cristales se crean 1) mediante un control estricto de los parámetros de formación de los cristales para crear cada clase de tamaño deseado de la partícula o 2) mediante creación de lotes de cristales con parámetros de formación de cristales poco controlados, seguido de clasificación de acuerdo con el tamaño y/o las longitudes de onda de emisión deseados. Dos ejemplos de referencias en las que se incluyen puntos cuánticos dentro de capas epitaxiales de silicio intrínsecas de dispositivos semiconductores emisores/detectores de luz son las Patentes de los Estados Unidos n.º 5.293.050 y 5.354.707 de Chapple Sokol, et al.  
40  
45  
50  
55

En determinadas realizaciones, los cebadores o sondas de oligonucleótidos pueden marcarse con uno o más colorantes emisores de luz o detectables de otro modo. La luz emitida por los colorantes puede ser luz visible o luz invisible, tal como luz ultravioleta o luz infrarroja. En realizaciones a modo de ejemplo, el colorante puede ser un colorante de transferencia de energía de resonancia de fluorescencia (FRET); un colorante de xanteno, tal como fluoresceína y rodamina; un colorante que tiene n grupo amino en la posición alfa o beta (tal como un colorante de naftilamina, 1-dimetilaminonaftil-5-sulfonato, sulfonato de 1-anilino-8-naftaleno y sulfonato de 2-p-touidín-6-naftaleno); un colorante que tiene 3-fenil-7-isocianatocoumarina; una acridina, tal como 9-isotiocianatoacridina y naranja de acridina; un pireno, un benzoxadiazol y estilbena; un colorante que tiene 3-( $\epsilon$ -carboxipentil)-3'-etil-5,5'-dimetiloxacarbocianina (CYA); 6-carboxi fluoresceína (FAM); 5 y 6-carboxirrodamina-110 (R110); 6-  
60  
65

carboxirrodamina-6G (R6G); N,N,N',N'-tetrametil-6-carboxirrodamina (TAMRA); 6-carboxi-X-rodamina (ROX); 6-carboxi-4',5'-dicloro-2',7'-dimetoxifluoresceína (JOE); ALEXA FLUOR™; Cy2; rojo Texas y rojo de rodamina; 6-carboxi-2',4,7,7'-tetraclorofluoresceína (TET); 6-carboxi-2',4,4',5',7,7'-hexaclorofluoresceína (HEX); 5-carboxi-2',4',5',7'-tetraclorofluoresceína (ZOE); NAN; NED; Cy3; Cy3.5; Cy5; Cy5.5; Cy7; y Cy7.5; IR800CW, ICG, Alexa Fluor 350; Alexa Fluor 488; Alexa Fluor 532; Alexa Fluor 546; Alexa Fluor 568; Alexa Fluor 594; Alexa Fluor 647; Alexa Fluor 680 o Alexa Fluor 750.

Los polinucleótidos y oligonucleótidos de AARS de la presente invención pueden usarse en cualquiera de las composiciones y métodos terapéuticos, diagnósticos, de investigación o de descubrimiento de fármacos descritos en el presente documento.

#### V. ANTICUERPOS

De acuerdo con otro aspecto, la presente invención proporciona además anticuerpos que muestran especificidad de unión por un polipéptido de AARS o su compañero de unión celular nativo (es decir, compañero de unión de receptor celular, lípido, carbohidrato, proteína o ácido nucleico), o un complejo los mismos y métodos de uso de los mismos. El término anticuerpo incluye las diversas variaciones de los mismos, tales como FAB, anticuerpos humanos, anticuerpos humanos modificados, cadenas individuales, anticuerpos no humanos y otros derivados del plegamiento de la inmunoglobulina que subyacen en el sistema inmunitario como ligandos para antígenos, tal como se describe en el presente documento y se conoce en la técnica. Los anticuerpos pueden usarse en cualquiera de los métodos y composiciones terapéuticas, diagnósticas, de descubrimiento de fármacos o de expresión/purificación de proteínas proporcionados en el presente documento.

Determinados anticuerpos de la presente invención difieren respecto de algunos anticuerpos producidos con anterioridad debido a que pueden distinguir entre los fragmentos de proteína AARS de las tablas 1-3 o las tablas 4-6 o las tablas 7-9 y sus correspondientes AARS de longitud completa, normalmente uniéndose con mayor afinidad a los fragmentos de proteína AARS que a las AARS de longitud completa correspondientes. En general, dichos anticuerpos pueden unirse a secuencias o estructuras únicas generadas o reveladas mediante variaciones de corte y empalme, proteólisis u otro procesamiento celular que genere un fragmento de proteína AARS de la invención (por ejemplo, procesamiento postraduccional, incluyendo, pero sin limitación, fosforilación y otras modificaciones que cambian la estructura de la proteína). En algunos aspectos, los anticuerpos pueden unirse a secuencias alrededor de una unión de corte y empalme única (por ejemplo, a una o más regiones de al menos 5 aminoácidos contiguos seleccionados entre las secuencias de unión de corte y empalme listadas en las tablas 2B, 5B u 8B o como alternativa, a cualquier secuencia de aminoácidos en C-terminal de este sitio de corte y empalme, por ejemplo, tal como se lista en las tablas 2B, 5B u 8B. Por ejemplo, dichos anticuerpos pueden tener especificidad de unión por una o más caras no expuestas a disolventes que se encuentran expuestas en el fragmento de proteína AARS pero no en la AARS de longitud completa o secuencias que no se encuentran o que son de otro modo inaccesibles en las AARS de longitud completa. Los anticuerpos también pueden unirse a estructuras tridimensionales únicas que surgen como resultado de diferencias en el plegamiento entre el fragmento de proteína AARS y la AARS de longitud completa. Dichas diferencias en el plegamiento pueden estar localizadas (por ejemplo, en un dominio o una región específica) o globalizadas. Como ejemplo, el plegamiento de fragmentos de proteína AARS pueden generar epítomos continuos o discontinuos únicos que no se encuentran en la AARS correspondiente o parental. Los ejemplos también incluyen anticuerpos que se unen específicamente a los extremos N o C-terminales generados por variaciones de corte y empalme, proteólisis u otro procesamiento celular; dichos extremos pueden ser únicos en comparación con la AARS de longitud completa o pueden no estar expuestos a la unión de anticuerpos en las versiones de longitud completa debido a que sus extremos están completa o parcialmente enterrados en la estructura general de la molécula parental de AARS de mayor tamaño.

En algunas realizaciones, los anticuerpos proporcionados en el presente documento no forman agregados, tienen una solubilidad deseada y/o tienen un perfil de inmunogenicidad que es adecuado para su uso en seres humanos, tal como se describe en el presente documento y se conoce en la técnica. También se incluyen anticuerpos que son adecuados para trabajos de producción, tal como para purificar los fragmentos de proteína AARS descritos en el presente documento. Preferentemente, pueden concentrarse los anticuerpos activos a al menos aproximadamente 10 mg/ml y opcionalmente formularse para usos bioterapéuticos.

En determinadas realizaciones, los anticuerpos son eficaces para modular una o más de las actividades no canónicas mediadas por un polipéptido de AARS de la invención. En determinadas realizaciones, por ejemplo, el anticuerpo es uno que se une a un polipéptido de AARS y/o su compañero de unión, inhibe su capacidad para interactuar entre sí y/o antagoniza la actividad no canónica del polipéptido de AARS. En determinadas realizaciones, por ejemplo, el anticuerpo se une al compañero de unión celular de un polipéptido de AARS e imita la actividad del polipéptido de AARS, tal como aumentando o agonizando la actividad no canónica mediada por el polipéptido de AARS. Por consiguiente, pueden usarse anticuerpos para diagnosticar, tratar o prevenir enfermedades, trastornos u otras afecciones que están mediadas por un polipéptido de AARS de la invención, tal como antagonizando o agonizando su actividad parcial o completamente.

Se dice que un anticuerpo o un fragmento de unión a antígeno del mismo "se une específicamente", "se une

inmunológicamente" y/o es "inmunológicamente reactivo" con un polipéptido de la invención en caso de que reaccione a un nivel detectable (en, por ejemplo, un ensayo ELISA) con el polipéptido y no reaccione detectablemente de un modo estadísticamente significativo con polipéptidos no relacionados en condiciones similares. En determinados casos, un agente de unión no interactúa significativamente con una versión de longitud completa del polipéptido de AARS.

La unión inmunológica, tal como se usa en este contexto, se refiere generalmente a las interacciones no covalentes del tipo que se produce entre una molécula de inmunoglobulina y un antígeno para el cual es específica la inmunoglobulina. La fuerza o afinidad de unión, tales como interacciones de unión inmunológica, pueden expresarse en términos de la constante de disociación ( $K_d$ ) de la interacción, en donde una  $K_d$  más pequeña representa una mayor afinidad. Las propiedades de unión inmunológica de los polipéptidos seleccionados pueden cuantificarse usando métodos bien conocidos en la técnica. Véase, *por ejemplo*, Davies et al. (1990) Annual Rev. Biochem. 59:439-473. En determinadas realizaciones ilustrativas, un anticuerpo tiene una afinidad por un fragmento de proteína AARS de al menos aproximadamente 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40 o 50 nM. En determinadas realizaciones, la afinidad del anticuerpo por un fragmento de proteína AARS es más fuerte que su afinidad por un polipéptido de AARS de longitud completa correspondiente, normalmente en aproximadamente 1,5x, 2x, 2,5x, 3x, 3,5x, 4x, 4,5x, 5x, 6x, 7x, 8x, 9x, 10x, 15x, 20x, 25x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x, 300x, 400x, 500x, 600x, 700x, 800x, 900x, 1000x o más (incluyendo todos los números enteros intermedios). En determinadas realizaciones, un anticuerpo tiene una afinidad por una proteína AARS de longitud completa correspondiente de la menos aproximadamente 0,05, 0,1, 0,25, 0,5, 0,75, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20  $\mu$ M. En determinadas realizaciones, un anticuerpo se une débilmente o de manera sustancialmente indetectable a una proteína AARS de longitud completa.

Un "sitio de unión a antígeno" o "porción de unión" de un anticuerpo, se refiere a la parte de la molécula de inmunoglobulina que participa en la unión al antígeno. El sitio de unión al antígeno está formado por restos de aminoácidos de las regiones N-terminales variables ("V") de las cadenas pesadas ("H") y ligeras ("L"). Tres series altamente divergentes dentro de las regiones V de las cadenas pesada y ligera se citan como "regiones hipervariables", que están interpuestas entre las series flanqueantes más conservadas, conocidas como "regiones marco conservadas" o "FR". Por lo tanto, el término "FR" se refiere a secuencias de aminoácidos que se encuentran de manera natural entre y adyacentes a regiones hipervariables en las inmunoglobulinas. En una molécula de anticuerpo, las tres regiones hipervariables de una cadena ligera y las tres regiones hipervariables de una cadena pesada están dispuestas relacionadas entre sí en un espacio tridimensional para formar una superficie de unión a antígeno. La superficie de unión a antígeno es complementaria a la superficie tridimensional de un antígeno unido y las tres regiones hipervariables de cada una de las cadenas pesadas y ligeras se citan como "regiones determinantes de la complementariedad" o "CDR".

Los anticuerpos pueden prepararse mediante cualquiera de diversas técnicas conocidas por los expertos en la materia. Véase, *por ejemplo*, Harlow y Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, 1988. Pueden prepararse anticuerpos monoclonales específicos para un polipéptido de interés, por ejemplo, usando la técnica de Kohler y Milstein, Eur. J. Immunol. 6:511-519, 1976 y las mejoras de la misma. También se incluyen métodos que utilizan animales transgénicos, tales como ratones, para expresar anticuerpos humanos. Véase, *por ejemplo*, Neuberger et al., Nature Biotechnology 14:826, 1996; Lonberg et al., Handbook of Experimental Pharmacology 113:49-101, 1994; y Lonberg et al., Internal Review of Immunology 13:65-93, 1995. Los ejemplos particulares incluyen la plataforma VELOCI mMUNE® de REGERNEREX® (véase, *por ejemplo*, la Patente de los Estados Unidos n.º 6.596.541). También pueden generarse o identificarse anticuerpos mediante el uso de bibliotecas de presentación en fagos o de presentación en levaduras (véase, *por ejemplo*, la Patente de los Estados Unidos n.º 7.244.592; Chao et al., Nature Protocols. 1:755-768, 2006). Los ejemplos no limitantes de bibliotecas disponibles incluyen bibliotecas clonadas o sintéticas, tales como la Human Combinatorial Antibody Library (HuCAL), en la que se representa la diversidad estructural del repertorio de anticuerpos humanos mediante siete genes de región variable de cadena pesada y siete de cadena ligera. La combinación de estos genes da lugar a 49 armazones en la biblioteca maestra. Al superponer casetes genéticos altamente variables (CDR = regiones determinantes de la complementariedad) sobre estos armazones, puede producirse el vasto repertorio de anticuerpos humanos. También se incluyen bibliotecas humanas diseñadas con fragmentos originarios del donante que codifican una región variable de cadena ligera, una CDR-3 de cadena pesada, ADN sintético que codifica la diversidad en la CDR-1 de cadena pesada y ADN sintético que codifica la diversidad en la CDR-2 de cadena pesada. Otras bibliotecas adecuadas para su uso serán evidentes para los expertos en la materia. Los polipéptidos de la presente invención pueden usarse en el proceso de purificación en, por ejemplo, una etapa de cromatografía de afinidad.

Puede producirse un fragmento "Fv" mediante escisión proteolítica preferencial de una IgM y en raras ocasiones de una molécula de inmunoglobulina IgG o IgA. Sin embargo, los fragmentos Fv se obtienen con más frecuencia usando técnicas recombinantes conocidas en la técnica. El fragmento Fv incluye un heterodímero no covalente de  $V_H::V_L$  que incluye un sitio de unión a antígeno que retiene gran parte de las capacidades de reconocimiento y unión al antígeno de la molécula de anticuerpo nativa. Véase, *por ejemplo*, Inbar et al. (1972) Proc. Nat. Acad. Sci. USA 69:2659-2662; Hochman et al. (1976) Biochem. 15:2706-2710; y Ehrlich et al.: (1980) Biochem 19:4091-4096.

Un polipéptido de Fv monocatenario ("sFv") es un heterodímero de  $V_H::V_L$  unido covalentemente que se expresa a partir de una fusión génica que incluye genes que codifican  $V_H$  y  $V_L$  unidos mediante un enlazador codificador de un péptido. Huston et al. (1988) PNAS USA. 85(16):5879-5883. Se han descrito una serie de métodos para discernir estructuras químicas para convertir las cadenas de polipéptido ligera y pesada agregadas de manera natural, pero

5 químicamente separadas de una región V de anticuerpo en una molécula sFv que se plegará en una estructura tridimensional sustancialmente similar a la estructura de un sitio de unión a antígeno. Véase, *por ejemplo*, las Patentes de los Estados Unidos n.º 5.091.513 y 5.132.405, de Huston et al.; y la Patente de los Estados Unidos n.º 4.946.778, de Ladner et al.

10 Cada una de las moléculas anteriormente descritas incluye un conjunto de CDR de cadena pesada y de cadena ligera, respectivamente interpuestos entre un conjunto de Fr de cadena pesada y de cadena ligera que proporciona soporte para las CDR y definen la relación espacial de las CDR unas respecto de otras. Tal como se usa en el presente documento, la expresión "conjunto de CDR" se refiere a tres regiones hipervariables de una región V de

15 cadena pesada o ligera. Procediendo desde el extremo N-terminal de una cadena ligera o pesada, estas regiones se anotan como "CDR1", "CDR2" y "CDR3", respectivamente. Un sitio de unión a antígeno, por lo tanto, incluye seis CDR, que comprenden el conjunto de CDR de cada una de las regiones V de cadena pesada y ligera. Un polipéptido que comprende una sola CDR (por ejemplo, una CDR1, CDR2 o CDR3) se cita en el presente documento como una "unidad de reconocimiento molecular". El análisis cristalográfico de una serie de complejos de antígeno-anticuerpo ha demostrado que los restos de aminoácido de las CDR forman un contacto extenso con el antígeno unido, en

20 donde el contacto con antígeno más extenso es con la CDR3 de cadena pesada. Por lo tanto, las unidades de reconocimiento molecular son responsables en primera instancia de la especificidad de un sitio de unión a antígeno.

Tal como se usa en el presente documento, la expresión "conjunto de FR" se refiere a las cuatro secuencias de aminoácidos flanqueantes que enmarcan las CDR de un conjunto de CDR de una región V de cadena pesada o

25 ligera. Algunos restos de FR pueden entrar en contacto con el antígeno unido; sin embargo, las FR son responsables principalmente del plegamiento de la región V en el sitio de unión a antígeno, en particular los restos de FR directamente adyacentes a las CDR. Dentro de las FR, determinados restos de aminoácido y determinadas características estructurales se encuentran muy altamente conservadas. A este respecto, todas las secuencias de región V contienen un bucle disulfuro de aproximadamente 90 restos de aminoácido. Cuando las regiones V se pliegan formando un sitio de unión, las CDR se muestran como motivos de bucle proyectados que forman una superficie de unión a antígeno. Generalmente se reconoce que hay regiones estructurales conservadas de las FR que tienen influencia sobre la forma plegada de los bucles de las CDR en determinadas estructuras "canónicas", independiente de la secuencia de aminoácidos específica de la CDR. Además, se sabe que determinados restos de FR participan en los contactos interdominio no covalentes que estabilizan la interacción de las cadenas pesada y

35 ligera de anticuerpo.

Determinadas realizaciones incluyen anticuerpos de un solo dominio (sdAb o "nanocuerpos"), que se refieren a un fragmento de anticuerpo que consiste en un solo dominio de anticuerpo variable monomérico (véase, *por ejemplo*, las Patentes de los Estados Unidos n.º 5.840.526; 5.874.541; 6.005.079, 6.765.087, 5.800.988; 5.874.541; y 6.015.695). Dichos sdAB tienen normalmente un peso molecular de aproximadamente 12-15 kDa. En determinados aspectos, los sdAB y otras moléculas de anticuerpo pueden derivarse o aislarse de los anticuerpos monocatenarios únicos de camellos y llamas inmunizados, normalmente citados como camélidos. Véase, *por ejemplo*, Conrath et al., JBC. 276:7346-7350, 2001.

45 Se ha descrito una serie de moléculas de anticuerpo "humanizadas" que comprenden un sitio de unión a antígeno derivado de una inmunoglobulina no humana, incluyendo anticuerpos quiméricos que tienen regiones V de roedor y sus CDR asociadas fusionadas a dominios constantes humanos (Winter et al. (1991) Nature 349:293-299; Lobuglio et al. (1989) Proc. Nat. Acad. Sci. USA 86:4220-4224; Shaw et al. (1987) J I mMunol. 138:4534-4538; y Brown et al. (1987) Cancer Res. 47:3577-3583), CDR de roedor injertadas en una FR de soporte humana antes de la fusión con un dominio constante de anticuerpo humano adecuado (Riechmann et al. (1988) Nature 332:323-327; Verhoeven et al. (1988) Science 239:1534-1536; y Jones et al. (1986) Nature 321:522-525), y CDR de roedor soportadas por FR chapadas de manera recombinante (Publicación de Patente Europea 519.596, publicada el 23 de diciembre de 1992). Estas moléculas "humanizadas" se diseñan para minimizar la respuesta inmunológica no deseada frente a moléculas de anticuerpo de roedor anti-humano, lo que limita la duración y la eficacia de las aplicaciones terapéuticas de estos restos en receptores humanos. Véase, *por ejemplo*, las Patentes de los Estados Unidos n.º 5.530.101; 5.585.089; 5.693.762; 6.180.370; y 7.022.500.

Los anticuerpos de la presente invención pueden usarse en cualquiera de los métodos y composiciones terapéuticas, diagnósticas, de descubrimiento de fármacos, de purificación de proteínas y analíticas descritas en el presente documento.

#### VI. ALTERNATIVAS A LOS ANTICUERPOS Y OTROS AGENTES DE UNIÓN

De acuerdo con otro aspecto, la presente divulgación proporciona además alternativas a anticuerpos u otros agentes de unión, tales como receptores solubles, adnectinas, péptidos, peptidomiméticos, moléculas pequeñas, aptámeros, etc., que muestran especificidad de unión por un polipéptido de AARS o su compañero de unión celular, tal como se

divulga en el presente documento o con una porción, variante o derivado de los mismos y composiciones y métodos para usar los mismos. Pueden usarse agentes de unión en cualquiera de los métodos y composiciones terapéuticas, diagnósticas, de descubrimiento de fármacos o de expresión/purificación de proteínas y analíticos descritos en el presente documento. Los agentes de unión de base biológica, tales como las adnectinas, los receptores solubles, los avimeros y las trinectinas son particularmente útiles.

En determinadas realizaciones, dichos agentes de unión son eficaces para modular una o más de las actividades no canónicas mediadas por un polipéptido de AARS de la invención. En algunas realizaciones, por ejemplo, el agente de unión es uno que se une a un polipéptido de AARS y/o su compañero de unión, inhibe su capacidad para interactuar entre sí y/o antagoniza la actividad no canónica del polipéptido de AARS. En determinadas realizaciones, por ejemplo, el agente de unión se une al compañero de unión celular de un polipéptido de AARS e imita la actividad del polipéptido de AARS, tal como aumentando o agonizando la actividad no canónica mediada por el polipéptido de AARS. Por consiguiente, dichos agentes de unión pueden usarse para diagnosticar, tratar o prevenir enfermedades, trastornos u otras afecciones que están mediadas por un polipéptido de AARS de la invención, tal como antagonizando o agonizando su actividad parcial o completamente.

Se dice que un agente de unión se "une específicamente" a un polipéptido de AARS de la invención o su compañero de unión celular, en caso de que reaccione a un nivel detectable (en, por ejemplo, un ensayo ELISA) con el polipéptido o su compañero de unión celular y no reaccione detectablemente de un modo estadísticamente significativo con polipéptidos no relacionados en condiciones similares. En determinados casos, un agente de unión no interactúa significativamente con una versión de longitud completa del polipéptido de AARS. En determinadas realizaciones ilustrativas, un agente de unión tiene una afinidad por un fragmento de proteínas AARS o su compañero de unión celular de al menos aproximadamente 0,01, 0,05, 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40 o 50 nM. En determinadas realizaciones, la afinidad del agente de unión por un fragmento de proteína AARS es más fuerte que su afinidad por un polipéptido de AARS de longitud completa correspondiente, normalmente en aproximadamente 1,5x, 2x, 2,5x, 3x, 3,5x, 4x, 4,5x, 5x, 6x, 7x, 8x, 9x, 10x, 15x, 20x, 25x, 30x, 40x, 50x, 60x, 70x, 80x, 90x, 100x, 200x, 300x, 400x, 500x, 600x, 700x, 800x, 900x, 1000x o más (incluyendo todos los números enteros intermedios). En determinadas realizaciones, un agente de unión tiene una afinidad por una proteína AARS de longitud completa correspondiente de la menos aproximadamente 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20  $\mu\text{M}$ .

Como se ha indicado anteriormente, los "péptidos" se incluyen como agentes de unión. El término péptido se refiere normalmente a un polímero de restos de aminoácidos y a variantes y análogos sintéticos del mismo. En determinadas realizaciones, el término "péptido" se refiere a polipéptidos relativamente cortos, incluyendo péptidos que consisten en aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45 o 50 aminoácidos, incluyendo todos los números enteros e intervalos (por ejemplo, 5-10, 8-12, 10-15) intermedios e interactúa con un polipéptido de AARS, su compañero de unión celular o ambos. Los péptidos pueden estar compuestos de aminoácidos de origen natural y/o aminoácidos de origen no natural, tal como se describen en el presente documento.

Además de los péptidos que consisten únicamente en aminoácidos de origen natural, también se proporcionan peptidomiméticos o análogos de péptidos. Los análogos de péptidos se usan comúnmente en la industria farmacéutica como fármacos no peptídicos con propiedades análogas a aquellas del péptido molde. Estos tipos de compuesto no peptídico se denominan "miméticos de péptidos" o "peptidomiméticos" (Luthman, et al., *A Textbook of Drug Design and Development*, 14:386-406, 2ª Ed., Harwood Academic Publishers (1996); Joachim Grante, *Angew. Chem. Int. Ed. Engl.*, 33:1699-1720 (1994); Fauchere, J., *Adv. Drug Res.*, 15:29 (1986); Veber y Freidinger *TINS*, p. 392 (1985); y Evans, et al., *J. Med. Chem.* 30:229 (1987)). Un peptidomimético es una molécula que imita la actividad biológica de un péptido pero su naturaleza química ya no es peptídica. Se conocen en la técnica compuestos peptidomiméticos y se describen, por ejemplo, en la Patente de los Estados Unidos n.º 6.245.886.

La presente invención también incluye peptoides. Los derivados de peptoides o péptidos representan otra forma de péptidos modificados que conservan los determinantes estructurales importantes para la actividad biológica, aunque eliminan los enlaces peptídicos, confiriendo de este modo resistencia a la proteólisis (Simon, et al., *PNAS USA*. 89:9367-9371, 1992). Los peptoides son oligómeros de glicinas N-sustituidas. Se ha descrito una serie de grupos de N-alquilo, correspondientes cada uno a la cadena lateral de un aminoácido natural. Los peptidomiméticos de la presente invención incluyen compuestos en los que al menos un aminoácido, unos pocos aminoácidos o todos los restos de aminoácidos están reemplazados por las glicinas N-sustituidas correspondientes. Se describen bibliotecas de peptoides, por ejemplo, en la Patente de los Estados Unidos n.º 5.811.387.

Un agente de unión también puede incluir una o más moléculas pequeñas. Una "molécula pequeña" se refiere a un compuesto orgánico que es de origen sintético o biológico (biomolécula), pero normalmente no es un polímero. Los compuestos orgánicos se refieren a una gran clase de compuestos bioquímicos cuyas moléculas contienen carbono, excluyendo normalmente aquellos que contienen únicamente carbonatos, óxidos de carbono sencillo o cianuros. Una "biomolécula" se refiere generalmente a una molécula orgánica que se produce por un organismo vivo, incluyendo moléculas poliméricas grandes (biopolímeros), tales como péptidos, polisacáridos y también ácidos

nucleicos y moléculas pequeñas, tales como metabolitos primarios o secundarios, lípidos, fosfolípidos, glucolípidos, esteroides, glicerolípidos, vitaminas y hormonas. Un "polímero" se refiere generalmente a una molécula o macromolécula grande compuesta de unidades estructurales repetitivas, que normalmente están conectadas mediante enlace químico covalente.

5 En determinadas realizaciones, una molécula pequeña tiene un peso molecular de menos de 1000-2000 Dalton, normalmente entre aproximadamente 300 y 700 Dalton e incluyendo aproximadamente 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000 o 2000 Dalton. En otras partes del presente documento se describen bibliotecas de molécula pequeña.

10 También se incluyen como agentes de unión los aptámeros (véase, *por ejemplo*, Ellington et al., Nature. 346, 818-22, 1990; y Tuerk et al., Science. 249, 505-10, 1990). Los ejemplos de aptámeros incluyen aptámeros de ácido nucleico (por ejemplo, aptámeros de ADN, aptámeros de ARN) y aptámeros peptídicos. Los aptámeros de ácido nucleico se refieren en general a especies de ácido nucleico que se han modificado a través de rondas repetidas de selección *in vitro* o un método equivalente, tal como SELEX (evolución sistemática de ligandos mediante enriquecimiento exponencial), para unir varias dianas moleculares, tales como moléculas pequeñas, proteínas, ácidos nucleicos e incluso células, tejidos y organismos. Véase, *por ejemplo*, las Patentes de los Estados Unidos n.º 6.376.190; y 6.387.620. Por lo tanto, se incluyen aptámeros de ácidos nucleicos que se unen a los polipéptidos de AARS descritos en el presente documento y/o sus compañeros de unión celulares.

20 Los aptámeros peptídicos incluyen normalmente un bucle de péptido variable unido en ambos extremos a un armazón de proteína, una restricción estructural doble que normalmente aumenta la afinidad de unión del aptámero peptídico a niveles comparables con los de un anticuerpo (por ejemplo, en el intervalo nanomolar). En determinadas realizaciones, la longitud del bucle variable puede estar compuesta de aproximadamente 10-20 aminoácidos (incluyendo todos los números enteros intermedios), y el armazón puede incluir cualquier proteína que tenga buenas propiedades de solubilidad y compacidad. Determinadas realizaciones a modo de ejemplo pueden utilizar la proteína bacteriana tioredoxina A como proteína de armazón, estando el bucle variable insertado dentro del sitio activo reductor (bucle -Cys-Gly-Pro-Cys- en la proteína silvestre), pudiendo las dos cadenas laterales de las cisteínas formar un puente disulfuro. Los métodos para identificar aptámeros peptídicos se describen, por ejemplo, en la Solicitud de los Estados Unidos n.º 2003/0108532. Por tanto, se incluyen aptámeros de péptidos que se unen a los polipéptidos de AARS descritos en el presente documento y/o sus compañeros de unión celulares. La selección de aptámeros de péptidos puede efectuarse usando diferentes sistemas conocidos en la técnica, incluyendo el sistema de dos híbridos de levadura.

35 También se incluyen ADNECTINS™, AVIMERS™, anafonas y anticalinas que se unen específicamente a un fragmento de proteína AARS de la invención. Las ADNECTINS™ se refieren a una clase de agentes biológicos dirigidos obtenidos de fibronectina humana, una proteína extracelular abundante que se une de manera natural a otras proteínas. Véase, *por ejemplo*, las Solicitudes de los Estados Unidos n.º 2007/0082365; 2008/0139791; y 2008/0220049. Las ADNECTINS™ consisten normalmente en un armazón de fibronectina natural, así como los dominios de direccionamiento múltiple de una porción específica de la fibronectina humana. Los dominios de direccionamiento pueden diseñarse por ingeniería genética para permitir que una ADNECTIN™ reconozca específicamente una diana terapéutica de interés, tal como un fragmento de proteína AARS de la invención.

45 Los AVIMERS™ se refieren a proteínas o péptidos de unión multiméricos diseñados por ingeniería genética usando reordenamiento de exones *in vitro* y presentación en fagos. Se unen múltiples dominios de unión, dando como resultado una mayor afinidad y especificidad en comparación con los dominios de inmunoglobulina de un solo epítipo. Véase, *por ejemplo*, Silverman et al., Nature Biotechnology. 23:1556-1561, 2005; la Patente de los Estados Unidos n.º 7.166.697; y las Solicitudes de los Estados Unidos n.º 2004/0175756, 2005/0048512, 2005/0053973, 2005/0089932 y 2005/0221384.

50 También se incluyen proteínas diseñadas de repetición de anquirina (DARPin), que incluyen una clase de proteínas no de inmunoglobulina que pueden ofrecer ventajas frente a los anticuerpos para la unión a la diana en el descubrimiento de fármacos y el desarrollo de fármacos. Entre otros usos, las DARPins son idealmente adecuadas para obtención de imágenes *in vivo* o para el suministro de toxinas u otras cargas terapéuticas debido a sus propiedades moleculares favorables, incluyendo un pequeño tamaño y una alta estabilidad. La producción de bajo coste en bacterias y la rápida generación de muchas DARPins específicas de diana hacen que la estrategia de DARPin sea útil para el descubrimiento de fármacos. Además, pueden generarse fácilmente DARPins en formatos multiespecíficos, ofreciendo el potencial de DARPin diana y efectoras para un órgano específico o para dirigirse a múltiples receptores con una molécula compuesta de varias DARPins. Véase, *por ejemplo*, Stump et al., Curr Opin Drug Discov Devel. 10:153-159, 2007; Solicitud de los Estados Unidos n.º 2009/0082274; y el documento PCT/EP2001/10454.

65 Determinadas realizaciones incluyen "monobodies", que normalmente utilizan el 10º dominio de fibronectina de tipo III de fibronectina humana (FNfn10) como armazón para presentar múltiples bucles superficiales para la unión a la diana. FNfn10 es una proteína pequeña (94 restos) con una estructura de β-sándwich similar al plegamiento de la inmunoglobulina. Es altamente estable sin enlaces disulfuro o iones metálicos y puede expresarse en la forma

plegada correcta a un alto nivel en bacterias. El armazón de FNfn10 es compatible con prácticamente cualquier tecnología de presentación. Véase, *por ejemplo*, Batori et al., Protein Eng. 15:1015-20, 2002; y Wojcik et al., Nat Struct Mol Biol., 2010; y la Patente de los Estados Unidos n.º 6.673.901.

5 Las anticalinas se refieren a una clase de miméticos de anticuerpo, que normalmente se sintetizan a partir de lipocalinas humanas, una familia de proteínas de unión con una región de bucle hipervariable soportada por un armazón estructuralmente rígido. Véase, *por ejemplo*, la Solicitud de los Estados Unidos n.º 2006/0058510. Las anticalinas tienen normalmente un tamaño de aproximadamente 20 kDa. Las anticalinas pueden caracterizarse por una estructura de barril formada por ocho hebras  $\beta$  antiparalelas (un armazón de barril  $\beta$  estable) que están  
10 conectadas por pares mediante cuatro bucles peptídicos y una hélice  $\alpha$  unida. En determinados aspectos, se practican desviaciones conformacionales en las regiones de bucle hipervariable para lograr una unión específica. Véase, *por ejemplo*, Skerra, FEBS J. 275:2677-83, 2008, incorporado al presente documento por referencia.

#### 15 VII. BIOENSAYOS Y ENSAYOS ANALÍTICOS PARA ENSAYOS DE LIBERACIÓN DE FÁRMACOS Y ESPECIFICACIONES DE PRODUCTO, DIAGNÓSTICOS, Y REACTIVOS

También se incluyen bioensayos que están relacionados con los fragmentos de proteína AARS y con agentes relacionados como reactivos terapéuticos y diagnósticos. Los ejemplos incluyen bioensayos y ensayos analíticos que miden la pureza, la actividad biológica, la afinidad, la solubilidad, el pH, los niveles de endotoxina, entre otros,  
20 muchos de los cuales se describen en el presente documento. También se incluyen ensayos que establecen curvas de respuesta a la dosis y/o proporcionan una o más bases para comparación entre diferentes lotes de agentes. Las comparaciones entre lotes pueden estar basadas en uno cualquiera o más de caracterización química, caracterización biológica y caracterización clínica. Para agentes proteínicos, también se incluyen métodos para evaluar la potencia, estabilidad, farmacocinética e inmunogenicidad de un agente seleccionado. Entre otros usos,  
25 pueden usarse estos y otros métodos para las pruebas de liberación de lotes de agentes biológicos o químicos, incluyendo los fragmentos de proteína AARS, anticuerpos, agentes de unión, polinucleótidos, tales como agentes antisentido y vectores y otros descritos en el presente documento.

Determinadas realizaciones incluyen el uso de ensayos de bioafinidad. Dichos ensayos pueden usarse para evaluar la afinidad de unión, por ejemplo, entre un fragmento de proteína AARS y un compañero de unión celular o entre un fragmento de proteína AARS y un anticuerpo. También puede medirse la afinidad de unión entre un fragmento de proteína AARS y un agente de unión alternativo, tal como un compuesto de ensayo candidato o líder (por ejemplo, un modulador de molécula pequeña de un AARS), o entre un compañero de unión celular de AARS y un compuesto de ensayo candidato o líder. Determinados ensayos de afinidad de unión a modo de ejemplo pueden utilizar ensayos  
35 ELISA, tal como se describe en el presente documento y se conoce en la técnica. Determinados ensayos utilizan cromatografía de unión a receptor de alto rendimiento (véase, *por ejemplo*, Roswall et al., Biologicals. 24:25-39, 1996). Otros ensayos de afinidad de unión a modo de ejemplo pueden utilizar tecnologías basadas en resonancia de plasmón superficial (SPR). Los ejemplos incluyen tecnologías BIAcore, algunas de las cuales integran tecnología SPR con un sistema de microfluidos para controlar las interacciones moleculares en tiempo real a concentraciones  
40 en el intervalo de pM a mM. También se incluyen los ensayos KinExa™, que proporcionan mediciones precisas de la especificidad de unión, la afinidad de unión y las constantes de cinética/velocidad de unión.

Determinadas realizaciones están relacionadas con inmunoensayos para evaluar u optimizar la inmunogenicidad de agentes proteínicos. Los ejemplos incluyen ensayos celulares humanos *ex vivo* y ensayos inmunoenzimáticos *in vitro* para proporcionar información útil acerca del potencial inmunogénico de una proteína terapéutica. Pueden usarse ensayos de respuesta celular *ex vivo*, por ejemplo, para reproducir la cooperación celular entre células presentadoras de antígenos (APC) y células T y de este modo medir la activación de células T tras entrar en contacto con una proteína de interés. Determinados ensayos enzimáticos *in vitro* pueden utilizar una colección de moléculas de HLA-DR recombinantes que abarcan una porción significativa de una población humana relevante y  
50 puede incluir ensayos inmunoenzimáticos automatizados para probar la unión de péptidos (que surgen de la fragmentación de la proteína terapéutica) con las moléculas de HLA-DR. También se incluyen métodos para reducir la inmunogenicidad de una proteína seleccionada, tal como mediante el uso de estos métodos y otros relacionados para identificar y después retirar o alterar uno o más epítomos de células T de un agente de proteína.

También se incluyen ensayos de liberación biológica (por ejemplo, ensayos basados en células) para medir parámetros tales como actividades biológicas específicas, incluyendo actividades no canónicas y citotoxicidad. Determinados ensayos biológicos específicos incluyen, por ejemplo, ensayos basados en células que utilizan un compañero de unión celular (por ejemplo, receptor de la superficie celular) de un fragmento de proteína AARS seleccionado, que está acoplado funcionalmente a un marcador de lectura, tal como un indicador fluorescente o  
60 luminiscente de una actividad biológica no canónica, tal como se describe en el presente documento. Por ejemplo, realizaciones específicas incluyen una célula que comprende un receptor de superficie celular o una porción extracelular del mismo que se une a un fragmento de proteína AARS, en donde la célula comprende un detector o marcador de lectura. También se incluyen ensayos biológicos *in vivo* para caracterizar la farmacocinética de un agente, tal como un polipéptido de AARS o un anticuerpo, utilizando normalmente ratones u otro mamífero modificados por ingeniería genética (véase, *por ejemplo*, Lee et al., The Journal of Pharmacology. 281:1431-1439, 1997). Los ejemplos de ensayos biológicos basados en la citotoxicidad incluyen ensayos de liberación (por ejemplo,  
65

ensayos de liberación de cromo o europio para medir la apoptosis; véase, por ejemplo, von Zons et al., Clin Diagn Lab Immunol.4:202-207, 1997), entre otros, que pueden evaluar la citotoxicidad de fragmentos de proteína AARS, ya sea para determinar curvas de respuesta a la dosis, prueba de lotes u otras propiedades relacionadas con la aprobación por diversas agencias regulatorias, tales como la Food and Drug Administration (FDA).

5 Dichos ensayos pueden usarse, por ejemplo, para desarrollar una curva de respuesta a la dosis para un fragmento de proteína AARS seleccionado u otro agente y/o para comparar las curvas de respuesta a la dosis de diferentes lotes de proteínas u otros agentes. Una curva de respuesta a la dosis es una gráfica X-Y que está relacionada con la magnitud de un agente inductor de estrés con la respuesta de un receptor; la respuesta puede ser una respuesta fisiológica o bioquímica, tal como una actividad biológica no canónica en una célula *in vitro* o en una célula o tejido *in vivo*, una cantidad terapéuticamente eficaz medida *in vivo* (por ejemplo, medida mediante la  $CE_{50}$ ), o muerte, ya se mida *in vitro* o *in vivo* (por ejemplo, muerte celular, muerte del organismo). La muerte se indica normalmente en forma de una  $DL_{50}$ , una dosis obtenida estadísticamente que es letal para el 50 % de una población modelo. Aunque puede estar indicada por la  $CL_{01}$  (dosis letal para un 1 % de la población de animales de ensayo),  $CL_{100}$  (dosis letal para el 100 % de la población de animales de ensayo), o  $CL_{LO}$  (dosis más baja que provoca letalidad). Puede caracterizarse prácticamente cualquier efecto o criterio de valoración deseado de este modo.

La dosis medida de una curva de respuesta a la dosis se representa normalmente en el eje X y la respuesta se representa en el eje Y. Más normalmente, se representa el logaritmo de la dosis en el eje X, generando con mayor frecuencia una curva sigmoidea con la porción con más pendiente en medio. El nivel de efecto no observable (NOEL) se refiere a la dosis experimental más baja para la que no se observa un efecto medible y la dosis umbral se refiere al primer punto en la gráfica que indica una respuesta por encima de cero. Por norma general, los fármacos más fuertes generan curvas de respuesta a la dosis más pronunciadas. Para muchos fármacos, se observan los efectos deseados a dosis ligeramente mayores que la dosis umbral, a menudo debido a que las dosis menores son relativamente ineficaces y las dosis mayores dan lugar a efectos secundarios no deseados. Para las curvas de respuesta a la dosis generadas *in vivo*, puede caracterizarse una curva por valores tales como  $\mu\text{g}/\text{kg}$ ,  $\text{mg}/\text{kg}$  o  $\text{g}/\text{kg}$  de peso corporal, si se desea.

Para las comparaciones por lotes, puede ser útil calcular el coeficiente de variación (CV) entre diferentes curvas de respuesta a la dosis de diferentes lotes (por ejemplo, entre diferentes lotes de fragmentos de proteína AARS, anticuerpos u otros agentes), en parte debido a que el CV permite la comparación entre conjuntos de datos con diferentes unidades o diferentes medios. Por ejemplo, en determinadas realizaciones a modo de ejemplo, dos o tres o más lotes diferentes de fragmentos de proteína AARS u otros agentes tienen un CV entre ellos de menos de aproximadamente un 15 %, 14 %, 13 %, 12 %, 11 %, 10 %, 9 %, 8 %, 7 %, 6 %, 5 %, 4 %, 3 %, 2 %, o 1 % para una curva de respuesta a la dosis de 4, 5, 6, 7 u 8 puntos. En determinadas realizaciones, la curva de respuesta a la dosis se mide en un ensayo basado en células y su lectura se relaciona con un aumento o una reducción en una actividad no canónica seleccionada del fragmento de proteína AARS. En determinadas realizaciones, la curva de respuesta a la dosis se mide en un ensayo de liberación celular o un modelo animal (por ejemplo, modelo de ratón), y su lectura se relaciona con la muerte celular o la muerte del animal. Otras variaciones serán evidentes para las personas expertas en la materia.

#### VIII. SISTEMAS DE EXPRESIÓN Y PURIFICACIÓN

Las realizaciones de la presente invención incluyen métodos y composiciones relacionadas para expresar y purificar los fragmentos de proteína AARS u otros agentes a base de polipéptidos de la invención. Dichos polipéptidos de AARS recombinantes pueden prepararse convenientemente usando protocolos convencionales, tal como se describen, por ejemplo, en Sambrook, *et al.*, (1989, anteriormente citado), en particular, las Secciones 16 y 17; Ausubel *et al.*, (1994, anteriormente citado), en particular, los Capítulos 10 y 16; y Coligan *et al.*, Current Protocols in Protein Science (John Wiley & Sons, Inc. 1995-1997), en particular, los Capítulos 1, 5 y 6. A modo de ejemplo general, pueden prepararse polipéptidos de AARS mediante un procedimiento que incluye una o más de las etapas de: (a) preparar una construcción que comprende una secuencia de polinucleótido que codifica un polipéptido de AARS y que está unida operablemente a un elemento regulador; (b) introducir la construcción en una célula hospedadora; (c) cultivar la célula hospedadora para que exprese el polipéptido de AARS; y (d) aislar el polipéptido de AARS de la célula hospedadora.

En otras partes del presente documento se describen polinucleótidos de AARS. Para expresar un polipéptido deseado, puede insertarse una secuencia de nucleótido que codifica el polipéptido o un equivalente funcional en un vector de expresión adecuado, es decir, un vector que contiene los elementos necesarios para la transcripción y traducción de la secuencia codificante insertada. Pueden usarse métodos bien conocidos por los expertos en la materia para construir vectores de expresión que contienen secuencias que codifican un polipéptido de interés y elementos de control transcripcional y traduccional adecuados. Estos métodos incluyen técnicas de ADN recombinante *in vitro*, técnicas sintéticas y recombinación genética *in vivo*. Dichas técnicas se describen en Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (1989), y Ausubel *et al.*, Current Protocols in Molecular Biology (1989).

Se conocen y pueden utilizarse diversos sistemas de vector de expresión/hospedador para que contengan y

expresen secuencias de polinucleótido. Estos incluyen, pero sin limitación, microorganismos, tales como bacterias transformadas con vectores de expresión recombinantes de ADN de bacteriófago, plásmido o cósmido; levaduras transformadas con vectores de expresión de levadura; sistemas de células de insecto infectadas con vectores de expresión de virus (por ejemplo, baculovirus); sistemas de células vegetales transformadas con vectores de expresión de virus (por ejemplo, virus del mosaico de la coliflor, CaMV; virus del mosaico del tabaco, TMV) o con vectores de expresión bacterianos (por ejemplo, plásmidos Ti o pBR322); o sistemas de células animales, incluyendo sistemas de células de mamífero y más específicamente, de células humanas.

Los "elementos de control" o "secuencias reguladoras" presentes en un vector de expresión son aquellas regiones no traducidas del vector (potenciadores, promotores, regiones 5' y 3' no traducidas) que interactúan con proteínas celulares del hospedador para llevar a cabo la transcripción y la traducción. Dichos elementos pueden variar en cuanto a su fuerza y especificidad. Dependiendo del sistema de vector y el hospedador utilizados, puede usarse cualquier número de elementos de transcripción y traducción adecuado, incluyendo promotores constitutivos e inducibles. Por ejemplo, cuando se clonan en sistemas bacterianos, pueden usarse promotores inducibles, tales como el promotor híbrido lacZ del fagémido PBLUESCRIPT (Stratagene, La Jolla, Calif.) o el plásmido PSPORT1 (Gibco BRL, Gaithersburg, Md.) y similares. En sistemas de células de mamífero, se prefieren generalmente promotores de genes de mamífero o de virus de mamífero. En caso de que sea necesario generar una línea celular que contenga múltiples copias de la secuencia que codifica un polipéptido, pueden usarse ventajosamente vectores basados en SV40 o EBV con un marcador de selección adecuado.

En sistemas bacterianos, puede seleccionarse un número de vectores de expresión dependiendo del uso previsto para el polipéptido expresado. Por ejemplo, cuando se necesitan grandes cantidades, pueden usarse vectores que dirigen la expresión de alto nivel de proteínas de fusión que se purifican fácilmente. Dichos vectores incluyen, pero sin limitación, los vectores de clonación y expresión en *E. coli* multifuncionales, tales como BLUESCRIPT (Stratagene), en los que puede ligarse en el vector la secuencia que codifica el polipéptido de interés en fase con secuencias para la Met amino-terminal y los 7 restos posteriores de  $\beta$ -galactosidasa, de tal forma que se produce una proteína híbrida; vectores pIN (Van Heeke y Schuster, J. Biol. Chem. 264:5503-5509 (1989)); y similares. Los vectores pGEX (Promega, Madison, Wis.) también pueden usarse para expresar polipéptidos exógenos en forma de proteínas de fusión con glutatión S-transferasa (GST). En general, dichas proteínas de fusión son solubles y pueden purificarse fácilmente de células lisadas mediante adsorción sobre perlas de glutatión-agarosa seguido de elución en presencia de glutatión libre. Las proteínas producidas en dichos sistemas pueden diseñarse para incluir sitios de escisión de heparina, trombina o de proteasa de factor XA, de tal forma que el polipéptido de interés clonado puede liberarse del resto de GST a voluntad.

Determinadas realizaciones pueden emplear sistemas de expresión basados en *E. coli* (véase, por ejemplo, Structural Genomics Consortium et al., Nature Methods. 5:135-146, 2008). Estas realizaciones y otras relacionadas pueden apoyarse parcial o totalmente en la clonación independiente de ligadura (LIC) para producir un vector de expresión adecuado. En realizaciones específicas, la expresión de proteínas puede controlarse mediante una ARN polimerasa T7 (por ejemplo, vector de la serie pET). Estas realizaciones y otras relacionadas pueden utilizar la cepa hospedadora de expresión BL21(DE3), un lisogen  $\lambda$ DE3 de BL21 que soporta la expresión mediada por T7 y que es deficiente para las proteasas lon y ompT para una estabilidad mejorada de la proteína diana. También se incluyen cepas de hospedador de expresión que portan plásmidos que codifican ARNt raramente usados en *E. coli*, tales como las cepas ROSETTA™ (DE3) y Rosetta 2 (DE3). También puede mejorarse la lisis celular y el manejo de la muestra usando reactivos comercializados bajo las marcas nucleasa BENZONASE® y reactivo de extracción de proteínas BUGBUSTER®. Para el cultivo celular, un medio auto-inductor puede mejorar la eficacia de muchos sistemas de expresión, incluyendo sistemas de expresión de alto rendimiento. Los medios de este tipo (por ejemplo, el sistema de auto-inducción OVERNIGHT EXPRESS™) desencadenan gradualmente la expresión de proteínas mediante un cambio metabólico sin la adición de agentes inductores artificiales, tales como IPTG. Realizaciones particulares emplean marcadores de hexahistidina (tales como aquellos comercializados bajo la marca fusiones HIS TAG®), seguido de purificación por cromatografía de afinidad de metal inmovilizado (IMAC) o técnicas relacionadas. En determinados aspectos, sin embargo, pueden aislarse proteínas de grado clínico a partir de cuerpos de inclusión de *E. coli*, con o sin el uso de marcadores de afinidad (véase, por ejemplo, Shimp et al., Protein Expr Purif. 50:58-67, 2006). A modo de ejemplo adicional, determinadas realizaciones pueden emplear un sistema de producción de alto rendimiento en *E. coli* inducido por choque frío, debido a que la sobreexpresión de proteínas en *Escherichia coli* a baja temperatura mejora su solubilidad y estabilidad (véase, por ejemplo, Qing et al., Nature Biotechnology. 22:877-882, 2004).

También se incluyen sistemas de fermentación bacterianos de alta densidad. Por ejemplo, el cultivo de alta densidad celular de *Ralstonia eutropha* permite la producción de proteína a densidades celulares de más de 150 g/l y la expresión de proteína recombinante a títulos que superan los 10 g/l.

En la levadura *Saccharomyces cerevisiae*, puede emplearse una serie de vectores que contienen promotores constitutivos o inducibles, tales como factor alfa, alcohol oxidasa y PGH. Para una revisión, véase Ausubel et al. anteriormente citado y Grant et al., Methods Enzymol. 153:516-544 (1987). También se incluyen sistemas de expresión de *Pichia pastoris* (véase, por ejemplo, Li et al., Nature Biotechnology. 24, 210 - 215, 2006; y Hamilton et al., Science, 301:1244, 2003). Determinadas realizaciones incluyen sistemas de levadura que están modificados por

ingeniería genética para glucosilar proteínas de manera selectiva, incluyendo levaduras que tienen rutas de N-glucosilación humanizadas, entre otras (véase, *por ejemplo*, Hamilton et al., Science. 313:1441-1443, 2006; Wildt et al., Nature Reviews Microbiol. 3:119-28, 2005; y Gerngross et al., Nature-Biotechnology. 22:1409 -1414, 2004; las Patentes de los Estados Unidos n.º 7.629.163; 7.326.681; y 7.029.872). Solo a modo de ejemplo, pueden cultivarse cultivos de levadura recombinantes en matraces Fernbach o fermentadores de 15L, 50L, 100L y 200L, entre otros.

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

También puede usarse un sistema de insecto para expresar un polipéptido de interés. Por ejemplo, en un sistema similar, se usa el virus de la poliedrosis nuclear de *Autographa californica* (AcNPV) como vector para expresar genes exógenos en células de *Spodoptera frugiperda* o en larvas de *Trichoplusia*. Las secuencias que codifican el polipéptido pueden clonarse en una región no esencial del virus, tal como el gen de poliedrina y colocarse bajo el control del promotor de poliedrina. La inserción exitosa de la secuencia codificante de polipéptido hará que el gen de poliedrina sea inactivo y produzca virus recombinante que carece de proteína de la envuelta. Entonces, pueden usarse los virus recombinantes para infectar, por ejemplo, células de *S. frugiperda* o larvas de *Trichoplusia* en las que puede expresarse el polipéptido de interés (Engelhard et al., Proc. Natl. Acad. Sci. U.S.A. 91:3224-3227 (1994)). También se incluyen sistemas de expresión de baculovirus, incluyendo aquellos que utilizan células SF9, SF21 y Tni (véase, *por ejemplo*, Murphy y Piwnica-Worms, Curr Protoc Protein Sci. Capítulo 5:Unidad5.4, 2001). Los sistemas de insecto pueden proporcionar modificaciones postraduccionales que son similares a los sistemas de mamífero.

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

Los ejemplos de células de mamífero útiles incluyen la línea CV1 de riñón de mono transformada mediante SV40 (COS-7, ATCC CRL 1651); línea celular de riñón embrionario humano (células 293 o 293 subclonadas para crecimiento en cultivo en suspensión, Graham et al., J. Gen Virol. 36:59 (1977)); células de riñón de cría de hámster (BHK, ATCC CCL 10); células de sertoli de ratón (TM4, Mather, Biol. Reprod. 23:243-251 (1980)); células de riñón de mono (CV1 ATCC CCL 70); células de riñón de mono verde africano (VERO-76, ATCC CRL-1587); células de carcinoma de cuello de útero humano (HELA, ATCC CCL 2); células de riñón canino (MDCK, ATCC CCL 34); células de hígado de rata búfalo (BRL 3A, ATCC CRL 1442); células de pulmón humano (W138, ATCC CCL 75); células de hígado humano (Hep G2, HB 8065); tumor mamario de ratón (mMT 060562, ATCC CCL51); células TR1 (Mather et al., Annals N.Y. Acad. Sci. 383:44-68 (1982)); células MRC 5; células FS4; y una línea celular de hepatoma humano (Hep G2). Otras líneas celulares de mamífero útiles incluyen células de ovario de hámster chino (CHO), incluyendo células CHO DHFR- (Urlaub et al., PNAS USA 77:4216 (1980)); y líneas celulares de mieloma, tales como NS0 y Sp2/0. Para una revisión de determinadas líneas celulares de mamífero hospedadoras adecuadas para la producción de anticuerpos, véase, *por ejemplo*, Yazaki y Wu, Methods in Molecular Biology, Vol. 248 (B. K.C Lo, ed., Humana Press, Totowa, N.J., 2003), págs. 255-268. Determinados sistemas de expresión en células de mamífero incluyen sistemas de expresión basados en células CHO y HEK293. Los sistemas de expresión de mamífero pueden utilizar líneas células unidas, por ejemplo, en matraces T, botellas rotatorias o fábricas de células o cultivos en suspensión, por ejemplo, en matraces rotatorios de 1L y 5L, biorreactores de tanque agitado de 5L, 14L, 40L, 100L y 200L o biorreactores WAVE de 20/50L y 100/200L, entre otros conocidos en la técnica.

También se incluye la expresión sin células de proteínas. Estas realizaciones y otras relacionadas utilizan normalmente ARN polimerasa purificada, ribosomas, ARNt y ribonucleótidos; estos reactivos pueden producirse por extracción de células o a partir de un sistema de expresión a base de células.

También pueden usarse señales de inicio específicas para lograr una traducción más eficaz de secuencias que codifican un polipéptido de interés. Dichas señales incluyen el codón de iniciación ATG y secuencias adyacentes. En los casos donde las secuencias que codifican el polipéptido, su codón de iniciación y secuencias cadena arriba se inserten en el vector de expresión adecuado, pueden no ser necesarias señales de control transcripcional o traduccional adicionales. Sin embargo, en los casos donde solo se inserte secuencia codificante o una porción de la misma, deben proporcionarse secuencias de control traduccional exógenas que incluyan el codón de iniciación ATG.

Además, el codón de iniciación debería encontrarse en la fase de lectura correcta para asegurar la traducción del inserto completo. Los elementos traducionales y los codones de iniciación exógenos pueden ser de diversos orígenes, tanto naturales como sintéticos. La eficacia de la expresión puede aumentarse mediante la inclusión de potenciadores que son adecuados para el sistema celular particular que se esté usando, tales como aquellos descritos en la bibliografía (Scharf. et al., Results Probl. Cell Differ. 20:125-162 (1994)).

Además, puede seleccionarse una cepa de célula hospedadora por su capacidad para modular la expresión de las secuencias insertadas o para procesar la proteína expresada del modo deseado. Dichas modificaciones del polipéptido incluyen, pero sin limitación, modificaciones postraduccionales, tales como acetilación, carboxilación, glucosilación, fosforilación, lipidación y acilación. También puede usarse un procesamiento postraducciona que escinde una forma "prepro" de la proteína para facilitar una correcta inserción, plegamiento y/o función. Pueden seleccionarse diferentes células hospedadoras, tales como de levadura, CHO, HeLa, MDCK, HEK293 y W138, además de células bacterianas, que tienen o incluso carecen de maquinaria celular específica y de mecanismos característicos para dichas actividades postraduccionales, para asegurar la correcta modificación y procesamiento de la proteína exógena.

Para la producción a largo plazo y de alto rendimiento de proteínas recombinantes, se prefiere generalmente la expresión estable. Por ejemplo, pueden transformarse líneas celulares que expresan de manera estable un polinucleótido de interés usando vectores de expresión que pueden contener orígenes de replicación víricos y/o elementos de expresión endógenos y un gen marcador de selección en el mismo o en un vector separado. Después de la introducción del vector, pueden dejarse crecer las células durante aproximadamente 1-2 días en un medio enriquecido antes de que se cambien a un medio selectivo. El fin del marcador de selección es conferir resistencia a la selección y su presencia permite el crecimiento y la recuperación de células que expresan exitosamente las secuencias introducidas. Los clones resistentes de las células transformadas de manera estable pueden proliferarse usando técnicas de cultivo tisular adecuadas para el tipo celular. También puede emplearse la producción transitoria, tal como mediante transfección o infección transitoria. Los sistemas de expresión de mamífero que son adecuados para la producción transitoria incluyen sistemas basados en HEK293 y CHO.

Puede usarse cualquier número de sistemas de selección para recuperar líneas celulares transformadas o transducidas. Estos incluyen, pero sin limitación, los genes de timidina cinasa del virus del herpes simple (Wigler et al., Cell 11:223-232 (1977)) y de adenina fosforibosiltransferasa (Lowy et al., Cell 22:817-823 (1990)) que pueden emplearse en células tk- o aprt-, respectivamente. Asimismo, puede emplearse la resistencia a antimetabolitos, antibióticos o herbicidas como base para la selección; por ejemplo, dhfr, que confiere resistencia al metotrexato (Wigler et al., Proc. Natl. Acad. Sci. U.S.A. 77:3567-70 (1980)); npt, que confiere resistencia a los aminoglucósidos, neomicina y G-418 (Clobere-Garapin et al., J. Mol. Biol. 150:1-14 (1981)); y als o pat, que confiere resistencia al clorsulfurón y a la fosfotricina acetiltransferasa, respectivamente (Murry, anteriormente citado). Se han descrito genes de selección adicionales, por ejemplo, trpB, que permite a las células utilizar indol en lugar de triptófano o hisD, que permite a las células utilizar histinol en lugar de histidina (Hartman y Mulligan, Proc. Natl. Acad. Sci. U.S.A. 85:8047-51 (1988)). El uso de marcadores visibles se ha vuelto popular con marcadores tales como la proteína fluorescente verde (GFP) y otras proteínas fluorescentes (por ejemplo, RFP, YFP), antocianinas,  $\beta$ -glucuronidasa y su sustrato, GUS y luciferasa y su sustrato, luciferina, que se usan ampliamente no solo para identificar transformantes, sino también para cuantificar la cantidad de expresión de proteína transitoria o estable atribuible a un sistema de vector específico (véase, *por ejemplo*, Rhodes et al., Methods Mol. Biol. 55:121-131 (1995)).

Las realizaciones de la presente invención también incluyen sistemas de producción de proteínas de alto rendimiento o sistemas de microproducción. Determinados aspectos pueden utilizar, por ejemplo, marcadores de fusión de hexa-histidina para la expresión y purificación de proteínas en superficies de portaobjetos modificadas con quelato o partículas MagneHis Ni (véase, *por ejemplo*, Kwon et al., BMC Biotechnol. 9:72, 2009; y Lin et al., Methods Mol Biol. 498:129-41, 2009)). También se incluyen sistemas de expresión de proteínas sin células de alto rendimiento (véase, *por ejemplo*, Sitaraman et al., Methods Mol Biol. 498:229-44, 2009). Pueden usarse estas y otras realizaciones relacionadas, por ejemplo, para generar micromatrices de fragmentos de proteínas AARS, que después pueden usarse para explorar bibliotecas para identificar agentes que interactúan con los fragmentos de proteína AARS.

Se conocen en la técnica una serie de protocolos para detectar y medir la expresión de productos codificados por polinucleótidos usando agentes de unión o anticuerpos, tales como anticuerpos policlonales o monoclonales específicos para el producto. Los ejemplos incluyen ensayo inmunosorbente ligado a enzimas (ELISA), inmunotransferencias de Western, radioinmunoensayos (RIA), y clasificación celular activada por fluorescencia (FACS). Estos y otros ensayos se describen, entre otros lugares, en Hampton et al., Serological Methods, a Laboratory Manual (1990) y Maddox et al., J. Exp. Med. 158:1211-1216 (1983).

Los expertos en la materia conocen una gran variedad de marcadores y técnicas de conjugación y pueden usarse en diversos ensayos de ácidos nucleicos y aminoácidos. Los medios para producir sondas de hibridación o PCR marcadas para detectar secuencias relacionadas con polinucleótidos incluyen marcaje de oligos, traducción de cortes, marcaje terminal o amplificación PCR usando un nucleótido marcado. Como alternativa, pueden clonarse las secuencias o cualquier porción de las mismas en un vector para la producción de una sonda de ARNm. Dichos

5 vectores se conocen en la técnica, se encuentran disponibles comercialmente y pueden usarse para sintetizar sondas de ARN *in vitro* mediante la adición de una ARN polimerasa adecuada, tal como T7, T3 o SP6 y nucleótidos marcados. Estos procedimientos pueden llevarse a cabo usando una serie de kits disponibles comercialmente. Las moléculas indicadoras o marcadores que pueden usarse incluyen radionúclidos, enzimas, agentes fluorescentes, quimioluminiscentes o cromogénicos, así como sustratos, cofactores, inhibidores, partículas magnéticas y similares.

10 Las células hospedadoras transformadas con una secuencia de polinucleótido de interés pueden cultivarse en condiciones adecuadas para la expresión y recuperación de la proteína del cultivo celular. Determinadas realizaciones específicas utilizan sistemas de expresión en células sin suero. Los ejemplos incluyen células HEK293 y células CHO que pueden crecer en medio libre de suero (véase, *por ejemplo*, Rosser et al., Protein Expr. Purif. 40:237-43, 2005; y la Patente de los Estados Unidos n.º 6.210.922).

15 La proteína producida por una célula recombinante puede secretarse o estar contenida intracelularmente, dependiendo de la secuencia y/o el vector usado. Como apreciarán los expertos en la materia, los vectores de expresión que contienen polinucleótidos de la invención pueden diseñarse para contener secuencias de señal que dirigen la secreción del polipéptido codificado a través de una membrana celular procariota o eucariota. Pueden usarse otras construcciones recombinantes para unir secuencias que codifican un polipéptido de interés a una secuencia de nucleótidos que codifica un dominio de polipéptido que facilitará la purificación y/o la detección de proteínas solubles. Los ejemplos de dichos dominios incluyen marcadores de purificación por afinidad y epitópicos escindibles y no escindibles, tales como avidina, marcadores FLAG, marcadores de poli-histidina (por ejemplo, 6xHis), marcadores cMyc, marcadores V5, marcadores de glutatión S-transferasa (GST) y otros.

25 La proteína producida por una célula recombinante puede purificarse y caracterizarse según diversas técnicas conocidas en la materia. Los sistemas a modo de ejemplo para llevar a cabo la purificación de proteínas y analizar la pureza de la proteína incluyen cromatografía líquida rápida de proteínas (FPLC) (por ejemplo, los sistemas de FPLC AKTA y Bio-Rad), cromatografía líquida de alta presión (HPLC) (por ejemplo, HPLC Beckman y Waters). Las químicas a modo de ejemplo para la purificación incluyen cromatografía de intercambio iónico (por ejemplo, Q, S), cromatografía de exclusión por tamaños, gradientes de sal, purificación por afinidad (por ejemplo, Ni, Co, FLAG, maltosa, glutatión, proteína A/G), filtración en gel, fase reversa, cromatografía de intercambio iónico cerámica HYPERD® y columnas de interacción hidrófoba (HIC), entre otros conocidos en la técnica. También se incluyen métodos analíticos, tales como SDS-PAGE (por ejemplo, Coomassie, tinción de plata), inmunotransferencia, Bradford y ELISA, que pueden utilizarse durante cualquier etapa del proceso de producción o purificación, normalmente para medir la pureza de la composición de proteína.

35 También se incluyen métodos para concentrar fragmentos de proteína AARS y composiciones que comprenden proteínas solubles concentradas. En diferentes aspectos, dichas soluciones concentradas de polipéptidos de AARS pueden comprender proteínas a una concentración de aproximadamente 5 mg/ml; o aproximadamente 8 mg/ml; o aproximadamente 10 mg/ml; aproximadamente 15 mg/ml; o aproximadamente 20 mg/ml.

40 En un aspecto, dichas composiciones pueden ser sustancialmente monodispersas, lo que significa que las composiciones de polipéptido de AARS existen principalmente (es decir, al menos aproximadamente un 90 %, o más) en una forma de peso molecular aparente cuando se evalúa, por cromatografía de exclusión por tamaños, dispersión de luz dinámica o ultracentrifugación analítica.

45 En otro aspecto, dichas composiciones tienen una pureza (basándose en la proteína) de al menos aproximadamente el 90 %, o en algunos aspectos, una pureza de al menos el 95 % o en algunas realizaciones, una pureza de al menos el 98 %. La pureza puede determinarse mediante cualquier método analítico rutinario, tal como se conoce en la técnica.

50 En otro aspecto, dichas composiciones tienen un contenido de agregados de alto peso molecular de menos de aproximadamente el 10 %, en comparación con la cantidad total de proteína presente o en algunas realizaciones, dichas composiciones tienen un contenido de agregados de alto peso molecular de menos de aproximadamente el 5 %, o en algunos aspectos, dichas composiciones tienen un contenido de agregados de alto peso molecular de menos de aproximadamente el 3 %, o en algunas realizaciones, un contenido de agregados de alto peso molecular de menos de aproximadamente el 1 %. El contenido de agregados de alto peso molecular puede determinarse mediante diversas técnicas analíticas que incluyen, por ejemplo, por cromatografía de exclusión por tamaños, dispersión de luz dinámica o ultracentrifugación analítica.

60 En determinadas realizaciones, tal como se indica en el presente documento, las composiciones de polipéptido de AARS tienen un contenido de endotoxina de menos de aproximadamente 10 UE/mg de polipéptido de AARS o de menos de aproximadamente 5 UE/mg de polipéptido de AARS, menos de aproximadamente 3 UE/mg de polipéptido de AARS o menos de aproximadamente 1 UE/mg de polipéptido de AARS.

65 Los ejemplos de estrategias de concentración contempladas en el presente documento incluyen liofilización, que se emplea normalmente cuando la solución contiene pocos componentes solubles aparte de la proteína de interés. La liofilización se lleva a cabo normalmente después de un ciclo de HPLC y puede retirar la mayor parte de o todos los

componentes volátiles de la mezcla. También se incluyen técnicas de ultrafiltración, que normalmente emplean una o más membranas permeables selectivas para concentrar una solución de proteína. La membrana permite que pasen a su través agua y moléculas pequeñas y retiene la proteína; la solución puede forzarse contra la membrana mediante una bomba mecánica, presión de gas o centrifugación, entre otras técnicas.

5 En determinadas realizaciones, los reactivos, fragmentos de proteína AARS o agentes relacionados (por ejemplo, anticuerpos) tienen una pureza de al menos aproximadamente el 90 %, medida de acuerdo con técnicas rutinarias en la técnica. En determinadas realizaciones, tales como composiciones diagnósticas o determinadas  
10 composiciones terapéuticas, las composiciones de AARS de la presente invención tienen una pureza de al menos aproximadamente el 95 %. En realizaciones específicas, tales como composiciones terapéuticas o farmacéuticas, las composiciones de AARS de la presente invención tienen una pureza de al menos aproximadamente el 97 % o el 98 % o el 99 %. En otras realizaciones, tales como cuando se usan como reactivos de referencia o de investigación, los fragmentos de proteína AARS pueden tener una menor pureza y pueden tener una pureza de al menos  
15 aproximadamente el 50 %, 60 %, 70 %, u 80 %. La pureza puede medirse en general o en relación con componentes seleccionados, tales como otras proteínas, *por ejemplo*, pureza basándose en una proteína.

Los fragmentos de proteína AARS purificados también pueden caracterizarse según sus características biológicas. Los ejemplos incluyen afinidad de unión o cinética de unión a un ligando seleccionado (por ejemplo, un compañero de unión celular del fragmento de proteína ARS, tal como un receptor de la superficie celular o un dominio  
20 extracelular del mismo), y la presencia o niveles de una o más actividades biológicas canónicas o no canónicas, tal como se describe en el presente documento. La afinidad de unión y la cinética de unión pueden medirse según diversas técnicas conocidas en la materia, tales como BIACORE® y tecnologías relacionadas que utilizan resonancia de plasmón superficial (SPR), un fenómeno óptico que permite la detección de agentes que interactúan no marcados en tiempo real. Pueden usarse biosensores basados en SPR para la determinación de la concentración  
25 de agente activo, la exploración y la caracterización en términos tanto de afinidad como cinética. La presencia o los niveles de una o más actividades canónicas o no canónicas pueden medirse de acuerdo con ensayos basados en células, incluyendo aquellos que utilizan un compañero de unión celular (por ejemplo, receptor de la superficie celular) de un fragmento de proteína AARS seleccionado, que está acoplado funcionalmente a un marcador de lectura o un indicador, tal como un indicador fluorescente o luminiscente de una actividad biológica no canónica, tal  
30 como se describen en el presente documento.

En determinadas realizaciones, tal como se ha indicado anteriormente, las composiciones de polipéptido de AARS se encuentran aproximadamente sustancialmente libres de endotoxina, incluyendo, por ejemplo, aproximadamente  
35 un 95 % libres de endotoxina, preferentemente aproximadamente un 99 % libre de endotoxina y más preferentemente, aproximadamente un 99,99 % libre de endotoxina. La presencia de endotoxinas puede detectarse de acuerdo con técnicas rutinarias en la materia, tal como se describen en el presente documento. En realizaciones específicas, las composiciones de AARS se preparan a partir de una célula eucariota, tal como una célula de mamífero o humana en medio sustancialmente libre de suero.

40 En determinadas realizaciones, las composiciones de polipéptido de AARS comprenden menos de aproximadamente un 10 % p/p de agregados de alto peso molecular o menos de aproximadamente un 5 % p/p de agregados de alto peso molecular o menos de aproximadamente un 2 % p/p de agregados de alto peso molecular o menos de aproximadamente un 1 % p/p de agregados de alto peso molecular.

45 También se incluyen ensayos analíticos y métodos basados en proteínas, que pueden usarse para evaluar, por ejemplo, la pureza de proteínas, el tamaño, la solubilidad y el grado de agregación, entre otras características. La pureza de proteínas puede evaluarse mediante una serie de formas. Por ejemplo, la pureza puede evaluarse basándose en la estructura primaria, la estructura de orden mayor, el tamaño, la carga, la hidrofobia y la glucosilación. Los ejemplos de métodos para evaluar la estructura primaria incluyen secuenciación N y C-terminal y mapeo peptídico (véase, *por ejemplo*, Allen et al., *Biologicals*. 24:255-275, 1996)). Los ejemplos de métodos para  
50 evaluar la estructura de mayor orden incluyen dicroísmo circular (véase, *por ejemplo*, Kelly et al., *Biochim Biophys Acta*. 1751:119-139, 2005), espectroscopía de fluorescencia (véase, *por ejemplo*, Meagher et al., *J. Biol. Chem.* 273:23283-89, 1998), FT-IR, cinética de intercambio de hidrogenodeuteruro de amida, calorimetría de barrido diferencial, espectroscopía RMN, inmunorreactividad con anticuerpos conformacionalmente sensibles. La estructura de orden mayor también puede evaluarse en función de diversos parámetros, tales como pH, temperatura o sales añadidas. Los ejemplos de métodos para evaluar las características de proteínas, tales como el tamaño, incluyen ultracentrifugación analítica y HPLC de exclusión por tamaño (SEC-HPLC), y los métodos a modo de ejemplo para medir la carga incluyen cromatografía de intercambio iónico e isoelectroenfoque. La hidrofobia puede evaluarse, por  
60 ejemplo, mediante HPLC en fase reversa y cromatografía HPLC de interacción hidrófoba. La glucosilación puede afectar a la farmacocinética (por ejemplo, eliminación), la conformación o la estabilidad, la unión a receptor y la función de proteínas y pueden expresarse, por ejemplo, mediante espectrometría de masas y espectroscopía de resonancia magnética nuclear (RMN).

65 Como se ha indicado anteriormente, determinadas realizaciones incluyen el uso de SEC-HPLC para evaluar características de proteínas, tales como pureza, tamaño (por ejemplo, homogeneidad de tamaño) o grado de agregación y/o para purificar proteínas, entre otros usos. SEC, que también incluye cromatografía de filtración en gel

(GFC) y cromatografía de permeación en gel (GPC), se refiere a un método cromatográfico en el que se separan moléculas en solución en un material poroso basándose en su tamaño o más específicamente, su volumen hidrodinámico, el coeficiente de difusión y/o propiedades superficiales. El proceso se usa generalmente para separar moléculas biológicas y para determinar pesos moleculares y distribuciones de peso molecular de polímeros. normalmente, se carga una muestra biológica o de proteínas (tal como un extracto de proteína producido de acuerdo con los métodos de expresión de proteínas proporcionados en el presente documento y conocidos en la técnica) en una columna de exclusión por tamaños seleccionada con una fase estacionaria definida (el material poroso), preferentemente una fase que no interactúa con las proteínas en la muestra. En determinados aspectos, la fase estacionaria está compuesta por partículas inertes empaquetadas en una matriz tridimensional densa dentro de una columna de vidrio o acero. La fase móvil puede ser agua pura, un tampón acuoso, un disolvente orgánico o una mezcla de los mismos. Las partículas de la fase estacionaria tienen normalmente poros pequeños y/o canales que permiten que entren solo moléculas por debajo de un determinado tamaño. Por lo tanto, se excluyen las partículas grandes de estos poros y canales y su interacción limitada con la fase estacionaria hace que estas se eluyan en forma de un pico "totalmente excluido" al comienzo del experimento. Las moléculas más pequeñas, que pueden caber en los poros, se retiran de la fase móvil fluida y el tiempo que pasan inmovilizadas en los poros de la fase estacionaria depende, en parte, de la profundidad con la que penetran en los poros. Su retirada del flujo de la fase móvil hace que tarden más en eluir de la columna y da como resultado una separación entre las partículas basándose en diferencias en su tamaño. Una columna de exclusión por tamaños dada tiene un intervalo de pesos moleculares que puede separarse. En su conjunto, las moléculas mayores que el límite superior no quedarán atrapadas por la fase estacionaria, las moléculas menores que el límite inferior entrarán completamente en la fase sólida y se eluirán como una sola banda y las moléculas dentro del intervalo eluirán a diferentes velocidades, definidas por sus propiedades, tales como el volumen hidrodinámico. Para ejemplos de estos métodos en la práctica con proteínas farmacéuticas, véase Bruner et al., *Journal of Pharmaceutical and Biomedical Analysis*. 15: 1929-1935, 1997.

La pureza de proteínas para aplicaciones químicas se discute también, por ejemplo, por Anicetti et al. (*Trends in Biotechnology*. 7:342-349, 1989). Las técnicas más recientes para analizar la pureza de proteínas incluyen, sin limitación, el LabChip GXII, una plataforma automatizada para un rápido análisis de proteínas y ácidos nucleicos, que proporciona un análisis de alto rendimiento del título, el tamaño y el análisis de pureza de proteínas. En determinadas realizaciones no limitantes, pueden obtenerse proteínas de grado clínico, tales como fragmentos de proteínas y anticuerpos utilizando una combinación de materiales cromatográficos en al menos dos etapas ortogonales, entre otros métodos (véase, *por ejemplo*, *Therapeutic Proteins: Methods and Protocols*. Vol. 308, Eds., Smales y James, Humana Press Inc., 2005). normalmente, los agentes proteínicos (por ejemplo, fragmentos de proteína AARS, anticuerpos, agentes de unión) y otros agentes (por ejemplo, antisentido, iARN, moléculas pequeñas) se encuentran sustancialmente libres de endotoxinas, según se mide de acuerdo con técnicas conocidas en la técnica y descritas en el presente documento.

También se incluyen ensayos de solubilidad de proteínas. Dichos ensayos pueden utilizarse, por ejemplo, para determinar las condiciones de crecimiento y purificación óptimas para la producción recombinante, para optimizar la selección de tampones, y para optimizar la selección de fragmentos de proteína AARS o variantes de la misma. La solubilidad o la agregación pueden evaluarse de acuerdo con diversos parámetros, incluyendo la temperatura, el pH, las sales y la presencia o ausencia de otros aditivos. Los ejemplos de ensayos de exploración de solubilidad incluyen, sin limitación, métodos basados en microplacas para medir la solubilidad de proteínas usando la turbidez u otra medida como criterio de valoración, ensayos de alto rendimiento para el análisis de la solubilidad de proteínas recombinantes purificadas (véase, *por ejemplo*, Stenvall et al., *Biochim Biophys Acta*. 1752:6-10, 2005), ensayos que usan complementación estructural de una proteína marcadora genética para controlar y medir el plegamiento y la solubilidad de proteínas *in vivo* (véase, *por ejemplo*, Wigley et al., *Nature Biotechnology*. 19:131-136, 2001), y exploración electroquímica de la solubilidad de proteínas recombinantes en *Escherichia coli* usando microscopía electrónica de barrido (SECM) (véase, *por ejemplo*, Nagamine et al., *Biotechnology and Bioengineering*. 96:1008-1013, 2006), entre otros. Los fragmentos de proteína AARS con solubilidad aumentada (o agregación reducida) pueden identificarse o seleccionarse de acuerdo con técnicas rutinarias en la técnica, incluyendo ensayos *in vivo* sencillos para solubilidad de proteínas (véase, *por ejemplo*, Maxwell et al., *Protein Sci*. 8:1908-11, 1999).

La solubilidad y agregación de proteínas también pueden medirse mediante técnicas de dispersión de luz dinámica. La agregación es un término general que abarca varios tipos de interacciones o características, incluyendo interacciones y características solubles/insolubles, covalentes/no covalentes, reversibles/irreversibles y nativas/desnaturalizadas. Para agentes terapéuticos proteínicos, la presencia de agregados se considera normalmente no deseada debido a la preocupación de que los agregados puedan causar una reacción inmunogénica (por ejemplo, agregados pequeños), o pueden causar acontecimientos adversos al administrarlos (por ejemplo, partículas). La dispersión de luz dinámica se refiere a una técnica que puede usarse para determinar el perfil de distribución por tamaños de pequeñas partículas en suspensión o polímeros como proteínas en solución. Esta técnica, también citada como espectroscopía de correlación fotónica (PCS) o dispersión de luz cuasielástica (QELS), usa luz dispersada para medir la velocidad de difusión de las partículas de proteína. Las fluctuaciones de la intensidad de dispersión pueden observarse debido al movimiento Browniano de las moléculas y partículas en solución. Estos datos de movimiento pueden procesarse convencionalmente para derivar una distribución por tamaños para la muestra, en donde el tamaño se proporciona por el radio de Stokes o el radio hidrodinámico de la

partícula de proteína. El tamaño hidrodinámico depende tanto de la masa como de la forma (conformación). La dispersión dinámica puede detectar la presencia de cantidades muy pequeñas de proteína agregada (<0,01 % en peso), incluso en muestras que contienen una gran cantidad de masas. También puede usarse para comparar la estabilidad de diferentes formulaciones, incluyendo, por ejemplo, aplicaciones que se basan en el control en tiempo real de los cambios a altas temperaturas. Por consiguiente, determinadas realizaciones incluyen el uso de dispersión de luz dinámica para analizar la solubilidad y/o la presencia de agregados en una muestra que contiene un fragmento de proteína AARS, un anticuerpo u otro agente de la invención.

#### IX. MÉTODOS DIAGNÓSTICOS Y COMPOSICIONES

Los agentes de AARS, tales como fragmentos de proteína AARS, polinucleótidos de AARS y anticuerpos y otros agentes de unión descritos en el presente documento pueden usarse en ensayos diagnósticos y en composiciones diagnósticas. Se incluyen métodos y composiciones bioquímicas, histológicas y basadas en células, entre otros.

Estas y otras realizaciones relacionadas incluyen la detección de las secuencias de polinucleótido de AARS o las secuencias o porciones de polipéptido de AARS correspondientes de uno o más fragmentos de proteína AARS recientemente identificados, también citados como polipéptidos de AARS. Por ejemplo, determinados aspectos incluyen la detección de las secuencias de polinucleótido de AARS o las secuencias de polipéptido correspondientes o porciones del mismo de una o más variantes de corte y empalme de AARS recientemente identificadas y/o una o más uniones de corte y empalme de aquellas variantes de corte y empalme. En determinadas realizaciones, el polinucleótido o las secuencias de polipéptido correspondientes de al menos una de las uniones de corte y empalme son únicos para esa variante de corte y empalme de AARS particular.

También se incluye la detección directa de fragmentos de proteína AARS, incluyendo variantes de corte y empalme, fragmentos proteolíticos y otros. En determinadas realizaciones, la presencia o los niveles de uno o más fragmentos de proteína AARS recientemente identificados se asocian o correlacionan con uno o más tipos celulares o estados celulares. Por tanto, puede usarse la presencia o niveles de un polipéptido o polinucleótido de AARS para distinguir entre diferentes tipos celulares en diferentes estados celulares. Pueden detectarse la presencia o los niveles de fragmentos de proteína AARS o sus polinucleótidos relacionados de acuerdo con técnicas diagnósticas basadas en polinucleótidos y/o polipéptidos, tal como se describe en el presente documento y se conoce en la técnica.

Determinados aspectos pueden emplear los fragmentos de proteína AARS, anticuerpos o polinucleótidos de AARS como parte de un método diagnóstico compañero, normalmente para evaluar si un sujeto o una población de sujetos responderá favorablemente a un tratamiento médico específico. Por ejemplo, un agente terapéutico de AARS dado (por ejemplo, fragmento de proteína, antisentido, iARN, anticuerpo, agente de unión) podría identificarse como adecuado para un sujeto o determinadas poblaciones de sujetos basándose en si los sujetos tienen uno o más biomarcadores de selección para una enfermedad o afección dada. Los ejemplos de biomarcadores incluyen marcadores séricos/tisulares así como marcadores que pueden identificarse mediante técnicas de obtención de imágenes médicas. En determinadas realizaciones, un fragmento de proteína AARS de origen natural (o su polinucleótido correspondiente) puede proporcionar por sí mismo un biomarcador sérico y/o tisular que puede utilizarse para medir el resultado farmacológico o evaluar la idoneidad del uso del fármaco en un sujeto específico o una población de sujetos específica. En determinados aspectos, la identificación de un polipéptido de AARS o una secuencia de polinucleótido de referencia puede incluir la caracterización de la expresión diferencial de esa secuencia, ya sea en un sujeto seleccionado, un tejido seleccionado u otro, tal como se describe en el presente documento y se conoce en la técnica.

Algunos de los métodos proporcionados en el presente documento se basan en la expresión diferencial de un polipéptido o polinucleótido de AARS para caracterizar la afección o el estado de una célula, tejido o sujeto y para distinguirlo de otra célula, tejido o sujeto. Los ejemplos no limitantes incluyen métodos para detectar la presencia o los niveles de un polipéptido o polinucleótido de AARS en una muestra biológica para distinguirlo entre células o tejidos de diferentes especies, células de diferentes tejidos u órganos, estados de desarrollo celular, tales como neonatal y adulto, estados de diferenciación celular, afecciones, tales como sano, enfermo y tratado, fracciones intracelulares y extracelulares, además de cultivos de células primarias y otros cultivos celulares, tales como cultivos de células inmortalizadas.

La expresión diferencial incluye una diferencia estadísticamente significativa en uno o más niveles de expresión génica de un polinucleótido de AARS o una secuencia de polipéptido de referencia en comparación con los niveles de expresión de la misma secuencia en un control adecuado. La diferencia estadísticamente significativa puede estar relacionada con un aumento o una reducción en los niveles de expresión, medidos por los niveles de ARN, los niveles de proteína, la función de proteína o cualquier otra medida relevante de expresión génica, tales como aquellas descritas en el presente documento. También se incluye una comparación entre un polinucleótido o polipéptido de AARS de la invención y una secuencia de AARS citosólica o mitocondrial de longitud completa o de tipo silvestre, normalmente del mismo tipo o uno correspondiente. La expresión diferencial puede detectarse mediante diversas técnicas en la materia y descritas en el presente documento, incluyendo técnicas basadas en polinucleótidos y polipéptidos, tales como PCR en tiempo real, hibridación substractiva, matrices de polinucleótidos y polipéptidos y otros.

Normalmente, un resultado se cita como estadísticamente significativo en caso de que sea improbable que se haya producido por azar. El nivel de significación de una prueba o resultado se relaciona tradicionalmente con un concepto de prueba de una hipótesis estadística frecuentista. En los casos sencillos, la significación estadística puede definirse como la probabilidad de tomar la decisión de rechazar la hipótesis nula cuando la hipótesis nula es de hecho verdadera (una decisión conocida como error de tipo I o "determinación de falso positivo"). Normalmente, esta decisión se toma usando el valor de  $p$ : en caso de que el valor de  $p$  sea menor que el nivel de significación, se rechaza la hipótesis nula. Cuanto menor sea el valor de  $p$ , más significativo será el resultado. También pueden usarse factores de Bayes para determinar la significación estadística (véase, *por ejemplo*, Goodman S., Ann Intern Med 130:1005-13, 1999).

En los casos más complicados pero de importancia práctica, el nivel de significación de una prueba o un resultado puede reflejar un análisis en el que la probabilidad de tomar la decisión de rechazar la hipótesis nula cuando la hipótesis nula es de hecho verdadera es meramente la probabilidad indicada. Este tipo de análisis posibilita aplicaciones en las que la probabilidad de decidir el rechazo pueda ser mucho más pequeña que el nivel de significación para algunos conjuntos de suposiciones abarcadas dentro de la hipótesis nula.

En determinadas realizaciones a modo de ejemplo, la expresión diferencial estadísticamente significativa puede incluir situaciones en donde el nivel de expresión de una secuencia de AARS dada proporciona una diferencia al menos 1,2X, 1,3X, 1,4X, 1,5X, 1,6X, 1,7X, 1,8X, 1,9X, 2,0X, 2,2X, 2,4X, 2,6X, 2,8X, 3,0X, 4,0X, 5,0X, 6,0X, 7,0X, 8,0X, 9,0X, 10,0X, 15,0X, 20,0X, 50,0X, 100,0X o más en la expresión (es decir, una expresión diferencial que puede ser una expresión mayor o menor) en una muestra biológica sospechosa en comparación con un control adecuado, incluyendo todos los números enteros y puntos decimales intermedios (por ejemplo, 1,24X, 1,25X, 2,1X, 2,5X, 60,0X, 75,0X, etc.). En determinadas realizaciones, la expresión diferencial estadísticamente significativa puede incluir situaciones en donde el nivel de expresión de una secuencia de AARS dada proporciona aproximadamente un 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 por ciento (%) o más de diferencia en la expresión (es decir, una expresión diferencial que puede ser mayor o menor) en una muestra biológica sospechosa en comparación con un control adecuado, incluyendo todos los números enteros y puntos decimales intermedios.

A modo de ejemplo adicional, también puede determinarse la expresión diferencial efectuando una prueba de Z, es decir, calculando una puntuación absoluta de Z, tal como se describe en el presente documento y se conoce en la técnica (véase el ejemplo 1). La prueba de Z se utiliza normalmente para identificar diferencias significativas entre una media de muestra y una media de población. Por ejemplo, en comparación con una tabla normal estándar (por ejemplo, un tejido de control), a un intervalo de confianza del 95% (es decir, a un nivel de significación del 5%), una puntuación de Z con un valor absoluto mayor de 1,96 indica falta de aleatoriedad. Para un intervalo de confianza del 99%, en caso de que la Z absoluta sea mayor de 2,58, significa que  $p < 0,01$  y la diferencia es incluso más significativa, la hipótesis nula puede rechazarse con mayor confianza. En estas realizaciones y otras relacionadas, una puntuación de Z absoluta de 1,96, 2, 2,58, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más, incluyendo todos los puntos decimales intermedios (por ejemplo, 10,1, 10,6, 11,2, etc.), pueden proporcionar una medición fuerte de la significación estadística. En determinadas realizaciones, una puntuación de Z absoluta de más de 6 puede proporcionar una significación estadística excepcionalmente alta.

Una similitud sustancial se refiere generalmente a la ausencia de una diferencia estadísticamente significativa en los niveles de expresión entre la muestra biológica y el control de referencia. Los ejemplos de niveles de expresión sustancialmente similares pueden incluir situaciones en donde el nivel de expresión de un SSCIGS dado proporciona una diferencia de menos de aproximadamente 0,05X, 0,1X, 0,2X, 0,3X, 0,4X, 0,5X, 0,6X, 0,7X, 0,8X, 0,9X, 1,0X, 1,1X, 1,2X, 1,3X o 1,4X en la expresión (es decir, una expresión diferencial que puede ser una expresión mayor o menor) en una muestra biológica sospechosa en comparación con una muestra de referencia, incluyendo todos los puntos decimales intermedios (por ejemplo, 0,15X, 0,25X, 0,35X, etc.). En determinadas realizaciones, la expresión diferencial puede incluir situaciones en donde el nivel de expresión de una secuencia de AARS dada proporciona una diferencia de menos de aproximadamente un 0,25, 0,5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50 por ciento (%) en la expresión (es decir, una expresión diferencial que puede ser mayor o menor) en una muestra biológica sospechosa en comparación con una muestra de referencia, incluyendo todos los puntos decimales intermedios.

En determinadas realizaciones, tal como cuando se usa una micromatriz Affymetrix para medir los niveles de expresión de un polinucleótido de AARS o una secuencia de polipéptido de referencia, la expresión diferencial puede determinarse también mediante el valor de expresión media resumido por el programa Affymetrix Microarray Suite 5 (Affymetrix, Santa Clara, CA), u otro programa informático similar, normalmente con un valor de expresión media escalada de 1000.

Las realizaciones de la presente invención incluyen métodos para detectar la presencia o los niveles de un polinucleótido de AARS o una secuencia de polipéptido de referencia o una porción de la misma para distinguir entre células u otra muestra biológica de un organismo o especie diferente, en donde la presencia o los niveles de esa secuencia se asocia con un organismo o especie seleccionada. Los ejemplos generales incluyen métodos para distinguir entre seres humanos y cualquier combinación de bacterias, hongos, plantas y otros animales no humanos.

Entre los animales se incluyen métodos para distinguir entre seres humanos y cualquier combinación de vertebrados e invertebrados, incluyendo vertebrados, tales como peces, anfibios, reptiles, aves y mamíferos no humanos e invertebrados, tales como insectos, moluscos, crustáceos y corales. Entre los mamíferos no humanos se incluyen métodos para distinguir entre humanos y cualquier combinación de mamíferos no humanos del orden *Afrosoricida*, *Macroscelidea*, *Tubulidentata*, *Hyracoidea*, *Proboscidea*, *Sirenia*, *Cingulata*, *Pilosa*, *Scandentia*, *Dermoptera*, *Primates*, *Rodentia*, *Lagomorpha*, *Erinaceomorpha*, *Soricomorpha*, *Chiroptera*, *Pholidota*, *Cetacea*, *Carnivora*, *Perissodactyla* o *Artiodactyla*. En el orden *Primate* se incluyen monos, simios, gorilas y chimpancés, entre otros conocidos en la técnica. Por consiguiente, la presencia o los niveles de un polinucleótido de AARS o una secuencia de polipéptido de referencia o una variante, como se describen en el presente documento, puede usarse para identificar la fuente de una muestra biológica dada, tal como una célula, tejido u órgano, distinguiendo entre cualquier combinación de estos organismos o distinguiendo entre seres humanos y uno cualquiera o más de estos organismos, tal como un panel de organismos. En determinadas realizaciones, también puede determinarse la fuente de una muestra biológica dada comparando la presencia o los niveles de una secuencia de AARS o una porción de la misma con un valor predeterminado.

Las realizaciones de la presente invención incluyen métodos para detectar la presencia o los niveles de un polinucleótido de AARS o una secuencia de polipéptido de referencia o una porción de la misma para distinguir entre células u otras muestras biológicas que se originan en diferentes tejidos u órganos. Los ejemplos no limitantes incluyen métodos para distinguir entre una célula u otra muestra biológica que se origina de cualquier combinación de piel (por ejemplo, dermis, epidermis, capa subcutánea), folículos pilosos, sistema nervioso (por ejemplo, cerebro, médula espinal, nervios periféricos), sistema auditivo u órganos del equilibrio (por ejemplo, oído interno, oído medio, oído externo), sistema respiratorio (por ejemplo, nariz, tráquea, pulmones), tejidos gastroesofágicos, el sistema gastrointestinal (por ejemplo, boca, esófago, estómago, intestino delgado, intestino grueso, recto), sistema vascular (por ejemplo, corazón, vasos sanguíneos y arterias), hígado, vesícula biliar, sistema linfático/inmunitario (por ejemplo, ganglios linfáticos, folículos linfoides, bazo, timo, médula ósea), sistema urogenital (por ejemplo, riñones, uréter, vejiga, uretra, cuello de útero, trompas de Falopio, ovarios, útero, vulva, próstata, glándulas bulbouretrales, epidídimo, próstata, vesículas seminales, testículos), sistema musculoesquelético (por ejemplo, músculos esqueléticos, músculos lisos, hueso, cartílago, tendones, ligamentos), tejido adiposo, mamario y el sistema endocrino (por ejemplo, hipotálamo, pituitaria, tiroides, páncreas, glándulas suprarrenales). Por tanto, basándose en la asociación de un polinucleótido de AARS o una secuencia de polipéptido tal como se describe en el presente documento, pueden usarse estos métodos para identificar o caracterizar el tejido u órgano del que se deriva una célula u otra muestra biológica.

Las realizaciones de la presente invención incluyen métodos para detectar la presencia o los niveles de un polinucleótido de AARS o una secuencia de polipéptido de referencia o una porción de la misma para distinguir entre o caracterizar el estado de desarrollo o diferenciación de la célula. También se incluyen métodos para diferenciar entre células germinales, células madre y células somáticas. Los ejemplos de estados de desarrollo incluyen neonatales y adultos. Los ejemplos de estados de diferenciación celular incluyen todos los estados discretos e identificables entre una célula totipotente, una célula pluripotente, una célula madre multipotente progenitora y una célula madura, completamente diferenciada.

Una célula totipotente tiene un potencial total, surge normalmente durante la reproducción sexual y asexual e incluye esporas y cigotos, aunque en determinados casos las células pueden desdiferenciarse y recuperar la totipotencia. Una célula pluripotente incluye una célula madre que tiene el potencial de diferenciar en cualquiera de las tres capas germinales, incluyendo el endodermo (revestimiento interior del estómago, tracto gastrointestinal, los pulmones), el mesodermo (músculo, hueso, sangre, urogenital), y el ectodermo (tejidos epidérmicos y sistema nervioso). Las células progenitoras multipotentes son normalmente capaces de diferenciarse en un número limitado de tipos de tejidos. Los ejemplos de células multipotentes incluyen, sin limitación, células madre hematopoyéticas (células madre adultas) de la médula ósea que dan lugar a células inmunitarias, tales como glóbulos rojos, glóbulos blancos y plaquetas, células madre mesenquimales (células madre adultas) de la médula ósea que dan lugar a células estromales, adipocitos y diversos tipos de células óseas, células madre epiteliales (células progenitoras) que dan lugar a los diversos tipos de células de la piel y células satelitales musculares (células progenitoras) que contribuyen al tejido muscular diferenciado. Por consiguiente, la presencia o los niveles de un polinucleótido de AARS particular o de una secuencia de polipéptido (por ejemplo, unión de corte y empalme de una variante de corte y empalme de AARS, fragmento proteolítico de AARS), puede usarse para distinguir entre o caracterizar los estados de diferenciación celular anteriormente indicados, en comparación con un control o un nivel predeterminado.

En la medida que la invención se refiere a células embrionarias o cigotos, estos no son humanos.

Las realizaciones de la presente invención incluyen métodos para detectar la presencia o los niveles de un polinucleótido de AARS o una secuencia de polipéptido de referencia para caracterizar o diagnosticar la afección o una célula, tejido, órgano o sujeto, en el que dicha afección puede caracterizarse como sana, enferma, en riesgo de enfermedad o tratada. Para dichos fines diagnósticos, el término "diagnóstico" o "diagnosticado" incluyen identificar la presencia o la naturaleza de una afección patológica, caracterizar el riesgo de desarrollar dicha afección y/o medir el cambio (o no cambio) de una afección patológica en respuesta a la terapia. Los métodos diagnósticos pueden diferir en cuanto a su sensibilidad y especificidad. En determinadas realizaciones, la "sensibilidad" de un ensayo

- diagnóstico se refiere al porcentaje de células, tejidos o sujetos enfermos que dan positivo en la prueba (porcentaje de "auténticos positivos"). Las células, tejidos o sujetos enfermos no detectados por el ensayo se citan normalmente como "falsos negativos". Las células, tejidos o sujetos que no están enfermos y que dan negativo en el ensayo pueden denominarse "auténticos negativos". En determinadas realizaciones, la "especificidad" de un ensayo
- 5 diagnóstico puede definirse como uno (1) menos la tasa de falsos positivos, donde la tasa de "falsos positivos" se define como la proporción de aquellas muestras o sujetos son la enfermedad y que dan positivo. Aunque un método diagnóstico particular puede no proporcionar un diagnóstico definitivo de una afección, es suficiente si el método proporciona una indicación positiva que ayuda en el diagnóstico.
- 10 En determinados casos, puede diagnosticarse la presencia o el riesgo de desarrollar una afección patológica comparando la presencia o los niveles de uno o más polinucleótidos de AARS seleccionados o secuencias de polipéptido de referencia o porciones de las mismas que se correlacionan con la afección, ya sea por niveles aumentados o reducidos, en comparación con un control adecuado. Un "control adecuado" o "control apropiado" incluye un valor, nivel, rasgo, característica o propiedad determinada en una célula u otra muestra biológica de
- 15 tejido u organismo, *por ejemplo*, una célula, tejido u organismo de control o normal que muestra, por ejemplo, rasgos normales, tales como la ausencia de la afección. En determinadas realizaciones, un "control adecuado" o "control apropiado" es un valor, nivel, rasgo, característica o propiedad predefinida. Otros controles adecuados serán evidentes para los expertos en la materia. Los ejemplos de enfermedades y afecciones se describen en otras partes del presente documento.
- 20 Las realizaciones de la presente invención incluyen técnicas de detección basadas en polinucleótidos o ácidos nucleicos de AARS, que ofrecen ciertas ventajas debido a la sensibilidad de la detección. Por tanto, determinadas realizaciones se refieren al uso o la detección de polinucleótidos de AARS como parte de un método o ensayo diagnóstico. Puede medirse la presencia y/o los niveles de polinucleótidos de AARS mediante cualquier método
- 25 conocido en la técnica, incluyendo ensayos de hibridación, tales como transferencia de Northern, reacción en cadena de la polimerasa (PCR) cuantitativa o cualitativa, PCR de retrotranscriptasa (RT-PCR) cuantitativa o cualitativa, micromatriz, transferencias de puntos o ranuras o hibridación *in situ*, tal como hibridación *in situ* de fluorescencia (FISH), entre otros. Algunos de estos métodos se describen con más detalle más adelante.
- 30 Los polinucleótidos de AARS, tales como ADN y ARN, pueden recogerse y/o generarse a partir de sangre, fluidos biológicos, tejidos u órganos, líneas celulares u otra muestra relevante usando técnicas conocidas en la materia, tales como aquellas descritas en Kingston. (2002 Current Protocols in Molecular Biology, Greene Publ. Assoc. Inc. y John Wiley & Sons, Inc., NY, NY (véase, *por ejemplo*, como se describe por Nelson et al. Proc Natl Acad Sci U S A, 99: 11890-11895, 2002) y en otras partes. Además, diversos kits comerciales para construir ARN es útil para
- 35 producir el ARN para su uso en la presente invención. El ARN puede construirse a partir de órganos/tejidos/células obtenidas de sujetos sanos normales; sin embargo, la presente invención también contempla la construcción de ARN de sujetos enfermos. Algunas realizaciones contemplan el uso de cualquier tipo de órgano de cualquier tipo de sujeto o animal. Para las muestras de ensayo, el ARN puede obtenerse de un individuo (por ejemplo, cualquier animal, incluyendo mamíferos) con o sin enfermedad visible y de muestras de tejidos, fluidos biológicos (por ejemplo, sangre entera) o similares.
- 40 En determinadas realizaciones, puede ser útil la amplificación o construcción de secuencias de ADNc para aumentar las capacidades de detección. La presente divulgación, así como la técnica, proporciona el nivel de detalle requerido para llevar a cabo dichas tareas. En una realización ilustrativa, se usa sangre completa como fuente de ARN y por consiguiente, se usan opcionalmente reactivos estabilizantes del ARN, tales como tubos PAX, tal como se describe, por ejemplo, en Thach et al., J. Immunol. Methods. Dec 283(1-2):269-279, 2003 y Chai et al., J. Clin. Lab Anal. 19(5):182-188, 2005 (ambos incorporados por referencia). Pueden generarse bibliotecas de ADN complementario (ADNc) usando técnicas conocidas en la materia, tales como aquellas descritas en Ausubel et al. (2001 Current Protocols in Molecular Biology, Greene Publ. Assoc. Inc. y John Wiley & Sons, Inc., NY, NY); Sambrook et al. (1989
- 50 Molecular Cloning, segunda edición, Cold Spring Harbor Laboratory, Plainview, NY); Maniatis et al. (1982 Molecular Cloning, Cold Spring Harbor Laboratory, Plainview, NY) y en otras partes. Además, diversos kits comerciales para construir bibliotecas de ADNc son útiles para producir las bibliotecas de ADNc de la presente invención. Las bibliotecas pueden construirse a partir de órganos/tejidos/células obtenidas de sujetos normales sanos.
- 55 Determinadas realizaciones pueden emplear métodos de hibridación para detectar secuencias de polinucleótido de AARS. Los métodos para llevar a cabo ensayos de hibridación de polinucleótidos se han desarrollado bien en la técnica. Los procedimientos y las condiciones de los ensayos de hibridación variarán dependiendo de la aplicación y se seleccionan de acuerdo con los métodos de unión generales conocidos, incluyendo aquellos citado en: Maniatis et al. Molecular Cloning: A Laboratory Manual (2ª Ed. Cold Spring Harbor, N.Y., 1989); Berger and Kimmel Methods in Enzymology, Vol. 152, Guide to Molecular Cloning Techniques (Academic Press, Inc., San Diego, Calif., 1987); Young y Davis, PNAS. 80: 1194 (1983). Los métodos y aparatos para llevar a cabo reacciones de hibridación repetidas y controladas se han descrito en las Patentes de los Estados Unidos n.º 5.871.928, 5.874.219, 6.045.996 y 6.386.749, 6.391.623, incorporándose cada una de ellas al presente documento por referencia.
- 60
- 65 Determinadas realizaciones pueden emplear métodos de amplificación de ácidos nucleicos para detectar secuencias de polinucleótido de AARS. El término "amplificación" o la expresión "amplificación de ácidos nucleicos" se refieren a

la producción de múltiples copias de un ácido nucleico diana que contiene al menos una porción de la secuencia de ácido nucleico diana específica prevista. Las múltiples copias pueden citarse como amplicones o productos de amplificación. En determinadas realizaciones, la diana amplificada contiene menos de la secuencia génica diana completa (intrones y exones) o una secuencia de gen diana expresada (transcrito cortado y empalmado de exones y secuencias no traducidas flanqueantes). Por ejemplo, pueden producirse amplicones específicos amplificando una parte del polinucleótido diana usando cebadores de amplificación que hibridan con e inician la polimerización a partir de, posiciones internas del polinucleótido diana. Preferentemente, la porción amplificada contiene una secuencia diana detectable que puede detectarse usando cualquiera de una serie de métodos de sobra conocidos.

5  
10 La "amplificación selectiva" o la "amplificación específica", tal como se usa en el presente documento, se refiere a la amplificación de una secuencia de ácido nucleico diana de acuerdo con la presente invención, en donde la amplificación de la secuencia diana está limitada sustancialmente a la amplificación de la secuencia diana aportada por una muestra de ácido nucleico de interés que se está ensayando y no se aporta por la secuencia de ácido nucleico diana aportada por cualquier otra fuente de muestra, *por ejemplo*, la contaminación presente en los reactivos usados durante las reacciones de amplificación o en el ambiente en el que se efectúan las reacciones de amplificación.

15  
20 La expresión "condiciones de amplificación" se refiere a las condiciones que permiten la amplificación de ácidos nucleicos de acuerdo con la presente invención. Las condiciones de amplificación pueden, en algunas realizaciones, ser menos rigurosas que las "condiciones de hibridación rigurosas" descritas en el presente documento. Los oligonucleótidos usados en las reacciones de amplificación de la presente invención hibridan con sus dianas previstas en condiciones de amplificación, pero pueden hibridar o no en condiciones de hibridación rigurosas. Por otro lado, las sondas de detección de la presente invención hibridan normalmente en condiciones de hibridación rigurosas. Las condiciones aceptables para llevar a cabo las amplificaciones de ácidos nucleicos de acuerdo con la presente invención pueden determinarse fácilmente por alguien normalmente versado en la materia, dependiendo del método de amplificación concreto empleado.

25  
30 Muchos métodos de sobra conocidos para la amplificación de ácidos nucleicos requieren de ciclos térmicos para, de manera alterna, desnaturalizar los ácidos nucleicos bicatenarios e hibridar los cebadores; sin embargo, otros métodos de amplificación de ácidos nucleicos son isotermos. La reacción en cadena de la polimerasa (Patentes de los Estados Unidos n.º 4.683.195; 4.683.202; 4.800.159; 4.965.188), normalmente citada como PCR, usa múltiples ciclos de desnaturalización, hibridación de pares de cebador a hebras complementarias y extensión de cebador para aumentar exponencialmente el número de copias de la secuencia diana. En una variación denominada RT-PCR, se usa retrotranscriptasa (RT) para producir un ADN complementario (ADNc) a partir de ARNm y después el ADNc se amplifica mediante la PCR para producir múltiples copias de ADN.

35  
40 Como se ha indicado anteriormente, el término "PCR" se refiere a múltiples ciclos de amplificación que amplifican de manera selectiva una especie de ácido nucleico diana. Se incluyen la PCR cuantitativa (qPCR), la PCR en tiempo real (rtPCR), la PCR de retrotranscripción (RT-PCR) y la PCR de retrotranscripción cuantitativa (qRT-PCR) se encuentra bien descrita en la técnica. El término "qPCR" se refiere a la reacción en cadena de la polimerasa cuantitativa y el término "qRT-PCR" se refiere a la reacción en cadena de la polimerasa de retrotranscripción cuantitativa. La qPCR y la qRT-PCR pueden usarse para amplificar y cuantificar simultáneamente una molécula de ADNc diana. Esta posibilidad tanto la detección como la cuantificación de una secuencia específica en una agrupación de ADNc, tal como un gen o transcrito de AARS seleccionado.

45  
50 La "PCR en tiempo real" puede usar colorante de unión a ADN para todos los ADN bicatenarios en la PCR, provocando la fluorescencia del tinte. Por lo tanto, un aumento en el producto de ADN durante la PCR da lugar a un aumento en la intensidad de fluorescencia y se mide en cada ciclo, permitiendo de este modo cuantificar las concentraciones de ADN. Sin embargo, los colorantes de ADNbc, tales como SYBR Green, se unirán a todos los productos de ADNbc de la PCR. La fluorescencia se detecta y se mide en un termociclador de PCR en tiempo real y su aumento genómico correspondiente al aumento exponencial en el producto se usa para determinar el ciclo umbral ("Ct") en cada reacción.

55 La expresión "puntuación de Ct" se refiere al número del ciclo umbral, que es el ciclo en el cual la amplificación PCR ha sobrepasado un nivel umbral. En caso de que haya una mayor cantidad de ARNm para un gen particular en una muestra, traspasará el umbral antes que un gen poco expresado ya que hay más ARN de inicio para amplificar. Por lo tanto, una baja puntuación de Ct indica una alta expresión génica en una muestra y una alta puntuación de Ct indica baja expresión génica.

60 Determinadas realizaciones pueden emplear la reacción en cadena de la ligasa (Weiss, Science. 254: 1292, 1991), normalmente citada como LCR, que usa dos conjuntos de oligonucleótidos de ADN complementario que hibridan con regiones adyacentes del ácido nucleico diana. Los oligonucleótidos de ADN se unen covalentemente mediante una ADN ligasa en ciclos repetidos de desnaturalización térmica, hibridación y ligadura para producir un producto oligonucleotídico ligado bicatenario detectable.

65

Otro método es la amplificación de desplazamiento de hebra (Walker, G. et al., 1992, Proc. Natl. Acad. Sci. USA 89:392-396; Patentes de los Estados Unidos n.º 5.270.184 y 5.455.166), normalmente citada como SDA, que usa ciclos de hibridación de pares de secuencias de cebador para hebras opuestas de una secuencia diana, extensión del cebador en presencia de una dNTPaS para producir un dúplex de producto de extensión del cebador hemifosforotioado, corte mediado por endonucleasa de un sitio de reconocimiento de endonucleasa de restricción hemimodificado y extensión del cebador mediada por la polimerasa desde el extremo 3' del corte para desplazar una hebra existente y producir una hebra para la siguiente ronda de hibridación del cebador, corte y desplazamiento de la hebra, dando como resultado la amplificación geométrica del producto. La SDA termófila (tSDA) usa endonucleasas y polimerasas termófilas a mayores temperaturas esencialmente en el mismo método (Patente Europea n.º 0 684 315).

Otros métodos de amplificación incluyen, por ejemplo: amplificación basada en secuencias de ácidos nucleicos (Patente de los Estados Unidos n.º 5.130.238), normalmente citada como NASBA; una que usa una ARN replicasa para amplificar la molécula de la sonda en sí (Lizardi, P. et al., 1988, BioTechnol. 6: 1197-1202), normalmente citada como Q $\beta$  replicasa; un método de amplificación basado en la transcripción (Kwoh, D. et al., 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177); replicación de secuencias autosostenida (Guatelli, J. et al., 1990, Proc. Natl. Acad. Sci. USA 87: 1874-1878); y, amplificación mediada por transcripción (Patentes de los Estados Unidos n.º 5.480.784 y 5.399.491), normalmente citada como TMA. Para una descripción adicional de métodos de amplificación conocidos, véase Persing, David H., 1993, "In Vitro Nucleic Acid Amplification Techniques" en Diagnostic Medical Microbiology: Principles and Applications (Persing et al., Eds.), págs. 51-87 (American Society for Microbiology, Washington, DC).

Los sistemas de amplificación basados en la transcripción ilustrativos de la presente invención incluyen TMA, que emplea una ARN polimerasa para producir múltiples transcritos de ARN de una región diana (Patentes de los Estados Unidos n.º 5.480.784 y 5.399.491). La TMA usa un "promotor-cebador" que hibrida con un ácido nucleico diana en presencia de una retrotranscriptasa y una ARN polimerasa para formar un promotor bicatenario a partir del cual la ARN polimerasa produce transcritos de ARN. Estos transcritos pueden convertirse en moldes para rondas posteriores de TMA en presencia de un segundo cebador capaz de hibridar con los transcritos de ARN. A diferencia de la PCR, la LCR u otros métodos que requieren la desnaturalización térmica, la TMA es un método isoterma que usa una actividad de RNasa H para digerir la hebra de ARN de un híbrido de ARN:ADN, haciendo de este modo que la hebra de ADN se encuentre disponible para la hibridación con un cebador o promotor-cebador. En general, se usa la actividad de la RNasa H asociada con la retrotranscriptasa proporcionada para la amplificación.

En un método de TMA ilustrativo, un cebador de amplificación un promotor-cebador oligonucleotídico que comprende una secuencia promotora que se hace funcional cuando es bicatenaria, ubicada en 5' de una secuencia de unión a la diana, que es capaz de hibridar con un sitio de unión de un ARN diana en una ubicación 3' respecto de la secuencia que se va a amplificar. Un promotor-cebador puede citarse como un "T7-cebador" cuando es específico para el reconocimiento de la ARN polimerasa T7. En determinadas circunstancias, puede modificarse el extremo 3' de un promotor-cebador o una subpoblación de dichos promotores-cebadores para bloquear o reducir la extensión del cebador. A partir de un promotor-cebador no modificado, la retrotranscriptasa crea una copia de ADNc del ARN diana, mientras que la actividad de RNasa H degrada el ARN diana. Después, un segundo cebador de amplificación se une al ADNc. Este cebador puede citarse como un "cebador no T7" para distinguirlo del "cebador T7". A partir de este segundo cebador de amplificación, la retrotranscriptasa crea otra hebra de ADN, dando como resultado un ADN bicatenario con un promotor funcional en un extremo. Cuando es bicatenaria, la secuencia del promotor es capaz de unirse a una ARN polimerasa para iniciar la transcripción de la secuencia diana con la que hibrida el promotor-cebador. Una ARN polimerasa usa esta secuencia de promotor para producir múltiples transcritos de ARN (es decir, amplicones), generalmente de aproximadamente 100 a 1.000 copias. Cada amplicón recientemente sintetizado puede hibridar con el segundo cebador de amplificación. La retrotranscriptasa puede crear entonces un ADN copia, mientras que la actividad de RNasa H degrada el ARN de este dúplex de ARN:ADN. Después, el promotor-cebador puede unirse al ADN recientemente sintetizado, permitiendo que la retrotranscriptasa cree un ADN bicatenario, a partir del cual la ARN polimerasa produce múltiples amplicones. Por lo tanto, puede lograrse una amplificación isotérmica de varios miles de millones de veces usando dos cebadores de amplificación.

En determinadas realizaciones, pueden usarse otras técnicas para evaluar transcritos de ARN de los transcritos de una biblioteca de ADNc particular, incluyendo análisis de micromatrices (Han, M., et al., Nat Biotechnol, 19: 631-635, 2001; Bao, P., et al., Anal Chem, 74: 1792-1797, 2002; Schena et al., Proc. Natl. Acad. Sci. USA 93:10614-19, 1996; y Heller et al., Proc. Natl. Acad. Sci. USA 94:2150-55, 1997) y SAGE (análisis seriado de la expresión génica). Al igual que MPSS, SAGE es digital y puede generar un gran número de secuencias de firma. (véase, por ejemplo, Velculescu, V. E., et al., Trends Genes, 16: 423-425., 2000; Tuteja R. y Tuteja N. Bioessays. agosto de 2004; 26(8):916-22), aunque con menores órdenes de magnitud que los disponibles a partir de técnicas tales como MPSS.

En determinadas realizaciones, el término "micromatriz" incluye una "micromatriz de ácido nucleico" que tiene una pluralidad de ácidos nucleicos unidos a sustrato, siendo detectable por separado la hibridación a cada uno de la pluralidad de ácidos nucleicos unidos. El sustrato puede ser sólido o poroso, planar o no planar, unitario o distribuido. Las micromatrices de ácidos nucleicos incluyen todos los dispositivos nombrados en Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); Nature Genet. 21(1) (supl.): 1-60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing

Company/BioTechniques Books Division (2000). Las micromatrices de ácidos nucleicos pueden incluir una pluralidad de ácidos nucleicos unidos a sustrato en las que la pluralidad de ácidos nucleicos se encuentran dispuestos sobre una pluralidad de perlas, en lugar de sobre un sustrato planar unitario, tal como se describe, por ejemplo, en Brenner et al., Proc. Natl. Acad. Sci. USA 97(4): 1665-1670 (2000). Pueden encontrarse ejemplos de micromatrices de ácidos nucleicos en las Patentes de los Estados Unidos n.º 6.391.623, 6.383.754, 6.383.749, 6.380.377, 6.379.897, 6.376.191, 6.372.431, 6.351.712, 6.344.316, 6.316.193, 6.312.906, 6.309.828, 6.309.824, 6.306.643, 6.300.063, 6.287.850, 6.284.497, 6.284.465, 6.280.954, 6.262.216, 6.251.601, 6.245.518, 6.263.287, 6.251.601, 6.238.866, 6.228.575, 6.214.587, 6.203.989, 6.171.797, 6.103.474, 6.083.726, 6.054.274, 6.040.138, 6.083.726, 6.004.755, 6.001.309, 5.958.342, 5.952.180, 5.936.731, 5.843.655, 5.814.454, 5.837.196, 5.436.327, 5.412.087 y 5.405.783, cuyas divulgaciones se incorporan por referencia.

Los ejemplos adicionales incluyen matrices de ácido nucleico que se encuentran disponibles comercialmente de Affymetrix (Santa Clara, Calif.) con el nombre comercial GENECHIP™. Se proporcionan métodos a modo de ejemplo adicionales para fabricar y usar matrices en, por ejemplo, las Patentes de los Estados Unidos n.º 7.028.629; 7.011.949; 7.011.945; 6.936.419; 6.927.032; 6.924.103; 6.921.642; y 6.818.394.

La presente invención, al estar relacionada con matrices y micromatrices, también contempla muchos usos para polímeros unidos a sustratos sólidos. Estos usos incluyen el control de la expresión génica, la elaboración de perfiles, la exploración de bibliotecas, el genotipado y los diagnósticos. Los métodos de control y elaboración de perfiles de la expresión génica y los métodos útiles para el control y la elaboración de perfiles de la expresión génica se muestran en las Patentes de los Estados Unidos n.º 5.800.992, 6.013.449, 6.020.135, 6.033.860, 6.040.138, 6.177.248 y 6.309.822. El genotipado y los usos del mismo se muestran en los n.º de Serie de los Estados Unidos 10/442.021, 10/013.598 (solicitud de los Estados Unidos n.º 2003/0036069), y las Patentes de los Estados Unidos n.º 5.925.525, 6.268.141, 5.856.092, 6.267.152, 6.300.063, 6.525.185, 6.632.611, 5.858.659, 6.284.460, 6.361.947, 6.368.799, 6.673.579 y 6.333.179. Otros métodos para la amplificación, marcaje y análisis de ácidos nucleicos que pueden usarse en combinación con los métodos divulgados en el presente documento se representan en las Patentes de los Estados Unidos n.º 5.871.928, 5.902.723, 6.045.996, 5.541.061 y 6.197.506.

Tal como será evidente para los expertos en la materia, algunas realizaciones pueden emplear oligonucleótidos, tales como cebadores o sondas, para la amplificación o detección, tal como se describen en el presente documento. Pueden producirse oligonucleótidos con una secuencia y estructura química definida mediante técnicas conocidas por los expertos en la materia, tales como mediante síntesis química o bioquímica y mediante expresión *in vitro* o *in vivo* a partir de moléculas de ácido nucleico recombinante, *por ejemplo*, vectores bacterianos o víricos. En determinadas realizaciones, un oligonucleótido no consiste únicamente en ADN cromosómico de tipo silvestre o los productos de transcripción *in vivo* del mismo.

Los oligonucleótidos o cebadores pueden modificarse de cualquier mod, en tanto que una modificación dada sea compatible con la función deseada de un oligonucleótido dado. Un experto en la materia podrá determinar fácilmente si una modificación dada es adecuada o deseada para cualquier oligonucleótido dado de la presente invención. Los oligonucleótidos de AARS relevantes se describen con más detalle en otras partes del presente documento.

Aunque el diseño y la secuencia de los oligonucleótidos depende de su función, tal como se describe en el presente documento, normalmente se tienen en cuenta diversas variables. Entre las más relevantes se encuentran: longitud, temperatura de fusión (T<sub>m</sub>), especificidad, complementariedad con otros oligonucleótidos en el sistema, contenido de G/Cm, tramos de polipirimidina (T, C) o de polipurina (A, G) y la secuencia del extremo 3'. El control de estas y otras variables es un aspecto convencional y de sobra conocido del diseño de oligonucleótidos y hay disponibles varios programas informáticos para explorar grandes números de oligonucleótidos potenciales respecto de los óptimos.

Por lo tanto, determinadas realizaciones incluyen métodos para detectar un polinucleótido diana de AARS en una muestra, comprendiendo el polinucleótido la secuencia de un polinucleótido de AARS de referencia, como se describe en el presente documento, que comprende a) hibridar la muestra con una sonda que comprende una secuencia complementaria al polinucleótido diana en la muestra e hibridando dicha sonda específicamente a dicho polinucleótido diana, en condiciones mediante las cuales se forma un complejo de hibridación entre dicha sonda y dicho polinucleótido diana o fragmentos del mismo y b) detectar la presencia o ausencia de dicho complejo de hibridación y opcionalmente, si está presente, la cantidad del mismo. También se incluyen métodos para detectar un polinucleótido de AARS diana en una muestra, comprendiendo el polinucleótido la secuencia de un polinucleótido de AARS de referencia, como se describen en el presente documento, que comprenden a) amplificar el polinucleótido diana o un fragmento del mismo y b) detectar la presencia o ausencia de dicho polinucleótido diana amplificado o un fragmento del mismo y, opcionalmente, si está presente, la cantidad del mismo. Las realizaciones específicas se refieren a la detección de variantes de corte y empalme de AARS, tal como detectando una unión de corte y empalme única de la variante de corte y empalme, ya sea por hibridación, amplificación u otro método de detección.

Las realizaciones de la presente invención incluyen una serie de técnicas de detección basadas en polipéptidos de AARS, incluyendo técnicas de detección basadas en anticuerpos. En estas realizaciones se incluye el uso de los polipéptidos de AARS para generar anticuerpos u otras moléculas de unión, que después pueden usarse en

métodos y composiciones diagnósticas para detectar o cuantificar polipéptidos de AARS seleccionados en una célula u otra muestra biológica, normalmente de un sujeto.

5 Determinadas realizaciones pueden emplear metodologías y detectores convencionales, tales como transferencia de Western e inmunoprecipitación, ensayos inmunoabsorbentes acoplados a enzimas (ELISA), citometría de flujo y ensayos de inmunofluorescencia (IFA), que utilizan un dispositivo para la obtención de imágenes. Estos métodos de  
10 sobre conocidos utilizan normalmente uno o más anticuerpos monoclonales o policlonales, tal como se describe en el presente documento, que se unen específicamente a un polipéptido de AARS seleccionado de la invención o a una región única de ese polipéptido de AARS y generalmente no se unen de manera significativa a otros polipéptidos de AARS, tal como un polipéptido de AARS de longitud completa. En determinadas realizaciones, la  
15 región única del polipéptido de AARS puede representar una estructura tridimensional única que es poseída por un fragmento de proteína recientemente identificado de un AARS.

15 Determinadas realizaciones pueden emplear "matrices", tales como "micromatrices". En determinadas realizaciones, una "micromatriz" también puede referirse a una "micromatriz de péptidos" o "micromatriz de proteínas" que tiene una colección o pluralidad de polipéptidos unidos al sustrato, siendo detectable por separado la unión de cada uno de la pluralidad de péptidos unidos. Como alternativa, la micromatriz de péptidos puede tener diversas moléculas de  
20 unión, incluyendo, pero sin limitación, anticuerpos monoclonales, anticuerpos policlonales, moléculas de unión de presentación en fagos, moléculas de unión de 2 híbridos de levadura y aptámeros, que pueden detectar específicamente la unión de los polipéptidos de AARS descritos en el presente documento. La matriz puede estar basada en la detección de autoanticuerpos de estos polipéptidos de AARS, tal como se describe, por ejemplo, en  
25 Robinson et al., Nature Medicine 8(3):295-301 (2002). Pueden encontrarse ejemplos de matrices de péptidos en los documentos WO 02/31463, WO 02/25288, WO 01/94946, WO 01/88162, WO 01/68671, WO 01/57259, WO 00/61806, WO 00/54046, WO 00/47774, WO 99/40434, WO 99/39210 y WO 97/42507 y en las Patentes de los Estados Unidos n.º 6.268.210, 5.766.960 y 5.143.854, incorporándose cada una de ellas por referencia.

30 Determinadas realizaciones pueden emplear EM u otros métodos basados en el peso molecular para detectar de manera diagnóstica secuencias de polipéptido de AARS. La espectrometría de masas (EM) se refiere generalmente a una técnica analítica para determinar la composición elemental de una muestra o molécula. La EM también puede usarse para determinar las estructuras químicas de las moléculas, tales como péptidos y otros compuestos  
35 químicos.

En general, el principio de la EM consiste en ionizar compuestos químicos para generar moléculas o fragmentos de moléculas cargados y después medir sus relaciones de masa a carga. En un procedimiento de EM ilustrativo, se  
35 carga una muestra en el instrumento de EM y se somete a vaporización, los componentes de la muestra se ionizan mediante uno de una serie de métodos (por ejemplo, haciéndoles impactar con un haz de electrones), lo que da como resultado la formación de partículas cargadas positivamente, después se aceleran los iones positivos mediante un campo magnético, se efectúan cálculos acerca de la relación de masa a carga ( $m/z$ ) de las partículas basándose en los detalles del movimiento de los iones a medida que transitan a lo largo de campos electromagnéticos y, se  
40 detectan los iones, que en la etapa anterior se clasificaron según su  $m/z$ .

Un instrumento de EM ilustrativo tiene tres módulos: una fuente de ionización, que convierte las moléculas de la muestra en fase gaseosa en iones (o, en el caso de ionización por electropulverización, mueven los iones que  
45 existen en solución a la fase gaseosa); un analizador de masas, que clasifica los iones por sus masas aplicando campos electromagnéticos; y un detector, que mide el valor de una cantidad indicadora y proporciona por lo tanto datos para calcular las abundancias de cada ion presente.

La técnica de la EM tiene usos tanto cualitativos como cuantitativos, incluyendo la identificación de compuestos desconocidos, determinar la composición isotópica de elementos en una molécula y determinar la estructura de un  
50 compuesto observando su fragmentación. Otros usos incluyen cuantificar la cantidad de un compuesto en una muestra o estudiar los fundamentos de la química de iones en fase gaseosa (la química de iones y moléculas neutras en el vacío). Se incluyen cromatografía de gases-espectrometría de masas (GC/MS o GC-MS), cromatografía líquida-espectrometría de masas (CL/EM o CL-EM), y la espectrometría de movilidad de iones/espectrometría de masas (EMI/EM o EMIS). Por consiguiente, pueden usarse técnicas de EM de acuerdo con  
55 cualquiera de los métodos proporcionados en el presente documento para medir la presencia o los niveles de un polipéptido de AARS de la invención en una muestra biológica y comparar estos niveles con una muestra de control o un valor predeterminado.

60 Determinadas realizaciones pueden emplear clasificación celular o visualización celular o dispositivos/técnicas para la obtención de imágenes para detectar o cuantificar la presencia o los niveles de polinucleótidos o polipéptidos de AARS. Los ejemplos incluyen citometría de flujo o FACS, análisis de inmunofluorescencia (IFA), y técnicas de hibridación *in situ*, tales como hibridación *in situ* de fluorescencia (FISH).

65 Determinadas realizaciones pueden emplear métodos biológicos, programas informáticos y sistemas convencionales para fines diagnósticos. Los productos de programas informáticos de la invención incluyen normalmente un medio legible por ordenador que tiene instrucciones que pueden ejecutarse por el ordenador para llevar a cabo las etapas

lógicas del método de la invención. Los medios legibles por ordenador adecuados incluyen discos flexibles, CD-ROM/DVD/DVD-ROM, discos duros, memorias flash, ROM/RAM, cintas magnéticas etc. Las instrucciones ejecutables por el ordenador pueden estar escritas en un lenguaje informático adecuado o una combinación de varios lenguajes. Los métodos básicos de biología computacional se describen en, por ejemplo, Setubal y Meidanis et al., *Introduction to Computational Biology Methods* (PWS Publishing Company, Boston, 1997); Salzberg, Searles, Kasif, (Ed.), *Computational Methods in Molecular Biology*, (Elsevier, Ámsterdam, 1998); Rashidi y Buchler, *Bioinformatics Basics: Application in Biological Science and Medicine* (CRC Press, Londres, 2000) y Ouelette y Bzevanis *Bioinformatics: A Practical Guide for Analysis of Gene and Proteins* (Wiley & Sons, Inc., 2ª ed., 2001). Véase la Patente de los Estados Unidos n.º 6.420.108.

Determinadas realizaciones pueden emplear varios productos de programas informáticos para diversos fines, tales como el diseño de sondas, la gestión de datos, el análisis y el manejo de instrumentos. Véase, las Patentes de los Estados Unidos n.º 5.593.839, 5.795.716, 5.733.729, 5.974.164, 6.066.454, 6.090.555, 6.185.561, 6.188.783, 6.223.127, 6.229.911 y 6.308.170.

El ensayo de muestreo de genoma completo (WGSA) se describe, por ejemplo, en Kennedy et al., *Nat. Biotech.* 21, 1233-1237 (2003), Matsuzaki et al., *Gen. Res.* 14: 414-425, (2004), y Matsuzaki, et al., *Nature Methods* 1:109-111 (2004). Los algoritmos para su uso con ensayos de mapeo se describen, por ejemplo, en Liu et al., *Bioinformatics.* 19: 2397-2403 (2003) y Di et al. *Bioinformatics.* 21:1958 (2005). Se divulgan métodos adicionales relacionados con WGSA y matrices útiles para WGSA y aplicaciones de WGSA, por ejemplo, en las Solicitudes de Patente de los Estados Unidos n.º 60/676.058, presentada el 29 de abril de 2005, 60/616.273 presentada el 5 de octubre de 2004, 10/912.445, 11/044.831, 10/442.021, 10/650.332 y 10/463.991. Los estudios de asociación de todo el genoma usando ensayos de mapeo se describen en, por ejemplo, Hu et al., *Cancer Res.*; 65(7):2542-6 (2005), Mitra et al., *Cancer Res.*, 64(21):8116-25 (2004), Butcher et al., *Hum Mol Genet.*, 14(10):1315-25 (2005), y Klein et al., *Science.* 308(5720):385-9 (2005).

Además, determinadas realizaciones pueden incluir métodos para proporcionar información genética por las redes, tales como Internet, tal como se muestra, por ejemplo, en las Solicitudes de los Estados Unidos n.º 10/197.621, 10/063.559 (Publicación de los Estados Unidos número 2002/0183936), 10/065.856, 10/065.868, 10/328.818, 10/328.872, 10/423.403 y 60/482.389.

#### X. AGENTES ANTISENTIDO Y DE IARN

Las realizaciones de la presente divulgación también incluyen oligonucleótidos antisentido y agentes de iARN que se dirigen a las secuencias de polinucleótido de AARS y a métodos de uso de los mismos para reducir la expresión de un transcrito y/o fragmento de proteína AARS. Determinadas realizaciones se refieren al direccionamiento de una o más uniones de corte y empalme (normalmente únicas) que generan una variante de corte y empalme, o un fragmento de proteína AARS de la presente invención. También se incluyen métodos de antisentido o inhibición de iARN que se dirigen a determinadas formas de corte y empalme, ya sea para alentar o desalentar el corte y empalme de un fragmento de proteína seleccionado. En determinadas realizaciones preferidas, las uniones de corte y empalme que generan los fragmentos de proteína AARS se sobreexpresan respecto de tejidos particulares y son únicos para esa variante de corte y empalme. En estas realizaciones y otras relacionadas, dichas variantes de corte y empalme no son la única fuente de actividad de AARS citosólica en el tipo celular diana. Por ejemplo, determinadas variantes de corte y empalme que se van a usar como diana pueden representar aproximadamente del 10 % al 50 % del número de copias total de las variantes de corte y empalme de ARN de AARS en una célula o tejido dado y preferentemente, aproximadamente el 1-10 % del número total de copias de las variantes de corte y empalme de ARN de AARS en una célula o tejido dado. También pueden usarse como diana las variantes de corte y empalme que representan aproximadamente < 1 % del número total de copias de las variantes de corte y empalme de ARN de AARS en una célula o tejido dado.

En determinadas realizaciones, la diana de iARN antisentido no se dirige a la proteína de longitud completa, debido a que dichas proteínas de longitud completa son responsables de una etapa clave en la síntesis de proteínas y de este modo se evita la letalidad que normalmente resulta de las supresiones génicas de AARS de tipo silvestre. Por lo tanto, algunos de los métodos descritos en el presente documento pueden usarse para evitar efectos no deseados, tales como toxicidades en tratamientos tanto crónicos como agudos y para modular selectivamente las actividades no canónicas del fragmento de proteína AARS. Sin embargo, algunas realizaciones pueden dirigirse genéricamente a secuencias de AARS, incluyendo secuencias de AARS de longitud completa, tal como para eliminar o trastornar sustancialmente la fisiología celular de una célula o tejido diana.

En determinadas realizaciones, la variante de corte y empalme de AARS que se va a usar como diana posee una actividad biológica no canónica. En algunas realizaciones, la variante de corte y empalme de AARS tiene una actividad canónica de AARS reducida o indetectable y el método relacionado con antisentido o iARN modula más específicamente su actividad no canónica. En determinadas realizaciones, los agentes relacionados con antisentido o iARN pueden combinarse con una estrategia dirigida o de suministro local para aminorar los efectos no deseados sistémicos a células o tejidos no diana. Entre otras descritas en el presente documento, las células o tejidos a modo de ejemplo que podrían usarse como diana de este modo incluyen células cancerosas y células para tejidos que los

hacen propicios para el direccionamiento localizado, tales como tumores o epitelios mediante aplicación tópica.

#### **A. Agentes antisentido**

5 Las expresiones "oligómero antisentido" o "compuesto antisentido" u "oligonucleótido antisentido" se usan de manera intercambiable y se refieren a una secuencia de subunidades cíclicas, que portan cada una un resto de emparejamiento de bases, unido mediante enlaces entre subunidades que periten que los restos de emparejamiento de bases hibriden con una secuencia diana en un ácido nucleico (normalmente un ARN) mediante emparejamiento de bases de Watson-Crick, para formar un heterodúplex de ácido nucleico:oligómero dentro de la secuencia diana e  
10 impidiendo normalmente de este modo la traducción de ese ARN. También se incluyen métodos de uso de los mismos para modular la expresión de un transcrito de AARS seleccionado, tal como una variante de corte y empalme o un fragmento proteolítico y/o su polipéptido correspondiente.

15 Los oligonucleótidos antisentido pueden contener entre aproximadamente 8 y 40 subunidades, normalmente aproximadamente 8-25 subunidades y preferentemente de 12 a 25 subunidades. En determinadas realizaciones, los oligonucleótidos pueden tener una complementariedad de secuencia exacta con la secuencia diana o una complementariedad próxima, como se define más adelante. En determinadas realizaciones, el grado de complementariedad entre la diana y la secuencia de direccionamiento antisentido es suficiente para formar un dúplex estable. La región de complementariedad de los oligómeros antisentido con la secuencia de ARN diana puede ser  
20 tan corta como de 8-11 bases, pero tiene preferentemente 12-15 bases o más, *por ejemplo*, 12-20 bases o 12-25 bases, incluyendo todos los números enteros entre estos intervalos. Un oligómero antisentido de aproximadamente 14-15 bases es generalmente lo suficientemente largo como para tener una secuencia complementaria única al dirigirse al gen de AARS seleccionado. En determinadas realizaciones, puede ser necesaria una longitud mínima de bases complementarias para lograr la Tm de unión requerida, tal como describe en el presente documento.

25 En determinadas realizaciones, pueden ser adecuados oligómeros antisentido tan largos como de 40 bases, donde al menos un número mínimo de bases, *por ejemplo*, 10-12 bases, son complementarias a la secuencia diana. En general, sin embargo, la captación facilitada o activa en células se optimiza con longitudes de oligómero de menos de aproximadamente 30. Para algunos oligómeros, descritos en más detalle más adelante, se produce un equilibrio  
30 óptimo de estabilidad de unión y captación generalmente con longitudes de 18-25 bases. Se incluyen oligómeros antisentido (por ejemplo, APN, ALN, 2'-OME, MOE) que consisten en aproximadamente 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 o 40 bases, en los que al menos aproximadamente 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 o 40 bases contiguas o no contiguas son complementarias a su secuencias diana de  
35 AARS o variantes de los mismos.

En determinadas realizaciones, los oligómeros antisentido pueden ser 100 % complementarios a una secuencia diana de ácido nucleico de AARS o puede incluir desemparejamientos, *por ejemplo*, para acomodar variantes, en tanto que un heterodúplex formado entre el oligómero y la secuencia diana de ácido nucleico de AARS sea lo  
40 suficientemente estable como para soportar la acción de las nucleasas celulares y otros modos de degradación que pueden producirse *in vivo*. La expresión "secuencia diana" se refiere a una porción del ARN diana contra el cual se dirige el oligonucleótido, esto es, la secuencia con la que hibridará el oligonucleótido mediante emparejamiento de bases de Watson-Crick de una secuencia complementaria. En determinadas realizaciones, la secuencia diana puede ser una región contigua de un ARNm de AARS (por ejemplo, una unión de corte y empalme única de un ARNm de  
45 AARS), o puede estar compuesta de regiones no contiguas del ARNm.

Más adelante se describen armazones de oligómeros que son menos susceptibles a la escisión por nucleasas. Los desemparejamientos, en caso de estar presentes, son menos desestabilizantes a medida que se aproximan a las regiones terminales del dúplex de híbrido que cuando se encuentran en el medio. El número de desemparejamientos  
50 permitido dependerá de la longitud del oligómero, del porcentaje de pares de bases G:C en el dúplex y de la posición de los desemparejamientos en el dúplex, de acuerdo con los principios de sobra conocidos acerca de la estabilidad de los dúplex. Aunque dicho oligómero antisentido no es necesariamente un 100 % complementario a la secuencia diana de ácido nucleico de AARS, es eficaz para unirse de manera estable y específica a la secuencia diana, de tal forma que se modula una actividad biológica del ácido nucleico diana, por ejemplo, la expresión de proteínas AARS.

55 La estabilidad del dúplex formado entre un oligómero y una secuencia diana es una función de la Tm de fusión y de la susceptibilidad del dúplex a la escisión enzimática celular. La Tm de un oligonucleótido antisentido con respecto al ARN de secuencia complementaria puede medirse por métodos convencionales, tales como aquellos descritos por Hames et al., *Nucleic Acid Hybridization*, IRL Press, 1985, págs. 107-108 o como se describe en Miyada C.G. y Wallace R.B., 1987, *Oligonucleotide hybridization techniques*, *Methods Enzymol.* Vol. 154 págs. 94-107. En determinadas realizaciones, el oligómero antisentido puede tener una Tm de unión, respecto de un ARN de secuencia complementaria, mayor de la temperatura corporal y preferentemente, mayor de 50 °C. Se prefieren las Tm en el intervalo de 60-80 °C o más. De acuerdo con principios bien conocidos, la Tm de un compuesto oligomérico, respecto de un híbrido de ARN de secuencia complementaria, puede aumentarse al aumentar la  
60 proporción de bases emparejadas de C:G en el dúplex y/o aumentando la longitud (en pares de bases) del heterodúplex. Al mismo tiempo, con el objetivo de optimizar la captación celular, puede ser ventajoso limitar el

tamaño del oligómero antisentido. Por este motivo, se prefieren generalmente los compuestos que muestran una  $T_m$  elevada (50 °C o más) con una longitud de 25 bases o menos frente a aquellos que requieren más de 25 bases para tener elevados valores de  $T_m$ .

5 Los oligómeros antisentido pueden diseñarse para que bloqueen o inhiban la traducción del ARNm o para inhibir el procesamiento de corte y empalme del pre-ARNm natural o para inducir la degradación de ARNm diana y puede decirse que están "dirigidos a" o "dirigidos contra" una secuencia diana con la que hibridan. En determinadas realizaciones, la secuencia diana puede incluir cualquier secuencia codificante o no codificante de un transcrito de ARNm de AARS y por lo tanto puede encontrarse dentro de un exón o dentro de un intrón. En determinadas realizaciones, la secuencia diana es relativamente única o excepcional entre las AARS (por ejemplo, una AARS de longitud completa) y es selectiva para reducir la expresión de un fragmento de proteína AARS seleccionado, tal como un fragmento proteolítico o una variante de corte y empalme. En determinadas realizaciones, el sitio diana incluye un sitio de corte y empalme 3' o 5' de un ARNm procesado previamente o un punto de ramificación. La secuencia diana para un sitio de corte y empalme puede incluir una secuencia de ARNm que tenga su extremo 5' a de 1 a aproximadamente 25 o aproximadamente 50 pares de bases cadena abajo de la unión aceptora de corte y empalme o cadena arriba de una unión donante de corte y empalme en un ARNm procesado previamente. En determinadas realizaciones, una secuencia diana puede incluir una unión de corte y empalme o un ARNm de AARS de corte y empalme alternativo, tal como una unión de corte y empalme que no se produce en la AARS de longitud completa o que es única o excepcional para ese transcrito, en tanto que no se produce o solo se produce rara vez en otras variantes de corte y empalme de AARS. Con mayor frecuencia se dice que un oligómero está "dirigido contra" una diana biológicamente relevante, tal como un polinucleótido de AARS de referencia, cuando está dirigido contra el ácido nucleico de la diana del modo descrito en el presente documento.

Normalmente, un oligonucleótido es complementario a una secuencia diana, tal como un ADN o ARN diana. Los términos "complementario" y "complementariedad" se refieren a polinucleótidos (es decir, una secuencia de nucleótidos) relacionados mediante las reglas de emparejamiento de bases. Por ejemplo, la secuencia "A-G-T" es complementaria a la secuencia "T-C-A". La complementariedad puede ser "parcial", en la que solo parte de las bases de los ácidos nucleicos están emparejadas de acuerdo con las reglas de emparejamiento de bases o puede haber una complementariedad "completa" o "total" (100 %) entre los ácidos nucleicos. El grado de complementariedad entre hebras de ácido nucleico tiene efectos significativos en la eficacia y la fuerza de la hibridación entre hebras de ácido nucleico. Aunque normalmente se desea una complementariedad perfecta, algunas realizaciones pueden incluir uno o más pero preferentemente 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 desemparejamiento con respecto a la secuencia diana. Se incluyen variaciones en cualquier ubicación dentro del oligómero. En determinadas realizaciones, se prefieren generalmente variaciones de secuencia próximas a los extremos de un oligómero frente a las variaciones en el interior y en caso de estar presentes, se encuentran a aproximadamente 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 nucleótido del extremo 5' y/o 3'.

La expresión "secuencia de direccionamiento" o en determinadas realizaciones "secuencia de direccionamiento antisentido" se refiere a la secuencia en un oligonucleótido que es complementaria (lo que significa, además, sustancialmente complementaria) a la secuencia diana en la molécula diana de ADN o ARN. La secuencia completa, o solo una porción, del compuesto antisentido puede ser complementaria a la secuencia diana. Por ejemplo, en un oligonucleótido que tiene 20-30 bases, aproximadamente 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 o 29 pueden ser secuencias de direccionamiento que son complementarias a la región diana. Normalmente, la secuencia de direccionamiento está formada por bases contiguas, pero como alternativa, puede estar formada por secuencias no contiguas que cuando se ponen juntas, *por ejemplo*, desde los extremos opuestos del oligonucleótido, constituyen una secuencia que abarca la secuencia diana.

Las secuencias diana y de direccionamiento se describen como "complementarias" entre sí cuando se produce hibridación en una configuración antiparalela. Una secuencia de direccionamiento puede tener una complementariedad "próxima" o "sustancial" con la secuencia diana y aun así cumplir el fin de la presente invención, esto es, puede seguir siendo funcionalmente "complementaria". En determinadas realizaciones, un oligonucleótido puede tener como máximo un desemparejamiento con la secuencia diana de cada 10 nucleótidos y preferentemente, como máximo un desemparejamiento de cada 20. Como alternativa, un oligonucleótido puede tener al menos aproximadamente un 80 %, 85 %, 90 % de homología de secuencia y preferentemente, al menos un 95 % de homología de secuencia, con una secuencia de polinucleótido de AARS de referencia descrita en el presente documento o su complemento.

Un oligonucleótido "hibrida específicamente" con un polinucleótido diana en caso de que el oligómero hibride con una diana (por ejemplo, un polinucleótido de AARS de referencia o su complemento) en condiciones fisiológicas, con una  $T_m$  sustancialmente mayor de 45 °C, preferentemente de al menos 50 °C y normalmente de 60 °C-80 °C o mayor. Dicha hibridación corresponde preferentemente a condiciones de hibridación rigurosas. A un pH y una fuerza iónica específicos, la  $T_m$  es la temperatura a la cual el 50 % de una secuencia diana hibrida con un polinucleótido complementario. De nuevo, dicha hibridación puede producirse con una complementariedad "próxima" o "sustancial" del oligómero antisentido respecto de la secuencia diana, así como con complementariedad exacta.

Las expresiones "se une específicamente" o "hibrida específicamente" se refieren en general a una sonda de

oligonucleótido o una secuencia de polinucleótido que no solo se une a su secuencia génica diana prevista en una muestra en condiciones de hibridación seleccionadas, sino que no se une significativamente a otras secuencias diana en la muestra y de este modo discrimina entre su diana prevista y todas las demás dianas en el grupo diana. Una sonda que hibrida específicamente con su secuencia diana prevista también puede detectar diferencias en la concentración en las condiciones de hibridación seleccionadas, tal como se describen en el presente documento.

Una molécula oligomérica (oligómero) "resistente a las nucleasas" se refiere a una cuyo armazón sea sustancialmente resistente a la escisión por nucleasas, en forma no hibridada o hibridada; por nucleasas extracelulares e intracelulares comunes en el organismo; esto es, el oligómero muestra poca o ninguna escisión por nucleasas en condiciones de nucleasas normales en el organismo en el que se encuentra expuesto el oligómero.

Un "heterodúplex" se refiere a un dúplex entre un oligonucleótido y la porción complementaria de un polinucleótido diana, tal como un ADN o ARN diana. Un "heterodúplex resistente a las nucleasas" se refiere a un heterodúplex formado por la unión de un oligómero a su diana complementaria, de tal forma que el heterodúplex es sustancialmente resistente a la degradación *in vivo* por nucleasas intracelulares y extracelulares, tales como RNasaH, que son capaces de cortar complejos bicatenarios de ARN/ARN o ARN/ADN.

Una "subunidad" de un oligonucleótido se refiere a una unidad de nucleótido (o análogo de nucleótido). El término puede hacer referencia a la unidad de nucleótido con o sin la unión intersubunidad unida, aunque, cuando se hace referencia a una "subunidad cargada", la carga reside normalmente entre el enlace intersubunidad (por ejemplo, un enlace de fosfato o fosforotioato o un enlace catiónico).

Las subunidades cíclicas de un oligonucleótido pueden estar basadas en ribosa u otro azúcar de pentosa o, en determinadas realizaciones, grupos alternativos o modificados. Los ejemplos de armazones oligonucleotídicos modificados incluyen, sin limitación, fosforotioatos, fosforotioatos quirales, fosforoditioatos, fosfotriésteres, aminoalquilfosfotriésteres, fosfonatos de metilo y otros alquilo que incluyen fosfonatos de 3'-alquileo y fosfonatos quirales, fosfinatos, fosforamidatos, incluyendo 3'-amino fosforamidato y aminoalquilfosforamidatos, tionofosforamidatos, tionalquilfosfonatos, tionalquilfosfotriésteres y boranofosfatos que tienen enlaces 3'-5' normales, análogos enlazados en 2'-5' de estos y aquellos que tienen la polaridad invertida, en donde los pares adyacentes de unidades de nucleósido están unidos de 3'-5' a 5'-3' o de 2'-5' a 5'-2'. También se contemplan ácidos peptidonucleicos (APN), ácidos nucleicos bloqueados (ALN), 2'-O-metil oligonucleótidos (2'-OMe), 2'-metoxietoxi oligonucleótidos (MOE), entre otros oligonucleótidos conocidos en la técnica.

El resto de emparejamiento de base de purina o pirimidina es normalmente adenina, citosina, guanina, uracilo, timina o inosina. También se incluyen bases tales como piridin-4-ona, piridin-2-ona, fenilo, pseudouracilo, 2,4,6-trimetiloxi benceno, 3-metil uracilo, dihidrouridina, naftilo, aminofenilo, 5-alquilcitolinas (por ejemplo, 5-metilcitolina), 5-alquiluridinas (por ejemplo, rubotimidina), 5-halouridina (por ejemplo, 5-bromouridina) o 6-azapirimidinas o 6-alquilpirimidinas (por ejemplo, 6-metiluridina), propina, quesosina, 2-tiouridina, 4-tiouridina, wibutosina, wibutoxosina, 4-acetiluridina, 5-(carboxihidroximetil)uridina, 5'-carboximetilaminometil-2-tiouridina, 5-carboximetilaminometiluridina, β-D-galactosilqueosina, 1-metiladenosina, 1-metilinosina, 2,2-dimetilguanosa, 3-metilcitolina, 2-metiladenosina, 2-metilguanosa, N6-metiladenosina, 7-metilguanosa, 5'-metoxiaminoetil-2-tiouridina, 5-metilaminometiluridina, 5-metilcarboniletiluridina, 5-metiloxiuridina, 5-metil-2-tiouridina, 2-metiltio-N6-isopentenniladenosina, β-D-manosilqueosina, ácido uridin-5-oxiacético, 2-tiocitolina, derivados de treonina y otros (Burgin et al., 1996, Biochemistry, 35, 14090; Uhlman y Peyman, anteriormente citado). Por "bases modificadas" en este aspecto se entienden bases de nucleótidos diferentes de adenina (A), guanina (G), citosina (C), timidina (T), y uracilo (U), tal como se ha ilustrado anteriormente; dichas bases pueden usarse en cualquier posición en la molécula antisentido. Los expertos en la materia apreciarán que dependiendo de los usos de los oligómeros, las T y las U son intercambiables. Por ejemplo, con otras químicas antisentido, tales como oligonucleótidos antisentido de 2'-O-metilo que son más similares al ARN, las bases de T pueden mostrarse como U.

Como se ha indicado anteriormente, determinados oligonucleótidos proporcionados en el presente documento incluyen ácidos peptidonucleicos (APN). Los ácidos peptidonucleicos (APN) son análogos del ADN en los que la cadena principal es estructuralmente homomorfa a una cadena principal de desoxirribosa, que consiste en unidades de N-(2-aminoetil) glicina a las que se unen bases pirimidínicas o púricas. Los APN que contienen bases pirimidínicas y púricas naturales hibridan con oligonucleótidos complementarios obedeciendo a las reglas de emparejamiento de bases de Watson-Crick e imitan al ADN en términos de reconocimiento de pares de bases (Egholm, Buchardt et al. 1993). La cadena principal de los APN está formada por enlaces peptídicos en lugar de enlaces fosfodiéster, haciendo que sean adecuados para aplicaciones antisentido véase la estructura a continuación). La cadena principal queda inalterada, dando como resultado dúplex de APN/ADN o APN/ARN que muestran una estabilidad térmica mayor de la normal. Los APN no son reconocidos por las nucleasas o proteasas.

Los APN pueden producirse sintéticamente usando cualquier técnica conocida en la materia. El APN es un análogo de ADN en el que un armazón de poliamida reemplaza al anillo de ribosa fosfato tradicional del ADN. A pesar de un cambio estructural radical en la estructura natural, El APN es capaz de unirse de manera específica de secuencia en forma de hélice al ADN o ARN. Las características del APN incluyen una elevada afinidad de unión a ADN o ARN complementario, un efecto desestabilizante causado por desemparejamientos de una sola base, resistencia a las

nucleasas y proteasas, hibridación con ADN o ARN independiente de la concentración de sal y formación de tripletes con ADN de homopurina. Panagene™ ha desarrollado sus monómeros de APN Bts registrados (Bts; grupo benzotiazol-2-sulfonilo) y un proceso de oligomerización registrado. La oligomerización de APN usando monómeros de APN Bts está compuesta de ciclos repetitivos de desprotección, acoplamiento y protección. Las patentes de Panagene para esta tecnología incluyen los documentos US 6969766, US 7211668, US 7022851, US 7125994, US 7145006 y US 7179896. Las Patentes de los Estados Unidos representativas que enseñan la preparación de compuestos de APN incluyen, pero sin limitación, las Patentes de los Estados Unidos n.º 5.539.082; 5.714.331; y 5.719.262, incorporándose cada una estas al presente documento por referencia. Pueden encontrarse enseñanzas adicionales acerca de compuestos de APN en Nielsen et al., Science, 1991, 254, 1497.

También se incluyen subunidades de "ácido nucleico bloqueado" (ALN). Las estructuras de los ALN se conocen en la técnica: por ejemplo, Wengel, et al., Chemical Communications (1998) 455; Tetrahedron (1998) 54, 3607 y Accounts of Chem. Research (1999) 32, 301; Obika, et al., Tetrahedron Letters (1997) 38, 8735; (1998) 39, 5401 y Bioorganic Medicinal Chemistry (2008)16, 9230.

Los oligonucleótidos pueden incorporar uno o más ALN; en algunos casos, los compuestos pueden estar compuestos totalmente por ALN. Los métodos para la síntesis de subunidades de nucleósidos de ALN individuales y su incorporación en oligonucleótidos son conocidos en la técnica: Patentes de los Estados Unidos 7.572.582; 7.569.575; 7.084.125; 7.060.809; 7.053.207; 7.034.133; 6.794.499; y 6.670.461. Los enlazadores intersubunidad típicos incluyen restos fosfodiéster y fosforotioato; como alternativa, pueden emplearse enlazadores que no contienen fósforo. Una realización preferida es un compuesto que contiene ALN en donde cada unidad de ALN está separada por una subunidad de ADN (es decir, un nucleótido de desoxirribosa). Los compuestos preferidos adicionales están compuestos de subunidades de ALN y ADN alternas en donde el enlazador intersubunidad es fosforotioato.

Determinados oligonucleótidos pueden comprender subunidades basadas en morfolino que portan restos de emparejamiento de bases, unidos por enlaces no cargados o sustancialmente no cargados. Los términos "oligómero de morfolino" o "PMO" (oligómero de morfolino de fosforamidato o fosforodiamidato) se refieren a un análogo de oligonucleótido compuesto de estructuras de subunidad de morfolino, donde (i) las estructuras están unidas entre sí mediante enlaces que contienen fósforo, de uno a tres átomos de longitud, preferentemente de dos átomos de longitud y preferentemente no cargados o catiónicos, que unen al nitrógeno de morfolino de una subunidad a un carbono exocíclico 5' de una subunidad adyacente e (ii) cada anillo de morfolino porta una purina o pirimidina o un resto de emparejamiento de bases equivalente eficaz para unirse, mediante enlace de hidrógeno específico de base, a una base en un polinucleótido.

Pueden efectuarse variaciones en este enlace en tanto que no interfieran con la unión o la actividad. Por ejemplo, el oxígeno unido al fósforo puede sustituirse con azufre (tiofosforodiamidato). El oxígeno 5' puede estar sustituido con amino o amino sustituido con alquilo inferior. El nitrógeno colgante unido al fósforo puede estar sin sustituir, monosustituido o disustituido con (opcionalmente sustituido) alquilo inferior. El resto de emparejamiento de base de purina o pirimidina es normalmente adenina, citosina, guanina, uracilo, timina o inosina. La síntesis, las estructuras y las características de unión de oligómeros de morfolino se detallan en las Patentes de los Estados Unidos n.º 5.698.685, 5.217.866, 5.142.047, 5.034.506, 5.166.315, 5.521.063 y 5.506.337 y las Solicitudes PCT n.º PCT/US07/11435 (enlaces catiónicos) y US08/012804 (síntesis mejorada), todas incorporadas al presente documento por referencia.

Las subunidades de morfolino también pueden unirse mediante enlaces intersubunidad no basados en fósforo, tal como se describe en más detalle a continuación, donde al menos un enlace se modifica con un grupo catiónico colgante tal como se ha descrito anteriormente. Podrían usarse otros enlaces análogos de oligonucleótidos que se encuentran no cargados en su estado no modificado pero que también podrían portar un sustituyente de amina colgante. Por ejemplo, podría emplearse un átomo de nitrógeno 5' en un anillo de morfolino en un enlace sulfamida o un enlace urea (en donde el fósforo se reemplaza con carbono o azufre, respectivamente) y se modifica de un modo análogo al átomo de nitrógeno 5' en la estructura (b3) anterior.

Determinadas realizaciones incluyen oligómeros de morfolino sustancialmente no cargados, tales como un oligómero de morfolino unido a fosforodiamidato sustancialmente no cargado. Un armazón que contiene fósforo sustancialmente no cargado en un análogo de oligonucleótido es uno en el que una mayoría de los enlaces de subunidades, *por ejemplo*, entre un 50-100 %, normalmente de al menos un 60 % a un 100 % o un 75 % o un 80 % de sus enlaces, están no cargados a pH fisiológico y contienen un solo átomo de fósforo. Los ejemplos de oligonucleótidos de morfolino que tienen enlaces de armazón que contienen fósforo incluyen oligonucleótidos de morfolino unidos con fosforoamidato y fosforodiamidato. Determinadas realizaciones pueden contener grupos cargados positivamente en preferentemente aproximadamente un 10 %-50 % de sus enlaces del armazón.

Las propiedades de las subunidades a base de morfolino incluyen, por ejemplo, la capacidad para unirse en una forma oligomérica mediante enlaces de armazón estables, no cargados o cargados positivamente, la capacidad de soportar una base de nucleótido (por ejemplo, adenina, citosina, guanina, timidina, uracilo e hipoxantina) de tal forma que el polímero formado pueda hibridar con un ácido nucleico diana con complementariedad de bases, incluyendo

ARN diana, valores de Tm por encima de aproximadamente 45 °C en oligonucleótidos relativamente cortos (por ejemplo, 10-15 bases), la capacidad del oligonucleótido de ser transportado activa o pasivamente en células de mamífero y la capacidad del heterodúplex de oligonucleótidos antisentido:ARN para resistir la degradación por RNasa y RNasaH, respectivamente.

5 En determinadas realizaciones, puede modificarse un oligonucleótido sustancialmente no cargado para que incluya enlaces cargados, *por ejemplo*, hasta aproximadamente 1 por cada 2-5 enlaces no cargados, tal como aproximadamente 4-5 por cada 10 enlaces no cargados. En determinadas realizaciones, puede observarse una mejora óptima en la actividad antisentido cuando es catiónico aproximadamente un 25 % de los enlaces del armazón. En determinadas realizaciones, puede observarse potenciación con un número pequeño, por ejemplo, un 10-20 % de enlaces catiónicos o donde el número de enlaces catiónicos está en el intervalo del 50-80 %, tal como aproximadamente el 60 %. En determinadas realizaciones, las cargas del armazón catiónico pueden potenciarse adicionalmente distribuyendo el volumen de las cargas próximas de los enlaces de armazón de la "región central" del oligonucleótido antisentido, *por ejemplo*, en un oligonucleótido 20-meto con 8 enlaces de armazón catiónicos, que tienen al menos un 70 % de estos enlaces cargados en los 10 enlaces más próximos al centro.

Los oligonucleótidos que se dirigen a una o más porciones de una secuencia de polinucleótido de AARS de referencia o su complemento pueden usarse en cualquiera de los métodos terapéuticos, diagnósticos o de exploración de fármacos descritos en el presente documento y evidentes para las personas expertas en la materia.

### 20 **B. Agentes de interferencia de ARN**

Determinadas realizaciones se refieren a agentes de interferencia de ARN (iARN) que se dirigen a uno o más transcritos de ARNm de un polinucleótido de referencia de aminoacil-ARNt sintetasa (AARS), incluyendo fragmentos y variantes de corte y empalme de los mismos. También se incluyen métodos de uso de los mismos para modular los niveles de un transcrito de AARS seleccionado, tal como una variante de corte y empalme de AARS o un fragmento proteolítico endógeno.

30 El término "bicatenario" significa dos hebras de ácido nucleico separadas que comprenden una región en la que al menos una parte de las hebras es suficientemente complementaria para formar enlaces de hidrógeno y forman una estructura de dúplex. El término "dúplex" o la expresión "estructura de dúplex" se refieren a la región de una molécula bicatenaria en donde las dos hebras separadas son sustancialmente complementarias y por lo tanto hibridan entre sí. "ARNbc" se refiere a una molécula de ácido ribonucleico que tiene una estructura de dúplex que comprende dos hebras de ácido nucleico complementarias y antiparalelas (es decir, las hebras sentido y antisentido). No todos los nucleótidos de un ARNbc deben mostrar pares de bases de Watson-Crick; las dos hebras de ARN pueden ser sustancialmente complementarias. Las hebras de ARN pueden tener el mismo número u otro diferente de nucleótidos.

40 En determinadas realizaciones, un ARNbc es o incluye una región que es al menos parcialmente complementaria al ARN diana. En determinadas realizaciones, el ARNbc es completamente complementario al ARN diana. No es necesario que haya una complementariedad perfecta entre el ARNbc y la diana, pero la correspondencia debe ser suficiente para permitir el ARNbc o un producto de escisión del mismo, para dirigir el silenciamiento específico de secuencia, tal como mediante escisión de iARN del ARN diana. La complementariedad o el grado de homología con la hebra diana, es normalmente más crítica en la hebra antisentido. Aunque normalmente se desea una complementariedad perfecta en la hebra antisentido, algunas realizaciones pueden incluir uno o más pero preferentemente 6, 5, 4, 3, 2 o menos desemparejamientos respecto del ARN diana. Los desemparejamientos se toleran mejor en las regiones terminales y en caso de estar presentes, se encuentran en una región o regiones terminales, *por ejemplo*, a 6, 5, 4 o 3 nucleótido del extremo 5' y/o 3'. La hebra complementaria solo necesita ser sustancialmente complementaria a la hebra antisentido para mantener el carácter bicatenario general de la molécula.

50 Tal como se usa en el presente documento, "ARNbc modificado" se refiere a una molécula de ARNbc que comprende al menos una alteración que la hace más resistente a las nucleasas (por ejemplo, proteína cinasa) que una molécula de ARNbc idéntica que antagoniza al mismo ARN diana. Los ARNbc bicatenarios pueden incluir un saliente de nucleótidos monocatenarios y/o al menos un nucleótido sustituido.

55 Tal como se usa en el presente documento, un "saliente de nucleótidos" se refiere al nucleótido o a los nucleótidos no emparejados que sobresalen de la estructura de dúplex cuando un extremo 3' de una hebra de ARN se prolonga más allá del extremo 5' de la otra hebra complementaria o viceversa. "Romo" o "extremo romo" significa que no hay nucleótidos no emparejados en ese extremo del ARNbc, es *decir*, ningún saliente de nucleótidos. Un ARNbc de "extremos romos" es un ARNbc que es bicatenario a lo largo de toda su secuencia, es *decir*, ningún saliente de nucleótidos en cualquiera de los dos extremos de la molécula.

65 La expresión "par de bases terminales", tal como se usa en el presente documento, se refiere al último par de bases de nucleótidos en un extremo de la región dúplex de una molécula bicatenaria. Por ejemplo, en caso de que un ARNbc u otra molécula tenga extremos romos (es decir, no tiene salientes de nucleótidos), el último par de bases de nucleótidos en ambos extremos de la molécula son pares de bases terminales. En los casos donde un ARNbc u otra

molécula tenga un saliente de nucleótidos en uno o ambos extremos de la estructura de dúplex, el último par de bases inmediatamente adyacente a los salientes de nucleótidos es el par de bases terminal en esos extremos de la molécula.

5 En determinadas realizaciones, los métodos proporcionados en el presente documento pueden utilizar moléculas de ácido ribonucleico bicatenario (ARNbc) como agentes moduladores, para reducir la expresión de un transcrito de AARS, tal como un fragmento o variante de corte y empalme seleccionado. Los ARNbc comprenden generalmente dos hebras individuales. Una hebra del ARNbc comprende una secuencia de nucleótidos que es sustancialmente idéntica a una parte del gen diana o de la región diana (la hebra con "sentido"), y la otra hebra (la hebra  
10 "complementaria" o "antisentido") comprende una secuencia que es sustancialmente complementaria a una parte de la región diana. Las hebras son lo suficientemente complementarias como para hibridar formando una estructura de dúplex. En determinadas realizaciones, la hebra de ARN complementario puede tener menos de 30 nucleótidos, menos de 25 nucleótidos de longitud o incluso de 19 a 24 nucleótidos de longitud. En determinados aspectos, la secuencia de nucleótidos complementaria puede tener 20-23 nucleótidos de longitud o 22 nucleótidos de longitud.

15 En determinadas realizaciones, al menos una de las hebras de ARN comprende un saliente de nucleótidos de 1 a 4 nucleótidos de longitud. En otras realizaciones, el ARNbc puede comprender adicionalmente al menos un nucleótido modificado químicamente. En determinados aspectos, un ARNbc que comprende un saliente monocatenario de 1 a 4 nucleótidos puede comprender una molécula en donde el nucleótido no emparejado del saliente monocatenario  
20 que se encuentra directamente adyacente al par de nucleótido terminal contiene una base púrica. En otros aspectos, los últimos pares de nucleótidos complementarios en ambos extremos de un ARNbc son un par G-C o, al menos dos de los cuatro pares de nucleótidos terminales son pares G-C.

25 Determinadas realizaciones de la presente invención pueden comprender microARN. Los microARN representan un gran grupo de ANR pequeños producidos de manera natural en los organismos, algunos de los cuales regulan la expresión de genes diana. Los microARN se forman a partir de un transcrito precursor en horquilla monocatenario de aproximadamente 70 nucleótidos mediante Dicer. (V. Ambros et al. Current Biology 13:807, 2003). Algunos microARN pueden transcribirse en forma de precursores de ARN en horquilla, que después se procesan en sus formas maduras mediante la enzima Dicer.

30 Determinadas realizaciones también pueden emplear ARN pequeños de interferencia (ARNpi). En determinadas realizaciones, la primera hebra del oligonucleótido bicatenario contiene dos restos de nucleósido más que la segunda hebra. En otras realizaciones, la primera hebra y la segunda hebra tienen el mismo número de nucleósidos; sin embargo, las hebras primera y segunda pueden estar desfasadas, de tal forma que dos nucleósidos terminales  
35 en las hebras primera y segunda no están emparejados con un resto en la hebra complementaria. En determinados casos, los dos nucleósidos que no están emparejados son restos de timidina.

También se incluyen ARN en horquilla corta (ARNhc) y microARN (miARN). Una estructura bicatenaria de un ARNhc se forma por una sola hebra de ARN autocomplementaria y la formación del dúplex de ARN puede iniciarse tanto  
40 dentro como fuera de la célula. Los microARN son ARN cortos no codificantes de 20-22 nucleótidos, normalmente cortados de estructuras precursoras de ARN replegado de ~70 nucleótidos, conocidos como premiARN.

45 En los casos donde el agente modulador comprende ARNpi, el agente debe incluir una región con homología suficiente con la región diana y tener una longitud suficiente en cuanto a los nucleótidos, de tal forma que el agente de ARNpi o un fragmento del mismo, puede mediar la regulación negativa del ARN diana. Se entenderá que el término "ribonucleótido" o "nucleótido" puede, en el caso de un ARN modificado o un nucleótido subrogado, referirse también a un nucleótido modificado o un resto de reemplazo subrogado en una o más posiciones. Por lo tanto, un agente de ARNbc es o incluye una región que es al menos parcialmente complementaria al ARN diana, tal como se describen en el presente documento.

50 Además, puede modificarse un agente modulador de ARNpi o este incluir nucleótidos subrogados. Las regiones monocatenarias de un agente de ARNpi pueden modificarse o incluir nucleósidos subrogados, *por ejemplo*, la región o las regiones no emparejadas de la estructura de horquilla, *por ejemplo*, una región que una dos regiones complementarias, puede tener modificaciones o nucleósidos subrogados. También son útiles las modificaciones para  
55 estabilizar uno o más extremos 3' o 5' de un agente de ARNpi, *por ejemplo*, contra exonucleasas o para favorecer la entrada del agente de ARNpi antisentido en RISC. Las modificaciones pueden incluir enlazadores amino C3 (o C6, C7, C12), enlazadores tiol, enlazadores carboxilo, espaciadores no nucleotídicos (C3, C6, C9, C12, abásicos, trietilenglicol, hexaetilenglicol), especialmente reactivos de biotina o fluoresceína que se presentan como fosforamiditas y que tienen otro grupo hidroxilo protegido con DMT, que permiten múltiples acoplamientos durante la  
60 síntesis de ARN.

Los agentes de ARNpi pueden incluir, por ejemplo, moléculas que son lo suficientemente largas para desencadenar la respuesta de interferón (que pueden escindirse por Dicer (Bernstein et al. 2001. Nature, 409:363-366) y entrar en un RISC (complejo de silenciamiento inducido por iARN)), además de moléculas que son lo suficientemente cortas  
65 como para desencadenar la respuesta de interferón (pudiéndose escindir dichas moléculas por Dicer y/o entrar en un RISC), *por ejemplo*, moléculas que tienen un tamaño que les permite entrar en un RISC, *por ejemplo*,

moléculas que se asemejan a productos escindibles por Dicer. Un agente modulador de ARNpi o un producto escindible del mismo, puede regular negativamente un gen diana, *por ejemplo*, induciendo iARN respecto de un ARN diana, preferentemente una diana de AARS, tal como una variante de corte y empalme seleccionada.

- 5 Cada hebra de un agente de ARNpi puede ser igual o menor de 35, 30, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16 o 15 nucleótidos de longitud. La hebra tiene preferentemente al menos 19 nucleótidos de longitud. Por ejemplo, cada hebra puede tener entre 21 y 25 nucleótidos de longitud. Los agentes de ARNpi preferidos tienen una región de dúplex de 17, 18, 19, 29, 21, 22, 23, 24 o 25 pares de nucleótidos y uno o más salientes, preferentemente uno o dos salientes 3', de 2-3 nucleótidos.

10

Además de la homología con el ARN diana y la capacidad para regular negativamente un gen diana, un agente de ARNpi puede tener una o más de las siguientes propiedades: puede, a pesar de las modificaciones, incluso en un gran número o en todos los nucleósidos, tener una hebra antisentido que puede presentar bases (o bases modificadas) en el armazón tridimensional adecuado para ser capaz de formar un emparejamiento de bases correcto y formar una estructura de dúplex con un ARN diana homólogo que es suficiente para permitir la regulación negativa de la diana, *por ejemplo*, mediante escisión del ARN diana; puede, a pesar de las modificaciones, incluso en un gran número o en todos los nucleósidos, seguir teniendo propiedades "similares al ARN", *es decir*, puede poseer las propiedades estructurales, químicas y físicas generales de una molécula de ARN, a pesar del contenido basado en ribonucleótidos de manera no exclusiva o incluso parcial. Por ejemplo, un agente de ARNpi puede contener, *por ejemplo*, una hebra con sentido y/o antisentido en la que todos los azúcares del nucleótido contienen, por ejemplo, 2' fluoro en lugar de 2' hidroxilo. Aún puede esperarse que este agente que contiene desoxirribonucleótidos muestre propiedades similares al ARN. Sin desear quedar ligados a ninguna teoría, el flúor, electronegativo, prefiere una orientación axial cuando se une a la posición C2' de la ribosa. Esta preferencia espacial del flúor puede, a su vez, forzar a que los azúcares adopten una conformación C<sub>3'</sub>-endo. Este es el mismo modo de estructura observado en las moléculas de ARN y da lugar a la hélice de tipo familia A característica del ARN. Además, ya que el flúor es un buen aceptor de enlace de hidrógeno, puede participar en las mismas interacciones de enlace de hidrógeno con moléculas de agua que se sabe que estabilizan las estructuras del ARN. En general, se prefiere que un resto modificado en la posición 2' del azúcar sea capaz de entrar en enlaces de H, lo que es más característico del resto OH de un ribonucleótido que del resto H de un desoxirribonucleótido.

20

- Un "agente de iARN monocatenario", tal como se usa en el presente documento, es un agente de iARN que está formado por una sola molécula. Pueden incluir una región de dúplex, formada por emparejamiento intracadena, *por ejemplo*, puede ser o incluir, una horquilla o estructura de mango de sartén. Los agentes moduladores de iARN monocatenarios son preferentemente antisentido respecto de la molécula diana. Un agente de iARN monocatenario debe ser lo suficientemente largo como para que pueda entrar en el RISC y participar en la escisión mediada por RISC de un ARNm diana. Un agente de iARN monocatenario tiene al menos 14 y más preferentemente al menos 15, 20, 25, 29, 35, 40 o 50 nucleótidos de longitud. Preferentemente tiene menos de 200, 100 o 60 nucleótidos de longitud.

30

- Los agentes moduladores de iARN en horquilla pueden tener una región de dúplex igual a o de al menos 17, 18, 19, 29, 21, 22, 23, 24 o 25 pares de nucleótidos. La región de dúplex puede ser preferentemente igual a o menor de 200, 100 o 50, de longitud. Algunos intervalos para la región dúplex son de 15-30, 17 a 23, 19 a 23 y de 19 a 21 pares de nucleótidos de longitud. La horquilla puede tener un saliente monocatenario o región desemparejada terminal, preferentemente la 3', y preferentemente del lado antisentido de la horquilla. En determinadas realizaciones, los salientes tienen 2-3 nucleótidos de longitud.

45

Algunos agentes moduladores utilizados de acuerdo con los métodos proporcionados en el presente documento pueden comprender oligonucleótidos de iARN, tales como oligonucleótidos quiméricos o "quimeras", que contienen dos o más regiones químicamente distintas, cada una formada por al menos una unidad de monómero, *es decir*, un nucleótido en el caso de un compuesto oligonucleotídico. Estos oligonucleótidos contienen normalmente al menos una región en donde el oligonucleótido se modifica para conferir en el oligonucleótido resistencia aumentada a la degradación por nucleasas, captación celular aumentada y/o afinidad de unión aumentada por el ácido nucleico diana. Por consiguiente, normalmente pueden obtenerse resultados comparables con oligonucleótidos más cortos cuando se usan oligonucleótidos quiméricos, en comparación con los oligodesoxinucleótidos de fosforotioato. Los oligonucleótidos quiméricos pueden estar formados como estructuras compuestas de dos o más oligonucleótidos, oligonucleótidos modificados, oligonucleótidos y/o miméticos de oligonucleótidos, tal como se ha descrito anteriormente. Dichos oligonucleótidos también se han citado en la técnica como híbridos o gapmeros. Las Patentes de los Estados Unidos representativas que enseñan la preparación de dichas estructuras híbridas incluyen, pero sin limitación, las Patentes de los Estados Unidos n.º 5.013.830; 5.149.797; 5.220.007; 5.256.775; 5.366.878; 5.403.711; 5.491.133; 5.565.350; 5.623.065; 5.652.355; 5.652.356; 5.700.922; y 5.955.589, incorporándose cada una estas al presente documento por referencia. En determinadas realizaciones, el oligonucleótido quimérico es de ARN-ADN, ADN-ARN, ARN-ADN-ARN, ADN-ARN-ADN o ARN-ADN-ARN-ADN, en donde el oligonucleótido tiene entre 5 y 60 nucleótidos de longitud.

50

- En un aspecto de la invención, los agentes de iARN se refieren a un oligonucleótido que comprende al menos un ligando anclado a una nucleobase alterada o no natural. Un gran número de compuestos puede funcionar como

65

base alterada. La estructura de la base alterada es importante en tanto que la base alterada no debe impedir sustancialmente la unión del oligonucleótido a su diana, *por ejemplo*, ARNm. En determinadas realizaciones, la base alterada es difluorotolilo, nitropirrolilo, nitroimidazolilo, nitroindolilo, naftalenilo, antraceno, piridinilo, quinolinilo, pirenilo o el radical divalente de una cualquiera de las nucleobases no naturales descritas en el presente documento.

5 En determinadas realizaciones, la nucleobase no natural es difluorotolilo, nitropirrolilo o nitroimidazolilo. En determinadas realizaciones, la nucleobase no natural es difluorotolilo. Se conocen en la técnica una gran variedad de ligandos y son adecuados para la presente invención. Por ejemplo, el ligando puede ser un esteroide, ácido biliar, lípido, ácido fólico, piridoxal, B12, riboflavina, biotina, compuesto aromático, compuesto policíclico, éter corona, intercalador, molécula escindible, agente de unión a proteínas o carbohidrato. En determinadas realizaciones, el  
10 ligando es un esteroide o un compuesto aromático. En determinados casos, el ligando es colesterilo.

En otras realizaciones, el agente de iARN es un oligonucleótido anclado a un ligando con el fin de mejorar el direccionamiento y la captación celular. Por ejemplo, puede anclarse un agente de iARN a un anticuerpo o fragmento de unión a antígeno del mismo. A modo de ejemplo adicional, puede anclarse un agente de iARN a una molécula de  
15 unión específica de ligando, tal como un polipéptido o fragmento de polipéptido que se une específicamente a un receptor específico de la superficie celular.

En otras realizaciones, el agente modulador comprende una nucleobase no natural, tal como se describen en el presente documento. En determinados casos, se reemplaza el resto de azúcar de ribosa de origen natural por un  
20 azúcar de hexosa. En determinados aspectos, el azúcar de hexosa es una alosa, altrosa, glucosa, manosa, gulosa, idosa, galactosa, talosa o un derivado de las mismas. En una realización preferida, la hexosa es una D-hexosa. En determinados casos, se reemplaza el resto de azúcar de ribosa de origen natural en los nucleósidos por un anillo de heteroalquilo policíclico o un grupo ciclohexenilo. En determinados casos, el grupo heteroalquilo policíclico es un  
25 anillo bicíclico que contiene un átomo de oxígeno en el anillo. En determinados casos, el grupo heteroalquilo policíclico es un biciclo[2.2.1]heptano, un biciclo[3.2.1]octano o un biciclo[3.3.1]nonano. Los ejemplos de agentes de iARN modificados también incluyen oligonucleótidos que contienen armazones modificados o enlaces internucleósido no naturales, tal como se describen en el presente documento.

La presente invención abarca además oligonucleótidos que emplean ribozimas. Las moléculas de ARN sintético y  
30 los derivados de las mismas que catalizan actividades de endorribonucleasa altamente específicas se conocen como ribozimas. (Véase, *por ejemplo*, la Patente de los Estados Unidos n.º 5.543.508 de Haseloff et al., y la Patente de los Estados Unidos n.º 5.545.729 de Goodchild et al.). Las reacciones de escisión están catalizadas por las moléculas de ARN en sí. En las moléculas de ARN de origen natural, los sitios de escisión autocatalizada se ubican dentro de regiones altamente conservadas de estructura secundaria de ARN (Buzayan et al., Proc. Natl. Acad. Sci. U.S.A.,  
35 1986, 83, 8859; Forster et al., Cell, 1987, 50, 9). Las moléculas de ARN autocatalíticas de origen natural se han modificado para generar ribozimas que pueden dirigirse a una molécula de ARN células o patógena con un alto grado de especificidad. Por lo tanto, las ribozimas cumplen el mismo fin general que los oligonucleótidos antisentido (es decir, la modulación de la expresión de un gen específico) y, al igual que los oligonucleótidos, son ácidos nucleicos que poseen porciones significativas monocatenarias.

40 En determinados casos, los agentes de iARN o los oligonucleótidos antisentido para su uso con los métodos proporcionados en el presente documento pueden formarse por un grupo no ligando. Se ha conjugado una serie de moléculas no ligando a oligonucleótidos para potenciar la actividad, la distribución celular, el direccionamiento celular o la captación celular del oligonucleótido y se encuentran disponibles en la bibliografía científica procedimientos para llevar a cabo dichas conjugaciones. Dichos restos no ligados tienen incluidos restos lipídicos, tales como colesterol (Letsinger et al., Proc. Natl. Acad. Sci. USA, 1989, 86:6553), péptidos ricos en arginina, ácido cólico (Manoharan et al., Bioorg. Med. Chem. Lett., 1994, 4:1053), un tioéter, *por ejemplo*, hexil-5-tritilol (Manoharan et al., Ann. N.Y. Acad. Sci., 1992, 660:306; Manoharan et al., Bioorg. Med. Chem. Lett., 1993, 3:2765), un tiocolesterol (Oberhauser et al., Nucl. Acids Res., 1992, 20:533), una cadena alifática, *por ejemplo*, restos de dodecandiol o undecilo (Saison-  
50 Behmoaras et al., EMBO J., 1991, 10:111; Kabanov et al., FEBS Lett., 1990, 259:327; Svinarchuk et al., Biochimie, 1993, 75:49), un fosfolípido, *por ejemplo*, di-hexadecil-rac-glicerol o 1,2-di-O-hexadecil-rac-glicero-3-H-fosfonato de trietilamino (Manoharan et al., Tetrahedron Lett., 1995, 36:3651; Shea et al., Nucl. Acids Res., 1990, 18:3777), una poliamina o una cadena de polietilenglicol (Manoharan et al., Nucleosides & Nucleotides, 1995, 14:969), o ácido adamantanoacético (Manoharan et al., Tetrahedron Lett., 1995, 36:3651), un resto de palmitilo (Mishra et al.,  
55 Biochim. Biophys. Acta, 1995, 1264:229), o un resto de octadecilamina o hexilamino-carbonil-oxicolesterol (Croke et al., J. Pharmacol. Exp. Ther., 1996, 277:923). Las Patentes de los Estados Unidos representativas que enseñan la preparación de dichos conjugados de oligonucleótidos se han listado anteriormente. Los protocolos de conjugación típicos implican la síntesis de oligonucleótidos que portan un enlazador de amino en una o más posiciones de la secuencia. Después, se hace reaccionar el grupo amino con la molécula que se está conjugando usando reactivos de acoplamiento o de activación adecuados. La reacción de conjugación puede llevarse a cabo con el oligonucleótido aún unido al soporte sólido o después de la escisión del nucleótido en la fase en solución. La purificación del conjugado de oligonucleótido mediante HPLC proporciona normalmente el conjugado puro.

65 Pueden encontrarse ejemplos adicionales de agentes de iARN en las Publicaciones de Solicitudes de los Estados Unidos n.º 2007/0275465, 2007/0054279, 2006/0287260, 2006/0035254, 2006/0008822, que se incorporan por referencia. También se incluyen sistemas de suministro de vector que son capaces de expresar las secuencias de

direccionamiento a AARS descritas en el presente documento. Se incluyen vectores que expresan ARNpi u otras moléculas de interferencia de ARN formadoras de dúplex.

5 Un sistema de vector o de construcción de ácido nucleico puede comprender un solo vector o plásmido, dos o más  
 10 vectores o plásmidos, que conjuntamente contienen el ADN total que se va a introducir en el genoma de la célula  
 hospedadora o un transposón. La selección del vector dependerá normalmente de la compatibilidad del vector con la  
 célula hospedadora en la que se va a introducir el vector. En el presente caso, el vector o la construcción de ácido  
 nucleico es preferentemente una que sea operablemente funcional en una célula de mamífero, tal como una célula  
 15 muscular. El vector también puede incluir un marcador de selección, tal como un gen de resistencia a antibióticos o  
 fármacos o un gen indicador (es decir, proteína fluorescente verde, luciferasa), que puede usarse para la selección o  
 identificación de transformantes o transfectantes adecuados. Los sistemas de suministro a modo de ejemplo pueden  
 incluir sistemas de vector vírico (es decir, transducción mediada por virus) que incluyen, pero sin limitación, vectores  
 retrovíricos (por ejemplo, lentivíricos), vectores adenovíricos, vectores víricos adenoasociados y vectores herpes  
 víricos, entre otros conocidos en la técnica.

#### 15 XI. DESCUBRIMIENTO DE FÁRMACOS

Determinadas realizaciones se refieren al uso de polipéptidos de AARS, anticuerpos o polinucleótidos para el  
 20 descubrimiento de fármacos, normalmente para identificar agentes que modulan una o más de las actividades no  
 canónicas del polipéptido AARS de referencia, *por ejemplo*, el fragmento de proteína AARS.

Por ejemplo, determinadas realizaciones incluyen métodos para identificar uno o más "compañeros de unión celular"  
 25 de un polipéptido de referencia de AARS, tal como una proteína celular, lípido, ácido nucleico u otra molécula del  
 hospedador que interactúe directa o físicamente con el polipéptido de AARS. Los ejemplos particulares incluyen, por  
 ejemplo, receptores de la superficie celular, tales como GPCR, dominios de interacción proteína-proteína y dominios  
 extracelulares o intracelulares de los mismos.

También se incluyen métodos para identificar moléculas del hospedador que participan en una o más actividades no  
 30 canónicas del polipéptido de AARS, incluyendo moléculas que interactúan directa o indirectamente con el  
 compañero de unión celular y bien regulan su papel en una actividad no canónica o están reguladas por el  
 compañero de unión. Dichas moléculas del hospedador incluyen componentes tanto aguas arriba como aguas abajo  
 de la vía no canónica, relacionados normalmente por aproximadamente 1, 2, 3, 4, 5 o más etapas identificables en la  
 vía, en relación a la interacción de compañero de unión celular/proteína AARS.

Determinados aspectos incluyen métodos para identificar un compuesto (por ejemplo, polipéptido) u otro agente que  
 35 agoniza o antagoniza la actividad no canónica de un polipéptido de referencia de AARS o una variante activa del  
 mismo, tal como mediante la interacción con el polipéptido de AARS y/o uno o más de sus compañeros de unión  
 celular. También se incluyen métodos para identificar agentes que modulan la expresión (por ejemplo, corte y  
 empalme) de las variantes de corte y empalme de AARS o modulan la actividad de proteasas que de otro modo  
 40 regulan la producción de fragmentos de proteína AARS endógenos (resectinas) a nivel de proteína.

Determinadas realizaciones incluyen, por lo tanto, métodos para identificar un compañero de unión de un polipéptido  
 45 de referencia de AARS, que comprenden a) combinar el polipéptido de AARS con una muestra en condiciones  
 adecuadas y b) detectar la unión específica del polipéptido de AARS a un compañero de unión, identificando de este  
 modo un compañero de unión que se une específicamente al polipéptido de referencia de AARS. También se  
 incluyen métodos para la exploración de un compuesto que se une específicamente a un polipéptido de referencia  
 de AARS o un compañero de unión del polipéptido de AARS, que comprenden a) combinar el polipéptido o el  
 50 compañero de unión con al menos un compuesto de ensayo en condiciones adecuadas y b) detectar la unión del  
 polipéptido o del compañero de unión al compuesto de ensayo, identificando de este modo un compuesto que se  
 une específicamente al polipéptido o su compañero de unión. En determinadas realizaciones, el compuesto es un  
 polipéptido o péptido. En determinadas realizaciones, el compuesto es una molécula pequeña u otro compuesto  
 químico (por ejemplo, no biológico). En determinadas realizaciones, el compuesto es un peptidomimético.

Puede emplearse cualquier método adecuado para detectar interacciones proteína-proteína para identificar  
 55 proteínas celulares que interactúan con un polipéptido de referencia de AARS, que interactúa con uno o más de sus  
 compañeros de unión celulares o ambos. Los ejemplos de métodos tradicionales que pueden emplearse incluyen co-  
 inmunoprecipitación, reticulación y co-purificación a través de gradientes o columnas cromatográficas de lisados  
 celulares o proteínas obtenidas de lisados celulares, principalmente para identificar proteínas en lisado que  
 interactúan con el polipéptido de AARS.

60 En estas realizaciones y otras relacionadas, puede dilucidarse al menos una porción de la secuencia de aminoácidos  
 de una proteína que interactúa con un polipéptido de AARS o su compañero de unión usando técnicas bien  
 conocidas por los expertos en la materia, tal como mediante la técnica de degradación de Edman. Véase, *por  
 ejemplo*, Creighton Proteins: Structures and Molecular Principles, W. H. Freeman & Co., N.Y., págs. 34 49, 1983. La  
 65 secuencia de aminoácidos obtenida puede usarse como guía para la generación de mezclas de oligonucleótidos que  
 pueden usarse para explorar secuencias génicas que codifican dichas proteínas. La exploración puede llevarse a

cabo, por ejemplo, mediante técnicas de hibridación o PCR convencionales, tal como se describe en el presente documento y se conoce en la técnica. Las técnicas para la generación de mezclas de oligonucleótidos y su exploración se conocen bien en la técnica. Véase, *por ejemplo*, Ausubel et al. Current Protocols in Molecular Biology Green Publishing Associates and Wiley Interscience, N.Y., 1989; e Innis et al., eds. PCR Protocols: A Guide to Methods and Applications Academic Press, Inc., Nueva York, 1990.

Además, pueden emplearse métodos en la identificación simultánea de genes que codifican el compañero de unión u otro polipéptido. Estos métodos incluyen, por ejemplo, sondado de bibliotecas de expresión, de un modo similar a la técnica bien conocida de sondado con anticuerpo de bibliotecas de lambda-gt11, usando proteína AARS marcada u otro polipéptido, péptido o proteína de fusión, *por ejemplo*, una variante de polipéptido AARS o un dominio de AARS fusionado a un marcador (por ejemplo, una enzima, flúor, proteína luminiscente o colorante), o un dominio Fc de Ig.

Un método que detecta interacciones de proteína *in vivo*, el sistema de dos híbridos, se describe en detalle con fines ilustrativos y no de manera limitante. Se ha descrito un ejemplo de este sistema (Chien et al., PNAS USA 88:9578 9582, 1991) y se encuentra disponible comercialmente de Clontech (Palo Alto, Calif.).

En resumen, utilizando dicho sistema, pueden construirse plásmidos que codifican dos proteínas híbridas: un plásmido consiste en nucleótidos que codifican el dominio de unión a ADN de una proteína activadora de la transcripción fusionado a una secuencia de nucleótido de referencia de AARS (o, en determinadas realizaciones, su compañero de unión), o una variante del mismo y el otro plásmido consiste en nucleótidos que codifican el dominio de activación de la proteína activadora de la transcripción fusionado a un ADNc (o colección de ADNc) que codifican una proteína desconocida que se ha recombinado en el plásmido como parte de una biblioteca de ADNc. El plásmido de fusión de dominio de unión a ADN y la biblioteca de ADNc activadora pueden transformarse en una cepa de la levadura *Saccharomyces cerevisiae* que contiene un gen indicador (por ejemplo, HBS o lacZ) cuya región reguladora contiene el sitio de unión al activador de transcripción. Cualquiera de las dos proteínas híbridas sola no podría actuar la transcripción del gen indicador: el híbrido de dominio de unión a ADN no puede debido a que no proporciona función de activación y el híbrido de dominio de activación no puede debido a que no puede localizarse en los sitios de unión del activador. La interacción de las dos proteínas híbridas reconstituye la proteína activadora funcional y da como resultado la expresión del gen indicador, que se detecta mediante un ensayo para el producto génico indicador.

Puede emplearse el sistema de dos híbridos u otra metodología similar para explorar bibliotecas de dominio de activación respecto de proteínas que interactúan con el producto génico "señuelo". A modo de ejemplo y no de limitación, puede usarse un polipéptido de referencia de AARS o una variante como producto génico "señuelo". También puede usarse un compañero de unión de AARS como producto génico "señuelo". Las secuencias genómicas totales o de ADNc se fusionan al ADN que codifica un dominio de activación. Se cotransforman esta biblioteca y un plásmido que codifica un híbrido de un producto génico de AARS "señuelo" al dominio de unión a ADN en una cepa indicadora de levadura y se exploran los transformantes resultantes respecto de aquellos que expresan el gen indicador.

Puede producirse una biblioteca de ADNc de la línea celular de la que se van a detectar proteínas que interactúan con productos génicos de AARS "señuelo" usando métodos practicados de manera rutinaria en la técnica. Por ejemplo, pueden insertarse los fragmentos de ADNc en un vector de tal forma que se fusionan traduccionalmente al dominio de activación transcripcional de GAL4. Esta biblioteca puede cotransformarse junto con el plásmido de fusión de gen "señuelo"-GAL4 en una cepa de levadura, que contiene un gen lacZ dirigido por un promotor que contiene la secuencia de activación de GAL4. Una proteína codificada por el ADNc, fusionada al dominio de activación transcripcional de GAL4, que interactúa con el producto génico "señuelo" reconstituirá una proteína GAL4 activa y de este modo dirigirá la expresión del gen HIS3. Las colonias, que expresan HIS3, pueden detectarse por su crecimiento sobre placas de Petri que contienen medio de agar semisólido que carece de histidina. Después, puede purificarse el ADNc a partir de estas cepas y usarse para producir y aislar la proteína que interactúa con el gen de AARS "señuelo" usando técnicas practicadas de manera rutinaria en la técnica.

También se incluyen sistema de tres híbridos, que permiten la detección de interacciones de ARN-proteína en levadura. Véase, *por ejemplo*, Hook et al., RNA. 11:227-233, 2005. Por consiguiente, pueden usarse estos métodos y otros relacionados para identificar un compañero de unión celular de un polipéptido de AARS y para identificar otras proteínas o ácidos nucleicos que interactúan con el polipéptido de AARS, el compañero de unión celular o ambos.

Determinadas realizaciones se refieren al uso de estrategias de exploración de interactoma. Los ejemplos particulares incluyen exploración basada en dominios de proteínas (véase, *por ejemplo*, Boxem et al., Cell. 134:534-545, 2008; y Yu et al., Science. 322:10-110, 2008).

Como se ha indicado anteriormente, una vez aislados, los compañeros de unión pueden identificarse y pueden, a su vez, usarse conjuntamente con técnicas estándar para identificar proteínas u otros compuestos con los que interactúan. Determinadas realizaciones se refieren por lo tanto a métodos de exploración para un compuesto que se

5 unes específicamente al compañero de unión de un polipéptido de referencia de AARS, que comprenden a) combinar el compañero de unión con al menos un compuesto de ensayo en condiciones adecuadas y b) detectar la unión del compañero de unión al compuesto de ensayo, identificando de este modo un compuesto que se une específicamente al compañero de unión. En determinadas realizaciones, el compuesto de ensayo es un polipéptido.

5 En determinadas realizaciones, el compuesto de ensayo es un compuesto químico, tal como un compuesto de molécula pequeña o un peptidomimético.

10 Determinadas realizaciones incluyen métodos para explorar respecto de un compuesto que modula la actividad de un polipéptido de referencia de AARS, que comprenden a) combinar el polipéptido con al menos un compuesto de ensayo en condiciones que permiten la actividad del polipéptido, b) evaluar la actividad del polipéptido en presencia del compuesto de ensayo y c) comparar la actividad del polipéptido en presencia del compuesto de ensayo con la actividad del polipéptido en ausencia del compuesto de ensayo, en donde un cambio en la actividad del polipéptido en presencia del compuesto de ensayo es indicativo de un compuesto que modula la actividad del polipéptido.

15 Determinadas realizaciones incluyen métodos para explorar respecto de un compuesto que modula la actividad de un compañero de unión de un polipéptido de referencia de AARS, que comprenden a) combinar el polipéptido con al menos un compuesto de ensayo en condiciones que permiten la actividad del compañero de unión, b) evaluar la actividad del compañero de unión en presencia del compuesto de ensayo y c) comparar la actividad del compañero de unión en presencia del compuesto de ensayo con la actividad del compañero de unión en ausencia del compuesto de ensayo, en donde un cambio en la actividad del compañero de unión en presencia del compuesto de ensayo es indicativo de un compuesto que modula la actividad del compañero de unión. normalmente, estas realizaciones y otras relacionadas incluyen evaluar una actividad no canónica seleccionada que se asocia con el polipéptido de AARS o su compañero de unión. Se incluyen condiciones *in vitro* e *in vivo*, tales como condiciones de cultivo celular.

25 Determinadas realizaciones incluyen métodos para explorar un compuesto respecto de su eficacia como agonista completo o parcial de un polipéptido de referencia de AARS o un fragmento activo o una variante del mismo, que comprenden a) exponer una muestra que comprende el polipéptido a un compuesto y b) detectar la actividad agonista en la muestra, normalmente midiendo un aumento en la actividad no canónica del polipéptido de AARS. Determinados métodos incluyen a) exponer una muestra que comprende un compañero de unión del polipéptido de

30 AARS a un compuesto y b) detectar la actividad agonista en la muestra, normalmente midiendo un aumento en la actividad no canónica seleccionada del polipéptido de AARS. Determinadas realizaciones incluyen composiciones que comprenden un compuesto agonista identificado mediante el método y un vehículo o excipiente farmacéuticamente aceptable.

35 También se incluyen métodos para explorar un compuesto respecto de su eficacia como antagonista completo o parcial de un polipéptido de referencia de AARS, que comprenden a) exponer una muestra que comprende el polipéptido a un compuesto y b) detectar la actividad antagonista en la muestra, normalmente midiendo una reducción en la actividad no canónica del polipéptido de AARS. Determinados métodos incluyen a) exponer una muestra que comprende un compañero de unión del polipéptido de AARS a un compuesto y b) detectar la actividad

40 antagonista en la muestra, normalmente midiendo una reducción en la actividad no canónica seleccionada del polipéptido de AARS. Determinadas realizaciones incluyen composiciones que comprenden un compuesto antagonista identificado mediante el método y un vehículo o excipiente farmacéuticamente aceptable.

45 En determinadas realizaciones, pueden diseñarse sistemas *in vitro* para identificar compuestos capaces de interactuar con o de modular una secuencia de referencia de AARS o su compañero de unión. Algunos de los compuestos identificados mediante dichos sistemas pueden ser útiles, por ejemplo, para modular la actividad de la vía y para elaborar componentes de la vía en sí. También pueden usarse en exploraciones para identificar compuestos que alteran interacciones entre componentes de la vía; o pueden alterar dichas interacciones de manera directa. Una estrategia a modo de ejemplo implica preparar una mezcla de reacción del polipéptido de AARS y un

50 compuesto de ensayo en condiciones y durante un tiempo suficiente para permitir que los dos interactúen y se unan, formando de este modo un complejo que puede retirarse de y/o detectarse en la mezcla de reacción.

Los ensayos de exploración *in vitro* pueden llevarse a cabo de diversos modos. Por ejemplo, un polipéptido de AARS, un compañero de unión celular o los compuestos de ensayo pueden anclarse sobre una fase sólida. En estas

55 realizaciones y otras relacionadas, pueden capturarse los complejos resultantes y detectarse sobre la fase sólida al final de la reacción. En un ejemplo de dicho método, el polipéptido de AARS y/o su compañero de unión se anclan sobre una superficie sólida y los compuestos de ensayo, que no están anclados, pueden marcarse, ya sea directa o indirectamente, de tal forma que pueda detectarse su captura mediante el componente sobre la superficie sólida. En otros ejemplos, se anclan los compuestos de ensayo a la superficie sólida y el polipéptido de AARS y/o su

60 compañero de unión, que no están anclados, se marcan de algún modo que sea detectable. En determinadas realizaciones, pueden utilizarse convenientemente placas de microtitulación como fase sólida. El componente anclado (o el compuesto de ensayo) puede inmovilizarse mediante uniones no covalentes o covalentes. Puede lograrse la unión no covalente simplemente recubriendo la superficie sólida con una solución de la proteína y secando. Como alternativa, puede usarse un anticuerpo inmovilizado, preferentemente un anticuerpo monoclonal,

65 específico para la proteína que se vaya a inmovilizar para anclar la proteína a la superficie sólida. Las superficies prepararse por adelantado y almacenarse.

Para llevar a cabo un ensayo a modo de ejemplo, normalmente se añade el componente no inmovilizado a la superficie recubierta que contiene el componente anclado. Después de que se complete la reacción, se retiran los componentes que no hayan reaccionado (por ejemplo, mediante lavado) en condiciones tales que permanezcan inmovilizados sobre la superficie sólida cualesquiera complejos específicos formados. La detección de complejos anclados sobre la superficie sólida puede lograrse de varios modos. Por ejemplo, en los casos donde esté pre-marcado el componente no inmovilizado previamente, la detección de marcador inmovilizado sobre la superficie indica que se formaron complejos. En los casos donde el componente no inmovilizado previamente no está marcado previamente, puede usarse un marcador indirecto para detectar complejos anclados sobre la superficie; *por ejemplo*, usando un anticuerpo marcado específico para el componente no inmovilizado previamente (el anticuerpo, a su vez, puede marcarse directamente o marcarse indirectamente con un anticuerpo anti-Ig marcad).

Como alternativa, puede determinarse la presencia o ausencia de unión de un compuesto de ensayo, por ejemplo, usando resonancia de plasmón superficial (SPR) y el cambio en el ángulo de resonancia como índice, en donde se inmoviliza un polipéptido de AARS o un compañero de unión celular sobre la superficie de una microplaca sensora disponible comercialmente (por ejemplo, fabricada por BIACORE™) mediante un método convencional, se pone en contacto el compuesto de ensayo con la misma y se ilumina la microplaca sensora con una luz de una longitud de onda concreta con un ángulo específico. También puede medirse la unión de un compuesto de ensayo detectando la aparición de un pico correspondiente al compuesto de ensayo mediante un método en donde se inmoviliza un polipéptido de AARS o un compañero de unión celular sobre la superficie de una placa de proteína adaptable a un espectrómetro de masas, se pone en contacto un compuesto de ensayo con la misma y se combina un método de ionización, tal como MALDI-EM, IEN-EM, FAB-EM y similares con un espectrómetro de masas (por ejemplo, espectrómetro de masas de doble enfoque, espectrómetro de masas cuádrupolar, espectrómetro de masas de tiempo de vuelo, espectrómetro de masas de transformación de Fourier, espectrómetro de masas de ion ciclotrón y similares).

En determinadas realizaciones, pueden usarse ensayos basados en células, ensayos basados en vesículas de membrana o ensayos basados en fracciones de membrana para identificar compuestos que modulan interacciones en la vía no canónica del polipéptido de AARS seleccionado. Para este fin, pueden usarse líneas celulares que expresan un polipéptido de AARS y/o un compañero de unión o una proteína de fusión que contiene un dominio o fragmento de dichas proteínas (o una combinación de las mismas) o líneas celulares (por ejemplo, células COS, células CHO, células HEK293, células HeLa, etc.) que se han modificado por ingeniería genética para que expresen dichas proteínas o proteínas de fusión. Pueden identificarse los compuestos de ensayo que influyen la actividad no canónica controlando un cambio (por ejemplo, un cambio estadísticamente significativo) en esa actividad en comparación con un control o una cantidad predeterminada.

Para las realizaciones que se refieren a agentes antisentido y de iARN, por ejemplo, también se incluyen métodos para explorar un compuesto respecto de su eficacia para alterar la expresión de un polinucleótido de AARS de referencia, que comprenden a) exponer una muestra que comprende el polinucleótido de AARS de referencia a un compuesto, tal como un oligonucleótido antisentido potencial y b) detectar la expresión alterada del polinucleótido de AARS. En determinados ejemplos no limitantes, pueden emplearse estas realizaciones y otras relacionadas en ensayos basados en células o en ensayos de traducción sin células, de acuerdo con técnicas rutinarias en la técnica. También se incluyen los agentes antisentido y de iARN identificados mediante dichos métodos.

También pueden usarse anticuerpos para fragmentos de proteína AARS en los ensayos de exploración, tal como para identificar un agente que se une específicamente a una AARS, confirmar la especificidad o afinidad de un agente que se une a un fragmento de proteína AARS o identificar el sitio de interacción entre el agente y el fragmento de proteína AARS. Se incluyen ensayos en los que se usa el anticuerpo como inhibidor competitivo del agente. Por ejemplo, un anticuerpo que se une específicamente a un fragmento de proteína AARS con una afinidad conocida puede actuar como inhibidor competitivo de un agente seleccionado y usarse para calcular la afinidad del agente por el fragmento de proteína AARS. Asimismo, pueden usarse uno o más anticuerpos que se unen específicamente a epítopos o sitios conocidos de un fragmento de proteína AARS como inhibidor competitivo para confirmar si el agente se une o no a ese mismo sitio. Otras variaciones serán evidentes para las personas expertas en la materia.

También se incluyen cualquiera de los métodos anteriores u otros métodos de exploración conocidos en la técnica, que se adaptan para exploración de alto rendimiento (HTS). La HTS usa normalmente automatización para ejecutar una exploración de un ensayo contra una biblioteca de compuestos candidatos, por ejemplo, un ensayo que mide un aumento o una reducción en una actividad no canónica, tal como se describe en el presente documento.

Cualquiera de los métodos de exploración proporcionados en el presente documento puede utilizar bibliotecas de molécula pequeña o bibliotecas generadas mediante química combinatoria. Las bibliotecas de mezclas químicas y/o biológicas, tales como extractos fúngicos, bacterianos o de algas, se conocen en la técnica y pueden explorarse con cualquiera de los ensayos de la invención. Los ejemplos de métodos para la síntesis de bibliotecas moleculares pueden encontrarse en: (Carell *et al.*, 1994a; Carell *et al.*, 1994b; Cho *et al.*, 1993; DeWitt *et al.*, 1993; Gallop *et al.*, 1994; Zuckermann *et al.*, 1994).

Las bibliotecas de compuestos pueden presentarse en solución (Houghten *et al.*, 1992) o sobre perlas (Lam *et al.*, 1991), sobre placas (Fodor *et al.*, 1993), bacterias, esporas (Ladner *et al.*, Patente de los Estados Unidos n.º 5.223.409, 1993), plásmidos (Cull *et al.*, 1992) o sobre fagos (Cwirla *et al.*, 1990; Devlin *et al.*, 1990; Felici *et al.*, 1991; Ladner *et al.*, Patente de los Estados Unidos n.º 5.223.409, 1993; Scott y Smith, 1990). Las realizaciones de la presente invención abarcan el uso de diferentes bibliotecas para la identificación de moduladores de molécula pequeña de uno o más fragmentos de proteína AARS, sus compañeros de unión celulares y/o sus actividades no canónicas relacionadas. Las bibliotecas útiles para los fines de la invención incluyen, pero sin limitación, (1) bibliotecas químicas, (2) bibliotecas de productos naturales y (3) bibliotecas combinatorias compuestas de péptidos aleatorios, oligonucleótidos y/o moléculas orgánicas.

Las bibliotecas químicas consisten en análogos estructurales de compuestos conocidos o compuestos que se identifican como "aciertos" o "líderes" mediante exploración de productos naturales. Las bibliotecas de productos naturales se obtienen de colecciones de microorganismos, animales, plantas u organismos marinos que se usan para crear mezclas para su exploración mediante: (1) fermentación y extracción de caldos del suelo, plantas o microorganismos marinos o (2) extracción de plantas u organismos marinos. Las bibliotecas de productos naturales incluyen policétidos, péptidos no ribosómicos y variantes (de origen no natural) de los mismos. Véase, *por ejemplo*, Cane *et al.*, Science 282:63-68, 1998. Las bibliotecas combinatorias pueden estar compuestas de grandes números de péptidos, oligonucleótidos o compuestos orgánicos en forma de una mezcla. Son relativamente fáciles de preparar mediante métodos de síntesis automatizada tradicionales, PCR, clonación o métodos sintéticos registrados.

Más específicamente, una biblioteca química combinatoria es una colección de diversos compuestos químicos generados mediante síntesis química o síntesis biológica, combinando un número de "bloques de construcción" químicos, tales como reactivos. Por ejemplo, se forma una biblioteca química combinatoria lineal, tal como una biblioteca de polipéptidos, combinando un conjunto de bloques de construcción químicos (aminoácidos) de cada forma posible para una longitud de compuesto dada (es decir, el número de aminoácidos en un compuesto polipeptídico). Pueden sintetizarse millones de compuestos químicos mediante dicho mezclado combinatorio de bloques de construcción químicos.

Para una revisión acerca de la química combinatoria y de las bibliotecas creadas a partir de la misma, véase, *por ejemplo*, Huc, I. y Nguyen, R. (2001) Comb. Chem. High Throughput Screen 4:53-74; Lepre, C A. (2001) Drug Discov. Today 6:133-140; Peng, S. X. (2000) Biomed. Chromatogr. 14:430-441; Bohm, H. J. y Stahl, M. (2000) Curr. Opin. Chem. Biol. 4:283-286; Barnes, C y Balasubramanian, S. (2000) Curr. Opin. Chem. Biol. 4:346-350; Lepre, Enjalbal, C, et al., (2000) Mass Spectrom Rev. 19:139-161; Hall, D. G., (2000) Nat. Biotechnol. 18:262-262; Lazo, J. S., y Wipf, P. (2000) J. Pharmacol. Exp. Ther. 293:705-709; Houghten, R. A., (2000) Ann. Rev. Pharmacol. Toxicol. 40:273-282; Kobayashi, S. (2000) Curr. Opin. Chem. Biol. (2000) 4:338-345; Kopylov, A. M. y Spiridonova, V. A. (2000) Mol. Biol. (Mosk) 34:1097-1113; Weber, L. (2000) Curr. Opin. Chem. Biol. 4:295-302; Dolle, R. E. (2000) J. Comb. Chem. 2:383-433; Floyd, C D., et al., (1999) Prog. Med. Chem. 36:91-168; Kundu, B., et al., (1999) Prog. Drug Res. 53:89-156; Cabilly, S. (1999) Mol. Biotechnol. 12:143-148; Lowe, G. (1999) Nat. Prod. Rep. 16:641-651; Dolle, R. E. y Nelson, K. H. (1999) J. Comb. Chem. 1:235-282; Czarnick, A. W. y Keene, J. D. (1998) Curr. Biol. 8:R705-R707; Dolle, R. E. (1998) Mol. Divers. 4:233-256; Myers, P. L., (1997) Curr. Opin. Biotechnol. 8:701-707; y Pluckthun, A. y Cortese, R. (1997) Biol. Chem. 378:443.

Se encuentran disponibles comercialmente dispositivos para la preparación de bibliotecas combinatorias (véase, *por ejemplo*, 357 MPS, 390 MPS, Advanced Chem Tech, Louisville Ky., Symphony, Rainin, Woburn, Mass., 433A Applied Biosystems, Foster City, Calif., 9050 Plus, Millipore, Bedford, Mass.). Además, se encuentran disponibles comercialmente numerosas bibliotecas combinatorias en sí (véase, *por ejemplo*, ComGenex, Princeton, N.J., Asinex, Moscú, Ru, Tripos, Inc., St. Louis, Mo., ChemStar, Ltd., Moscú, RU, 3D Pharmaceuticals, Exton, Pa., Martek Biosciences, Columbia, Md., etc.).

## XII. MÉTODOS DE USO

Las realizaciones de la presente invención incluyen métodos de tratamiento terapéuticos. Por consiguiente, los agentes de AARS descritos en el presente documento, incluyendo polipéptidos de AARS, polinucleótidos de AARS, vectores basados en polinucleótidos de AARS, células hospedadoras que expresan AARS, oligonucleótidos antisentido, agentes de iARN, así como agentes de unión, tales como péptidos, anticuerpos y fragmentos de unión a antígeno, peptidomiméticos y otras moléculas pequeñas, pueden usarse para tratar una serie no limitante de enfermedades o afecciones asociadas con las actividades no canónicas de un AARS de referencia. Los ejemplos de dichas actividades no canónicas incluyen modulación de la señalización extracelular, modulación de la proliferación celular, modulación de la migración celular, modulación de la diferenciación celular (por ejemplo, hematopoyesis, neurogénesis, miogénesis, osteogénesis y adipogénesis), modulación de la apoptosis u otras formas de muerte celular, modulación de la angiogénesis, modulación de la unión celular, modulación del metabolismo celular, modulación de la producción o actividad de citocinas, modulación de la actividad de receptor de citocinas, modulación de la captación o secreción celular, inmunomodulación, modulación de la inflamación, modulación de procesos metabólicos, tales como control de la glucosa y similares.

Se incluyen terapias a base de polinucleótidos, tales como terapias antisentido y terapias de interferencia de iARN,

que normalmente están relacionadas con la reducción de la expresión de una molécula diana, tal como un fragmento endógeno de una AARS o un compañero de unión celular de un polipéptido de AARS, que de otro modo contribuye a su actividad no canónica. Las terapias antisentido o de iARN antagonizan normalmente la actividad no canónica, tal como reduciendo la expresión del polipéptido de referencia de AARS. También se incluyen polipéptidos o péptidos, anticuerpos o fragmentos de unión a antígeno, peptidomiméticos u otras terapias basadas en moléculas pequeñas, que agonizan o antagonizan la actividad no canónica de un polipéptido de referencia de AARS, tal como interactuando directamente con el polipéptido de AARS, sus compañeros de unión celulares, o ambas cosas.

Estas realizaciones y otras relacionadas incluyen métodos para usar los agentes o las composiciones de AARS de la presente invención para tratar una célula, tejido o sujeto. Las células o tejidos que pueden tratarse o modularse mediante la presente invención son células o tejidos de mamífero o más preferentemente, células o tejidos humanos. Dichas células o tejidos pueden ser de un estado sano o de un estado patológico.

En determinadas realizaciones, por ejemplo, se proporcionan métodos para modular actividades celulares terapéuticamente relevantes que incluyen, pero sin limitación, metabolismo celular, diferenciación celular, proliferación celular, captación celular, secreción celular, muerte celular, movilización celular, migración celular, transcripción génica, traducción de ARNm, impedancia celular, respuestas inmunitarias, respuestas inflamatorias y similares, que comprenden poner en contacto una célula con un agente o composición de AARS tal como se describe en el presente documento. En determinadas realizaciones, la célula se encuentra en un sujeto. Por consiguiente, las composiciones de AARS pueden emplearse en el tratamiento de esencialmente cualquier célula o tejido o sujeto que pudiera beneficiarse de la modulación de una o más de dichas actividades.

Los agentes y las composiciones de AARS también pueden usarse en cualquiera de una serie de contextos terapéuticos que incluyen, por ejemplo, aquellos que se refieren al tratamiento o la prevención de enfermedades neoplásicas, enfermedades o afecciones del sistema inmunitario (por ejemplo, enfermedades autoinmunitarias e inflamación), enfermedades infecciosas, enfermedades metabólicas, enfermedades neuronales/neurológicas, enfermedades musculares/cardiovasculares, enfermedades asociadas con una hematopoyesis aberrante, enfermedades asociadas con una miogénesis aberrante, enfermedades asociadas con una neurogénesis aberrante, enfermedades asociadas con una adipogénesis aberrante, enfermedades asociadas con una osteogénesis aberrante, enfermedades asociadas con una angiogénesis aberrante, enfermedades asociadas con una supervivencia celular aberrante, enfermedades asociadas con una captación de lípidos aberrante, enfermedades asociadas con el envejecimiento (por ejemplo, pérdida auditiva, neuropatías periféricas o autónomas, demencia senil, retinopatía) y otras.

Por ejemplo, en determinadas realizaciones ilustrativas, las composiciones de AARS de la invención pueden usarse para modular la angiogénesis, *por ejemplo*, mediante la modulación de la proliferación y/o la señalización de células endoteliales. La proliferación y/o la señalización de células endoteliales puede controlarse usando una línea celular adecuada (por ejemplo, células endoteliales pulmonares microvasculares humanas (HMVEC-L) y células endoteliales de la vena umbilical humanas (HUVEC)), y usando un ensayo adecuado (por ejemplo, ensayos de migración de células endoteliales, ensayos de proliferación de células endoteliales, ensayos de formación de tubos, ensayos de tapón de matrigel, etc.), muchos de los cuales son conocidos y se encuentran disponibles en la técnica.

Por lo tanto, en realizaciones relacionadas, las composiciones de la invención pueden emplearse en el tratamiento de esencialmente cualquier célula o tejido que pudiera beneficiarse de la modulación de la angiogénesis. Por ejemplo, en algunas realizaciones, puede ponerse en contacto una célula o tejido que experimenta o es susceptible a la angiogénesis (por ejemplo, una afección angiogénica) con una composición adecuada de la invención para inhibir una afección angiogénica. En otras realizaciones, puede ponerse en contacto una célula o tejido que experimenta o es susceptible a una angiogénesis insuficiente (por ejemplo, una afección angiostática) con una composición adecuada de la invención para interferir con la actividad angiostática y/o promover la angiogénesis.

También se incluyen métodos para modular la hematopoyesis y afecciones relacionadas. Los ejemplos de procesos hematopoyéticos que pueden modularse por los polipéptidos de AARS de la invención incluyen, sin limitación, la formación de células mieloides (por ejemplo, células eritroides, mastocitos, monocitos/macrófagos, células dendríticas mieloides, granulocitos, tales como basófilos, neutrófilos y eosinófilos, megacariocitos, plaquetas) y células linfoides (por ejemplo, linfocitos citolíticos naturales, células dendríticas linfoides, células B y células T). Determinados procesos hematopoyéticos específicos incluyen eritropoyesis, granulopoyesis, linfopoyesis, megacariopoyesis, trombopoyesis y otros. También se incluyen métodos para modular el tráfico o la movilización de células hematopoyéticas, incluyendo células madre hematopoyéticas, células progenitoras, eritrocitos, granulocitos, linfocitos, megacariocitos y trombocitos.

Los métodos para modular la hematopoyesis pueden ponerse en práctica *in vivo*, *in vitro*, *ex vivo* o en cualquier combinación de los mismos. Estos métodos pueden ponerse en práctica en cualquier muestra biológica, cultivo celular o tejido que contenga células madre hematopoyéticas, células progenitoras hematopoyéticas u otras células madre o progenitoras que sean capaces de diferenciarse en todo el linaje hematopoyético (por ejemplo, células madre obtenidas de tejido adiposo). Para los métodos *in vitro* y *ex vivo*, pueden aislarse y/o diferenciarse células madre y células progenitoras, ya sean de origen hematopoyético u otro de acuerdo con las técnicas y características

descritas en el presente documento y conocidas en la técnica.

Las composiciones de la invención también pueden ser útiles como inmunomoduladores para tratar indicaciones anti o pro-inflamatorias mediante la modulación de células que median, ya sea directa o indirectamente, enfermedades, afecciones y trastornos autoinmunitarios y/o inflamatorios. Puede controlarse la utilidad de las composiciones de la invención como inmunomoduladores o moduladores de la inflamación usando cualquiera de una serie de técnicas conocidas y disponibles en la técnica que incluyen, por ejemplo, ensayos de migración (por ejemplo, usando leucocitos o linfocitos) o ensayos de viabilidad celular (por ejemplo, usando células B, células T, monocitos o células NK).

La "inflamación" se refiere en general a la respuesta biológica de los tejidos a estímulos dañinos, tales como patógenos, células dañadas (por ejemplo, heridas), e irritantes. La expresión "respuesta inflamatoria" se refiere a los mecanismos específicos mediante los cuales se consigue y regula la inflamación, incluyendo, solo a modo ilustrativo, activación o migración de células inmunitarias, producción de citocinas, vasodilatación, incluyendo liberación de cinina, fibrinólisis y coagulación, entre otras descritas entre otras descritas en el presente documento y conocidas en la técnica.

Los signos clínicos de la inflamación crónica dependen de la duración de la enfermedad, las lesiones inflamatorias, la causa y el área anatómica afectada. (Véase, *por ejemplo*, Kumar et al., Robbins Basic Pathology-8ª Ed., 2009 Elsevier, Londres; Miller, LM, Pathology Lecture Notes, Atlantic Veterinary College, Charlottetown, PEI, Canadá). La inflamación crónica se asocia con diversas patologías o enfermedades, incluyendo, por ejemplo, alergias, enfermedad de Alzheimer, anemia, estenosis de la válvula aórtica, artritis, tal como artritis reumatoide y artrosis, cáncer, insuficiencia cardíaca congestiva, fibromialgia, fibrosis, ataque cardíaco, insuficiencia renal, lupus, pancreatitis, ictus, complicaciones quirúrgicas, enfermedad inflamatoria pulmonar, enfermedad inflamatoria del intestino, aterosclerosis, trastornos neurológicos, diabetes, trastornos metabólicos, obesidad y psoriasis, entre otras descritas entre otras descritas en el presente documento y conocidas en la técnica. Por tanto, las composiciones de AARS pueden usarse para tratar o controlar la inflamación crónica, modular cualquiera de una o más de las respuestas inflamatorias crónicas o tratar una cualquiera o más enfermedades o afecciones asociadas con la inflamación crónica.

Los criterios para evaluar los signos y síntomas de las afecciones inflamatorias u otras, incluyendo con el fin de efectuar un diagnóstico diferencial y también para controlar tratamientos, tal como determinar si se ha administrado una dosis terapéuticamente eficaz durante el transcurso del tratamiento, *por ejemplo*, determinando la mejora de acuerdo con criterios clínicos aceptados, serán evidentes para los expertos en la materia y se ejemplifican por las enseñanzas de, por ejemplo, Berkow et al., eds., The Merck Manual, 16ª edición, Merck and Co., Rahway, N.J., 1992; Goodman et al., eds., Goodman and Gilman's The Pharmacological Basis of Therapeutics, 10ª edición, Pergamon Press, Inc., Elmsford, N.Y., (2001); Avery's Drug Treatment: Principles and Practice of Clinical Pharmacology and Therapeutics, 3ª edición, ADIS Press, Ltd., Williams and Wilkins, Baltimore, MD. (1987); Ebadi, Pharmacology, Little, Brown and Co., Boston, (1985); Osolci al., eds., Remington's Pharmaceutical Sciences, 18ª edición, Mack Publishing Co., Easton, PA (1990); Katzung, Basic and Clinical Pharmacology, Appleton and Lange, Norwalk, CT (1992).

En otras realizaciones, las composiciones de AARS de la invención pueden usarse para modular la proliferación y/o la supervivencia celular y, por consiguiente, para tratar o prevenir enfermedades, trastornos o afecciones caracterizadas por anomalías en la proliferación y/o supervivencia celular. Por ejemplo, en determinadas realizaciones, las composiciones de AARS pueden usarse para modular la apoptosis y/o para tratar enfermedades o afecciones asociadas con la apoptosis anormal. La apoptosis puede controlarse mediante cualquiera de una serie de técnicas disponibles conocidas y disponibles en la técnica, incluyendo, por ejemplo, ensayos que miden la fragmentación del ADN, alteraciones en la asimetría de la membrana, activación de caspasas apoptóticas y/o liberación de citocromo C y AIF.

El progreso de estas y otras terapias (por ejemplo, terapias *ex vivo*) puede controlarse fácilmente por métodos y ensayos convencionales y basándose en criterios conocidos por el médico u otras personas expertas en la materia.

### XIII. FORMULACIONES FARMACÉUTICAS, ADMINISTRACIÓN Y KITS

Las realizaciones de la presente invención incluyen polinucleótidos de AARS, polipéptidos de AARS, células hospedadoras que expresan polipéptidos de AARS, agentes de unión, agentes moduladores u otros compuestos descritos en el presente documento, formulados en soluciones farmacéuticamente aceptables o fisiológicamente aceptables para su administración a una célula o a un animal, ya sean solas o en combinación con otras una o más modalidades de terapia. También se entenderá que, si se desea, pueden administrarse las composiciones de la invención en combinación también con otros agentes, tales como, *por ejemplo*, otras proteínas o polipéptidos o diversos agentes farmacéuticamente activos. No hay prácticamente ningún límite en cuanto a otros componentes que puedan incluirse también en las composiciones, a condición de que los agentes adicionales no afecten de manera adversa a los efectos moduladores u otros que se desee conseguir.

5 En las composiciones farmacéuticas de la invención, la formulación de soluciones de excipientes y vehículos farmacéuticamente aceptables es de sobra conocida para los expertos en la materia, al igual que el desarrollo de regímenes de dosificación y tratamiento adecuados para usar las composiciones particulares descritas en el presente documento en diversos regímenes de tratamiento, que incluyen, por ejemplo, administración y formulación oral, parenteral, intravenosa, intranasal, subcutánea e intramuscular.

10 En determinadas aplicaciones, las composiciones farmacéuticas o terapéuticas de la invención no estimulan una reacción inmunitaria. En otras realizaciones, las composiciones farmacéuticas o terapéuticas de la invención, que normalmente comprenden uno o más polipéptidos o polinucleótidos de AARS, estimulan una reacción inmunitaria, tal como sirviendo como adyuvante en una vacuna o una composición relacionada o estando presente en una composición junto con un adyuvante o agente separado que estimula una respuesta inmunitaria.

15 En determinadas realizaciones, los agentes de AARS, tales como los polipéptidos de AARS, polinucleótidos de AARS y anticuerpos tienen una solubilidad que es deseable para el modo de administración particular, tal como administración intravenosa. Los ejemplos de solubilidades deseables incluyen al menos aproximadamente 1 mg/ml, al menos aproximadamente 10 mg/ml, al menos aproximadamente 25 mg/ml y al menos aproximadamente 50 mg/ml.

20 En determinadas aplicaciones, las composiciones farmacéuticas divulgadas en el presente documento pueden suministrarse por administración oral a un sujeto. Como tales, estas composiciones pueden formularse con un diluyente inerte o con un transportador comestible asimilable o pueden atraparse en cápsulas con vaina de gelatina dura o blanda o pueden comprimirse en comprimidos o pueden incorporarse directamente con el alimento de la dieta.

25 En determinadas circunstancias, será deseable suministrar las composiciones farmacéuticas divulgadas en el presente documento por vía parenteral, subcutánea, intravenosa, intramuscular, intraarterial, intratecal, intraparenquimal, intraventricular, intrauretral, intraesternal, intracraneal, intrasinoval o incluso intraperitoneal, tal como se describe, por ejemplo, en la Patente de los Estados Unidos n.º 5.543.158; la Patente de los Estados Unidos n.º 5.641.515 y la Patente de los Estados Unidos n.º 5.399.363 (cada una incorporada específicamente al presente documento por referencia en su totalidad). Los dispositivos adecuados para administración parenteral incluyen inyectores de aguja (incluyendo microaguja), inyectores sin agujas y técnicas de infusión.

30 Pueden prepararse soluciones de los compuestos activos en forma de base libre o de sales farmacéuticamente aceptables en agua mezcladas de manera adecuada con un tensioactivo, tal como hidroxipropilcelulosa. También pueden prepararse dispersiones en glicerol, polietilenglicoles líquidos y mezclas de los mismos y en aceites. En condiciones normales de almacenamiento y uso, estas preparaciones contienen un conservante para impedir el crecimiento de microorganismos.

35 Las formas farmacéuticas adecuadas para uso inyectable incluyen soluciones o dispersiones acuosas estériles y polvos estériles para la preparación extemporánea de soluciones o dispersiones inyectables estériles (Patente de los Estados Unidos n.º 5.466.468, incorporada al presente documento específicamente por referencia en su totalidad). En todos los casos, la forma debe ser estéril y debe ser fluida hasta el punto de que se pueda inyectar fácilmente con jeringa. Debe ser estable en las condiciones de fabricación y almacenamiento y debe conservarse contra la acción contaminante de los microorganismos, tales como bacterias y hongos. El vehículo puede ser un disolvente o medio de dispersión que contiene, por ejemplo, agua, etanol, poliol (por ejemplo, glicerol, propilenglicol y polietilenglicol líquido y similares), mezclas adecuadas de los mismos y/o aceites vegetales. Puede mantenerse una fluidez adecuada, por ejemplo, mediante el uso de un recubrimiento, tal como lecitina, manteniendo el tamaño de partículas adecuado en el caso de dispersiones y mediante el uso de tensioactivos. Puede facilitarse la prevención de la acción de microorganismos mediante diversos agentes antibacterianos y antifúngicos, por ejemplo, parabenos, clorobutanol, fenol, ácido sórbico, timerosal y similares. En muchos casos, será preferible incluir agentes isotónicos, por ejemplo, azúcares o cloruro de sodio. Puede lograrse la absorción prolongada de las composiciones inyectables mediante el uso en las composiciones de agentes que retrasan la absorción, por ejemplo, monoestearato de aluminio y gelatina.

40 Para administración parenteral en solución acuosa, por ejemplo, la solución debe tamponarse de manera adecuada en caso necesario y hacer que el diluyente líquido sea en primer lugar isotónico con suficiente suero salino o glucosa. Estas soluciones acuosas particulares son especialmente adecuadas para administración intravenosa, intramuscular, subcutánea e intraperitoneal. A este respecto, será conocido por los expertos en la materia un medio acuoso estéril que pueda emplearse a la luz de la presente divulgación. Por ejemplo, puede disolverse una dosis en 1 ml de solución isotónica de NaCl y añadirse a 1000 ml de fluido de hipodermolisis o inyectarse en el sitio de infusión propuesto (véase, *por ejemplo*, Remington's Pharmaceutical Sciences, 15ª Edición, págs. 1035-1038 y 1570-1580). Se producirá, necesariamente, cierta variación en la dosificación dependiendo de la afección del sujeto que se esté tratando. En cualquier caso, la persona responsable de la administración determinará la dosis adecuada para el sujeto individual. Asimismo, para administración humana, las preparaciones deben cumplir los estándares de esterilidad, pirogenicidad y los de seguridad y pureza generales requeridos por la Oficina de Estándares Biológicos de la FDA.

Las soluciones inyectables estériles pueden prepararse incorporando los compuestos activos en la cantidad requerida en el disolvente adecuado con los diversos otros ingredientes enumerados anteriormente, según sea necesario, seguido de esterilización por filtración. En general, las dispersiones se preparan incorporando los diversos principios activos esterilizados en un vehículo estéril que contiene el medio básico de dispersión y los demás  
 5 ingredientes necesarios de entre aquellos enumerados anteriormente. En el caso de polvos estériles para la preparación de soluciones inyectables estériles, los métodos preferidos de preparación son las técnicas secado al vacío y criodesecación que producen un polvo del principio activo más cualquier otro ingrediente adicional deseado a partir de una solución previamente esterilizada por filtración del mismo.

10 Las composiciones divulgadas en el presente documento pueden formularse en una forma neutra o de sal. Las sales farmacéuticamente aceptables incluyen las sales de adición de ácido (formadas con los grupos amino libres de la proteína) y que se forman con ácidos inorgánicos, tales como, por ejemplo, los ácidos clorhídrico y fosfórico o ácidos orgánicos, tales como acético, oxálico, tartárico, mandélico y similares. Las sales formadas con los grupos carboxilo  
 15 libres también pueden derivarse de bases inorgánicas, tales como, por ejemplo, hidróxidos de sodio, potasio, amonio, calcio o férricos y bases inorgánicas tales como isopropilamina, trimetilamina, histidina, procaína y similares. Tras su formulación, las soluciones se administrarán de un modo compatible con la formulación de dosificación y en una cantidad tal que sea terapéuticamente eficaz. Las formulaciones se administran fácilmente en diversas formas de dosificación, tales como soluciones inyectables, cápsulas de liberación de fármaco y similares.

20 Tal como se usa en el presente documento, "transportador" incluye cualquiera y todos los disolventes, medios de dispersión, vehículos, recubrimientos, diluyentes, agentes antibacterianos y antifúngicos, agentes isotónicos y de retraso de la absorción, tampones, soluciones transportadoras, suspensiones, coloides y similares. Se conoce bien en la técnica el uso de dichos medios y agentes para sustancias farmacéuticamente activas. Excepto en el caso de  
 25 que cualquier agente o medio convencional sea incompatible con el principio activo, se contempla su uso en las composiciones terapéuticas. También pueden incorporarse principios activos complementarios en las composiciones.

La expresión "farmacéuticamente aceptable" se refiere a entidades moleculares y composiciones que no producen una reacción alérgica o indeseada similar cuando se administra a un ser humano. La preparación de una  
 30 composición acuosa que contiene una proteína como ingrediente activo se entiende bien en la técnica. Normalmente, dichas composiciones se preparan como inyectables, tanto como soluciones o suspensiones líquidas; también pueden prepararse formas sólidas adecuadas para solución en o suspensión en líquido antes de la inyección. La preparación también puede emulsionarse.

35 En determinadas realizaciones, las composiciones farmacéuticas pueden suministrarse mediante pulverizadores intranasales, inhalación y/u otros vehículos de suministro de aerosol. Los métodos para suministrar genes, polinucleótidos y composiciones peptídicas directamente a los pulmones mediante pulverizadores de aerosol nasales se han descrito, por ejemplo, en la Patente de los Estados Unidos n.º 5.756.353 y la Patente de los Estados Unidos n.º 5.804.212 (cada una incorporada específicamente al presente documento por referencia en su totalidad).  
 40 Del mismo modo, también se conoce en la técnica farmacéutica el suministro de fármacos usando resinas de micropartículas intranasales (Takenaga *et al.*, 1998) y compuestos de lisofosfatidilglicerol (Patente de los Estados Unidos n.º 5.725.871, incorporada específicamente al presente documento por referencia en su totalidad). Del mismo modo, el suministro de fármacos transmucosal en forma de una matriz de soporte de politetrafluoroetileno se describe en la Patente de los Estados Unidos n.º 5.780.045 (incorporada al presente documento específicamente  
 45 por referencia en su totalidad).

Las composiciones farmacéuticas pueden formularse para que sean de liberación inmediata y/o sostenida. Las composiciones de liberación sostenida incluyen liberación retardada, modificada, pulsada, controlada, dirigida y programada. Por lo tanto, las composiciones pueden formularse en forma de una suspensión o en forma de un  
 50 sólido, un semisólido o un líquido tixotrópico para su administración en forma de un depósito implantado que proporciona la liberación sostenida de los polinucleótidos de AARS, polipéptidos de AARS, agentes de unión, agentes moduladores y otros agentes activos. Los ejemplos de dichas formulaciones incluyen, pero sin limitación, stents recubiertos de fármaco y semisólidos y suspensiones que comprenden vesículas lamelares o micropartículas de ácido poli(DL-láctico-co-glicólico) (PGLA), poli(DL-láctida-co-glicólido) (PLG) o poli(láctida) (PLA) cargadas de  
 55 fármaco, hidrogeles (Hoffman AS: Ann. N.Y. Acad. Sci. 944: 62-73 (2001)), sistemas de nanopartículas de poliaminoácidos, comercializados de la marca MEDUSA® desarrollados por Flamel Technologies Inc., sistemas de gel no acuosos comercializados de la marca ATRIGEL® desarrollados por Atrix, Inc., y formulaciones de liberación extendida de acetato isobutirato de sacarosa, comercializadas de la marca SABER®, desarrollados por Durect Corporation y sistemas a base de lípidos desarrollados por SkyePharma y comercializados de la marca DEPOFOAM®.  
 60

Se conocen en la técnica dispositivos de liberación sostenida capaces de suministrar dosis deseadas de las composiciones farmacéuticas durante periodos de tiempo prolongados. Por ejemplo, las Patentes de los Estados Unidos n.º 5.034.229; 5.557.318; 5.110.596; 5.728.396; 5.985.305; 6.113.938; 6.156.331; 6.375.978; y 6.395.292;  
 65 enseñan dispositivos impulsados por ósmosis capaces de suministrar una formulación de principio activo, tal como una solución o una suspensión, a una velocidad deseada durante un periodo de tiempo prolongado (es decir, un

periodo en el intervalo de más de una semana hasta un año o más). Otros dispositivos de liberación sostenida a modo de ejemplo incluyen bombas de tipo regulador que proporcionan un flujo constante, flujo ajustable o flujo programable de formulaciones de agentes beneficiosos, que se encuentran disponibles de Medtronic, incluyendo las bombas intratecales comercializadas de la marca SYNCHROMED INFUSION SYSTEM®, los sistemas de Johnson and Johnson comercializados de la marca CODMAN® division pumps y bombas de las tecnologías INSET®. Se describen ejemplos adicionales de dispositivos en las Patentes de los Estados Unidos n.º 6.283.949; 5.976.109; 5.836.935; y 5.511.355.

En determinadas realizaciones, el suministro puede producirse mediante el uso de liposomas, nanocápsulas, micropartículas, microesferas, partículas lipídicas, vesículas y similares, para la introducción de las composiciones de la presente invención en células hospedadoras adecuadas. En particular, las composiciones de la presente invención pueden formularse para su suministro ya sea encapsuladas en una partícula de lípido, un liposoma, una vesícula, una nanoesfera, una nanopartícula o similares. La formulación y el uso de dichos vehículos de suministro puede llevarse a cabo usando técnicas conocidas y convencionales.

En determinadas realizaciones, los agentes proporcionados en el presente documento pueden unirse a un sustrato sólido farmacéuticamente aceptable, incluyendo sustratos biocompatibles y biodegradables, tales como polímeros y matrices. Los ejemplos de dichos sustratos sólidos incluyen, sin limitación, poliésteres, hidrogeles (por ejemplo, poli(2-hidroxietilmetacrilato), o poli(alcohol vinílico), polilactidas (Patente de los Estados Unidos n.º 3.773.919), copolímeros de ácido L-glutámico y  $\gamma$ -etil-L-glutamato, acetato de etilenvinilo no degradable, copolímeros degradables de ácido láctico-ácido glicólico, tales como poli(ácido láctico-co-glicólico) (PLGA) y las LUPRON DEPOT™ (microesferas inyectables compuestas de copolímero de ácido láctico-ácido glicólico y acetato de leuprolida), ácido poli-D-(-)-3-hidroxibutírico, colágeno, metales, hidroxiapatita, vidrio biológico, aluminato, materiales biocerámicos y proteínas purificadas.

En una realización particular, el sustrato sólido comprende polímeros biodegradables, comercializados de la marca ATRIGEL™ (QLT, Inc., Vancouver, B.C.). El sistema de suministro de fármacos ATRIGEL® consiste en polímeros biodegradables disueltos en transportadores biocompatibles. Los agentes farmacéuticos pueden mezclarse en este sistema de suministro líquido en el momento de la fabricación o, dependiendo del producto, pueden añadirse posteriormente por el médico en el momento de uso. Cuando se inyecta el producto líquido en el espacio subcutáneo mediante una aguja de calibre fino o se coloca en sitios de tejido accesibles mediante una cánula, el agua en los fluidos del tejido hace que el polímero se precipite y atrape el fármaco en un implante sólido. El fármaco encapsulado dentro del implante se libera después de manera controlada a medida que la matriz de polímero se degrada biológicamente con el tiempo.

Las composiciones farmacéuticas para su uso en la presente invención también pueden administrarse por vía tópica, (intra)dérmica o transdérmica a la piel o mucosa. Las formulaciones típicas para este fin incluyen geles, hidrogeles, lociones, soluciones, cremas, pomadas, polvos para espolvorear, vendajes, espumas, películas, parches cutáneos, obleas, implantes, esponjas, fibras, vendas y microemulsiones. También pueden usarse liposomas. Los vehículos típicos incluyen alcohol, agua, aceite mineral, vaselina líquida, vaselina blanca, glicerina, polietilenglicol y propilenglicol. Pueden incorporarse potenciadores de la penetración; véase, *por ejemplo*, Finin y Morgan: J. Pharm. Sci. 88(10): 955-958, (1999). Otros medios de administración tópica incluyen suministro por electroporación, iontoforesis, fonoforesis, sonoforesis o inyección con microaguja o sin aguja, por ejemplo, usando los sistemas comercializados de las marcas POWDERJECT™ y BIOJECT™.

Se conocen bien en la técnica métodos de formulación y se divulgan, por ejemplo, en Remington: The Science and Practice of Pharmacy, Mack Publishing Company, Easton, Pa., 20ª edición, ISBN: 0683306472 (2000). Las composiciones y agentes proporcionados en el presente documento pueden administrarse de acuerdo con los métodos de la presente invención en cualquier régimen de dosificación terapéuticamente eficaz. La cantidad y la frecuencia de dosificación se seleccionan para crear un nivel eficaz del agente sin efectos perjudiciales. La cantidad eficaz de un compuesto de la presente invención dependerá de la ruta de administración, del tipo de animal de sangre caliente que se esté tratando y de las características físicas del animal de sangre caliente concreto que se esté considerando. Estos factores y su relación para determinar esta cantidad son de sobra conocidos por los expertos en la técnica médica. Esta cantidad y el método de administración pueden ajustarse para lograr una eficacia óptima pero dependerán de factores tales como el peso, la dieta, la medicación concurrente y otros factores que reconocerán los expertos en la técnica médica.

En realizaciones particulares, la cantidad de una composición o agente administrado variará normalmente de una dosis de aproximadamente 0,1 a aproximadamente 100 mg/kg/día y normalmente, de aproximadamente 0,1 a 10 mg/kg, ya se administre por vía oral o intravenosa. En realizaciones particulares, una dosis es de 5 mg/kg o 7,5 mg/kg. En diversas realizaciones, la dosis es de aproximadamente 50-2500 mg al día, 100-2500 mg/día, 300-1800 mg/día o 500-1800 mg/día. En una realización, la dosis es de entre aproximadamente 100 a 600 mg/día. En otra realización, la dosis es de entre aproximadamente 300 y 1200 mg/día. En realizaciones particulares, la composición o el agente se administra a una dosis de 100 mg/día, 240 mg/día, 300 mg/día, 600 mg/día, 1000 mg/día, 1200 mg/día o 1800 mg/día, en una o más dosis al día (es decir, lográndose con las dosis combinadas la dosis diaria deseada). En realizaciones relacionadas, una dosis es de 100 mg dos veces al día, 150 mg dos veces al día, 240 mg

dos veces al día, 300 mg dos veces al día, 500 mg dos veces al día o 600 mg dos veces al día. En diversas realizaciones, la composición o el agente se administra en una sola dosis o en dosis repetidas. La dosis inicial y las dosis posteriores pueden ser iguales o diferentes.

- 5 En determinadas realizaciones, se administra una composición o agente en una sola dosis de 0,1 a 10 mg/kg o de 0,5 a 5 mg/kg. En otras realizaciones, se administra una composición o agente en una dosis de 0,1 a 50 mg/kg/día, 0,5 a 20 mg/kg/día o de 5 a 20 mg/kg/día.

- 10 En determinadas realizaciones, se administra una composición o agente por vía oral o intravenosa, *por ejemplo*, por infusión durante un periodo de tiempo de aproximadamente, *por ejemplo*, 10 minutos a 90 minutos. En otras realizaciones relacionadas, se administra una composición o agente mediante infusión continua, *por ejemplo*, a una dosis de entre aproximadamente 0,1 a aproximadamente 10 mg/kg/h durante un periodo de tiempo. Aunque el periodo de tiempo puede variar, en determinadas realizaciones el periodo de tiempo puede ser de entre aproximadamente 10 minutos a aproximadamente 24 horas o de entre aproximadamente 10 minutos a aproximadamente tres días.

- 15 En realizaciones particulares, una cantidad eficaz o una cantidad terapéuticamente eficaz es una cantidad suficiente para lograr una concentración total de la composición o el agente en el plasma sanguíneo de un sujeto con una  $C_{m\acute{a}x}$  de entre aproximadamente 0,1  $\mu\text{g/ml}$  y aproximadamente 20  $\mu\text{g/ml}$  o de entre aproximadamente 0,3  $\mu\text{g/ml}$  y aproximadamente 20  $\mu\text{g/ml}$ . En determinadas realizaciones, una dosis oral es una cantidad suficiente para lograr una concentración en plasma sanguíneo ( $C_{m\acute{a}x}$ ) de entre aproximadamente 0,1  $\mu\text{g/ml}$  a aproximadamente 5  $\mu\text{g/ml}$  o de entre aproximadamente 0,3  $\mu\text{g/ml}$  a aproximadamente 3  $\mu\text{g/ml}$ . En determinadas realizaciones, una dosis intravenosa es una cantidad suficiente para lograr una concentración en plasma sanguíneo ( $C_{m\acute{a}x}$ ) de entre aproximadamente 1  $\mu\text{g/ml}$  a aproximadamente 10  $\mu\text{g/ml}$  o de entre aproximadamente 2  $\mu\text{g/ml}$  a aproximadamente 6  $\mu\text{g/ml}$ . En una realización relacionada, la concentración total de un agente en el plasma sanguíneo del sujeto tiene una concentración media constante de menos de aproximadamente 20  $\mu\text{g/ml}$  y/o una concentración en estado estacionario de menos de aproximadamente 20  $\mu\text{g/ml}$ . En una realización adicional, la concentración total de un agente en el plasma sanguíneo del sujeto tiene una concentración media constante de menos de aproximadamente 10  $\mu\text{g/ml}$  y/o una concentración en estado estacionario de menos de aproximadamente 10  $\mu\text{g/ml}$ .

- 20 En otra realización más, la concentración total de un agente en el plasma sanguíneo del sujeto tiene una concentración media constante de entre aproximadamente 1 ng/ml y aproximadamente 10  $\mu\text{g/ml}$  y/o una concentración en estado estacionario de entre aproximadamente 1 ng/ml y aproximadamente 10  $\mu\text{g/ml}$ . En una realización, la concentración total de un agente en el plasma sanguíneo del sujeto tiene una concentración media constante de entre aproximadamente 0,3  $\mu\text{g/ml}$  y aproximadamente 3  $\mu\text{g/ml}$  y/o una concentración en estado estacionario de entre aproximadamente 0,3  $\mu\text{g/ml}$  y aproximadamente 3  $\mu\text{g/ml}$ .

- 25 En realizaciones particulares, se administra una composición o agente en una cantidad suficiente para lograr en el mamífero una concentración en el plasma sanguíneo que tenga una concentración media constante de entre aproximadamente 1 ng/ml y aproximadamente 10  $\mu\text{g/ml}$  y/o una concentración en estado estacionario de entre aproximadamente 1 ng/ml y aproximadamente 10  $\mu\text{g/ml}$ . En realizaciones relacionadas, la concentración total del agente en el plasma sanguíneo del mamífero tiene una concentración media constante de entre aproximadamente 0,3  $\mu\text{g/ml}$  y aproximadamente 3  $\mu\text{g/ml}$  y/o una concentración en estado estacionario de entre aproximadamente 0,3  $\mu\text{g/ml}$  y aproximadamente 3  $\mu\text{g/ml}$ .

- 30 En realizaciones particulares de la presente invención, la cantidad eficaz de una composición o agente o la concentración en el plasma sanguíneo de la composición o agente se logra o se mantiene, *por ejemplo*, durante al menos 15 minutos, al menos 30 minutos, al menos 45 minutos, al menos 60 minutos, al menos 90 minutos, al menos 2 horas, al menos 3 horas, al menos 4 horas, al menos 8 horas, al menos 12 horas, al menos 24 horas, al menos 48 horas, al menos 3 días, al menos 4 días, al menos 5 días, al menos 6 días, al menos una semana, al menos 2 semanas, al menos un mes, al menos 2 meses, al menos 4 meses, al menos 6 meses, al menos un año, al menos 2 años o más de 2 años.

- 35 En determinadas realizaciones basadas en polipéptidos, la cantidad de polipéptido administrada se encontrará normalmente en el intervalo de aproximadamente 0,1  $\mu\text{g/kg}$  a aproximadamente 0,1 mg/kg a aproximadamente 50 mg/kg de peso corporal del paciente. Dependiendo del tipo y gravedad de la enfermedad, puede ser una dosis inicial candidata para su administración a un paciente de aproximadamente 0,1  $\mu\text{g/kg}$  a aproximadamente 0,1 mg/kg a aproximadamente 50 mg/kg de peso corporal (por ejemplo, aproximadamente 0,1-15 mg/kg/dosis) de polipéptido, ya sea, por ejemplo, mediante una o más administraciones separadas o mediante infusión continua. Por ejemplo, un régimen de dosificación puede comprender administrar una dosis de carga inicial de aproximadamente 4 mg/kg, seguida de una dosis de mantenimiento semanal de aproximadamente 2 mg/kg del polipéptido o de aproximadamente la mitad de la dosis de carga. Sin embargo, pueden ser útiles otras pautas de dosificación. Una dosis diaria típica puede variar de aproximadamente 0,1  $\mu\text{g/kg}$  a aproximadamente 1  $\mu\text{g/kg}$  a 100 mg/kg o más, dependiendo de los factores mencionados anteriormente. Para administraciones repetidas a lo largo de varios días o un periodo mayor, dependiendo de la afección, el tratamiento se mantiene hasta que se produce una supresión deseada de los síntomas de la enfermedad.

En realizaciones particulares, la dosis eficaz logra los niveles en el plasma sanguíneo o la concentración media constante de una composición o agente descrito en el presente documento. Estas pueden determinarse fácilmente usando procedimientos rutinarios.

5 Las realizaciones de la presente invención, en otros aspectos, proporcionan kits que comprenden uno o más recipientes rellenos con uno o más de los polipéptidos, polinucleótidos, anticuerpos, complejos de múltiples unidades, las composiciones de los mismos, etc., de la invención, tal como se describen en el presente documento. Los kits pueden incluir instrucciones escritas acerca de cómo usar dichas composiciones (por ejemplo, para modular la señalización celular, angiogénesis, cáncer, afecciones inflamatorias, diagnósticos, etc.).

10 Los kits del presente documento también pueden incluir uno o más agentes terapéuticos adicionales u otros componentes adecuados o deseados para la indicación que se esté tratando o para la aplicación diagnóstica deseada. Un agente terapéutico adicional puede estar contenido en un segundo recipiente, si se desea. Los ejemplos de agentes terapéuticos adicionales incluyen, pero sin limitación, agentes antineoplásicos, agentes antiinflamatorios, agentes antibacterianos, agentes antiviricos, agentes angiogénicos, etc.

Los kits del presente documento también pueden incluir una o más jeringuillas u otros componentes necesarios o deseados para facilitar un modo de suministro previsto (por ejemplo, stents, depósitos implantables, etc.).

20 Aunque la invención anterior se ha descrito con cierto detalle a modo de ilustración y ejemplo para propósitos de claridad de comprensión, será fácilmente evidente para un experto habitual en la materia a la luz de las enseñanzas de la presente invención que pueden efectuarse determinados cambios y modificaciones a la misma sin apartarse del alcance de las reivindicaciones adjuntas. Los siguientes ejemplos se proporcionan solo a modo ilustrativo y de ningún modo limitativo. Los expertos en la materia reconocerán fácilmente una serie de parámetros no críticos que podrían cambiarse o modificarse para proporcionar resultados esencialmente similares.

#### XIV. EJEMPLOS

30 **MÉTODOS GENERALES.** A menos que se indique de otro modo en los ejemplos a continuación, se usaron los siguientes métodos generales para la optimización génica, la expresión de proteínas a pequeña y gran escala, la purificación de proteínas, la elaboración de perfiles transcripcionales y la exploración para producir y caracterizar los polipéptidos de AARS descritos en los ejemplos a continuación.

#### SÍNTESIS GÉNICA Y CLONACIÓN EN VECTORES DE EXPRESIÓN

35 Se optimizaron por codones y se clonaron en vectores bacterianos las secuencias de polinucleótidos que codifican versiones marcadas epitópicamente de los polipéptidos de AARS usando los métodos listados a continuación.

40 En el método (1), el ADN optimizado por codones de *E. coli* (Welch et al., PLoS ONE 4(9): e7007 doi:10.1371/journal.pone.0007002) que codifica cada uno de los polipéptidos de AARS se sintetizó por DNA 2.0 (Menlo Park, CA), y se sintetizaron dos versiones de cada polipéptido de AARS, conteniendo un marcador epitópico combinado N-terminal o C-terminal que comprende tanto un marcador de seis histidinas como un marcador epitópico V5.

45 El ADN que codifica los polipéptidos de AARS marcados N-terminalmente se sintetiza con una extensión 5' que codifica en orientación de 5' a 3', un sitio de unión a ribosomas (rbs (subrayado más adelante)), un sitio de restricción de NdeI, un marcador de seis histidinas y un marcador epitópico V5, (AGGAGGTA<sup>AA</sup>ACATATGCATCATCATCACGGTAAGCCTATCCCTAA CCCTTTGCTCGGTCTCGATTCTACG) (**SEQ ID NO 1**), que está fusionado en fase a la fase abierta de lectura predicha del polipéptido de AARS. En los casos donde el polipéptidos de AARS comprende un resto de metionina (ATG) de iniciación nativo predicho o el primer resto de aminoácido del polipéptido de AARS predicho es Met, se eliminó. Al final de la fase abierta de lectura del polipéptido de AARS predicho, se añadieron dos codones de parada en un sitio de XhoI (TAATGACTCGAG) (**SEQ ID NO: 2**).

55 EL ADN que codificaba los polipéptidos de AARS marcados en C-terminal se sintetizó con una extensión 5' que codificaba un rbs (subrayado más adelante) y un sitio de restricción de NdeI que o bien recapitula el codón de inicio nativo predicho para el polipéptido de AARS o inserta un ATG en fase con la fase abierta de lectura del polipéptido de AARS predicho, (AGGAGATA<sup>AA</sup>ACATATG) (**SEQ ID NO 3**). En diferentes realizaciones, el sitio de unión a ribosomas puede comprender las secuencias "AGGAGGTA<sup>AA</sup>ACAT" (**SEQ ID NO 4**), "AGGAGATA<sup>AA</sup>ACAT" (**SEQ ID NO 5**), o GAAGGAGATATACAT (**SEQ ID NO 6**). En el extremo 3' de la fase abierta de lectura del polipéptido de AARS predicho, se sintetizó una extensión 3' que codificaba en orden de 5' a 3', un marcador epitópico V5, marcador de seis histidinas, dos codones de parada y un sitio de XhoI, (GGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGCACCACCATCA TCACCATTAATGACTCGAG) (**SEQ ID NO 7**), que está fusionado en fase a la fase abierta de lectura predicha del polipéptido de AARS. En caso de que el polipéptido de AARS incluyese un codón de parada nativo predicho, se eliminó.

Las secuencias de ADN sintetizadas que codificaban los polipéptidos de AARS se subclonaron en el vector pExpress411 (DNA 2.0). Después de secuenciar para confirmar la síntesis del producto correcto, los vectores de expresión se transformaron en bacterias para expresar las proteínas, tal como se describe en más detalle a continuación.

5 En el método (2), El ADN optimizado por codones de *E. coli* (Ermolaeva MD (2001) Curr. Iss. Mol. Biol. 3 (4) 91-7) que codifica cada polipéptido de AARS se sintetizó por GENEWIZ (South Plainfield, NJ). Cada secuencia de polinucleótido que codificaba el polipéptido de AARS se sintetizó con extensiones 5' y 3' cortas que comprendían sitios de restricción únicos para la posterior clonación.

10 Específicamente, se insertó un sitio de restricción de *Bab*HI en el extremo 5' de la fase abierta de lectura predicha. En los casos donde el polipéptido de AARS comprendía un resto de metionina (ATG) de iniciación nativo predicho, o el primer resto de aminoácido del polipéptido de AARS predicho fuese metionina, se eliminó. Además, se insertó un sitio de restricción de *Xho*I en el extremo 3' de la fase abierta de lectura predicha. En los casos donde el polipéptido de AARS comprendía un codón de parada nativo predicho, se eliminó.

15 Después de la digestión de restricción, las secuencias de ADN resultantes se subclonaron en vectores pET-24b modificados (EMD, Gibbstown, NJ) que contenían un marcador epitópico N-terminal (pET24b\_N-6XHis/V5), o C-terminal (pET24b\_C-V5/6XHis) combinado que comprendía un marcador epitópico tanto de seis histidinas como V5 (modificación del vector por GENEWIZ, (South Plainfield, NJ)).

20 Después de la digestión de restricción y de la clonación, se clonó el ADN que codifica el polipéptido de AARS marcado en el extremo N en el vector marcado en el extremo N (pET24b\_N-6XHis/V5), que comprende una secuencia de ADN 5' que codifica seis histidinas y un marcador epitópico V5, (CATATGCATCATCATCATCACGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTA CGGGATCC) (SEQ ID NO 8), en fase con un codón de iniciación (ATG) incluido dentro del sitio de restricción de *Nde*I. Esta extensión 5' se fusionó a la fase abierta de lectura del polipéptido de AARS predicho mediante un enlazador corto de 2 aminoácidos (GS).

30 En el extremo 3' de la fase abierta de lectura predicha, el ADN que codifica el polipéptido de AARS marcado en el extremo N comprendía una secuencia de ADN que codificaba una extensión de 2 aminoácidos (LE) seguida de dos codones de terminación (CTCGAGTAATGA) (SEQ ID NO 9).

35 Después de la digestión de restricción y de la clonación, el ADN que codificaba el polipéptido de AARS marcado en el extremo C se clonó en el vector marcado en el extremo C (pET24b\_C-V5/6XHis), que comprende una secuencia 5' que codifica un codón de iniciación (ATG) incluido dentro del sitio de restricción de *Nde*I que está fusionado a la fase abierta de lectura del polipéptido de AARS predicho mediante un enlazador corto de 2 aminoácidos (GS), (CATATGGGATCC) (SEQ ID NO 10).

40 En el extremo 3' de la fase abierta de lectura predicha, la secuencia de ADN que codifica el polipéptido de AARS marcado en el extremo C comprende una secuencia de ADN 3' que codifica un enlazador corto de 2 aminoácidos (LE) seguido de un marcador epitópico V5 seguido de seis histidinas y dos codones de parada.

CTCGAGGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGCACCAC  
CACCACCACCTAATGA (SEQ. ID. No. 11).

#### 45 EXPRESIÓN, PURIFICACIÓN Y CARACTERIZACIÓN BIOFÍSICA DEL POLIPÉPTIDO DE AARS

Se expresaron polipéptidos de AARS marcados con 6xHis en bacterias en un formato de medio rendimiento y/o en cultivos en matraces a mayor escala dependiendo de la cantidad de proteína requerida. Los polipéptidos de AARS se purificaron usando cromatografía de afinidad de intercambio iónico como se describe más adelante y según se especificase para experimentos específicos.

55 **Cultivos bacterianos:** Se transformaron 100 ng de vector de expresión que comprendía ADN optimizado por codones transformado en bacterias *E. coli* competentes BL21(DE3) (EMD Chemicals, n.º de cat. 69450) a 42 °C durante 30 segundos en placas de PCR. También se evaluaron las cepas C41(DE3) (Lucigen, n.º de cat. 60442), HMS174(DE3) (EMD chemicals, n.º de cat. 69453) y Origami2(DE3) (EMD chemicals, n.º de cat. 71345). Las placas se colocaron sobre hielo durante 2 minutos y se añadieron 100 µl de medio SOC, seguido de una incubación durante 1 hora a 37 °C. Se añadieron 5 ml de medio de autoinducción (EMD Chemicals, n.º de cat. 71491) complementado con canamicina (100 µg/ml) a cada pocillo de un bloque de 24 pocillos (Qiagen, n.º de cat. 19583). Las reacciones de transformación se añadieron a los pocillos individuales, se selló el bloque con película adhesiva (VWS, n.º de cat. 60941-078) y se incubó durante una noche a 250 rpm en un agitador a 37 °C. Cuando se usaron condiciones de baja temperatura (25 °C), se llevó a cabo la incubación durante 48 horas en su lugar.

Para expresión a mayor escala, se añadieron 200 ml de medio de autoinducción complementado con canamicina (100 µg/ml) en matraces Erlenmeyer de 500 ml con tapones ventilados (Corning, n.º de cat. 431401). Las reacciones de transformación se añadieron a los matraces individuales y se incubaron durante 30 horas a 250 en un agitador a 37 °C.

5 **Aislamiento de proteína:** Después de que el cultivo alcanzase la fase estacionaria (DO<sub>600</sub> típica de 3-6), se centrifugaron los bloques a 3600 x g durante 10 minutos. Se aspiró el medio cuidadosamente y se congelaron los bloques a -80 °C o -20 °C durante 10 minutos. Después, se dejaron descongelar los bloques a temperatura ambiente y se añadió a cada pocillo 1 ml de tampón de lisis (100 ml de Bugbuster complementado con 200 µl de lisonasa (EMD chemicals, n.º de cat. 71370) e inhibidores de proteasa "complete mini sin EDTA" (Roche, n.º de cat. 11 836 170 001)). Los sedimentos se resuspendieron pipeteando repetidas veces hasta que no había grumos visibles y se transfirieron a tubos Eppendorf, seguido de una incubación de 10-20 minutos en un agitador a temperatura ambiente. Después de la centrifugación a 16.000 g durante 10 minutos a 4 °C, se cargaron los lisados en una placa TurboFilter 96 incluida en el kit Ni-NTA Superflow 96 BiorRobot (Qiagen, n.º de cat. 969261) y se centrifugaron a 500 g durante 5-10 minutos.

20 Para expresión a mayor escala, se transfirió el cultivo en fase estacionaria a botellas de 500 ml y se centrifugaron a 6.000 g durante 10 minutos. Se decantó el medio y se almacenó el sedimento a -80 °C o -20 °C antes de su posterior procesamiento. Después se dejó descongelar el sedimento a temperatura ambiente y se añadieron 20 ml de este tampón a cada botella. Los sedimentos se resuspendieron pipeteando repetidas veces hasta que no había grumos visibles, seguido de una incubación de 20 minutos en un agitador a temperatura ambiente. Después de la centrifugación a 10.000 g durante 30 minutos a 4 °C, se transfirieron los lisados a tubos o botellas limpias. En caso de que se transfiriesen cantidades traza de restos durante la transferencia, se centrifugó la muestra de nuevo o se hizo pasar a través de una membrana de acetato de celulosa de 0,45 µm (Corning, n.º de cat. 430314) para su clarificación adicional.

30 **Purificación por afinidad:** Se cargó una placa QIAFilter 96 con 200 µl de suspensión Ni-NTA Superflow incluida en el kit Ni-NTA Superflow 96 BioRobot y se equilibró la resina añadiendo 600 µl de tampón de unión (fosfato de sodio 20 mM, cloruro de sodio 500 mM e imidazol 10 mM, pH 7,5). Se aplicó un vacío de -0,5 bar hasta que todo el tampón había pasado a través de la resina. Los lisados celulares aclarados de la etapa anterior se cargaron en la placa QIAFilter® 96 y se dejaron unir durante 5 minutos. Se aplicó un vacío de -0,1 bar durante aproximadamente 5 minutos hasta que todas las muestras habían pasado a través de la resina. Después, se lavó la resina con 1 ml de tampón de unión, seguido de dos lavados con 1 ml de tampón de unión que contenía Triton X-100 al 0,1 %. Después, se lavó la resina 10 veces con 1 ml de tampón de unión sin Triton X-100. Los polipéptidos de AARS marcados con 6xHis unidos se eluyeron con 450 µl de tampón de elución (fosfato de sodio 20 mM, cloruro de sodio 500 mM e imidazol 500 mM, pH 7,5) y se almacenaron a 4 °C.

40 Para expresión a mayor escala, una columna desechable vacía "Poly-Prep" (Bio-Rad, n.º de cat. 731-1550) se cargó con 1 ml de suspensión Ni-NTA Superflow (Qiagen, n.º de cat. 30450) y se equilibraron los 0,5 ml de la resina añadiendo 5 ml de tampón de unión. El lisado celular aclarado de la etapa anterior se cargó en la columna y se dejó pasar a su través por gravedad. En primer lugar, se lavó la resina con 50 ml de tampón de unión más Triton X-100 al 0,1 %, después, se lavó con 50 ml de tampón de unión sin Triton X-100. Los polipéptidos de AARS marcados con 6xHis unidos se eluyeron con 2 ml de tampón de elución y se almacenaron a 4 °C.

45 **Etapas de desalado y pulido:** Para los polipéptidos de AARS con una masa molecular > 10 kDa, se enjuagó la membrana Omega 10K de una placa de filtro AcroPrep 96 (Pall, n.º de cat. 5034) con 20 µl de PBS 1X y se colocó la placa sobre un colector de vacío (>0,33 bar) hasta que todo el líquido pasó a su través. Los eluatos de la etapa anterior (Ni-NTA) se dispensaron en cada pocillo y se aplicó vacío hasta que todo el líquido pasó a su través. Estas etapas se repitieron hasta que el volumen total de eluato (450 µl) se había procesado. Los polipéptidos de AARS se recuperaron añadiendo 180 µl de PBS 1X a pH 7,4 a cada pocillo, pipeteando hacia arriba y hacia abajo 10 veces con cuidado y después se transfirieron a un bloque limpio. Esta etapa se repitió para dar un volumen total de 360 µl por pocillo y se almacenó el bloque a 4 °C. Para los polipéptidos de AARS con una masa molecular <10 kDa, se cargaron los eluatos de Ni-NTA en una unidad de filtrado por centrifugación Amicon Ultra-15 con membrana Ultracel-3 (Millipore, n.º de cat. UFC900308), seguido de la adición de 10 ml de PBS 1X y una centrifugación a 3.600 g durante 10-30 minutos hasta que el volumen era menor de 360 µl. Se recuperaron las muestras y se añadió PBS 1X hasta un volumen final de 360 µl.

60 Para retirar las endotoxinas, una placa de filtro AcroPrep Advance con membrana Mustang Q (Pall, n.º de cat. 8171) se enjuagó con 300 µl de PBS 1X y se centrifugó a 1.000 g durante 5 minutos para retirar el tampón. Los polipéptidos de AARS desalados (360 µl/pocillo) se añadieron a la placa de filtro y se incubaron sobre un agitador durante 5-10 minutos. Después, se centrifugó la placa a 1.000 g durante 5-10 minutos y se recogieron las fracciones de flujo pasante que contenían los polipéptidos de AARS y se almacenaron a 4 °C.

65 Para expresión a mayor escala, se cargaron los eluatos de Ni-NTA en una unidad de filtrado por centrifugación Amicon Ultra-15 con membrana Ultracel-3 Ultracel-10 (Millipore, n.º de cat. UFC900308 o UFC901008) dependiendo del peso molecular del polipéptido de AARS y después se centrifugaron a 3.600 g durante 10-30 hasta que el

volumen se redujo a 250 µl. Las muestras se mezclaron en 10 ml de PBS 1X, pH 7,4 y se centrifugaron de nuevo a 3.600 g durante 10-30 minutos hasta que el volumen era de aproximadamente 250 µl. Esta etapa se repitió una vez más, se recuperaron los sobrenadantes y se añadió PBS 1X hasta un volumen final de 1,5 ml.

- 5 Para retirar las endotoxinas, una membrana de intercambio aniónico fuerte Sartobind Q 5 (Sartorius, n.º de cat. Q5F) se enjuagó con 1 ml de PBS 1X y se hicieron pasar lentamente los polipéptidos de AARS a través de la membrana usando una jeringuilla de plástico. La fracción del flujo pasante que contenía los polipéptidos de AARS se recogió en un bloque de 96 pocillos profundos que se selló y almacenó a 4 °C.
- 10 Los polipéptidos de AARS marcados con 6xHis expresados en bacterias y encontrados en los cuerpos de inclusión se purificaron usando cromatografía de afinidad y una serie de etapas de replegamiento, como se describe más adelante.

- 15 **Cultivos bacterianos:** Se transformaron 100 ng de plásmido que codificaba cada polipéptido de AARS en bacterias *E. coli* BL21(DE3) (EMD Chemicals, n.º de cat. 69450) o C41(DE3) (Lucigen, n.º de cat. 60442) competentes a 42 °C durante 30 segundos en placas para PCR. Las placas se colocaron sobre hielo durante 2 minutos y se añadieron 100 µl de medio SOC, seguido de una incubación durante 1 hora a 37 °C. Se añadieron 5 ml de medio de autoinducción (EMD Chemicals, n.º de cat. 71491) complementado con canamicina (100 µg/ml) a cada pocillo de un bloque de 24 pocillos (Qiagen, n.º de cat. 19583). Las reacciones de transformación se añadieron a los pocillos
- 20 individuales, se selló el bloque con película adhesiva (VWS, n.º de cat. 60941-078) y se incubó durante una noche a 250 rpm en un agitador a 37 °C.

- Para expresión a mayor escala, se añadieron 200 ml de medio de autoinducción complementado con canamicina (100 µg/ml) en matraces Erlenmeyer de 500 ml con tapones ventilados (Corning, n.º de cat. 431401). Las reacciones
- 25 de transformación se añadieron a los matraces individuales y se incubaron durante 30 horas a 250 en un agitador a 37 °C.

- 30 **Aislamiento:** Después de que los cultivos alcanzasen la fase estacionaria (DO<sub>600</sub> típica de 3-6), se centrifugaron los bloques a 3.600 x g durante 10 minutos. Se aspiró el medio cuidadosamente y se congelaron los bloques a -80 °C o -20 °C durante 10 minutos. Después, se dejaron descongelar los bloques a temperatura ambiente y se añadió a cada pocillo 1 ml de tampón de lisis (100 ml de Bugbuster complementado con 200 µl de lisonasa (EMD chemicals, n.º de cat. 71370) e inhibidores de proteasa "complete mini sin EDTA" (Roche, n.º de cat. 11 836 170 001)). Los sedimentos se resuspendieron pipeteando repetidas veces hasta que no había grumos visibles y se transfirieron a tubos Eppendorf, seguido de una incubación de 10-20 minutos en un agitador a temperatura ambiente. Después
- 35 de la centrifugación a 16.000 x g durante 10 minutos a 4 °C, se desecharon los lisados solubles y se resuspendieron exhaustivamente los cuerpos de inclusión en tampón de unión desnaturalizante (fosfato de sodio 20 mM, cloruro de sodio 500 mM, clorhidrato de guanidinio 6 M, imidazol 10 mM, pH 7,5). Las muestras se centrifugaron a 16.000 g durante 10 minutos y se cargaron los sobrenadantes en una placa TurboFilter 96 incluida en el kit Ni-NTA Superflow 96 BioRobot (Qiagen, n.º de cat. 969261) seguido de centrifugación a 500 g durante 5-10 minutos. Los filtrados se
- 40 recogieron en un bloque limpio de 96 pocillos (Greiner, n.º de cat. 780286).

- Para expresión a mayor escala, se transfirió el cultivo en fase estacionaria a botellas de 500 ml y se centrifugaron a 6.000 g durante 10 minutos. Se decantó el medio y se almacenó el sedimento a -80 °C o -20 °C antes de su
- 45 posterior procesamiento. Después se dejó descongelar el sedimento a temperatura ambiente y se añadieron 20 ml de este tampón a cada botella. Los sedimentos se resuspendieron pipeteando repetidas veces hasta que no había grumos visibles, seguido de una incubación de 20 minutos en un agitador a temperatura ambiente. Después de la centrifugación a 10.000 g durante 30 minutos a 4 °C, se desecharon los lisados solubles y se resuspendieron exhaustivamente los cuerpos de inclusión en tampón de unión desnaturalizante.

- 50 **Purificación por afinidad:** Se cargó una placa QIAFilter 96 con 200 µl de suspensión Ni-NTA Superflow incluida en el kit Ni-NTA Superflow 96 BioRobot y se equilibró la resina añadiendo 600 µl de tampón de unión desnaturalizante (véase más arriba). Se aplicó un vacío de -0,5 bar hasta que todo el tampón había pasado a través de la resina. Las muestras desnaturalizadas aclaradas de la etapa anterior se cargaron en la placa QIAFilter® 96 y se dejaron unir durante 5 minutos. Se aplicó un vacío de aproximadamente -0,1 bar durante aproximadamente 5 minutos hasta que
- 55 todas las muestras habían pasado a través de la resina. Después, se lavó la resina con 1 ml de tampón de unión desnaturalizante, seguido de cinco lavados con 1 ml de tampón de unión desnaturalizante que contenía Triton X-100 al 0,1 %. Después, se lavó la resina 15 veces con 1 ml de tampón de unión desnaturalizante sin Triton X-100. Los polipéptidos de AARS marcados con 6xHis unidos se eluyeron después con 450 µl de tampón de elución desnaturalizante (fosfato de sodio 20 mM, cloruro de sodio 500 mM, clorhidrato de guanidinio 6 M e imidazol
- 60 500 mM, pH 7,5) y se almacenaron a 4 °C.

- Para expresión a mayor escala, una columna desechable vacía "Poly-Prep" (Bio-Rad, n.º de cat. 731-1550) se cargó con 1 ml de suspensión Ni-NTA Superflow (Qiagen, n.º de cat. 30450) y se equilibraron los 0,5 ml de la resina
- 65 añadiendo 5 ml de tampón de unión desnaturalizante (véase lo anterior). Los cuerpos de inclusión desnaturalizados de la etapa anterior se cargaron entonces en la columna y se dejaron pasar a su través por gravedad. En primer lugar, se lavó la resina con 50 ml de tampón de unión desnaturalizante más Triton X-100 al 0,1 %, después, se lavó

con 50 ml de tampón de unión desnaturante sin Triton X-100. Los polipéptidos de AARS marcados con 6xHis unidos se eluyeron con 2 ml de tampón de elución desnaturante y se almacenaron a 4 °C.

5 **Replegamiento:** Para polipéptidos de AARS de >10 kDa, se enjuagó la membrana Omega 10K de una placa de filtro AcroPrep 96 (Pall, n.º de cat. 5034) con 20 µl de PBS 1X y se colocó la placa sobre un colector de vacío (>0,33 bar) hasta que todo el líquido pasó a su través. Los eluatos de la etapa anterior (Ni-NTA) se dispensaron en cada pocillo y se aplicó vacío hasta que todo el líquido pasó a su través. Estas etapas se repitieron hasta que el volumen total de eluato (450 µl) se había procesado. Los polipéptidos de AARS se recuperaron añadiendo 200 µl de tampón de replegado que contenía Tris 50 mM, cloruro de sodio 250 mM, cloruro de potasio 10 mM, cloruro de magnesio 2 mM, cloruro de calcio 2 mM, sacarosa 400 mM, arginina 500 mM, DTT 1 mM y polisorbato 80 al 0,01 %, pH 7,4) a cada pocillo, pipeteando hacia arriba y hacia abajo 10 veces con cuidado y después se transfirieron a un bloque limpio. Esta etapa se repitió para dar un volumen total de 400 µl por pocillo y se colocó el bloque sobre el agitador durante una noche a 4 °C. Para polipéptidos de AARS <10 kDa, se cargaron los eluatos de Ni-NTA en una unidad de filtrado por centrifugación Amicon Ultra-15 con membrana Ultracel-3 (Millipore, n.º de cat. UFC900308), seguido de la adición de 10ml de tampón de replegamiento y una centrifugación a 3.600 g durante 10-30 minutos hasta que el volumen era menor de 400 µl. Se recuperaron las muestras y se añadió tampón de replegamiento hasta un volumen final de 400 µl. Las muestras se transfirieron a un bloque de 96 pocillos, se selló con película y se colocó sobre un agitador durante una noche a 4 °C.

20 Para cultivos a mayor escala, se cargaron los eluatos de Ni-NTA en una unidad de filtrado por centrifugación Amicon Ultra-15 con membrana Ultracel-3 Ultracel-10 (Millipore, n.º de cat. UFC900308 o UFC901008) dependiendo del peso molecular del polipéptido de AARS y después se centrifugaron a 3.600 g durante 10-30 hasta que el volumen se redujo a aproximadamente 500 µl. Para los polipéptidos de AARS con un pI > 7, las muestras se diluyeron 20 veces en el siguiente tampón: acetato de sodio 50 mM, cloruro de sodio 10 mM, cloruro de potasio 0,4 mM, EDTA 1 mM, sacarosa 400 mM, arginina 500 mM, DTT 1 mM y polisorbato 80 al 0,01 %, pH 6,0. Para los polipéptidos de AARS con un pI < 7, las muestras se diluyeron 20 veces en el siguiente tampón: Tris 50 mM, cloruro de sodio 250 mM, cloruro de potasio 10 mM, cloruro de magnesio 2 mM, cloruro de calcio 2 mM, sacarosa 400 mM, arginina 500 mM, DTT 1 mM y polisorbato 80 al 0,01 %, pH 8,0. Las muestras se incubaron en un agitador a 4 °C durante una noche.

30 **Etapas de desalado y pulido:** Después de cultivar durante toda la noche, se centrifugó el bloque de 96 pocillos a 3.600 g para retirar cualquier agregado potencial. Después se sometió a los sobrenadantes a intercambio de tampón con PBS 1X (Invitrogen, n.º de cat. 10010). Para polipéptidos de AARS de >10 kDa, se enjuagó la membrana Omega 10K de una placa de filtro AcroPrep 96 con 20 µl de PBS 1X y se colocó la placa sobre un colector de vacío (>0,33 bar) hasta que todo el líquido pasó a su través. Las muestras en el tampón de replegamiento se dispensaron en cada pocillo y se aplicó el vacío hasta que todo el líquido había pasado a su través. Estas etapas se repitieron hasta que el volumen total de muestra (400 µl) se había procesado. Los polipéptidos de AARS se recuperaron añadiendo 180 µl de PBS 1X a pH 7,4 a cada pocillo, pipeteando hacia arriba y hacia abajo 10 veces con cuidado y después se transfirieron a un bloque limpio. Esta etapa se repitió para dar un volumen total de 360 µl por pocillo y se almacenó el bloque a 4 °C. Para los polipéptidos de AARS de <10 kDa, se cargaron las muestras replegadas en una unidad de filtrado por centrifugación Amicon Ultra-15 con membrana Ultracel-3 (Millipore, n.º de cat. UFC900308) seguido de la adición de 1 ml de PBS 1X y centrifugación a 3.600 g durante 10-30 minutos hasta que el volumen era menor de 360 µl. Se recuperaron las muestras y se añadió PBS 1X hasta un volumen final de 360 µl.

45 Para retirar las endotoxinas, una placa de filtro AcroPrep Advance con membrana Mustang Q (Pall, n.º de cat. 8171) se enjuagó con 300 µl de PBS 1X y se centrifugó a 1.000 g durante 5 minutos para retirar el tampón. Los polipéptidos de AARS (360 µl/pocillo) se añadieron a la placa de filtro y se incubaron sobre un agitador durante 5-10 minutos. Después, se centrifugó la placa a 1.000 g durante 5-10 minutos y se recogieron las fracciones de flujo pasante que contenían los polipéptidos de AARS y se almacenaron a 4 °C.

50 Para cultivos a mayor escala, después de incubar durante una noche, se centrifugaron las muestras replegadas a 10.000 g durante 10 minutos para retirar cualquier agregado insoluble. El sobrenadante se cargó sobre una unidad de filtro de centrifugación Amicon Ultra-15 y se centrifugó a 3.600 g hasta que el volumen se había reducido a 250 µl. Las muestras se mezclaron en 10 ml de PBS 1X y se centrifugaron de nuevo a 3.600 g durante 10-30 minutos hasta que el volumen era de aproximadamente 250 µl. Obsérvese que el pH del PBS 1X se ajustó para que coincidiese con el pH del tampón de replegamiento, a pH 6,0 o pH 8,0. Esta etapa se repitió una vez más, se recuperaron los sobrenadantes y se añadió PBS 1X hasta un volumen final de 1,5 ml.

60 Para retirar las endotoxinas, una membrana de intercambio aniónico fuerte Sartobind Q 5 (Sartorius, n.º de cat. Q5F) se enjuagó con 1 ml de PBS 1X y se hicieron pasar lentamente los polipéptidos de AARS a través de la membrana usando una jeringuilla de plástico. La fracción del flujo pasante que contenía los polipéptidos de AARS se recogió en un bloque de 96 pocillos profundos que se selló y almacenó a 4 °C.

65 **CHARACTERIZACIÓN BIOFÍSICA:** Todos los polipéptidos de AARS purificados se analizaron mediante SDS-PAGE, determinándose su concentración basándose en la A<sub>280</sub> y el coeficiente de extinción calculado (ProtParam en el servidor ExpASY). Se midieron los niveles de endotoxina mediante el ensayo de criterio de valoración cromogénico

LAL QCL-1000 (Lonza, n.º de cat. 50-648U) según las instrucciones del fabricante.

**Dispersión de luz dinámica:** Se calentaron un instrumento DynaPro 99 de Wyatt Technology y el controlador de la temperatura (20 °C) durante 15 minutos antes del experimento seguido de la conexión del programa informático Dynamics al instrumento. El tiempo de adquisición se ajustó a 10 segundos para múltiples adquisiciones y se ajustó la potencia del láser al 100 %. Se lavó exhaustivamente la cubeta de cuarzo con agua desionizada y metanol antes de la adición de la muestra de proteína (15 µl a una concentración de aproximadamente 1 mg/ml en PBS). Se retiraron las burbujas de aire golpeando suavemente la cubeta antes de insertarla en el soporte con el lado congelado hacia la izquierda. En caso de que la intensidad fuese demasiado elevada (se muestra un mensaje de advertencia en la pantalla), se diluye adicionalmente la muestra con PBS hasta que la intensidad se reduce hasta un intervalo normal. Los datos recogidos incluyen el radio hidrodinámico, polidispersidad, peso molecular medio predicho, porcentaje de intensidad y porcentaje de masa.

**Cromatografía de exclusión por tamaños:** La muestra de proteína se diluyó hasta una concentración de aproximadamente 5-10 mg/ml en PBS antes de cargarla en un asa de muestra de 100 µl en el dispositivo de FPLC AKTA de General Electric. Para la separación, se usó la columna de exclusión por tamaños Superdex 200 10/300 GL (General Electric, n.º de cat. 17-5175-01). En primer lugar, la columna se equilibró con 1,5 volúmenes de columna (VC) de tampón PBS 1X, seguido de la inyección de la muestra. La columna se corrió con 1 VC de tampón PBS 1x (flujo isocrático) controlando la absorbancia a 280 nm. Se integró el área máxima y se calculó el porcentaje con el programa informático Unicorn. Se usó el volumen de elución para estimar el peso molecular basándose en la comparación con kits de calibración de filtración en el (General Electric, n.º de cat. 28-4038-41 y 28-4038-42).

**Recuperación de proteína tras su almacenamiento a alta concentración:** se transfirieron 10 µl de los polipéptidos de AARS concentrados a  $\geq 10$  mg/ml usando una unidad de filtro Amicon Ultra-15 (Millipore, n.º de cat. UFC901024 o UFC900324 dependiendo del peso molecular) a un tubo de microcentrifugación limpio. La muestra se almacenó a temperatura ambiente durante una semana seguido de centrifugación a 16.000 g durante 10 minutos para aglomerar cualquier precipitado. Se determinó la concentración del sobrenadante mediante un ensayo de proteína Bradford y se comparó con la concentración medida antes de la exposición de una semana a temperatura ambiente. La recuperación se expresó como porcentaje de la concentración inicial.

**Caracterización de los polipéptidos de AARS por CL-EM:** Se diluyeron polipéptidos de AARS purificados (1 mg/ml) a 1:10 en ácido fórmico al 0,1 % y se cargaron 0,6 µg de proteína con un automuestreador Dionex en una columna capilar C4. La columna capilar se preparó cortando 150 mM de tubos de sílice fusionado (0,36 mM de DE por 0,1 mM de DI, Polymicro Technologies, n.º de cat. 2000023). El capilar se estiró desde un extremo con un instrumento láser para estirar fibras de Suter y se cortó con un cortador de sílice fusionado para generar una punta de 5 µm. El capilar se empaquetó hasta una longitud de 75 mM con resina C4 (5µm, 300Å, Michrom, n.º de cat. PM5/64300/00) usando una bomba de presión. Se llevó a cabo el análisis CL-EM en un espectrómetro de masas de trampa iónica ThermoFisher LTQ acoplado a un sistema HPLC Dionex Ultimate3000. El analito se eluyó de la columna usando un gradiente de 35 minutos de acetonitrilo al 5-70 % en ácido fórmico al 0,1 % a un caudal de 0,9 µl/min. El LTQ se hizo funcionar en un modo de escaneo de EM completo (300-2.000  $m/z$ ) con un voltaje de pulverización de 2,5 kV.

Recogida y análisis de datos: los datos de espectrometría de masas en bruto se almacenaron en archivos RAW generados mediante XCalibur ejecutado en el espectrómetro de masas LTQ XL. Los espectros de masas de los picos principales en el cromatógrafo se analizaron adicionalmente con el algoritmo de desconvolución ProMass de ThermoFisher para obtener los pesos moleculares del polipéptido de AARS.

## ANÁLISIS FUNCIONAL DE POLIPÉPTIDOS DE AARS

### Elaboración de perfiles transcripcionales

**Antecedentes y relevancia terapéutica:** Además de las técnicas de identificación de dianas tradicionales, han surgido recientemente las herramientas genómicas como estrategias importantes para ayudar en la dilucidación del mecanismo de acción de los polipéptidos de AARS y pueden proporcionar una perspectiva directa acerca de la relevancia terapéutica de manera temprana en el proceso de descubrimiento de fármacos. Para facilitar una comprensión de la utilidad terapéutica potencial, se cultivaron tipos celulares humanos primarios con polipéptidos de AARS y se evaluaron los perfiles de transcripción en dos instantes separados después de la incubación con polipéptidos de AARS.

Los tipos celulares seleccionados para la elaboración de perfiles transcripcionales estaban basados en las capacidades pluripotentes de las células en cuestión y su potencial para identificar polipéptidos de AARS con un valor terapéutico directo. Por ejemplo, las células madre mesenquimales (MSC) pueden diferenciarse en linajes osteogénicos, adipogénicos, condrogénicos, miocárdicos o neurales cuando se exponen a estímulos específicos, haciendo que sean atractivas para comprender la relevancia potencial de los polipéptidos de AARS para una gran variedad de tipos celulares y enfermedades.

Además de apoyar a las células hematopoyéticas, también puede inducirse a las células estromales de la médula para que se diferencien en células de diferentes linajes de tejido conectivo, tales como hueso, cartílago y grasa. El potencial de las células madre mesenquimales humanas (hMSC) para mantener la multipotencia y proliferar de manera extensa *in vitro* proporciona nuevas vías de terapia basada en células para restaurar el tejido dañado o enfermo. Artículos recientes indican también que las hMSC son capaces de atravesar los límites de la capa germinal del destino celular. Además de diferenciarse en múltiples linajes del mesodermo, estas células también pueden diferenciarse en neuronas de origen ectodérmico y en células similares a hepatocitos de origen endodérmico. Durante el proceso de diferenciación, estas células pueden modificar los patrones de expresión de determinados transcritos específicos de linaje.

Por consiguiente, la capacidad de los polipéptidos de AARS específicos para modular patrones específicos de genes en las hMSC de una manera dependiente del tiempo demuestra que estas proteínas desempeñan potencialmente papeles significativos en una gran variedad de vías de diferenciación, así como enfermedades y trastornos que son el resultado de la disfunción o del deterioro de estos procesos o los correspondientes tipos celulares. Además, los polipéptidos de AARS con la capacidad de modular la transcripción génica en las MSC tienen una utilidad terapéutica significativa para permitir la modulación *in vitro* o *in vivo* de la hematopoyesis, neurogénesis, miogénesis, osteogénesis y adipogénesis, así como en una gran variedad de trastornos y enfermedades, que incluyen, por ejemplo, respuestas inflamatorias, autoinmunidad, cáncer, degeneración neuronal, distrofia muscular, osteoporosis y lipodistrofia.

Las células musculares esqueléticas humanas (HskMC) pueden sufrir diferenciación para mostrar miofilamentos de actina y miosina y se han usado en el estudio de enfermedades musculares genéticas, tales como la hipertermia maligna<sup>1</sup>. Las HskMC también tienen el potencial de actuar como injerto cardíaco, reparando el daño al corazón. Recientemente, se han usado células musculares esqueléticas humanas cultivadas en experimentos de microgravedad para estudiar los efectos de los ambientes con baja gravedad en el músculo esquelético.

Por consiguiente, la capacidad de los polipéptidos de AARS específicos para modular patrones específicos de genes en las HskMC de una manera dependiente del tiempo demuestra que estas proteínas desempeñan potencialmente papeles significativos en los procesos de miogénesis, así como enfermedades y trastornos que son el resultado de la disfunción o el deterioro de estos procesos, así como el desarrollo o metabolismo de las células musculares. Por consiguiente, los polipéptidos de AARS con la capacidad de modular la transcripción génica en células musculares tienen utilidad terapéutica en una gran variedad de enfermedades que incluyen, por ejemplo, el tratamiento de enfermedades metabólicas, caquexia, varias afecciones de emaciación muscular, así como enfermedades musculoesqueléticas.

**Métodos:** Se evaluó la capacidad de los polipéptidos de AARS para modular la expresión génica usando una estrategia de PCR cuantitativa en tiempo real (RT-qPCR) de microfluidos de alto rendimiento (Fluidigm Corporation). (Véase Petriv et al., (2010) PNAS (doi/10.1073/pnas.1009320107) en células estromales de médula humanas (hMSC) y células musculares esqueléticas humanas (HskMC). En los experimentos comunicados en el presente documento, se adquirieron HskMC humanas (n.º de cat. 150-05f) y hMSC (n.º de cat. 492-05f) de Cell Applications. Las células hMSC se criopreservan en el segundo pase y pueden cultivarse y propagarse hasta 10 duplicaciones de la población. En este caso, se usaron hMSC en el 6º pase. Las células musculares esqueléticas humanas (HskMC) se criopreservan en el segundo pase y pueden cultivarse y propagarse hasta al menos 15 duplicaciones de la población. En los experimentos comunicados en el presente documento se usaron HskMC en el pase 6 después de la recogida de un donante humano normal.

En ambos casos, las células se sembraron a 50000 células/ml en un volumen de 100 µl de medio de crecimiento y se expusieron a polipéptidos de AARS a una concentración de 250 nM o como se indica a continuación de otro modo, durante 24 horas y 72 horas. Los controles incluyeron medio de diferenciación con un cóctel estándar para promover (1) la adipogénesis, (2) la osteogénesis, (3) la condrogénesis y (4) la formación de miotubos de músculo esquelético. Los controles adicionales incluyeron pocillos no tratados que contenían únicamente medio de crecimiento. Se analizaron dos pocillos para cada control de diferenciación. Controles: todos los medios se elaboraron utilizando DMEM como medio basal. Se siguió la bibliografía estándar y se adquirió el medio de diferenciación de Cell Applications. Según el vendedor, el medio de diferenciación contenía los siguientes aditivos: Cóctel de diferenciación de músculos esqueléticos: FBS, insulina, glutamina, FGF, EGF; Cóctel de adipogénesis: insulina, dexametasona e IBMX; Cóctel de osteogénesis: FBS, dexametasona, ascorbato 2 fosfato, beta-glicerofosfato; Cóctel de condrogénesis: insulina, ascorbato -2 fosfato y TGF-β1.

Se utilizaron procedimientos estándar para usar un kit TAQMAN® Gene Expression Cells-to-CT™ de ABI (Applied Biosystems, artículo n.º AM 1728) para lisar las células y recoger el material genómico. Se usó una mezcla ABI Pre-Amp (Applied Biosystems, artículo n.º 4391128) para iniciar la pre-amplificación. Se crearon cebadores específicos de genes usando un programa Primer 3 adquirido de IDT Technologies. Se usaron matrices de elaboración de perfiles Fluidigm (artículo n.º BMK-M-96.96) para la PCR cuantitativa real con reactivos de carga de Fluidigm convencionales y dispositivos de pipeteado. La tabla E1 a continuación lista los genes de los que se elaboraron los perfiles.

Tabla E1 Lista de genes evaluados en la elaboración de perfiles transcripcionales			
Lista única compilada	refseq_nt	Nombre completo	Sinónimos
ABCA1	NM_005502	Casete de unión a ATP, subfamilia A (ABC1), miembro 1	ABC-1 ABC1 CERP FLJ14958  HDLDT1  MGC164864 MGC165011 TGD
ACTB	NM_001101	actina, beta	PS1TP5BP1
ACTG1	NM_001614	actina, ga mMa 1	ACT ACTG DFNA20 DFNA26
ACVR2B	NM_001106	receptor de activina A, tipo IIB	ACTRIIB ActR-IIB MGC116908
APOA1	NM_000039	apolipoproteína A-I	MGC117399
ARNT	NM_178427	traslocador nuclear del receptor de hidrocarburos de arilo	HIF-1beta HIF1B HIF1BETA  TANGO  bHLHe2
BAD	NM_032989	agonista de la muerte celular asociado a BCL2	BBC2 BCL2L8
BCL2	NM_000657	CLL/linfoma de células B 2	Bc1-2
BMP2	NM_001200	proteína morfogenética ósea 2	BMP2A
BMP4	NM_130851	proteína morfogenética ósea 4	BMP2B BMP2B1 MCOPS6 OFC11   ZYME
C3AR1	NM_004054	receptor 1 del componente de complemento 3a	AZ3B C3AR HNFAG09
CASP3	NM_032991	caspasa 3, cisteína peptidasa relacionada con la apoptosis	CPP32 CPP32B SCA-1
CAV1	NM_001753	caveolina 1, proteína caveolae, 22kDa	BSCL3 CGL3 MSTP085 VIP21
CDH5	NM_001795	cadherina 5, tipo 2 (endotelio vascular)	7B4 CD144 FLJ17376
CFLAR	NM_003879	regulador de la apoptosis similar a CASP8 y FADD	CASH CASP8AP1 CLARP Casper  FLAME FLAME-1 FLAME1 FLIP  I-FLICE MRIT  c-FLIP c-FLIPL  c-FLIPR c-FLIPS
COMP	NM_000095	proteína de matriz oligomérica de cartílago	EDM1 EPD1 MED MGC131819  MGC149768  PSACH THBS5
CSF1	NM_172212	factor estimulante de colonias 1 (macrófagos)	MCSF MGC31930
CTGF	NM_001901	factor de crecimiento de tejido conectivo	CCN2 HCS24 IGFBP8 MGC10283 9  NOV2
CTNNB1	NM_001904	catenina (proteína asociada a cadherina), beta 1, 88kDa	CTNNB DKFZp686D02253  FLJ25606  FLJ37923
DAAM1	NM_014992	activador de la morfogénesis 1 asociado a dishevelled	FLJ41657 KIAA0666
ELN	NM_001081755	elastina	FLJ38671 FLJ43523 SVAS WBS W S
ENO1	NM_001428	enolasa 1, (alfa)	ENO1L1 MPB1 NNE PPH
FABP3	NM_004102	proteína de unión a ácidos grasos 3, músculo y corazón (inhibidor del crecimiento derivado de mama)	FABP11 H-FABP MDGI O-FABP
FAK	NM_001199649	cinasa de adhesión focal	fak1

ES 2 638 311 T3

FGF4	NM_002007	factor de crecimiento de fibroblastos 4	HBGF-4 HST HST-1 HSTF1 K-FGF  KFGF
FIGF	NM_004469	factor de crecimiento inducido por c-fos (factor de crecimiento endotelial vascular D)	VEGF-D VEGFD
FLT1	NM_002019	tirosina cinasa 1 relacionada con fms (factor de crecimiento endotelial vascular/receptor de factor de permeabilidad vascular)	FLT VEGFR1
FOXA1	NM_004496	caja A1 de forkhead	HNF3A MGC33105 TCF3A
GAPDH	NM_002046	gliceraldehído-3-fosfato deshidrogenasa	G3PD GAPD MGC88685
GFAP	NM_002055	proteína ácida gliofibrilar	FLJ45472
SLC2A4	NM_001042	familia de transportador de soluto 2 (transportador de glucosa facilitado), miembro 4	GLUT4
HAND1	NM_004821	derivados de la cresta neural y cardiaca expresados 1	Hxt Thing1 bHLHa27 eHand
HIF1A	NM_181054	factor inducible por hipoxia 1, subunidad alfa (factor de transcripción de hélice-bucle-hélice básico)	HIF-1alfa HIF1 HIF1-ALFA MOP1 PASD8 bHLHe78
HK2	NM_000189	hexocinasa 2	DKFZp686M1669 HKII HXK2
HMGB1	NM_002128	caja 1 de grupo de alta movilidad	DKFZp686A04236 HMG1 HMG3  SBP-1
HNF4A	NM_178850	factor nuclear de hepatocitos 4, alfa	FLJ39654 HNF4 HNF4a7 HNF4a8  HNF4a9  HNF4alpha MODY MODY1 NR2A 1  NR2A21 TCF TCF14
HPRT1	NM_000194	hipoxantina fosforibosiltransferasa 1	HGPRT HPRT
HSPB1	NM_001540	proteína 1 de choque térmico de 27kDa	CMT2F DKFZp586P1322 HMN2B  HS.76067 HSP27 HSP28 Hsp25 SRP27
ICAM1	NM_000201	molécula de adhesión intercelular 1	BB2 CD54 P3.58
IFNG	NM_000619	interferón, ga mMa	IFG IFI
IGF1	NM_001111285	factor de crecimiento insulínico 1 (somatomedina C)	IGF-I IGF1A IGFI
IGF2	NM_001127598	factor de crecimiento insulínico 2 (somatomedina A)	C11orf43 FLJ22066 FLJ44734 INSI GF  pp9974.
IGFBP3	NM_001013398	proteína de unión 3 al factor de crecimiento insulínico	BP-53 IBP3
IGFBP5	NM_000599	proteína de unión 5 al factor de crecimiento insulínico	IBP5
IKKB	NM_001556	inhibidor del potenciador de gen de polipéptido ligero kappa en células B, cinasa beta	FLJ33771 FLJ36218 FLJ38368 FLJ 40509  IKK-beta IKK2 IKKB MGC131801  NFKBIKB

ES 2 638 311 T3

IL10	NM_000572	interleucina 10	CSIF IL-10 L10A MGC126450  MGC126451 TGIF
IL1B	NM_000576	interleucina 1, beta	IL-1 IL1-BETA IL1F2
IL3	NM_000588	interleucina 3 (factor estimulante de colonias, múltiple)	IL-3 MCGF MGC79398 MGC79399  MULTI-CSF
IL4	NM_172348	interleucina 4	BCGF-1 BCGF1 BSF1 IL-4  MGC79402
IL5	NM_000879	interleucina 5 (factor estimulante de colonias, eosinófilos)	EDF IL-5 TRF
IL6R	NM_181359	receptor de interleucina 6	CD126 IL-6R-1 IL-6R-alpha IL6RA MGC104991
IL8	NM_000584	interleucina 8	CXCL8 GCP1 GCP1 LECT LUCT  LYNAP MDNCF MONAP NAF NAP-1 NAP1
ITGA5	NM_002205	integrina, alfa 5 (receptor de fibronectina, polipéptido alfa)	CD49e FNRA VLA5A
KDR	NM_002253	receptor del dominio de inserción de cinasa (una tirosina cinasa receptora de tipo III)	CD309 FLK1 VEGFR VEGFR2
LEP	NM_000230	leptina	FLJ94114 OB OBS
LPL	NM_000237	lipoproteína lipasa	HDLCQ11 LIPD
MAPK11	NM_002751	proteína cinasa 11 activada por mitógeno	P38B P38BETA2 PRKM11 SAPK2  SAPK2B  p38-2 p38Beta
mMP1	NM_002421	metalopeptidasa de matriz 1 (colagenasa intersticial)	CLG CLGN
mMP3	NM_002422	metalopeptidasa de matriz 3 (estromelisinina 1, progelatinasa)	CHDS6 MGC126102 MGC126103  MGC126104  mMP-3 SL-1 STMY STMY1 STR1
MYH1	NM_005963	miosina, cadena pesada 1, músculo esquelético, adulto	MGC133384 MYHSA1 MYHa  MyHC-2X/D MyHC-2x
MYH11	NM_022844	miosina, cadena pesada 11, músculo liso	AAT4 DKFZp686D10126  DKFZp686D19237  FAA4 fLJ35232 MGC126726  MGC32963  SMHC S mMHC
MYH7	NM_000257	miosina, cadena pesada 7, músculo cardíaco, beta	CMD1S CMH1 DKFZp451F047  MGC138376  MGC138378 MPD1 MYHCB  SPMD SP mM
MYOD1	NM_002478	diferenciación miogénica 1	MYF3 MYOD PUM bHLHc1
NFATC1	NM_172390	factor nuclear de células T activadas, citoplasmático, dependiente de calcineurina 1	MGC138448 NFATC NFAT2 NFATc
NFATC2	NM_173091	factor nuclear de células T activadas, citoplasmático, dependiente de calcineurina 2	NFAT1 NFATP
NFKB1	NM_003998	factor nuclear del potenciador de gen de polipéptido ligero kappa en células B 1	DKFZp686C01211 EBP-1 KBF1  MGC54151  NF-kappa-B NF-kappaB NFKB-p105  NFKB-p50 p105 p50
NOS2	NM_000625	sintasa de óxido nítrico 2, inducible	HEP-NOS INOS NOS NOS2A
NOTCH1	NM_017617	notch 1	TAN1 hN1

NR3C1	NM_0010240 94	subfamilia 3 de receptor nuclear, grupo C, miembro 1 (receptor de glucocorticoides)	GCCR GCR GR GRL
NRP2	NM_201279	neuropilina 2	MGC126574 NP2 NPN2 PRO2714  VEGF165R2
PAX7	NM_013945	caja emparejada 7	FLJ37460 HUP1 PAX7B RMS2
PDGFB	NM_033016	polipéptido beta de factor de crecimiento derivado de plaquetas (homólogo de oncogén de sarcoma viral de simio (v-sis))	FLJ12858 PDGF2 SIS SSV c-sis
PDK4	NM_002612	piruvato deshidrogenasa cinasa, isozima 4	FLJ40832
PLA2G1 B	NM_000928	fosfolipasa A2, grupo IB (páncreas)	MGC119834 MGC119835 PLA2  PLA2A  PPLA2
PLIN1	NM_002666	proteína asociada a microgotas lipídicas	perilipina
PPARG	NM_138712	receptor ga mMa activado por proliferador de peroxisomas	CIMT1 GLM1 NR1C3 PPARG1  PPARG2  PPARga mMa
QARS	NM_005051	glutaminil-ARNt sintetasa	GLNRS PRO2195
RHOA	NM_001664	familia de genes homólogos de ras, miembro A	ARH12 ARHA RHO12 RHOH12
RUNX1	NM_001754	factor de transcripción 1 relacionado con runt	AML1 AML1-EVI-1  AMLCR1 CBFA2 EVI-1  PEBP2aB
RXRA	NM_002957	receptor X de retinoides, alfa	FLJ00280 FLJ00318 FLJ16020  FLJ16733  MGC102720 NR2B1
SERPINE1	NM_0011654 13	inhibidor de serpin peptidasa, clado E (nexina, inhibidor del activador de plasminógeno de tipo 1), miembro 1	PAI PAI-1  PAI1 PLANH1
SMAD2	NM_005901	miembro 2 de la familia de SMAD	JV18 JV18-1 MADH2 MADR2  MGC22139  MGC34440 hMAD-2 hSMAD2
SMAD4	NM_005359	miembro 4 de la familia de SMAD	DPC4 JIP MADH4
TERT	NM_198255	retrotranscriptasa de telomerasa	EST2 TCS1 TP2 TRT hEST2
TGFB1	NM_000660	factor de crecimiento transformante, beta 1	CED DPD1 LAP TGFB TGFbeta
TGFB3	NM_003239	factor de crecimiento transformante, beta 3	ARVD FLJ16571 TGF-beta3
THBS4	NM_003248	trombospondina 4	TSP4
TNF	NM_000594	factor de necrosis tumoral	DIF TNF-alfa TNFA TNFSF2
TUBB	NM_178014	tubulina, beta	M40 MGC117247 MGC16435  OK/SW-c1.56 TU13B1  TUBB5
TUBB1	NM_030773	tubulina, beta 1	isoforma beta (1) de tubulina
TUBG1	NM_001070	tubulina, ga mMa 1	GCP-1 TUBG TUBGCP1
VCAM1	NM_080682	molécula de adhesión celular vascular 1	CD106 DKFZp779G2333 INCAM-100 MGC99561
VEGFA	NM_003376	factor de crecimiento endotelial vascular A	MGC70609 MVCD1 VEGF VPF
VIM	NM_003380	vimentina	FLJ36605

WISP1	NM_080838	proteína de vía de señalización 1 inducible por WNT1	CCN4 WISP1c WISP1i WISP1tc
WNT1	NM_005430	familia de sitio de integración mMTV de tipo wntless, miembro 1	INT1

**Análisis bioinformático:** Los datos recopilados en formato .csv de la máquina Biomark por Fluidigm se convirtieron en un formato tabulado que incluía muestra, ARNm e información de replicado junto con el valor de fluorescencia en bruto. Las reacciones PCR que fallaron se marcaron como ausentes. Se combinaron múltiples experimentos después de normalizar a la expresión total de las especies de ARNm. Toda la expresión medida de ARNm se filtró basándose en el requisito de detección en al menos 2 de todos los replicados biológicos ensayados. Los presentes inventores evaluaron la media técnica, biológica y de desviación ajustada en todo el conjunto de datos.

Para el análisis de los datos, se normalizaron en primer lugar los valores de Ct para todos los genes a los valores de Ct promediados para genes constitutivos de la muestra correspondiente para obtener valores de  $\Delta Ct$  ( $\Delta Ct = Ct$  del gen - Ct medio de genes constitutivos). Después se normalizaron los genes de cada muestra al mismo gen en un control no tratado para obtener valores de  $\Delta\Delta Ct$  ( $\Delta\Delta Ct = \Delta Ct$  de muestra de control -  $\Delta Ct$  de muestra tratada).

Para obtener los valores de factor de cambio, se someten los genes regulados positivamente (es decir,  $\Delta\Delta Ct$  mayores de 0) al siguiente cálculo: Factor de cambio =  $2^{\Delta\Delta Ct}$ . Para genes regulados negativamente (es decir,  $\Delta\Delta Ct$  menores de 0): Factor de cambio =  $-(2^{|\Delta\Delta Ct|})$ .

#### ENSAYOS DE PROLIFERACIÓN CELULAR (ENSAYOS A1-A11 EN LAS TABLAS DE DATOS MÁS ADELANTE)

**Antecedentes y relevancia terapéutica:** La capacidad de modular la velocidad de proliferación celular y de apoptosis de diferentes tipos celulares representa una propiedad fundamental de muchos compuestos terapéuticos y tiene una relevancia directa en el tratamiento y la prevención de una gran variedad de enfermedades y trastornos.

Por consiguiente, los polipéptidos de AARS con capacidad para modular la velocidad de la proliferación celular y/o de la apoptosis tienen una utilidad terapéutica significativa en una gran variedad de enfermedades, incluyendo, como factores de crecimiento y factores de diferenciación para células madre y en regímenes de tratamiento para potenciar o suprimir la proliferación de tipos celulares de interés *in vivo* o *in vitro*, que incluyen, por ejemplo, células hematopoyéticas, células inmunomoduladoras, cáncer y para el tratamiento y prevención de enfermedades asociadas con el envejecimiento, que incluyen, por ejemplo, neurodegeneración, neuropatía periférica y pérdida del tono muscular y de tejidos blandos.

**Métodos:** Se evaluaron los efectos de los polipéptidos de AARS en la proliferación celular usando uno o más métodos listados a continuación y tal como se detalla de manera más específica en los métodos a continuación.

**Hoechst 33432.** Los recuentos celulares estándar para evaluar la proliferación se efectuaron usando Hoechst 33432, que es una contratinción nuclear que permea las células que emite fluorescencia azul cuando se une a ADNbc. Se encuentra disponible en forma de solución (Invitrogen, n.º de cat. H-3570) que se usa a una concentración final de 1 ug/ml en medio o en PBS. Las células se cultivaron en placas de 96 pocillos en presencia de polipéptidos de AARS durante un tiempo de cultivo estándar de 48 horas o más, dependiendo del tipo celular y tal como se describe en los ejemplos más adelante.

**ATP-lite.** Los niveles de ATP celular se correlacionan con la salud celular y pueden determinarse fácilmente usando una serie de kits disponibles comercialmente. ATP-lite (Perkin-Elmer, n.º de cat. 6016947, Boston, MA 02481) que es una mezcla homogénea de solución de lisis y reactivo de detección de ATP, se premezcla antes de su uso y se usa a una relación 1:1 v:v con células cultivadas. Las placas se incubaron durante 5 minutos para promover la lisis y se midieron las placas usando un lector de placas de luminiscencia. Las células se cultivaron en placas de 96 pocillos en presencia de polipéptidos de AARS durante un tiempo de cultivo estándar de 48 horas o más, dependiendo del tipo celular y tal como se describe en los ejemplos más adelante.

**ALAMARBLUE®** (Resazurina) es un indicador de la viabilidad celular que está basado en el estado redox de las células. La resazurina, el principio activo, es un compuesto no tóxico permeable a las células que es de color azul y prácticamente no fluorescente cuando está presente en su forma oxidada. Sin embargo, tras entrar en las células viables normales, la resazurina se reduce rápidamente en resorufina, que produce una señal de fluorescencia roja. Las células viables convierten de manera continua la resazurina en resorufina, generando de este modo una medición cuantitativa de la viabilidad y la toxicidad. La ausencia de toxicidad permite la exposición de larga duración de las células a la resazurina sin un impacto negativo; se observó que las células cultivadas en presencia de resazurina producen números similares de células viables como células de control, determinados por análisis de citometría de flujo.

**Se efectuaron mediciones añadiendo** una solución de resazurina/ALAMARBLUE® a las células, incubándolas durante 1-4 horas y leyendo la fluorescencia o la absorbancia. La cantidad de fluorescencia o absorbancia es proporcional al número de células vivas y corresponde a la actividad metabólica de las células. Las células dañadas y no viables tienen una actividad metabólica innata menor y por lo tanto, generan una señal proporcionalmente menor que las células sanas. Después de la incubación con ALAMARBLUE®, las muestras pueden medirse fácilmente en instrumentos para fluorescencia y absorbancia. Para las lecturas de fluorescencia: se usaron configuraciones de filtro de 530 nm de excitación y 590 nm de emisión.

Las células se cultivaron en placas de 96 pocillos en presencia de polipéptidos de AARS durante un tiempo de cultivo estándar de 48 horas o más, dependiendo del tipo celular y tal como se describe en los ejemplos más adelante.

#### **CAPTACIÓN DE LDL ACETILADAS EN HEPATOCITOS HUMANOS HEPG2C3A. (ENSAYO B1 N LAS TABLAS DE DATOS MÁS ADELANTE)**

**Antecedentes y relevancia terapéutica:** Las LDL son el transportador principal de colesterol en la sangre, suponiendo más del 60 % del colesterol en plasma. En seres humanos, el receptor de LDL hepático es responsable de la eliminación de aproximadamente el 70 % de las LDL en plasma de la circulación. Las LDL internalizadas se degradan en colesterol libre y aminoácidos en el lisosoma. El hígado es el órgano más importante para el catabolismo de las LDL y la actividad del receptor de LDL en seres humanos. Las LDL que no se internalizan y permanecen en la circulación pueden transportarse por células endoteliales al interior de la pared vascular, dando como resultado la formación de placas ateroscleróticas. Las LDL en circulación también pueden captarse por macrófagos y esto puede contribuir también a la formación de placas. Se cree que el aumento de la captación de LDL al tejido hepático es beneficioso para la salud humana y la búsqueda de agentes terapéuticos seguros y eficaces que puedan regular positivamente este proceso puede proporcionar nuevas terapias para enfermedades cardiovasculares y metabólicas. Para investigar si las propiedades únicas de los polipéptidos de AARS pueden regular la captación de LDL acetiladas, se emplea un ensayo convencional para medir la captación de LDL acetiladas en células HepG2C3a.

Por consiguiente, los polipéptidos de AARS con la capacidad para modular la captación de LDL tienen una utilidad terapéutica significativa en una gran variedad de enfermedades que incluyen, por ejemplo, el tratamiento de la hipercolesterolemia, hiperlipidemia, diabetes de tipo 1 y 2, síndrome metabólico y enfermedades cardiovasculares, incluyendo la aterosclerosis.

**Métodos:** Se mantienen células HEPG2C3a (ATCC n.º CRL-10741) en medio esencial mínimo de Eagle (EMEM) complementado con FBS al 10 % (HyClone, n.º de cat. SH30910.03), 50u/ml de penicilina/50µg/ml de estreptomycin, (Invitrogen) en 15 ml de medio en matraces de 75 ml. Las células se cultivan a 37 °C, CO<sub>2</sub> al 5 %, en un ambiente humidificado y se utilizaron en campanas certificadas para cultivo tisular BSL2 usando técnicas estériles y equipamiento de protección personal adecuado que incluye gafas, guantes y batas de laboratorio. HEPG2C3a expresan el receptor de LDL y son competentes para la captación de LDL acetiladas cuando se cultivan sobre placas de fondo transparente recubiertas con colágeno. Se sembró un volumen de 100 µl de células sobre placas recubiertas con colágeno (Invitrogen n.º de cat. A11428) durante una noche en medio completo (anterior) a una densidad de 50.000 células/ml. Las células se lavan una vez con PBS (Invitrogen, n.º de cat. 10010) y se añaden 80 µl de EMEM sin suero a cada pocillo. Se añaden los polipéptidos de AARS a una concentración final de 250nM por pocillo en un volumen consistente en PBS a cada pocillo. En cada pocillo se coloca un solo polipéptido de AARS. Se priva de suero a las células y se las expone a los polipéptidos de AARS durante 16 horas. Después de la incubación de 16 horas, se recoge el sobrenadante y se mide ICAM soluble usando un kit de ELISA estándar de R&D Systems (n.º de cat. DY643), y se añade medio libre de suero complementado con 5µg/ml de ac-LDL (marcado con Alexa Fluor 488, n.º de cat. L23380, Invitrogen) a cada pocillo. Después de una incubación de 2 horas a 37 °C con CO<sub>2</sub> al 5 %, se lavan dos veces las células con PBS estéril antes de añadir 100 µl de PBS a cada pocillo para la cuantificación. Se analizó la intensidad de fluorescencia total de las placas usando una lectura inferior en un lector de placas de fluorescencia Victor X5 (Perkin Elmer) a una longitud de onda de excitación centrada en aproximadamente 485 nm y una longitud de onda de emisión centrada en aproximadamente 535 nm. Las células se tiñen con colorante Hoechst y se lee la intensidad de fluorescencia a 405nm de excitación/450 nm de emisión para confirmar el número de células total por toda la placa.

#### **REGULACIÓN DE LA EXPLOSIÓN OXIDATIVA Y LA PRODUCCIÓN DE ELASTASA DE NEUTRÓFILOS HUMANOS (ENSAYOS C1-C3 EN LAS TABLAS DE DATOS MÁS ADELANTE)**

##### **EXPLOSIÓN OXIDATIVA DE NEUTRÓFILOS**

**Antecedentes y relevancia terapéutica:** La fagocitosis por neutrófilos polimorfonucleares y monocitos constituye un arma esencial de las defensas del hospedador contra infecciones por microorganismos que incluyen bacterias y hongos. El proceso fagocítico puede separarse en varias etapas principales: quimiotaxis (migración de los fagocitos a los sitios de inflamación), unión de partículas a la superficie celular de los fagocitos, ingestión (fagocitosis) y eliminación intracelular por mecanismos dependientes de oxígeno (explosión oxidativa) e independientes de

oxígeno. Se observa una actividad explosiva reducida o ausente en defectos congénitos, tales como la enfermedad de la granulomatosis crónica (CGD). La CGF es un grupo heterogéneo de trastornos heredados que normalmente se manifiestan durante los dos primeros años de vida. La enfermedad se caracteriza por infecciones repetitivas y potencialmente letales causadas por organismos bacterianos y fúngicos. Estas infecciones consisten normalmente en neumonía, linfadenitis o abscesos que implican a los ganglios linfáticos, pulmones e hígado. La NADPH oxidasa es el sistema enzimático responsable de producir el anión superóxido, que rápidamente se convierte en peróxido de hidrógeno y radicales hidroxilo. Las anomalías en los péptidos constituyentes del sistema de enzima NADPH oxidasa dan lugar a las disfunciones características de la CGD. Los neutrófilos de los pacientes de CGD no logran producir una explosión oxidativa significativa después de la estimulación. Se han descrito diferentes formas de CGD (CGD clásica ligada a X y patrones autosómicos recesivos). La explosión oxidativa de los granulocitos se ve impedida en trasplantes, estadios tardíos de la infección por VIH y en los ancianos, haciendo que estas poblaciones sean más susceptibles a infecciones secundarias y a exacerbaciones de la enfermedad inflamatoria. Varios inmunomoduladores (por ejemplo, citocinas (GM-CSF, G-CSF, TNF) o fármacos) también parecen tener efectos en la explosión oxidativa. Existe el potencial de proteínas con la capacidad de regular positivamente o de regular negativamente la explosión oxidativa de un modo terapéutico para que sea útil para una serie de diferentes patologías.

**Métodos:** El ligando de proteína cinasa C, 12-miristato 13-acetato de forbol (PMA) puede utilizarse en este ensayo como agonista del proceso de explosión oxidativa. Se mezcla sangre completa heparinizada con dextrano estéril (concentración final del 0,6 %) durante 1 hora y se deja que se separe en capas. La capa inferior contiene neutrófilos, monocitos y glóbulos rojos sanguíneos. Se utiliza una etapa de lisis con cloruro de amonio para retirar todos los RBC y una población pura al 97 % de neutrófilos permaneciendo una contaminación por neutrófilos de aproximadamente un 3 % después de la etapa de lisis. Tras la estimulación, los granulocitos y monocitos producen metabolitos de oxígeno reactivos (anión superóxido, peróxido de hidrógeno, ácido hipocloroso) que destruyen a las bacterias dentro del fagosoma. La formación de los oxidantes reactivos durante la explosión oxidativa puede controlarse por la adición y oxidación de Amplex Red. Después, se analiza el porcentaje de células que han producido radicales de oxígeno reactivos así como su intensidad media de fluorescencia usando un lector de placas de fluorescencia. El transcurso de tiempo típico para esta reacción es de 10 minutos, observándose una explosión evidente a los 2 minutos y observándose una reducción de la señal a los 20 minutos. Este ensayo puede ejecutarse en modo agonista en ausencia de PMA o en modo antagonista, con la administración concomitante de polipéptidos de AARS y PMA a una concentración que se encuentra por debajo de la CE50 para este compuesto.

## REGULACIÓN DE LA PRODUCCIÓN DE ELASTASA DE NEUTRÓFILOS HUMANOS

**Antecedentes y relevancia terapéutica:** La elastasa de neutrófilos es una serina proteasa que se ha implicado como poseedora de un papel específico en el desarrollo de una gran variedad de enfermedades humanas, incluyendo trastornos inflamatorios del pulmón y el sistema cardiovascular. Aunque su papel fisiológico clave es en la defensa innata del hospedador, también puede participar en el remodelado de tejidos y posee acciones de secretagogo que en la actualidad se reconocen como importantes para las señales inflamatorias locales. La actividad de elastasa de neutrófilos se ha relacionado con el desarrollo del enfisema durante varias décadas, aunque hace relativamente poco se ha adscrito una función patógena a esta serina proteasa en situaciones donde se produce un exceso de deposición de matriz extracelular. El uso de modelos animales manipulados genéticamente está comenzando a descubrir vías potenciales en las que sus acciones podrían influenciar la reparación pulmonar fibrótica. Cada vez más pruebas sugieren que la implicación de las vías celulares con efectos más directos en la generación de mediador fibrogénico y la síntesis de colágeno parece remarcar las acciones de la elastasa de neutrófilos para promover la acumulación de matriz en los pulmones. La elastasa de neutrófilos humanos también está presente dentro de las placas ateroscleróticas, donde contribuye a la degradación de la matriz y al debilitamiento de la pared del vaso asociada con las complicaciones de la formación de aneurismas y la ruptura de la placa. En estas acciones está unida a otras proteasas extracelulares, pero la gran variedad de sustratos y la potencia de esta enzima, junto con la actividad asociada a la desgranulación de los neutrófilos hacen que esta proteasa disruptiva sea única como diana terapéutica en la enfermedad aterosclerótica.

**Métodos:** Este ensayo usa el kit de ensayo de elastasa ENZCHEK® (Invitrogen, n.º de catálogo E-12056). Los neutrófilos se preparan a partir de sangre humana fresca usando una solución de dextrano al 6 % y los glóbulos rojos sanguíneos se lisan antes de sembrar las células en medio RPMI (el medio no debe complementarse ni con suero, ni con antibióticos). Se prepara una solución madre a 1,0 mg/ml del sustrato de elastina DQ añadiendo 1,0 ml de agua desionizada (dH2O) directamente a uno de los tres viales que contienen el sustrato liofilizado y mezclando para disolver. Se prepara el tampón de reacción 1X diluyendo 6 ml del tampón de reacción 10X en 54 ml de dH2O. Se prepara una solución de trabajo a 100 µg/ml del sustrato de elastina DQ diluyendo la solución madre de elastina DQ en tampón de reacción 1X. Se prepara solución madre de elastasa pancreática porcina preparando una solución madre a 100 U/ml en dH2O. Para ensayar respecto de la actividad elastasa, se pipetea 50 µl de tampón de reacción 1X en cada pocillo de ensayo que contiene 500.000 neutrófilos/ml en un volumen de 30 µl. Se añaden 8 µl de cada polipéptido de AARS por pocillo y se incuba la muestra durante 20 minutos a 37 °C. Se añaden 50 µl de la solución de trabajo de elastina DQ a 100 µg/ml a cada pocillo y se mezcla. Las muestras se incuban a temperatura ambiente, protegidas de la luz, durante 30 minutos. Puede medirse la intensidad de fluorescencia en un lector de microplacas de fluorescencia equipado con filtros para fluorescencia de fluoresceína estándar (ex 485/em 535) en

múltiples instantes.

## UNIÓN A RECEPTORES DE TIPO TOLL Y ACTIVACIÓN DE NFKB (ENSAYOS D1-D4 EN LAS TABLAS DE DATOS MÁS ADELANTE)

5 **Antecedentes y relevancia terapéutica:** Los macrófagos son piezas clave del sistema inmunitario innato y expresan un vasto repertorio de diferentes clases de receptores de reconocimiento de patrones (PRR), incluyendo la familia de receptores de tipo Toll (TLR), que son potentes reguladores y controladores de la respuesta inmunitaria.

10 La estimulación de TLR por patógenos microbianos y ligandos endógenos inicia cascadas de señalización que inducen la secreción de citocinas proinflamatorias y citocinas efectoras que dirigen las respuestas inmunitarias adaptativas aguas abajo. Los ligandos endógenos, así como los componentes microbianos, son reconocidos por y pueden activar a los TLR, suscitando la posibilidad de que estos receptores puedan ser dianas críticas para el desarrollo de nuevas terapias para múltiples enfermedades.

15 Por consiguiente, los polipéptidos de AARS que modulan la actividad del receptor de TLR, tienen utilidad terapéutica en una gran variedad de enfermedades y trastornos que incluyen, por ejemplo, enfermedades y trastornos inflamatorios, enfermedades autoinmunitarias, trasplante de tejidos/rechazo de órganos, prevención o tratamiento del cáncer, la modulación de la hematopoyesis e infecciones.

### 20 **Medición de la activación de TLR en células RAW-BLUE**

Los macrófagos de ratón comercializados con la marca células RAW-BLUE™ (Invivogen, código de catálogo: raw-sp) expresan todos los TLR, salvo TLR5 e incluyen un gen de fosfatasa alcalina secretada (SEAP) que es inducible por los factores de transcripción NF-κB y AP-1. Tras la estimulación de TLR, las células RAW-BLUE™ activan a NF-κB y/o AP-1, dando lugar a la secreción de SEAP, que es medible cuando se usa medio de detección de SEAP.

25 **Métodos:** Las células RAW-BLUE™ se lavan dos veces con PBS, se tripsinizan y se resuspenden en medio de crecimiento reciente (medio de crecimiento: DMEM, 4,5 g/l de glucosa, suero bovino fetal inactivado por calor al 10 % (30 minutos a 56 °C), 100 mg/ml de ZEOCIN™, L-glutamina 2 mM). Las células se siembran a una concentración de 50.000 células/pocillo en una placa de 96 pocillos en un volumen total de 100 µl y los polipéptidos de AARS, los controles o los polipéptidos de AARS (+LPS) se añaden a cada pocillo a las concentraciones mostradas en los experimentos indicados a continuación. Las células se incuban a 37 °C en una incubadora con CO<sub>2</sub> al 5 % durante 18 horas. En el día 2 del experimento, se prepara el medio de detección de SEAP (QUANTI-BLUE™) (Invivogen, código de catálogo: rep-qb1) siguiendo las instrucciones y se añaden 120 µl por pocillos a una placa de fondo plano transparente de 96 pocillos y se añade sobrenadante de células (20 µl). Las muestras se incuban a 37 °C durante aproximadamente 30 minutos durante hasta 2 horas. Los niveles de SEAP se determinan usando un espectrofotómetro y leyendo la absorbancia a 650 nm.

40 Para detectar polipéptidos de AARS que bloquean específicamente la activación de TLR, este ensayo puede modificarse para identificar potenciales antagonistas de TLR. En este caso, se añaden polipéptidos de AARS a las células a una concentración final de aproximadamente 250nM por pocillo, (o como se especifique de otro modo en los ejemplos a continuación) 1 hora antes de añadir 50 ng/ml de LPS. Las células se incuban y se detecta SEAP como se ha descrito anteriormente. Se usan pocillos de control de PBS sin LPS o solo polipéptido de AARS añadido para encontrar el nivel basal de estimulación de TLR en el momento de la medición. Se pretratan los pocillos de control con PBS y agonistas y antagonistas de TLR conocidos. Se usa la proporción de fondo restado [PBS más señal de LPS] a [polipéptido de AARS más señal de LPS] para determinar el porcentaje de antagonismo.

### 50 **Exploración de TLR humano en células Hek293**

Se modificaron genéticamente células HEK293 humanas comercializadas con la marca células HEK-BLUE™ TLR (Invivogen). Las versiones de TLR2 y TLR4 de este tipo celular expresan de manera selectiva todas TLR2 o TLR4 e incluyen un gen de fosfatasa alcalina embrionaria secretada (SEAP) bajo el control de un promotor mínimo de IFN-beta que está fusionado a cinco sitios de unión de factor de transcripción NF-κB y AP-1. Usando agonistas específicos de TLR 2 o 4 (respectivamente), las células Hek-Blue™ TLR2 y HEK-BLUE™ TLR4 activan a NF-κB y/o AP-1, dando lugar a la secreción de SEAP, que es medible cuando se usa reactivo de detección de SEAP. Las células Hek-Blue™ TLR2 se cotransfectan con la proteína correceptora de LPS, CD14, para potenciar la capacidad de respuesta de TLR2 y mejorar la calidad de la señal. La célula progenitora expresa niveles endógenos de TLR1, 3, 5, 6 y también NOD1.

60 **Métodos:** Las células HEK-BLUE™ -TLR2 o HEK-BLUE™ -TLR4 se lavan dos veces con PBS, se tripsinizan y se resuspenden en medio de crecimiento reciente (medio de crecimiento: DMEM, 4,5 g/l de glucosa, suero bovino fetal inactivado por calor al 10 % (30 minutos a 56 °C), 100 mg/ml de ZEOCIN™, L-glutamina 2 mM). Las células se siembran a una concentración de 50.000 células/pocillo en una placa de 96 pocillos en un volumen total de 100 µl y los polipéptidos de AARS, los controles o los polipéptidos de AARS (+LPS) se añaden a cada pocillo a las concentraciones mostradas en los experimentos indicados a continuación. Las células se incuban a 37 °C en una

incubadora con CO<sub>2</sub> al 5 % durante 18 horas. En el día 2 del experimento, se prepara el medio de detección de SEAP (QUANTI-BLUE™) (Invivogen, código de catálogo: rep-qb1) siguiendo las instrucciones y se añaden 120 µl por pocillos a una placa de fondo plano transparente de 96 pocillos y se añade sobrenadante de células (20 µl). Las muestras se incuban a 37 °C durante aproximadamente 30 minutos durante hasta 2 horas. Los niveles de SEAP se determinan usando un espectrofotómetro y leyendo la absorbancia a 650 nm. Los pocillos de control se pretratan con PBS y agonistas de TLR conocidos tales como LPS ultrapuro (TLR-4) o PAM3CSK4 (TLR-2). Se usa la proporción de fondo restado [PBS más señal de LPS] a [polipéptido de AARS más señal de LPS] para determinar el porcentaje de agonismo.

## 10 LIBERACIÓN DE CITOCINAS (ENSAYOS E1-E16 EN LAS TABLAS DE DATOS MÁS ADELANTE)

**Antecedentes y relevancia terapéutica:** Las citocinas son un conjunto diverso de moléculas proteínicas de señalización celular pequeñas que se usan profusamente para la comunicación intercelular y desempeñan papeles significativos en la homeostasia corporal normal, incluyendo la modulación y regulación inmunitaria. Por consiguiente, los polipéptidos de AARS que modulan la liberación o las actividades biológicas de las citocinas, tienen utilidad terapéutica en una gran variedad de enfermedades y trastornos que incluyen, por ejemplo, enfermedades y trastornos inflamatorios, enfermedades autoinmunitarias, trasplante de tejidos/rechazo de órganos, prevención o tratamiento del cáncer, la modulación de la hematopoyesis e infecciones.

### 20 *Liberación de citocinas de células en cultivo*

**Métodos:** Se siembran células de ensayo en una placa de 24 pocillos a una densidad de aproximadamente 1 millón de células/pocillo en 1 ml de medio de crecimiento. Las células se tratan con polipéptido de ARS (a las concentraciones mostradas en los ejemplos más adelante) o en un volumen igual de PBS y se incuban durante una noche a 37 °C con CO<sub>2</sub> al 5 %. Después del tratamiento de las células, las muestras se centrifugan a 4 °C en una centrifugadora de cubeta oscilante a 2.000 x g durante 5 minutos. El medio se retira cuidadosamente para no alterar el sedimento de células y se transfiere a un nuevo tubo. Las muestras se ensayan inmediatamente o se conservan instantáneamente en nitrógeno líquido para su posterior análisis. La liberación de citocinas (incluyendo las citocinas MIF, IL-8, IL-10, Serpina E1, GM-CSF, GRO, IL-1 alfa, IL-1beta, IL-1ra, IL-6, MCP-1, MIP-1, RANTES y TNFalfa) se determina usando kits disponibles comercialmente (R&D Systems, Inc, MN, EE.UU.) o mediante una organización de investigación subcontratada (MD Biosciences (St. Paul, MN).

### *Liberación de citocinas de sangre humana completa*

**Métodos:** Se obtiene sangre humana completa de donantes humanos normales y se recoge con heparina en tubos de recogida estándar. La sangre se usa el mismo día que se recoge para asegurar una salud celular adecuada. La sangre se mezcla cuidadosamente y se siembra en un volumen de 100 µl en placas de fondo en V de policarbonato de 96 pocillos. Los polipéptidos de AARS se añaden y se mezclan lentamente en 2X de sangre usando una pipeta multicanal ajustada a 50 µl. Se usan puntas de filtro para todos los experimentos y se emplea PPE completa. Todos los experimentos se producen en una campana de bioseguridad dedicada que es adecuada para la experimentación con sangre humana. La sangre se incuba durante una noche a 37 °C con CO<sub>2</sub> al 5 %. Después del tratamiento de las células, las muestras se centrifugan en una centrifugadora de cubeta oscilante a 2.000 x g durante 5 minutos. El sobrenadante se recoge para los ELISA de citocinas. Los ELISA se llevan a cabo del modo descrito anteriormente.

### 45 *Liberación de citocinas de PBMC*

**Métodos:** Para aislar células mononucleares de sangre periférica, se dispone en capas sangre completa humana recientemente aislada sobre HISTOPAQUE®-1077 de Sigma a una proporción de 1:1 en tubos cónicos de 50 ml a temperatura ambiente. Las muestras dispuestas en capas se centrifugan a 400 x g en una centrifugadora clínica de cubeta oscilante durante 30 minutos a temperatura ambiente sin freno. La capa celular blanca en la interfaz entre el plasma y el gradiente de densidad se retira posteriormente con una pipeta. Estas células mononucleares de sangre periférica se lavan dos veces con RPMI-1640 (Invitrogen, n.º 22400-105) mediante dilución y centrifugación durante 10 minutos a 250 x g. Las PBMC lavadas se resuspendieron en RPMI-1640 + FBS al 10 % y se emplacaron a razón de 1x10<sup>6</sup> células/ml.

### *Liberación de citocinas de sinoviocitos humanos*

**Antecedentes y relevancia terapéutica:** Un gran número de estudios ha demostrado que la IL-6 y la IL-8 se sobreproducen en varias enfermedades y por lo tanto, pueden desempeñar un papel fundamental en la patogénesis de enfermedades inflamatorias. La IL-6 activa la producción de células endoteliales, dando lugar a la liberación de IL-8 y de proteína quimioatrayente de monocitos, la expresión de moléculas de adhesión y el reclutamiento de leucocitos a los sitios de inflamación. Estas citocinas se expresan en tipos celulares asociados con enfermedades inflamatorias, incluyendo células implicadas en la patogénesis de la artritis juvenil sistémica, el lupus eritematoso sistémico, la enfermedad de Crohn y la artritis reumatoide. Una de las acciones sistémicas más importantes de la producción de citocinas es la inducción de la respuesta de fase aguda. Las proteínas de fase aguda se producen principalmente por el hígado e incluyen proteínas que promueven la respuesta inmunitaria mediante la activación del

complemento, la inducción de citocinas proinflamatorias y la estimulación de la quimiotaxis de neutrófilos. Como alternativa, la respuesta de fase aguda puede ser útil y las proteínas de fase aguda, tales como antagonistas de proteinasas, opsoninas y procoagulantes, ayudar a limitar la destrucción de tejidos aliviando la inflamación. En particular, la IL-6 puede estimular la proliferación de sinoviocitos y la activación de osteoclastos, dando lugar a la formación y reparación del pannus sinovial. La IL-6 actúa con la IL-1 para aumentar la producción de metaloproteinasas de matriz, que pueden contribuir a la destrucción de articulaciones y cartílago. Sin embargo, la IL-6 también puede tener efectos protectores en las articulaciones, tal como se sugiere por el hallazgo de que esta citocina induce la expresión del inhibidor tisular de metaloproteinasas y estimula la síntesis de proteoglicano cuando se inyecta en las articulaciones de ratones con artritis inducida por antígeno. Los sinoviocitos similares a fibroblastos humanos de artritis reumatoide (HFLS-RA) se aíslan de los tejidos sinoviales obtenidos de pacientes con artritis reumatoide (RA). Estas se criopreservan en el segundo pase y pueden cultivarse y propagarse hasta al menos 5 duplicaciones de la población. Las HFLS se conocen desde hace tiempo debido a su papel en la destrucción de las articulaciones mediante la producción de citocinas y metaloproteinasas que contribuyen a la degradación del cartílago.

Por consiguiente, los polipéptidos de AARS con la capacidad de modular el crecimiento, la diferenciación o el perfil de liberación de citocinas de los sinoviocitos similares a fibroblastos de artritis reumatoide (HFLS-RA) tiene utilidad terapéutica en una gran variedad de enfermedades que incluyen, por ejemplo, el tratamiento de enfermedades y trastornos inflamatorios que incluyen la artritis juvenil sistémica, el lupus eritematoso sistémico, la enfermedad de Crohn y la artritis reumatoide.

**Métodos:** HFLS-RA, células adultas (Cell Applications, n.º de cat. 408RA-05a) se mantienen en medio de crecimiento de sinoviocitos (Cell Applications, n.º de cat. 415-50) en 15 ml de medio en matraces de 125 ml durante 1 pase antes de su uso. Las células se mantienen a 37 °C, CO<sub>2</sub> al 5 %, en un ambiente humidificado y se utilizaron en campanas certificadas para cultivo tisular BSL2 usando técnicas estériles y equipamiento de protección personal adecuado que incluye gafas, guantes y batas de laboratorio. Se sembró un volumen de 80 µl de células durante una noche en medio de crecimiento completo a una densidad de aproximadamente 50.000 células/ml. Se añaden los polipéptidos de AARS a una concentración final de 250 nM por pocillo (o como se indique de otro modo en los ejemplos más adelante) en PBS estéril a cada pocillo después de la adherencia durante una noche. Los pocillos de control contenían células no tratadas y se incubaron con un volumen equivalente de PBS. Las células se expusieron a las proteínas o PBS en medio basal (Cell Applications, n.º de cat. 310-470) durante 24 horas. Se retira el sobrenadante y se efectúan los ensayos ELISA para IL-8, IL-6 y TNFα según las instrucciones del fabricante (R&D Systems, n.º de cat. DY206 y DY-208, DY-210 Duo-set kits). La proliferación se evalúa con resazurina como se ha descrito anteriormente añadiendo medio reciente que contiene resazurina a las placas seguido de la retirada del sobrenadante e incubando durante tres horas a 37 °C. Las placas se leen en un lector de placas de fluorescencia y se expresa la viabilidad/proliferación en función de la fluorescencia asociada a resorufina de los pocillos tratados con polipéptido de AARS dividida entre la fluorescencia asociada a resorufina de los pocillos tratados solo con PBS.

#### **Proliferación de astrocitos humanos y producción de citocinas inflamatorias**

**Antecedentes y relevancia terapéutica:** Los astrocitos humanos (HA) se obtienen de la corteza cerebral humana. Estos se criopreservan en el segundo pase y pueden cultivarse y propagarse hasta 10 duplicaciones de la población. Los HA son las células más abundantes del sistema nervioso central y llevan a cabo muchas funciones, tales como proporcionar soporte mecánico y nutrientes a las neuronas y retirar los desechos de las neuronas. Además de desempeñar un papel crítico de soporte para un funcionamiento neuronal óptimo, también proporcionan soporte bioquímico para las células endoteliales que forman la barrera hematoencefálica. Estudios recientes han demostrado que los astrocitos son capaces de regular la neurogénesis instruyendo a las células madre para que adopten un destino celular y controlar la función de las sinapsis sencillas, participan activamente en la transferencia y el almacenamiento de información en el cerebro. El reconocimiento de la importancia de los astrocitos en el funcionamiento del sistema nervioso va en aumento, los HA pueden servir como modelo *in vitro* útil para explorar la diversidad de funciones de los astrocitos. Se ha demostrado que los astrocitos proliferan en respuesta a la IL6 y el TNFα. Además, estas células son capaces de producir sus propios IL6 y TNFα. Por lo tanto, los polipéptidos de AARS que modulan la proliferación y la producción de citocinas en los HA tienen utilidad terapéutica en diversas enfermedades neurológicas, incluyendo la inflamación neurológica, la neurodegeneración, la tumorigénesis del cerebro y la isquemia y reparación cerebral.

**Métodos:** Se mantienen astrocitos humanos (HA) de Cell Applications (n.º de cat. 882K-05f) en medio de crecimiento celular de HA de Cell Applications (n.º de cat. 821-500) según las instrucciones del fabricante. Las células se mantienen a 37 °C, CO<sub>2</sub> al 5 %, en un ambiente humidificado y se utilizaron en campanas certificadas para cultivo tisular BSL2 usando técnicas estériles y equipamiento de protección personal adecuado que incluye gafas, guantes y batas de laboratorio. Se sembró un volumen de 80 µl de células sobre placas recubiertas con colágeno durante una noche en medio completo (anterior) a una densidad de 50.000 células/ml. Las células se lavan una vez con PBS y se añaden 80 µl de medio de crecimiento sin suero a cada pocillo. Se añaden los polipéptidos de AARS a una concentración final de 250 nM por pocillo (o como se describa de otro modo en los ejemplos más adelante) en un volumen consistente en PBS estéril a cada pocillo. Las células se exponen a los polipéptidos de AARS durante 48 horas y se retira el medio gastado para la evaluación de citocinas (como se ha descrito

anteriormente). Las células se expusieron a las proteínas o PBS en medio basal (Cell Applications, n.º de cat. 310-470) durante 48 horas. Se retira el sobrenadante y se efectúan los ensayos ELISA para IL-8 e IL-6 según las instrucciones del fabricante (R&D Systems, n.º de cat. DY206 y DY-208, DY-210 Duo-set kits). La proliferación se evalúa con resazurina como se ha descrito anteriormente añadiendo medio reciente que contiene resazurina a las placas seguido de la retirada del sobrenadante e incubando durante tres horas a 37 °C. Las placas se leen en un lector de placas de fluorescencia y se expresa la viabilidad/proliferación en función de la fluorescencia asociada a resorufina de los pocillos tratados con polipéptido de AARS dividida entre la fluorescencia asociada a resorufina de los pocillos tratados solo con PBS.

#### 10 **Proliferación de células endoteliales microvasculares pulmonares humanas (HLMVEC) y producción de citocinas inflamatorias**

**Antecedentes y relevancia terapéutica:** La vasculatura pulmonar tiene una significación fisiológica/patológica enorme. En la actualidad se reconoce que es un tejido compuesto de células metabólicamente activas, que responden funcionalmente, que interactúan con sustratos en circulación y elementos formados de maneras que regulan la composición de la sangre arterial sistémica, afectan a las funciones de órganos diana y contribuyen a la trombosis, la hemostasia y las reacciones inmunitarias, así como a la metástasis tumoral. Las células endoteliales microvasculares pulmonares humanas (HLMVEC) muestran una expresión elevada de citocinas quimioatrayentes y de moléculas de adhesión celular que proporcionan señales críticas para la migración dirigida de leucocitos al pulmón durante una lesión pulmonar aguda. Este tipo celular primario puede una herramienta útil para estudiar varios aspectos de la patología y biología de la microvasculatura pulmonar *in vitro*. Se cree que la alteración de la estructura y la función de la microvasculatura en respuesta a estímulos inflamatorios es un factor clave en el daño orgánico y en las condiciones adecuadas, puede proporcionar un estímulo para la reparación. Una causa significativa de estas alteraciones vasculares es la inducción de una reacción inflamatoria que implica la infiltración de leucocitos. Una serie de estudios centrados en la adhesión de granulocitos al endotelio ha revelado que el reclutamiento y la emigración de leucocitos implica una cascada de adhesión bien orquestada. La cascada de adhesión comienza cuando el granulocito se une al endotelio y comienza a rodar en dirección del flujo de fluido a una baja velocidad. A medida que rueda el granulocito, se activa, posteriormente se adhiere firmemente al endotelio y migra a través del endotelio al espacio extravascular. Estos eventos de adhesión están mediados, en parte, por interacciones moleculares que se producen entre CAM sobre la superficie de los granulocitos y glucoproteínas afines presentes sobre el endotelio. Una serie de estudios ha revelado que la molécula de adhesión celular endotelial, E-selectina, puede interactuar con los ligandos de granulocitos presentadores de glucano de tipo SLex para mediar las etapas de unión y rodadura de la cascada de adhesión. Las etapas aguas abajo de la cascada implican la interacción de la molécula de adhesión intercelular expresada endotelial con las integrinas CD 18 expresadas por granulocitos.

Por lo tanto, los polipéptidos de AARS que modulan la proliferación y/o la producción de citocinas de células endoteliales microvasculares pulmonares humanas tienen utilidad terapéutica en una serie de enfermedades vasculares y pulmonares que incluyen las enfermedades pulmonares inflamatorias y obstructivas, incluyendo, por ejemplo, hipertensión pulmonar, enfermedad pulmonar obstructiva crónica, fibrosis pulmonar idiopática y asma.

**Métodos:** Las HLMVEC (Cell Applications, n.º de catálogo 540-05) se mantienen en medio de crecimiento de células endoteliales de Cell Applications (n.º de cat. 111-500). Para un crecimiento adecuado, se usa una solución de factor de unión que contiene colágeno (Cell Applications, n.º de catálogo 123-100), para recubrir las placas y los matraces antes de sembrar las células. Las células se mantienen a 37 °C, CO<sub>2</sub> al 5 %, en un ambiente humidificado y se utilizaron en campanas certificadas para cultivo tisular BSL2 usando técnicas estériles y equipamiento de protección personal adecuado que incluye gafas, guantes y batas de laboratorio. Se sembró un volumen de 80 µl de células sobre placas recubiertas con colágeno durante una noche en medio completo (anterior) a una densidad de 50.000 células/ml. Las células se lavan una vez con PBS y se añaden 80 µl de medio de crecimiento sin suero a cada pocillo. Se añaden los polipéptidos de AARS a una concentración final de 250 nM por pocillo (o como se describa de otro modo en los ejemplos más adelante) en un volumen consistente en PBS estéril a cada pocillo. Las células se exponen a los polipéptidos de AARS durante 48 horas y se retira el medio gastado para el ELISA para moléculas de adhesión celular y la evaluación de citocinas (como se ha descrito anteriormente). Las moléculas de adhesión celular que incluyen VCAM soluble y/o ICAM se miden usando un kit de ELISA estándar de R&D Systems (n.º de cat. DY643 y DY720, respectivamente). La proliferación se evalúa con resazurina como se ha descrito anteriormente añadiendo medio reciente que contiene resazurina a las placas seguido de la retirada del sobrenadante e incubando durante tres horas a 37 °C. Las placas se leen en un lector de placas de fluorescencia y se expresa la viabilidad/proliferación en función de la fluorescencia asociada a resorufina de los pocillos tratados con polipéptido de AARS dividida entre la fluorescencia asociada a resorufina de los pocillos tratados solo con PBS.

#### 60 **ADHESIÓN CELULAR (ENSAYOS F1-F7 EN LAS TABLAS DE DATOS MÁS ADELANTE)**

**Antecedentes y relevancia terapéutica:** Las moléculas de adhesión celular (CAM) son proteínas ubicadas sobre la superficie celular que están implicadas en la unión a otras células o con la matriz extracelular (ECM) en el proceso denominado adhesión celular. Estas proteínas son normalmente receptores transmembrana y están compuestas de tres dominios: un dominio intracelular que interactúa con el citoesqueleto, un dominio transmembrana y un dominio

extracelular que interactúa cono tras CAM del mismo tipo (unión homófila) o con otras CAM o la matriz extracelular (unión heterófila). La mayoría de las CAM pertenecen a cuatro familias de proteínas: superfamilia de Ig (inmunoglobulina) (CAM de IgSF), las integrinas, las cadherinas y las selectinas. Las moléculas de adhesión de la superfamilia de inmunoglobulina (IgSF) son glucoproteínas transmembrana independientes de calcio, que incluyen:

- 5 moléculas de adhesión de células neurales (NCAM), moléculas de adhesión celular intercelulares (ICAM), moléculas de adhesión de células vasculares (VCAM), molécula de adhesión de plaquetas-células endoteliales (PECAM-1), molécula de adhesión selectiva de células endoteliales (ESAM), molécula de adhesión de unión (JAM), nectinas y otras moléculas de adhesión celular.
- 10 Las moléculas de adhesión celular son glucoproteínas de la superficie que son críticas para la adhesión de los leucocitos al endotelio sinusoidal y la transmigración y citotoxicidad en una serie de enfermedades hepáticas inflamatorias. ICAM-1 desempeña un importante papel en la inflamación y la expresión aumentada de ICAM-1 en células endoteliales se refleja en la activación de las células endoteliales. ICAM-1 tiene una importancia clave debido a que media la adhesión endotelial firme y facilita la transmigración de leucocitos. Los estudios han demostrado que
- 15 se produce una regulación positiva de ICAM-1 tanto en células sinusoidales como en hepatocitos en afecciones inflamatorias hepáticas, tales como la infección vírica por la hepatitis B, trastornos hepáticos autoinmunitarios, hepatitis alcohólica y rechazo de aloinjerto de hígado.

20 Por lo tanto, los polipéptidos de AARS que modulan la producción de moléculas de adhesión celular y la adhesión celular a células endoteliales tienen utilidad terapéutica en diversas enfermedades inflamatorias que incluyen, por ejemplo, enfermedades cardiovasculares, aterosclerosis, autoinmunidad e hipertensión pulmonar.

**Métodos:** Las células de la vena umbilical humana (ATCC, n.º de cat. CRL-2873) (HUVEC) se siembran a una concentración de aproximadamente  $1,2 \times 10^5$  células/pocillo en placas de 12 pocillos recubiertas con solución de unión de fibronectina humana en el medio y con los complementos sugeridos por la ATCC y se cultivaron de acuerdo con las instrucciones del fabricante. Las células se estimularon con polipéptidos de AARS a sus concentraciones indicadas o solo PBS y se incubaron durante una noche en medio de crecimiento. Las células de leucemia monocítica aguda humana (THP-1 (TIB-202)), se resuspenden en medio de BSA al 0,1 %/RPMI libre de suero con calceína AM (6 µl/ml; Invitrogen, n.º de cat. C1430) y se incubaron durante 30 minutos. Las células marcadas se recogen y resuspenden en medio RPMI que contiene FBS al 10 % y se ajusta la densidad a  $2 \times 10^6$  células/ml.

30 Se colocan 100µl ( $2 \times 10^5$ ) de células THP-1 marcadas en cada pocillo de la monocapa de HUVEC e 1 ml de medio de crecimiento y se incubaron durante 15 minutos. Las células se lavaron dos veces con PBS para retirar las células no unidas y después se leyeron las células mediante un lector de placas de fluorescencia con una longitud de onda de excitación de 488 nm y una longitud de onda de emisión de 530 nm.

## DIFERENCIACIÓN CELULAR (ENSAYOS G1-G4 EN LAS TABLAS DE DATOS MÁS ADELANTE)

### *Diferenciación y proliferación de adipocitos en células preadipocitarias humanas.*

40 **Antecedentes y relevancia terapéutica:** Tanto la obesidad como la lipodistrofia se asocian comúnmente con patologías que incluyen la diabetes y enfermedades cardiovasculares. En la actualidad se reconoce que el tejido adiposo es un órgano endocrino que secreta una gran variedad de factores y su secreción desregulada afecta a la adipogénesis así como a la homeostasia de la glucosa/insulina en todo el organismo. El exceso de tejido adiposo que da lugar a la obesidad se ha convertido en una grave amenaza para la salud pública. El desarrollo del tejido adiposo puede verse afectado por el origen genético, el equilibrio hormonal, la dieta y la actividad física. La masa de tejido adiposo puede aumentar cuando las células grasas aumentan de tamaño debido a la acumulación de triacilglicerol. Además, también puede producirse un aumento en el número de células grasas, que surgen de la diferenciación de células precursoras en adipocitos, incluso en adultos, según se observa en la obesidad humana

45 grave y en roedores alimentados con una dieta alta en carbohidratos o diente alta en grasas. Se cree que los adipocitos surgen específicamente de células mesenquimales que sufren el proceso de especialización y diferenciación, la adipogénesis. Las líneas celulares de preadipocitos pueden sufrir diferenciación en adipocitos tras su tratamiento con agentes adipogénicos formados por glucocorticoide sintético, dexametasona (DEX), isobutilmetilxantina (IBMX), e insulina, que han sido valiosos en estos estudios. Se ha determinado de manera firme

50 que la familia de factores de transcripción del receptor y activado por proliferador de peroxisomas (PPAR $\gamma$ ) y la proteína de unión potenciadora de CCAAT (C/EBP) desempeña papeles críticos en la diferenciación de los adipocitos. De manera temprana durante la diferenciación de los adipocitos, C/EBP $\beta$  y C/EBP $\delta$  se inducen por DEX e IBMX, respectivamente, que después conjuntamente inducen a PPAR $\gamma$  y C/EBP $\alpha$  para que activen diversos marcadores de adipocitos que son necesarios para la función de los adipocitos. También se ha comunicado que

60 otros factores de transcripción regulan positiva o negativamente la adipogénesis y varios factores de crecimiento y hormonas pueden afectar a la diferenciación de adipocitos regulando la expresión de factores de transcripción adipogénicos. De hecho, además de ser el principal sitio para el almacenamiento de energía en los mamíferos mediante el almacenamiento de triacilglicerol y la liberación de ácidos grasos en los momentos de necesidad, el tejido adiposo secreta una gran variedad de moléculas que están implicadas en diversos procesos fisiológicos, incluyendo la respuesta inmunitaria, la función vascular y la homeostasia de la energía. Las citocinas tales como el

65 TNF- $\alpha$  y la IL-6 se secretan por los adipocitos. Algunos de estos factores también pueden afectar al crecimiento y

desarrollo del tejido adiposo mediante una acción autocrina/paracrina.

Por lo tanto, los polipéptidos de AARS que tengan la capacidad de modular la diferenciación y/o la proliferación de preadipocitos humanos normales tienen utilidad terapéutica en una gran variedad de enfermedades que incluye, por ejemplo, el tratamiento y la prevención de la enfermedad metabólica, enfermedades cardiovasculares, obesidad y lipodistrofias, así como las complicaciones a largo plazo de la diabetes.

**Métodos:** Se mantiene HPAd (preadipocitos humano) Cell Applications, n.º de cat. 803sD) de acuerdo con las instrucciones del proveedor. Para el cultivo, se descongelan rápidamente las células y se transfieren inmediatamente a 15ml de medio de crecimiento de adipocitos (Cell Application, n.º de cat. 811M-250) y se sembraron en un matraz tratado para cultivo tisular estéril convencional. El medio se reemplaza con medio de crecimiento de adipocitos reciente cada dos días hasta que las células alcanzan una confluencia > 60 %. Las células se cultivan a 37 °C, CO<sub>2</sub> al 5 %, en un ambiente humidificado y se utilizaron en campanas certificadas para cultivo tisular BSL2 usando técnicas estériles y equipamiento de protección personal adecuado que incluye gafas, guantes y batas de laboratorio. Las células se siembran en placas de ensayo tratadas para cultivo tisular de 96 pocillos de paredes negras y fondo redondo para su diferenciación a una concentración de aproximadamente 50.000 células/ml. Los polipéptidos de AARS a una concentración final de 250nM por pocillo (o como se indique de otro modo en los ejemplos más adelante) se añaden a cada pocillo de ensayo. Todas las células se mantienen en medio de crecimiento durante 2 días con la excepción de los controles positivos, que se estimulan con medio de diferenciación adipogénico (Cell Applications, n.º de cat. 811D-250). Las células se exponen a polipéptidos de AARS durante 48 horas. Las moléculas de adhesión celular que incluyen VCAM soluble y/o ICAM se miden usando un kit de ELISA estándar de R&D Systems (n.º de cat. DY643 y DY720, respectivamente). La proliferación se evalúa con resazurina como se ha descrito anteriormente añadiendo medio reciente que contiene resazurina a las placas seguido de la retirada del sobrenadante e incubando durante tres horas a 37 °C. Las placas se leen en un lector de placas de fluorescencia y se expresa la viabilidad/proliferación en función de la fluorescencia asociada a resorufina de los pocillos tratados con polipéptido de AARS dividida entre la fluorescencia asociada a resorufina de los pocillos tratados solo con PBS. Se añade medio fresco y se mantiene la diferenciación durante 16 días después del cambio inicial de medio, cambiándose por medio fresco cada dos días para mantener la salud celular. En el día 15, se colocaron las células en medio libre de suero. En el día 16, se evaluó la diferenciación en adipocitos maduros con tinción de rojo Nilo (Invivogen, concentración final de 3 µM) se cuantificó con un lector de placas de fluorescencia con las longitudes de onda adecuadas. Para llevar a cabo este ensayo, las células se fijaron con paraformaldehído al 10 %, se lavaron en PBS y se permeabilizaron en PBS que contenía BSA al 0,5 % y Triton X-100 al 0,1 %. La proliferación celular se evalúa con una medición de la intensidad en un lector de fluorescencia con colorante Hoechst 33432 a una concentración final de 1µg/ml, como se ha descrito anteriormente. La adipogénesis se expresa como la intensidad de la señal de rojo Nilo. La señal del colorante Hoechst se usa para evaluar el número de células.

#### ***Diferenciación y proliferación de células musculares esqueléticas humanas.***

**Antecedentes y relevancia terapéutica:** El desarrollo del músculo esquelético es un proceso en múltiples etapas que implica la determinación de las células mesodérmicas pluripotentes para dar lugar a mioblastos, la retirada de los mioblastos del ciclo celular y la su diferenciación en células musculares y finalmente, el crecimiento y la maduración de las fibras musculares esqueléticas. La diferenciación del músculo esquelético implica el alineamiento de mioblastos, la elongación y la fusión en miotubos multinucleados, junto con la inducción de genes reguladores y estructurales específicos del músculo. A nivel molecular, la diferenciación miogénica y la expresión génica específica del músculo implican a la familia de proteínas MyoD de hélice-bucle-hélice (bHLH) específica de músculos esqueléticos, que incluye MyoD, miogenina, myf-5 y MRF4 y el factor de unión 2 potenciador de miocitos (MEF2). La actividad de unión a ADN de la familia de proteínas MyoD se atenúa por Id, que forma complejos con productos génicos de E2a en células en proliferación y se regula negativamente cuando se las induce para que se diferencien. La decisión de diferenciarse en miotubos está influenciada negativamente por varios factores. El tratamiento de mioblastos con suero bovino fetal, factor de crecimiento de fibroblastos básico 2 o factor de crecimiento transformante β1 es conocido por inhibir la diferenciación de los mioblastos. La miogénesis también está regulada negativamente por oncogenes tales como c-myc, c-jun, c-fos, H-ras y E1a. Hay muy poca información relativa a la señalización que se desencadena en el mioblasto tras la privación de suero, lo que da lugar a la inducción de la expresión de genes de la familia de MyoD y a la diferenciación muscular. La diferenciación miogénica parece depender de la activación de integrinas presentes en la membrana plasmática de los mioblastos, lo que sugiere la operación de una vía bioquímica "de fuera hacia dentro" en la que la integrina es la especie molecular aguas arriba. Las interacciones del factor de crecimiento insulínico (IGF)-I y -II con sus receptores también son reguladores positivos de la diferenciación del músculo esquelético.

Por consiguiente, los polipéptidos de AARS con la capacidad de modular el desarrollo muscular tienen utilidad terapéutica en una gran variedad de enfermedades que incluye, por ejemplo, el tratamiento de enfermedades metabólicas, caquexia, varias afecciones de emaciación muscular, así como enfermedades musculoesqueléticas, donde la atrofia muscular desempeña un papel en la patogénesis y la sintomatología. Las células musculares esqueléticas humanas (HSkMC) pueden sufrir diferenciación para mostrar miofilamentos de actina y miosina. Las HSkMC se han usado en el estudio de enfermedades musculares genéticas, tales como la hipertermia maligna. Las HSkMC también tienen el potencial de actuar como injerto cardíaco, reparando el daño en el corazón y por lo tanto,

los polipéptidos de AARS con la capacidad de modular el desarrollo muscular también tienen utilidad como reguladores *in vitro* e *in vivo* de la miogénesis.

**Métodos:** Para evaluar el papel potencial de los polipéptidos de AARS en este proceso, se empleó un ensayo estándar de diferenciación de células musculares esqueléticas. Para este ensayo, las células musculares esqueléticas adultas humanas (HskMC, Cell Applications, n.º de cat. 150-05f) se aíslan de donantes humanos sanos de músculos esqueléticos de las extremidades. Las células se mantienen en medio de crecimiento de HskMC (Cell Applications, n.º de cat. 151-500). Estas células pueden cultivarse y propagarse durante al menos 15 duplicaciones de la población. Para la diferenciación, las células se mantienen en medio de crecimiento durante un pase y después se siembran a 50.000 células por ml de medio en placas tratadas para TC de paredes negras y fondo transparente de 96 pocillos tratadas con colágeno a 100 µl por pocillo. Se dejan adherir las células durante una noche. Los polipéptidos de AARS en PBS o solo PBS, se añaden a cada pocillo a una concentración final de 250nM de proteína (o como se indique de otro modo en los ejemplos más adelante). Los pocillos de control recibieron el mismo volumen de medio de diferenciación (Cell Applications, n.º de cat. 151D-250) en ese instante. Las células se incuban con proteína o medio de diferenciación durante 48 horas. A las 48 horas, se recoge el sobrenadante de cultivo de todos los pocillos y se añade medio de diferenciación a un volumen de 150 µl para toda la placa con la excepción de los pocillos de control, que se mantuvieron solo en medio de crecimiento. El sobrenadante se utilizó para evaluar la producción de citocinas incluyendo IL6 e IL8 como se ha descrito anteriormente. La proliferación se evalúa con resazurina como se ha descrito anteriormente añadiendo medio fresco que contenía resazurina a las placas después de la retirada del sobrenadante y de incubar durante tres horas a 37 °C. Las células se controlan bajo el microscopio y se intercambia el medio por medio de diferenciación fresco cada 2 días. En el día 10, se retira el medio y se fijan las células con paraformaldehído al 10 % durante 30 minutos. Las células se permeabilizan con triton X-100 al 0,1 % en PBS durante 15 minutos y se tiñen las células con faloidina marcada con TR y Hoechst 33432 (como se ha descrito anteriormente) para definir la actina y los núcleos, respectivamente. La intensidad nuclear se usa para determinar la proliferación celular en cada pocillo y la intensidad de la faloidina se usa para determinar el contenido total de actina. Las células se tiñen también con anticuerpo para alfa actina de músculo esquelético (GenTex, n.º de cat. GTX101362). Se efectúan fotografías digitales usando un microscopio de fluorescencia así como inspecciones visuales y puntuaciones de todos los pocillos.

### 30 **Diferenciación y proliferación de células madre mesenquimales de médula ósea humana.**

**Antecedentes y relevancia terapéutica:** Las células madre mesenquimales (MSC) son células madre multipotentes que pueden diferenciarse en una serie de tipos celulares, incluyendo osteoblastos, condrocitos, miocitos, adipocitos, células de islotes pancreáticos beta y potencialmente, células neuronales. Muchos eventos diferentes contribuyen a la diferenciación de las MSC en otros linajes que incluyen la coordinación de una red compleja de factores de transcripción, cofactores e intermedios de señalización de numerosas vías. Las MSC tienen un gran interés terapéutico debido a que representan una población de células con potencial para tratar una gran variedad de enfermedades agudas y degenerativas.

Además, los polipéptidos de AARS con la capacidad de modular la diferenciación de las MSC en diferentes vías de desarrollo tienen una utilidad terapéutica significativa para posibilitar la modulación *in vitro* o *in vivo* de la hematopoyesis, neurogénesis, miogénesis, osteogénesis y adipogénesis, así como en una gran variedad de trastornos y enfermedades, que incluyen, por ejemplo, respuestas inflamatorias, autoinmunidad, cáncer, degeneración neuronal, distrofia muscular, osteoporosis y lipodistrofia. Las MSC humanas son inmuno-privilegiadas y representan un tipo celular ventajoso para el trasplante alogénico, reduciendo el riesgo de rechazo y complicaciones del trasplante. Recientemente, también se han producido avances significativos en el uso de células madre mesenquimales autólogas para regenerar tejidos humanos, incluyendo cartílago y menisco, tendones y fracturas óseas. Muchos estudios han investigado también el uso de MSC para terapia génica, que incluye el trasplante de MSC transfectadas con factor de crecimiento endotelial vascular para la mejora de la función cardíaca después de un IM en ratas, MSC como vehículos para el suministro de interferón-β en tumores en ratones y terapia génica con MSC que expresan BMP para promover la formación de hueso. Por consiguiente, debido al gran interés en las MSC como agentes terapéuticos directos y modificados, así como el potencial de los polipéptidos de AARS para actuar como agentes terapéuticos para regular la diferenciación de las MSC *in vivo*, se ensayaron los polipéptidos de AARS como inductores potenciales de la proliferación y diferenciación de MSC.

**Métodos:** Se mantienen hMSC (células estromales de médula humanas) (Cell Applications, n.º de cat 492-05f) según las instrucciones del vendedor. Para el cultivo, se descongelan rápidamente las células y se transfieren inmediatamente a 15ml de medio de crecimiento de células estromales de médula (Cell Application, n.º de cat. 419-500) y se sembraron en un matraz tratado para cultivo tisular estéril convencional. El medio se reemplaza con medio de crecimiento de células estromales de médula reciente cada dos días hasta que las células alcanzan una confluencia > 60 %. Las células se cultivan a 37 °C, CO<sub>2</sub> al 5 %, en un ambiente humidificado y se utilizaron en campanas certificadas para cultivo tisular BSL2 usando técnicas estériles y equipamiento de protección personal adecuado que incluye gafas, guantes y batas de laboratorio. Las células se siembran en placas de ensayo tratadas para cultivo tisular de 96 pocillos de paredes negras y fondo transparente a una concentración de 50.000 células/ml. Se añaden las proteínas derivadas de ARNt sintetasa a una concentración final de 250 nM por pocillo (o como se especifique de otro modo en los ejemplos más adelante) a cada pocillo de ensayo. Todas las células se mantienen

en medio de crecimiento durante 2 días con la excepción de los controles positivos, que se estimularon con medio de diferenciación osteogénica o condrogénica (StemPro, Invitrogen, n.º de cat. A10072-01 y A10071-01, respectivamente). Las células se exponen a polipéptidos de AARS durante 48 horas. Se mide el VCAM soluble usando un kit de ELISA convencional de R&D Systems (n.º de cat. DY643). La proliferación se evalúa con resazurina como se ha descrito anteriormente añadiendo medio reciente que contiene resazurina a las placas seguido de la retirada del sobrenadante e incubando durante tres horas a 37 °C. Las placas se leen en un lector de placas de fluorescencia y se expresa la viabilidad/proliferación en función de la fluorescencia asociada a resorufina de los pocillos tratados con polipéptido de AARS dividida entre la fluorescencia asociada a resorufina de los pocillos tratados solo con PBS. Después de una evaluación de la viabilidad celular, se retira la resazurina con dos intercambios de medio y se añade medio de diferenciación 0,5X a todos los pocillos. La diferenciación se controla mediante inspecciones visuales de todos los pocillos durante 10 días después del intercambio de medio, cambiándose por medio fresco cada dos días para mantener la salud celular. La diferenciación se evaluó con tinción de fosfatasa alcalina usando tinción ELF-97 (Invitrogen, n.º de cat. E6601) en el día 10 después del primer intercambio de diferenciación. (Yang et al, Nature Protocols (6) 187-213 (2011) doi:10.1038/nprot.2010.189).

### ***Proliferación y diferenciación de células musculares lisas de la arteria pulmonar humana.***

**Antecedentes y relevancia terapéutica:** Las células musculares lisas de la arteria pulmonar (PASMC) en los vasos sanguíneos pulmonares de adultos humanos se encuentran en su mayoría quiescentes, no migratorias y están en su mayoría dedicadas a ejecutar su función contráctil en el pulmón. Sin embargo, las PASMC no están diferenciadas terminalmente y poseen la capacidad de modular su fenotipo y salir de su estado quiescente en respuesta a cambios en marcadores ambientales locales. Este estado de diferenciación puede producirse en el desarrollo, lesiones tisulares y la remodelación de vasos en respuesta a cambios en la demanda de tejido. La hipertensión pulmonar (PH) está asociada con una serie de afecciones subyacentes que incluyen un aumento en la resistencia vascular pulmonar periférica como resultado de un aumento del tono vascular y de la contractilidad y remodelado vascular de las PASMC. El remodelado vascular implica el crecimiento de PASMC, la síntesis de material de matriz y alteraciones en las interacciones célula-célula y célula-matriz en las paredes de las arterias pulmonares pequeñas (PA), que dan lugar a un engrosamiento del componente de músculo liso de la pared del vaso y a una musculación anormal de las normalmente no musculadas, PA distales. Este proceso contribuye a un diámetro del lumen reducido y a un aumento de la resistencia periférica. Aunque hay controversia acerca del papel preciso de las PASMC en la causa inicial de la enfermedad, los cambios que se producen desempeñan un papel clave en las consecuencias clínicas de la enfermedad. Una etapa crucial en el estudio de la diferenciación celular es la identificación de un grupo de genes específicos de células o selectivos de células que contribuyen a las funciones diferenciadas de la célula. Se ha identificado una serie de genes de células musculares lisas (SMC) que sirven como marcadores útiles del estado de diferenciación o maduración relativo de las SMC vasculares, tales como alfa-actina de SM, CMH de SM, h1-calpopina, SM22-alfa, desmina, metavinculina, smoothelina y otros. El marcador más ampliamente usado es alfa-actina de SM, parcialmente debido a la disponibilidad comercial de una serie de anticuerpos de muy alta afinidad y altamente selectivos para esta proteína. Los cambios en las PASMC como resultado de sus características inherentes o de la desregulación de los eventos moleculares que gobiernan el crecimiento de las PASMC sigue siendo una cuestión sin resolver. Sin embargo, la determinación de los marcadores reguladores y la gestión de la desregulación potencial proporciona una perspectiva terapéutica significativa para la gestión de una serie de enfermedades vasculares y pulmonares que incluyen hipertensión pulmonar, enfermedades vasculares.

Por lo tanto, los polipéptidos de AARS que tienen la capacidad de modular la diferenciación y/o proliferación de PASMC humanas normales obtenidas de humanos adultos tienen utilidad terapéutica en una serie de enfermedades vasculares y pulmonares que incluyen enfermedades pulmonares inflamatorias y obstructivas que incluyen, por ejemplo, hipertensión pulmonar, enfermedad pulmonar obstructiva crónica, fibrosis pulmonar idiopática y asma.

**Métodos:** Las HPASMC (Cell Applications, n.º de cat. 352-05a) se mantienen en medio de crecimiento de HPASMC (Cell Applications, n.º de cat. 352-05a) en 15 ml de medio en matraces de 125 ml durante 1 pase antes de su uso. Las células se mantienen a 37 °C, CO<sub>2</sub> al 5 %, en un ambiente humidificado y se utilizaron en campanas certificadas para cultivo tisular BSL2 usando técnicas estériles y equipamiento de protección personal adecuado que incluye gafas, guantes y batas de laboratorio. Se siembra un volumen de 80µl de células sobre colágeno recubierto sobre una noche en medio de crecimiento a una densidad celular de 50.000 células/ml. Los polipéptidos de AARS se añadieron en PBS estéril a cada pocillo a una concentración final de 250 nM (o como se indique de otro modo en los ejemplos más adelante). Los pocillos de control contenían únicamente un volumen equivalente de PBS. Las muestras de control positivo se incubaron con medio de diferenciación de HPASMC suministrado por el vendedor (Cell Applications, n.º de cat. 311D-250). Las células se exponen a los polipéptidos de AARS o a PBS en medio basal (Cell Applications, n.º de cat. 310-470) durante 48 horas seguido de un intercambio de medio a medio de diferenciación para toda la placa. El sobrenadante se recoge y se utiliza para evaluar la producción de citocinas incluyendo IL6 e IL8 como se ha descrito anteriormente. La proliferación se evalúa con resazurina como se ha descrito anteriormente añadiendo medio fresco que contenía resazurina a las placas después de la retirada del sobrenadante y de incubar durante tres horas a 37 °C. Las células se controlan durante 10 días con intercambio de medio cada dos días. La diferenciación se evalúa después de la fijación, como se ha descrito anteriormente y la permeabilización con Triton X-100 al 0,1 %, cuantificando la tinción para tinción de alfa-actina de músculo liso usando un anticuerpo anti-SMA-alfa (GeneTex, n.º de cat GTX101362) y un anticuerpo secundario conjugado a

Alexa 405. La proliferación se evalúa con tinción Hoechst después de fijar las células en formaldehído al 10 % durante 30 minutos. Después se lee el colorante Hoechst usando un lector de placas de lectura de fondo con una longitud de onda de excitación (Ex) de 405 nm y una longitud de onda de emisión (Em) de 450 nm. La tinción total de actina se evalúa mediante el uso de un tinte de faloidina marcado con Alexa-488 (Invitrogen, n.º de cat. A12379).

5 **Análisis de la unión de polipéptidos de AARS a células (Ensayos H1-H10 en las tablas de datos más adelante)**

10 **Antecedentes y relevancia terapéutica:** La unión de polipéptidos de AARS a tipos celulares específicos demuestra que el tipo celular en cuestión expresa receptores específicos para el polipéptido de AARS en cuestión. Dependiendo del tipo de célula en cuestión, la unión celular implica un papel potencial para el polipéptido de AARS en la regulación de la actividad o el comportamiento de la célula o de tipos de célula similares, *in vivo*. Los ejemplos específicos de dichos papeles reguladores incluyen, por ejemplo, la unión y modulación de las células B y las células T (inmunomodulación/quimiotaxis/autoinmunidad/inflamación); células HepG2 (control del metabolismo, captación o metabolismo del colesterol); THP-1, jurkat, células Raji (inmunomodulación/quimiotaxis/autoinmunidad/inflamación), plaquetas (trombopoyesis), adipocitos 3T3L1 (lipogénesis/metabolismo), y mioblastos de ratón C2C12 (miogénesis, osteogénesis).

20 **Unión a células sanguíneas**

**Métodos:** Se recoge sangre en tubos de EDTA de donantes sanos. Se colocan 2 ml de sangre completa en un tubo para FACS Falcon de 5 ml. Se añaden 2 ml de tampón de tinción (PBS + FBS al 2 %), se agita vorticialmente durante 3-5 segundos, se centrifuga durante 5 minutos a 300 x g. Se aspira el sobrenadante, se repite el lavado y se resuspende el sedimento en 2 ml de tampón de tinción.

25 Se transfieren 100µl de sangre lavada a tubos para muestras de FACS limpios de 5 ml. Los polipéptidos de AARS marcados con His6 o V5-His6 se añadieron a los tubos y se indican las concentraciones en los experimentos específicos indicados más adelante y se incubaron sobre hielo durante 45 minutos. Después de la incubación, los anticuerpos para los diferentes marcadores de la superficie celular (BD Pharmigen, n.º de cat. 560910, 555398, 555415, 340953, 560361), y anticuerpo anti-marcador V5 marcado con FITC (V5-FITC, Invitrogen, n.º de cat. R96325) o anticuerpo anti-His6 marcado con FITC (AbCam n.º de cat. ab1206) se añaden a los tubos., se incuban en la oscuridad sobre hielo 30 minutos. Después de la incubación se añadieron 2 ml de solución de lisis BD FACS (n.º de cat. 349202) a los tubos. Las muestras se agitan vorticialmente y se colocan sobre hielo durante 15 minutos. Las muestras se lavan con 1 x 2 ml de PBS y se resuspenden en 2 ml de formaldehído al 2 % en PBS antes del análisis FACS. Los polipéptidos de AARS que se unen a más de un 25 % de una población celular, donde el anticuerpo solo no tiene una señal significativa, se considera un acierto.

30 **Ensayos de unión de plaquetas:** Se transfieren 50 µl de sangre lavada a tubos para muestra de FACS de 5 ml, se añaden polipéptidos de AARS marcados con His6 o V5-His6 a los tubos a las concentraciones indicadas en los experimentos específicos indicados a continuación y se colocan los tubos sobre hielo durante 45 minutos. Se añaden a cada tubo 20 µl de anticuerpo CD61 para todas las plaquetas (BD Pharmigen, n.º de cat. 555754) y 0,5 µl de anticuerpo anti-V5-marcado para FITC (Invitrogen, R96325) o anticuerpo anti His6 marcado para FITC (AbCam, n.º de cat. ab1206). Los tubos se colocan sobre hielo y se protegen de la luz durante 30 minutos. Las muestras se enrasan hasta un volumen total en 2 ml de formaldehído al 1 % en PBS y se analizan por citometría de flujo antes de 45 24 horas. Los polipéptidos de AARS que se unen a más de un 25 % de una población celular, donde el anticuerpo solo no tiene una señal significativa, se considera un acierto.

50 **Unión a células en cultivo:** Se colocan aproximadamente  $1 \times 10^6$  células en 100 µl de medio RPMI completo en tubos para FACS de 5ml. se añaden polipéptidos de AARS marcados con His6 o V5-His6 a los tubos a las concentraciones indicadas en los experimentos específicos indicados a continuación y se colocan los tubos sobre hielo durante 45 minutos. Las muestras de células se lavan dos veces con 1 ml de tampón de tinción (PBS + FBS al 2 %), y después se añaden 0,5µl de anticuerpo anti-V5-FITC (Invitrogen R96325) o de anticuerpo anti-His6 marcado para FITC (AbCam, n.º de cat. ab1206) en tampón de tinción con 200µg/ml de IgG humana, y las muestras se incuban sobre hielo, protegidas de la luz, durante 30 minutos. Las muestras lavan dos veces con 1 ml de tampón de 55 tinción y después se enrasan hasta un volumen total en 2 ml de formaldehído al 1 % en PBS y se analizan por citometría de flujo antes de 24 horas. Los polipéptidos de AARS que se unen a más de un 25 % de una población celular, donde el anticuerpo solo no tiene una señal significativa, se considera un acierto.

## 60 ESTUDIOS EN ANIMALES: MODULACIÓN DE A HEMATOPOYESIS Y LAS CITOCINAS EN CIRCULACIÓN

**Antecedentes y relevancia terapéutica:** La hematopoyesis (como alternativa, hemopoyesis) es la formación de componentes celulares de la sangre. Todos los componentes celulares de la sangre proceden de células madre hematopoyéticas (HSC) que residen en la médula del hueso (médula ósea) y tienen la capacidad única de dar lugar a todos los diferentes tipos celulares de la sangre maduros. Las HSC se auto-renuevan: cuando proliferan, al menos una parte de sus células hijas permanecen como HSC, de tal forma que no se agota la reserva de células madre. Las otras hijas de las HSC (células progenitoras mieloides y linfoides), sin embargo, cada una puede comprometerse

a cualquiera de las vías de diferenciación alternativas que dan lugar a la producción de uno o más tipos específicos de células sanguíneas, pero no pueden renovarse por sí mismas. Un cambio en los componentes de la sangre en respuesta a la exposición a un polipéptido de AARS sugiere, por tanto, que el polipéptido de AARS es capaz de modular la hematopoyesis y de regular el desarrollo de las células madre hematopoyéticas.

5

Todas las células sanguíneas pueden dividirse en tres linajes; células eritroides, linfocitos y mielocitos.

Las células eritroides son los glóbulos rojos de la sangre que transportan el oxígeno. Tanto los reticulocitos como los eritrocitos son funcionales y se liberan a la sangre. Por consiguiente, un recuento de reticulocitos estima la tasa de eritropoyesis y un cambio en el recuento de glóbulos rojos de la sangre sugiere que un polipéptido de AARS modula la eritropoyesis.

10

Los linfocitos son la piedra angular del sistema inmunitario adaptativo. Se obtienen de los progenitores linfoides comunes. El linaje linfóide está compuesto principalmente de células T y células B (tipos de glóbulos blancos). Por consiguiente, un cambio en el recuento o la composición de glóbulos blancos en respuesta a la exposición a un polipéptido de AARS sugiere que el polipéptido de AARS modula la linfopoyesis.

15

Los monocitos, que incluyen granulocitos, megacariocitos y macrófagos y se obtienen a partir de progenitores mieloides comunes, están implicados en una serie de roles, incluyendo la inmunidad innata, la inmunidad adaptativa y la coagulación sanguínea. Por consiguiente, un cambio en el recuento o la composición de células mieloides en respuesta a la exposición a un polipéptido de AARS sugiere que el polipéptido de AARS modula la mielopoyesis. Puede usarse el mismo razonamiento para determinar si los polipéptidos de AARS modulan la granulopoyesis, midiendo los cambios en el número de granulocitos en respuesta a la exposición a los polipéptidos de AARS. Puede inferirse un papel para el polipéptido de AARS en la modulación de la megacariocitopoyesis por un cambio en la composición o el número de megacariocitos o plaquetas en la sangre.

20

25

La liberación de citocinas en ratones de tipo silvestre o en varios sistemas de modelo animal de inflamación, proporciona una evaluación inicial de la capacidad de potenciar de los polipéptidos de AARS para modular respuestas inflamatorias. El papel de los polipéptidos de AARS para modular los procesos inflamatorios agudos o crónicos, por ejemplo, puede evaluarse fácilmente usando un modelo de ratón de obesidad inducida por la dieta (DIO). El modelo DIO se centra en poner a los roedores en una dieta rica en grasa durante varios meses, lo que conduce a una obesidad aumentada, resistencia a la insulina y disfunción del sistema inmunitario. Una consecuencia particular de esta desregulación del sistema inmunitario es como resultado una producción aumentada de citocinas proinflamatorias en animales DIO, lo que da lugar a una afección de inflamación sistémica crónica. Cada vez hay más pruebas que sugieren que una inflamación de bajo grado contribuye al desarrollo y el mantenimiento de la obesidad y un fenotipo diabético que se observa de manera similar en la afección humana denominada síndrome metabólico. Como tal, la capacidad de los polipéptidos de AARS de modular el sistema inmunitario y de restaurar el equilibrio homeostático hacia una resolución de este estado inflamatorio crónico podría ser particularmente beneficioso en numerosas enfermedades que incluyen, pero sin limitación, el tratamiento y la prevención de los síntomas y efectos secundarios de la enfermedad metabólica, diabetes, enfermedades cardiovasculares, aterosclerosis, obesidad, así como varias enfermedades y trastornos autoinmunitarios, que incluyen, por ejemplo, esclerosis múltiple, trastornos vasculares y alérgicos.

30

35

40

**Métodos:** Se adquieren ratones de control de tipo silvestre (C57BL/6) o con obesidad inducida por la dieta (C57BL/6NHsd) macho de Harlan (Indianapolis, IN) y se alojan individualmente. Se administra a los ratones DIO una dieta rica en grasas (n.º de cat TD.06414-60 % de las kcal procedentes de grasa) y a los ratones de control se les da una dieta normal (n.º de cat. 2018S-18 % de las kcal procedentes de grasa). Se inicia la dieta rica en grasas de los ratones DIO comenzando a las 6 semanas de edad durante un total de 10 semanas. Se permite que los ratones tanto DIO como de control se alimenten y beban a voluntad. A las 16 semanas de edad, se clasifican los ratones y se asignan aleatoriamente a grupos de 5 animales basándose en el peso. En el día 2, se pesa a los ratones y se extrae sangre por la vena caudal (100µl) para un análisis del hemograma completo (CBC) antes del tratamiento. En el día 1, se pesa a los ratones y se les inyecta por vía intravenosa a través de la vena caudal vehículo (PBS) o polipéptidos de AARS individuales a 10mg/kg. Cuatro horas después de la inyección, se extrae sangre de los ratones por la vena facial (150-200µl) para el análisis posterior de citocinas. En los días 2, 3, y 4, se dosificó a los ratones por vía intravenosa como en el día 1. En el día 5, se pesó a los ratones, se los sacrificó y se recogió sangre mediante punción cardíaca para un hemograma completo (análisis CBC) (plasma-EDTA) y el examen de citocinas (suero).

45

50

55

**CBC y análisis de citocinas:** Los recuentos de sangre completa se analizan de extracciones sanguíneas que preceden a las inyecciones (día -2) y 24 horas después de la inyección final (día 5). Los valores del CBC se evalúan respecto de recuentos de glóbulos blancos totales y morfología general de glóbulos rojos. Los glóbulos blancos se caracterizan adicionalmente por el porcentaje total y fraccional de neutrófilos, linfocitos, monocitos, eosinófilos, y basófilos. El análisis fraccional de glóbulos rojos incluyó mediciones de la hemoglobina (dL), hematocrito (%), volumen corpuscular medio (fL), hemoglobina corpuscular media, concentración de hemoglobina corpuscular media (%), y recuento total de plaquetas ( $10^3/\mu\text{l}$ ). El análisis CBS se lleva a cabo por Antech Diagnostics (Fishers, IN).

60

65

Los niveles de citocinas en circulación se evalúan a las 4 horas después de la inyección (día 1) y 24 horas después de la inyección final (día 5). Se aísla el suero, se congela inmediatamente y se envía a Rules Based Medicine (Austin, TX) para la elaboración de perfiles de múltiples analitos. Las muestras de suero se analizan usando el panel RodentMap que abarca 59 biomarcadores únicos que incluyen Apo A-1, CD40, CD40-L, CRP, ET-1, eotaxina, EGF, Factor VII, fibrinógeno, FGF-9, FGF-básico, GST- $\alpha$ , GCP-2, GM-CSF, KC/GRO $\alpha$ , haptoglobina, IgA, IFN $\gamma$ , IP-10, IL-1 $\alpha$ , IL-1 $\beta$ , IL-10, IL-11, IL-12p70, IL-17A, IL-18, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, LIF, linfotactina, M-CSF-1, MIP-1 $\alpha$ , MIP-1 $\beta$ , MIP-1 $\gamma$ , MIP-2, MIP-3 $\beta$ , MDC, mMP-9, MCP-1, MCP-3, MCP-5, MPO, mioglobina, SAP, SGOT, SCF, RANTES, TPO, factor tisular, TIMP-1, TNF- $\alpha$ , VCAM-1, VEGF-A y vWF. Un cambio en los niveles de citocinas se contó como un acierto en caso de que la citocina aumentase en al menos un factor 2 o se redujese en al menos un 50 % en comparación con los controles de vehículo.

### Ejemplo 1

#### IDENTIFICACIÓN DE FRAGMENTOS PROTEOLÍTICOS Y PRODUCTOS DEL CORTE Y EMPALME ALTERNATIVO DE AARS USANDO TOPOGRAFÍA DE PROTEÍNAS Y PLATAFORMA DE ANÁLISIS DE MIGRACIÓN

Para identificar fragmentos de AARS de las líneas celulares, medios condicionados y tejidos, se preparan muestras de los siguientes modos:

*Macrófagos de ratón (RAW 264.7), citosol y medios condicionados:* Las células se tratan con medio DMEM libre de suero a una densidad de  $15 \times 10^6$  células/matraz. Después de 48 horas, se recogen los medios condicionados y los sedimentos celulares y se procesan. Se separan 200  $\mu$ g de proteína de las fracciones proteómicas secretadas y citosólicas mediante SDS-PAGE y se preparan secciones de gel para su análisis mediante espectrometría de masas.

*Tejido pancreático de ratón:* Se cortan los páncreas de tres ratones, se homogeneizan con Dounce y se someten a ultrasonidos en PBS con inhibidores de proteasa. El proteoma citosólico se aísla por centrifugación y se separan 200  $\mu$ g de proteína mediante SDS-PAGE y las secciones de gel se preparan para su análisis por espectrometría de masas.

*Tejido hepático de ratón:* Se cortan tres hígados de ratón, se homogeneizan con Dounce y se someten a ultrasonidos en PBS con inhibidores de proteasa. El proteoma citosólico se aísla por centrifugación y se separan 200  $\mu$ g de proteína mediante SDS-PAGE y las secciones de gel se preparan para su análisis por espectrometría de masas.

Las digestiones en gel se analizan por LTQ XL, espectrómetro de masas de trampa iónica (ThermoFisher) equipado con el sistema Ultimate 3000  $\mu$ IC (Dionex). En primer lugar, se cargan las muestras en PrepTrap (Michrom) durante 10 min con acetonitrilo al 5 % en ácido fórmico al 0,1 % usando un automuestreador Dionex. Después, se analizan las muestras con una columna capilar de sílice fundido de 100  $\mu$ m (diámetro interno) que contenía 10 cm de resina C18 (Michrom). Los péptidos se eluyen de la columna en el espectrómetro de masas con un caudal de 0,45  $\mu$ l/min usando un gradiente lineal de acetonitrilo al 5-33,5 % en ácido fórmico al 0,1 % en 110 min.

El LTQ funciona en modo de barrido dependiente de datos de tal forma que un barrido EM completo va seguido de siete barridos EM/EM de los siete iones más abundantes. Se posibilita la exclusión dinámica con cuentas repetidas iguales a 1, duraciones de repetición iguales a 20 segundos, el tamaño de la lista de exclusión es de 300 y la duración de la exclusión es de 60 segundos.

Después del análisis CL-EM/EM, se efectúan búsquedas de los datos en bruto con BioWorks3.3.1 (SEQUEST) usando una variante de diana/señuelo concatenada de la base de datos IPI de ratón. Los datos de SEQUEST se filtran y clasifican con DTASelect. Las tablas 1, 4 y 7 muestran secuencias identificadas de este modo.

### Ejemplo 2

#### IDENTIFICACIÓN DE VARIANTES DE CORTE Y EMPALME USANDO SECUENCIACIÓN PROFUNDA

Las variantes de corte y empalme de la aminoacil ARNt sintetasa se identifican usando secuenciación de alto rendimiento de bibliotecas de ADNc enriquecidas respecto de transcritos de aminoacil ARNt sintetasa. Los moldes de ADNc se preparan a partir de extractos de ARN total de tejidos, tales como cerebros humanos adultos y fetales y se enriquecen respecto de transcritos de aminoacil ARNt sintetasa usando secuencias de cebador específicas para todos los exones anotados de todas las aminoacil ARNt sintetastas humanas anotadas y sus proteínas asociadas.

Los ARN totales humanos se obtienen de Clontech. Para la línea celular y las muestras de tejido de ratón, se extraen los ARN totales usando el kit RNA Extract II (MN). El ADN genómico se digiere en las muestras de ARN total mediante DNasa I. Para obtener ARN mensajeros maduros (ARNm), se enriquecen las muestras de ARN dos veces por unión a poliA+ARN y digestión del ARN sin el tapón 5' mediante exonucleasa dependiente de 5'-fosfato. El ADN

- complementario (ADNc) se sintetiza a partir de los ARN maduros usando cebadores que se hibridan a secuencias exónicas de los genes de aminoacil ARNt sintetasa. Se amplifica un transcriptoma enriquecido respecto de genes de aminoacil ARNt sintetasa mediante PCR multiplexada usando el ADNc específico de exones de aminoacil ARNt sintetasa y diferentes combinaciones de cebadores para exones de aminoacil ARNt sintetasa. Los productos bicatenarios de la PCR del transcriptoma enriquecido respecto de aminoacil ARNt sintetasa se reparan enzimáticamente en ambos extremos antes de añadir salientes de A a los extremos 3' de los fragmentos reparados. Después se añaden adaptadores de secuenciación y secuencias de índice a los productos de la PCR de transcriptoma enriquecidos respecto de aminoacil ARNt sintetasa para generar bibliotecas de ADNc para secuenciación profunda con el kit de secuenciación multiplexada de Illumina. En resumen, los productos de la PCR de transcriptoma enriquecidos respecto de aminoacil ARNt sintetasa con salientes de 3'-A se ligan en los oligonucleótidos adaptadores InPE proporcionados en los kits. Se añaden secuencias de índice a los productos de la PCR con adaptadores InPE. Para obtener suficientes fragmentos de ADN para secuenciación profunda, los productos de la PCR con secuencias de índice se amplifican adicionalmente mediante PCR. Las bibliotecas de ADNc enriquecidas respecto de aminoacil ARNt sintetasa con diferentes índices se agrupan y secuencian usando una máquina de secuenciación de ADN de Illumina para obtener lecturas de los 50 pares de bases terminales. Las lecturas de secuenciación se mapean en el genoma humano o de ratón para la identificación de acontecimientos de corte y empalme alternativo. Se usa el programa informático "Splicemap" (disponible para su descarga pública en <http://www-stat.stanford.edu/~kinfai/SpliceMap/>) para identificar las uniones de corte y empalme.
- Se lleva a cabo la secuenciación profunda de estos ADNc para generar aproximadamente 1 millón de lecturas de secuenciación de aproximadamente 50 nucleótidos de longitud. Las secuencias específicas para exones de las aminoacil ARNt sintetatas se buscan contra uniones exónicas anotadas y se identifican nuevas uniones exónicas como eventos de corte y empalme alternativo.
- Las columnas en las **tablas 2, 5 y 8** marcadas como "exón 5'" y "exón 3'" indican, cuando están presentes, qué exones están fusionados entre sí en la secuencia de ADNc. Las **tablas 2, 5 y 8** muestran secuencias que se identificaron para eventos de corte y empalme alternativo, transcritos que contienen dichos eventos de corte y empalme y los polipéptidos expresados por estos transcritos. Las variantes de corte y empalme alternativo identificadas por secuenciación profunda se identifican en las **tablas 2, 5 y 8** como aquellas en las que hay números mayores de cero en las columnas marcadas como "Lecturas de secuenciación" en el cerebro de humano adulto o fetal.

### Ejemplo 3

#### IDENTIFICACIÓN DE POLIPÉPTIDOS DE AARS USANDO BIOINFORMÁTICA

- Los fragmentos de proteína AARS (péptidos de resectina o apendacrina) se identifican usando bioinformática. Las secuencias de aminoácidos de la aminoacil ARNt sintetasa humana de longitud completa se alinean con la secuencia de aminoácidos de longitud completa de su ortólogo de la bacteria *Escherichia coli* usando un programa tal como Fasta (disponible en la página web [http://fasta.bioch.virginia.edu/fasta\\_www2/fasta\\_www.cgi](http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi)) o el programa BLASTP del NCBI (disponible en la página web [http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PROGRAM=blastp&BLAST\\_PROGRAMS=blastp&PAGE\\_TYPE=BlastSearch&SHOW\\_DEFAULTS=on&LINK\\_LOC=blasthom](http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PROGRAM=blastp&BLAST_PROGRAMS=blastp&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthom)). Las secuencias de resectinas de las proteínas humanas se identifican como secuencias que abarcan regiones donde se producen huecos en la secuencia bacteriana en el alineamiento o regiones con baja homología entre las dos especies. El péptido y las secuencias de ADN correspondientes en las tablas 3, 6 y 9 incluyen ejemplos identificados de este modo.

### Ejemplo 4

#### EXPRESIÓN DIFERENCIAL DE POLIPÉPTIDOS DE AARS IDENTIFICADOS POR ESPECTROMETRÍA DE MASAS

- Se usa la técnica PROTOMAP como se describe en el ejemplo 1 para comparar la expresión diferencial de treonil ARNt sintetatas en diferentes tejidos/tipos celulares (véanse las **tablas 1, 4 y 7** para secuencias y comparaciones): La expresión de resectinas de aminoacil-ARNt sintetasa se compara entre el tejido de hígado de ratón y el tejido de páncreas de ratón. La expresión de resectinas de aminoacil-ARNt sintetatas se compara entre el citosol de RAW264.7 y el medio condicionado de células RAW264.7 recogidas después de 48 horas de privación de suero.

### Ejemplo 5

#### EXPRESIÓN DIFERENCIAL DE POLIPÉPTIDOS DE AARS IDENTIFICADOS MEDIANTE SECUENCIACIÓN PROFUNDA

- Para comprobar la expresión diferencial de eventos de corte y empalme, se efectúa la secuenciación profunda para ADNc preparados a partir de diferentes tejidos.

La expresión de eventos de corte y empalme alternativo para aminoacil ARNt sintetasa es inesperada e indica una importancia biológica. La variación en el número relativo de lecturas observadas en la secuenciación profunda de diferentes muestras de transcriptoma indica que los eventos de corte y empalme alternativo de las aminoacil ARNt sintetetas están regulados de manera diferencial y no son solo faltos positivos debidos al manejo de la muestra.

5

### Ejemplo 6

#### EXPLORACIÓN DE ANTICUERPOS

10 Para facilitar el descubrimiento de anticuerpo que muestren unión preferencial a fragmentos específico de aminoacil ARNt sintetasa (por ejemplo, una afinidad >10 veces mayor en comparación con la enzima parental de longitud completa), se exploró una biblioteca de presentación en fagos de anticuerpos humanos por AbD Serotec (una división de MorphoSys™, Martinsried/Planegg, Alemania) usando técnicas de enriquecimiento por afinidad (paseo).  
 15 Los anticuerpos enriquecidos después de múltiples rondas de exploración con los fragmentos de aminoacil ARNt sintetasa se caracterizan posteriormente mediante ELISA respecto de su reactividad con los fragmentos y con la enzima parental, de longitud completa. Los clones que demuestran unión preferencial (por ejemplo, una afinidad > 10 veces mayor) por los fragmentos de aminoacil ARNt sintetasa se caracterizan adicionalmente.

20 En caso de que no se logre la especificidad necesaria al final de este proceso, se usan estrategias de sustracción, tales como etapas de preadsorción con la enzima de longitud completa y/o contratinción, para eliminar los anticuerpos con reactividad cruzada y dirigir el proceso de selección hacia los epítomos únicos en los fragmentos de aminoacil ARNt sintetasa.

### Ejemplo 7

25

#### IDENTIFICACIÓN DE VARIANTES DE CORTE Y EMPALME USANDO PCR SISTEMÁTICA

Los moldes de ADNc para las reacciones PCR se retrotranscriben a partir de extractos de ARN total de tejidos o células (por ejemplo, cerebro humano, IMR-32 y HEK293T). Las reacciones PCR se llevan a cabo usando cebadores específicos para aminoacil ARNt sintetasa, emparejando un cebador directo (FP1) diseñado para hibridarse con la región 5' no traducida o los exones en la mitad 5' del gen con un cebador inverso (RP1) diseñado para hibridarse con los exones en la mitad 3' del gen o la 3'UTR. Los productos de ADN amplificados se analizan mediante electroforesis en gel de agarosa para identificar productos de la PCR que tienen un tamaño diferente al del fragmento amplificado a partir de los transcritos canónicos. Estos diferentes productos de la PCR se cortan y purifican del gel y se ligan en un vector de clonación estándar para el análisis de la secuencia de ADN. Las variantes de corte y empalme alternativo se identifican como diferentes secuencias a partir de los transcritos canónicos. Las variantes de corte y empalme identificadas mediante esta estrategia sistemática de PCR se muestran en las **tablas 2, 5 y 8**.

### Ejemplo 8

#### OPTIMIZACIÓN POR CODONES DE POLINUCLEÓTIDOS DE AARS SELECCIONADOS

Los polipéptidos de AARS representativos (resumidos en la **tabla E2**) se seleccionan para su caracterización bioquímica, biofísica y funcional adicional basándose en uno o más de los siguientes criterios, i) la identificación de fragmentos proteolíticos de polipéptidos de AARS, ii) la identificación de variantes de corte y empalme de polipéptidos de AARS, iii) la identificación de polipéptidos de AARS mediante análisis bioinformático, iv) evidencias de expresión diferencial de polipéptidos de AARS específicos, v) la estructura de dominios de la proteína AARS, vi) el tamaño del polipéptido de AARS y vii) la minimización de secuencias similares duplicadas.

50

Nombre del polipéptido de AARS	SEQ ID NO para los polipéptidos de AARS marcados epitópicamente	SEQ ID NO para los polinucleótidos de AARS	Restos de la proteína AARS	Ubicación del marcador epitópico	Método de clonación/síntesis usado
ThrRS1 <sup>N1</sup>	SEQ ID NO: 35	SEQ ID NO: 45	1-322	N-terminal	2
ThrRS1 <sup>N1</sup>	SEQ ID NO: 36	SEQ ID NO: 45	1-322	C-terminal	2
ThrRS1 <sup>N2</sup>	SEQ ID NO: 37	SEQ ID NO: 46	1-146	N-terminal	1
ThrRS1 <sup>N2</sup>	SEQ ID NO: 38	SEQ ID NO: 46	1-146	C-terminal	1

ThrRS1 <sup>N3</sup>	SEQ ID NO: 39	SEQ ID NO: 47	1-328 + 1aa	N-terminal	1
ThrRS1 <sup>N3</sup>	SEQ ID NO: 40	SEQ ID NO: 47	1-328 + 1aa	C-terminal	1
ThrRS1 <sup>N4</sup>	SEQ ID NO: 41	SEQ ID NO: 48	1-109 + 9aa	N-terminal	2
ThrRS1 <sup>N4</sup>	SEQ ID NO: 42	SEQ ID NO: 48	1-109 + 9aa	C-terminal	2
ThrRS1 <sup>N2</sup>	SEQ ID NO: 43	SEQ ID NO: 49	1-146	N-terminal	2
ThrRS1 <sup>N2</sup>	SEQ ID NO: 44	SEQ ID NO: 49	1-146	C-terminal	2
ThrRS1 <sup>C1</sup>	SEQ ID NO: 97	SEQ ID NO: 109	307-723	N-terminal	1
ThrRS1 <sup>C1</sup>	SEQ ID NO: 98	SEQ ID NO: 109	307-723	C-terminal	1
ThrRS1 <sup>C2</sup>	SEQ ID NO: 99	SEQ ID NO: 110	346-723	N-terminal	2
ThrRS1 <sup>C2</sup>	SEQ ID NO: 100	SEQ ID NO: 110	346-723	C-terminal	2
ThrRS1 <sup>C3</sup>	SEQ ID NO: 101	SEQ ID NO: 111	432-723	N-terminal	1
ThrRS1 <sup>C3</sup>	SEQ ID NO: 102	SEQ ID NO: 111	432-723	C-terminal	1
ThrRS1 <sup>C4</sup>	SEQ ID NO: 103	SEQ ID NO: 112	380-723	N-terminal	2
ThrRS1 <sup>C4</sup>	SEQ ID NO: 104	SEQ ID NO: 112	380-723	C-terminal	2
ThrRS1 <sup>C5</sup>	SEQ ID NO: 105	SEQ ID NO: 113	294-723	N-terminal	1
ThrRS1 <sup>C5</sup>	SEQ ID NO: 106	SEQ ID NO: 113	294-723	C-terminal	1
ThrRS1 <sup>C6</sup>	SEQ ID NO: 107	SEQ ID NO: 114	600-723	N-terminal	2
ThrRS1 <sup>C6</sup>	SEQ ID NO: 108	SEQ ID NO: 114	600-723	C-terminal	2
ThrRS1 <sup>I1</sup>	SEQ ID NO: 123	SEQ ID NO: 125	147-620	N-terminal	1
ThrRS1 <sup>I1</sup>	SEQ ID NO: 124	SEQ ID NO: 125	147-620	C-terminal	1

Los polinucleótidos que codifican los polipéptidos de AARS seleccionados listados en la tabla E2, junto con el marcador epitópico N o C-terminal adecuado, se sintetizan y clonan como se describe en la sección de Materiales y métodos generales usando la metodología de síntesis génica en la tabla E2.

5

### Ejemplo 9

#### EXPRESIÓN Y PURIFICACIÓN BACTERIANA A PEQUEÑA ESCALA

- 10 Los polipéptidos de AARS listados en la **tabla E2** se expresan en *E. coli* como se describe en la sección de Materiales y Métodos generales. La expresión relativa de polipéptidos de AARS solubles y localizados en cuerpos de inclusión se resume en la **tabla E3** a continuación.

Tabla E3 Resumen de características de expresión bacteriana de polipéptidos de AARS			
Polipéptido de AARS	Ubicación del marcador epitópico	Cantidad de proteína recuperada de la fracción soluble	Cantidad de proteína recuperada de los cuerpos de inclusión
ThrRS1 <sup>N1</sup>	N-terminal	++	ND

ThrRS1 <sup>N1</sup>	C-terminal	+++	ND
ThrRS1 <sup>N2</sup>	N-terminal	+++	ND
ThrRS1 <sup>N2</sup>	C-terminal	+++	ND
ThrRS1 <sup>N3</sup>	N-terminal	++	ND
ThrRS1 <sup>N3</sup>	C-terminal	++	ND
ThrRS1 <sup>N4</sup>	N-terminal	++	ND
ThrRS1 <sup>N4</sup>	C-terminal	+	ND
ThrRS1 <sup>C1</sup>	N-terminal	++	ND
ThrRS1 <sup>C1</sup>	C-terminal	++	ND
ThrRS1 <sup>C2</sup>	N-terminal	+	ND
ThrRS1 <sup>C2</sup>	C-terminal	+	ND
ThrRS1 <sup>C3</sup>	N-terminal	+	ND
ThrRS1 <sup>C3</sup>	C-terminal	+	ND
ThrRS1 <sup>C4</sup>	N-terminal	+	ND
ThrRS1 <sup>C4</sup>	C-terminal	+	ND
ThrRS1 <sup>C5</sup>	N-terminal	+	ND
ThrRS1 <sup>C5</sup>	C-terminal	++	ND
ThrRS1 <sup>C6</sup>	N-terminal	+	ND
ThrRS1 <sup>C6</sup>	C-terminal	+	ND
ThrRS1 <sup>I1</sup>	N-terminal	+	+
ThrRS1 <sup>I1</sup>	C-terminal	+	+
"+" representa expresión de polipéptidos de AARS de 0-1 mg/l "++" representa expresión de polipéptidos de AARS de 1-5 mg/l; "+++" representa expresión de polipéptidos de AARS de 5-10 mg/l; "++++" representa expresión de polipéptidos de AARS de 10-15 mg/l; "+++++" representa expresión de polipéptidos de AARS de ≥15 mg/l;			
ND: no determinado L: niveles de expresión de indetectables a bajos M: niveles de expresión medios H: niveles de expresión altos			

5 Sorprendentemente, los datos de expresión de proteínas demuestran la existencia de al menos tres dominios de proteína que muestran un alto nivel de expresión de proteína soluble cuando se expresan en *E. coli*. Específicamente, los datos demuestran que los polipéptidos de AARS ThrRS1<sup>N1</sup>, (aminoácidos 1-322), y ThrRS1<sup>N3</sup> (aminoácidos 1-328+1 aa), definen un primer límite de un primer nuevo dominio de proteína que está altamente expresado en *E. coli*; ThrRS1<sup>N2</sup> (aminoácidos 1-146), define un segundo límite de un segundo nuevo dominio de proteína que está altamente expresado en *E. coli*; y ThrRS1<sup>C1</sup> (aminoácidos 307-723) y ThrRS1<sup>C5</sup> (aminoácidos 294-723), definen un tercer límite de un tercer nuevo dominio de proteína que está altamente expresado en *E. coli*;

**Ejemplo 10**

**PRODUCCIÓN A GRAN ESCALA DE POLIPÉPTIDOS DE AARS**

- 5 Los polipéptidos de AARS representativos se preparan en cantidades mayores para permitir la caracterización funcional y biofísica adicional. Los polipéptidos de AARS listados en la **tabla E4** se expresan en *E. coli* en cultivo a gran escala como se describe en la sección de Materiales y Métodos generales. Los rendimientos y las características biofísicas especificadas, para cada proteína soluble expresada se resumen a continuación en la **tabla E4**.

10

Tabla E4  
Resumen del rendimiento y la caracterización biofísica de polipéptidos de AARS representativos

Polipéptido de AARS	Ubicación del marcador epitópico	Rendimiento [mg/l] <sup>(1)</sup>	Pureza [%]	Endotoxina [UE/mg]	Peso molecular	Concentración madre de trabajo [mg/ml]	Estabilidad [porcentaje de recuperación] <sup>(2)</sup>	Agregación [DLS]
ThRS1 N3	C-terminal	4,4	99	1,2	C: 39.660 D: 39.666; 39.711	13,1	52	+++++
ThRS1 C1	C-terminal	5,2	99	0,1	C:51,154 D: 51.159	11,2	82	+

Notas  
 (1): Rendimiento determinado midiendo la recuperación de proteína después de la última etapa de purificación  
 (2): Determinado como porcentaje de recuperación después de 1 semana a 25°C  
 (3): Medido después de la etapa de concentración después de 1 semana a 25°C  
 (4): Probable que represente el PM sin la metionina N-terminal,  
 C: Calculado  
 D: Determinado

Clave:  
 "+" representa menos de un 1 % de agregados de proteína de alto peso molecular  
 "++" representa menos de un 2% de agregados de proteína de alto peso molecular  
 "+++" representa menos de un 5% de agregados de proteína de alto peso molecular  
 "++++" representa menos de un 10 % de agregados de alto peso molecular  
 "+++++" representa más de un 10 % de agregados de proteína de alto peso molecular  
 ND: No determinado

Los resultados de este estudio establecen que las proteínas de AARS representativas de las familias ThrRS1<sup>N3</sup> y ThrRS1<sup>C1</sup> de las proteínas AARS, muestran rendimientos de expresión de proteínas y características de solubilidad favorables.

5 **Ejemplo 11**

**ELABORACIÓN DE PERFILES TRANSCRIPCIONALES DE POLIPÉPTIDOS DE AARS REPRESENTATIVOS**

10 Para ensayar respecto de la capacidad de los polipéptidos de AARS para modular la expresión génica, se incubaron polipéptidos de AARS seleccionados con células madre mesenquimales o células de músculo esquelético humanas durante los tiempos y a las concentraciones mostradas en la **tabla E5**.

Tabla E5 Elaboración de perfiles transcripcionales de polipéptidos de AARS representativos en células madre mesenquimales (MSC) o células musculares esqueléticas humanas (HskMC)						
Descripción de la muestra de ensayo		Tipo de célula y tiempo de exposición				
Polipéptidos de AARS	Ubicación del marcador epitópico	Concentración nM	MSC 24 horas	MSC 72 horas	HskMC 24 horas	HskMC 72 horas
ThrRS1 <sup>C1</sup>	N-terminal	70	1	7	7	5
ThrRS1 <sup>N2</sup>	N-terminal	250	2	4	4	5
ThrRS1 <sup>N2</sup>	C-terminal	250	1	6	6	4
Controles						
Media entre todos los polipéptidos de AARS explorados			3	5	6	7
Cóctel de osteogénesis			17	20	11	16
Cóctel de condrogénesis			17	19	14	19
Cóctel de adipogénesis			19	15	16	18
Control de SKMC Pos			11	8	5	4
No tratadas			0	0	1	1

15 **En la tabla E5**, los números en cada columna representan el número de genes que se modularon, ya sea positiva o negativamente en al menos un factor de 4 en comparación con los genes de control, tal como se describe en la sección de métodos generales. Los datos muestran que las formas específicas de los polipéptidos de AARS ensayadas tienen la capacidad sorprendente de regular la transcripción y por tanto potencialmente la de modular el destino de desarrollo o el estado de diferenciación cuando se añaden a células madre mesenquimales (MSC) y/o células musculares esqueléticas humanas (HskMC). Las celdas sombreadas con números en negrita en la tabla representan ejemplos donde el polipéptido de AARS muestra un impacto significativo en la regulación de la transcripción génica en las líneas celulares y los tiempos indicados en la tabla.

20 Se llega a la conclusión de que ThrRS1<sup>C1</sup> parece ser el regulador principal de la expresión génica de las células madre mesenquimales.

25 **Ejemplo 12**

**ELABORACIÓN DE PERFILES FUNCIONALES DE POLIPÉPTIDOS DE AARS**

30 Para ensayar respecto de la capacidad de los polipéptidos de AARS para modular una serie de procesos fenotípicos, se incubaron polipéptidos de AARS seleccionados con los tipos celulares y las condiciones proporcionadas en la sección de métodos generales y las **tablas E5 y E6**.

<b>Tabla E6</b>	
<b>Clave para los ensayos y criterios para indicar un acierto</b>	
<b>Ensayos de proliferación</b>	
<b>Fuente y tipo celular</b>	<b>Número de ensayo</b>
Células de leucemia megacariocítica humana / Mo7e	A1
Células de leucemia promielocítica aguda humanas / HL60	A2
Linfoblastos humanos (línea celular de cáncer) / RPMI8226	A3
Células madre mesenquimales humanas / hMSC	A4
Astroцитos humanos	A5
Células de aspirado de médula ósea humanas / células de médula ósea	A6
Células de aspirado de médula ósea humanas / células de médula ósea (cultivo de larga duración)	A7
Sinoviocitos humanos / HFLS-SynRA	A8
Pre-adipocitos humanos / hPAD	A9
Células musculares lisas de la arteria pulmonar humanas / hPASC	A10
Células de músculo esquelético humanas / hSKMC	A11
El análisis de los datos para los ensayos de proliferación se llevó a cabo dividiendo el valor numérico en el pocillo de ensayo entre el valor medio de PBS para la placa de ensayo. Se consideró que los polipéptidos de AARS eran proliferativos en caso de haber más de 3 DT respecto del valor de PBS en la dirección positiva. Se consideró que un polipéptido de AARS derivado de ARNt sintetasa era citotóxico en caso de haber más de 3 DT respecto del valor de PBS en la dirección negativa. Se utilizó un compuesto citotóxico como control negativo y el valor medio para este era siempre mayor de 3 DT respecto del valor medio de PBS.	
Ensayos de diferenciación celular y fenotipado	
Descripción del ensayo	Número de ensayo
Captación de LDL acetiladas por hepatocitos humanos (células HepG2C3a)	B1
El análisis de los datos para el ensayo de captación de ac-LDL se llevó a cabo dividiendo el valor numérico en el pocillo de ensayo entre el valor medio de PBS para la placa de ensayo. Se consideró que los polipéptidos de AARS eran un modulador de la captación de ac-LDL en caso de haber más de 2 DT respecto del valor de PBS en la dirección positiva o negativa. Se efectuó una comprobación visual para confirmar los resultados del lector de placas usando un microscopio de fluorescencia.	
Ensayos de neutrófilos humanos	
Descripción del ensayo	Número de ensayo
Elastasa de neutrófilos	C1
Explosión oxidativa de neutrófilos (agonista)	C2
Explosión oxidativa de neutrófilos (antagonista)	C3
El análisis de los datos para los ensayos de neutrófilos se llevó a cabo dividiendo el valor numérico en el pocillo de ensayo entre el valor medio de PBS para la placa de ensayo. Los polipéptidos de AARS se consideraron como un modulador de la producción de elastasa de neutrófilos o de la biología de la explosión oxidativa en caso de haber más de 2 DT respecto del valor de PBS en la dirección positiva o negativa.	
Modulación de receptores de tipo Toll (TLR)	

ES 2 638 311 T3

Descripción del ensayo	Número de ensayo
Activación de TLR en células RAW BLUE	D1
Antagonismo de TLR en células RAW BLUE	D2
Activación de hTLR2	D3
Activación de hTLR4	D4
Descripción del ensayo	Número de ensayo
El análisis de los datos para los ensayos de modulación de TLR se llevó a cabo dividiendo el valor numérico en el pocillo de ensayo entre el valor medio de PBS para la placa de ensayo. Se consideró que los polipéptidos de AARS eran un modulador de la biología específica de TLR en caso de haber más de 3 DT respecto del valor de PBS en la dirección positiva o negativa. Los controles positivos, que incluyen LPS y reactivo de detección, fueron siempre significativamente distintos y a > 3 DT respecto del valor medio de PBS.	
Liberación de citocinas	
Descripción del ensayo	Número de ensayo
Producción de citocinas por sinoviocitos humanos (liberación de IL6)	E1
Producción de citocinas por células musculares lisas de la arteria pulmonar humanas (hPASMC) (liberación de IL6)	E2
Producción de citocinas por células de músculo esquelético humanas (hSKMC) (liberación de IL6)	E3
Producción de citocinas por astrocitos humanos (liberación de IL6)	E4
Liberación de IL6 por sangre completa	E5
Producción de citocinas por células musculares lisas de la arteria pulmonar humanas (hPASMC) (liberación de IL8) 72 h de incubación	E6
Producción de IL8	
Producción de citocinas por sinoviocitos humanos (liberación de IL8)	E7
Producción de citocinas por células musculares lisas de la arteria pulmonar humanas (hPASMC) (liberación de IL8)	E8
Producción de citocinas por células de músculo esquelético humanas (hSKMC) (liberación de IL8)	E9
Producción de citocinas por astrocitos humanos (liberación de IL8)	E10
Liberación de IL8 por hepatocitos humanos (células HepG2C3a)	E11
Células de leucemia promielocítica aguda humanas / HL60 (liberación de IL8)	E12
Linfoblastos humanos (línea celular de cáncer) / RPMI8226 (liberación de IL8)	E13
Producción de TNF alfa	
Producción de citocinas por sinoviocitos humanos (liberación de TNF alfa)	E14
Liberación de TNF alfa por sangre completa	E15
Liberación de IL10	

ES 2 638 311 T3

Células de leucemia promielocítica aguda humanas / HL60 (liberación de IL10)	E16
Células mononucleares primarias de sangre humana (liberación de IL10)	E17
El análisis de los datos para los ensayos de liberación de citocinas se llevó a cabo dividiendo el valor numérico en el pocillo de ensayo entre el valor medio de PBS para la placa de ensayo. Los polipéptidos de AARS se consideraron como un modulador de la producción de citocinas o de la biología relacionada con las citocinas en caso de obtener un valor medido mayor de 2 DT respecto del valor de PBS en la dirección positiva o negativa. Se corrió un patrón de proteínas (específico para cada kit de ensayo) en cada placa para asegurar la buena calidad del ensayo. Solo los ensayos con curvas patrón de proteínas que tenían un valor de R <sup>2</sup> de > de 0,9 se seleccionaron para el análisis de datos.	
Adhesión celular y quimiotaxis	
Descripción del ensayo	Número de ensayo
Adhesión celular de monocitos THP 1 / células endoteliales de la vena umbilical humana (HUVEC)	F1
Hepatocitos humanos (células HepG2C3a) (liberación de ICAM)	F2
Descripción del ensayo	Número de ensayo
Regulación de la adhesión celular de células endoteliales microvasculares pulmonares humanas (HLMVEC) (liberación de ICAM)	F3
Regulación de la adhesión celular de células endoteliales de la vena umbilical humana (HUVEC) (liberación de VCAM)	F4
Regulación de la adhesión celular de células madre mesenquimales humanas (hMSC) (liberación de VCAM)	F5
Regulación de la adhesión de células de músculo esquelético humanas (hSKMC) (liberación de VCAM)	F6
Regulación de la adhesión celular de células musculares lisas de la arteria pulmonar humanas (hPASMC) (liberación de VCAM)	F7
El análisis de los datos para los ensayos de regulación de la adhesión celular se llevó a cabo dividiendo el valor numérico en el pocillo de ensayo entre el valor medio de PBS para la placa de ensayo. Los polipéptidos de AARS se consideraron un modulador de la adhesión celular o un modulador de la biología relacionada con la adhesión celular en caso de obtener un valor medido de más de 2 DT respecto del valor de PBS en la dirección positiva o negativa. En el caso de los ensayos ELISA, se corrió un patrón de proteínas (específico para cada kit de ensayo) en cada placa para asegurar la buena calidad del ensayo. Solo los ensayos con curvas patrón de proteínas que tenían un valor de R <sup>2</sup> de > de 0,9 se seleccionaron para el análisis de datos.	
Diferenciación celular	
Diferenciación celular de preadipocitos humanos (hPAD)	G1
Diferenciación celular de células de músculo esquelético humano (hSKMC)	G2
Diferenciación celular de células madre mesenquimales humanas (hMSC)	G3
Diferenciación de células musculares lisas de la arteria pulmonar humanas (hPASMC)	G4

El análisis de los datos para los ensayos de diferenciación celular se llevó a cabo dividiendo el valor numérico en el pocillo de ensayo entre el valor medio de PBS para la placa de ensayo. Los ensayos de diferenciación se puntuaron basándose en la intensidad de fluorescencia de anticuerpos particulares, tal como se describe en la sección de métodos. Se consideró que los polipéptidos de AARS eran un modulador de la diferenciación celular en caso de obtener una lectura de intensidad de más de 2 DT respecto del valor de PBS en la dirección positiva o negativa. Para el análisis de hSKMC, se tomaron fotografías digitales de todos los pocillos y se puntuaron las fotografías de un modo ciego por tres personas usando un sistema de puntuación de 4 puntos, donde 4 indicó una intensa tinción de actina de músculo esquelético y una formación evidente de miotubos. Se usó el valor medio de la puntuación visual y solo se consideraron como aciertos los pocillos con un valor medio > 3. Los pocillos tratados con control de diferenciación en este ensayo normalmente tuvieron puntuaciones > 2, mientras que los pocillos tratados con PBS tuvieron puntuaciones < 2.

Unión celular	
Descripción del ensayo	Número de ensayo
PBMC	H1
Célula T primaria	H2
Célula B primaria	H3
Monocito primario	H4
HepG2	H5
3T3L1	H6
C2C12	H7
THP1	H8
Jurkat	H9
Raji	H10
Descripción del ensayo	Número de ensayo

Se consideró que los polipéptidos de AARS se unían a un tipo celular particular en caso de que la intensidad de fluorescencia media de células unidas fuese mayor de 2 DT respecto de los valores de reactivo de control para ese tipo celular.

<b>Tabla E7</b> <b>Resultados de los estudios de elaboración de perfiles funcionales de polipéptidos de AARS</b>			
Polipéptidos de AARS	Ubicación del marcador epitópico	Concentración [nM]	Aciertos de ensayo
ThrRS1 <sup>N1</sup>	N-terminal	120	(Proliferación) C2 (Activación de neutrófilos) D1 (Modulación de receptores de tipo Toll) G2 (Diferenciación celular)
ThrRS1 <sup>N1</sup>	C-terminal	250	A7 (Proliferación) E8 (Liberación de citocinas) G1, G2 (Diferenciación celular)
ThrRS1 <sup>N2</sup>	N-terminal	250	A7 (Proliferación) B1 (Captación de LDL acetiladas) C2 (Activación de neutrófilos) D2 (Modulación de receptores de tipo Toll) G2 (Diferenciación celular)
ThrRS1 <sup>N2</sup>	C-terminal	250	A7 (Proliferación) C2 (Activación de neutrófilos) D2 (Modulación de receptores de tipo Toll) E11 (Liberación de citocinas) G2 (Diferenciación celular)

ThrRS1 <sup>N3</sup>	N-terminal	124	A7 (Proliferación) B1 (Captación de LDL acetiladas) C2 (Activación de neutrófilos) E4, E14 (Liberación de citocinas) G1 (Diferenciación celular)
ThrRS1 <sup>N3</sup>	C-terminal	250	A4, A7 (Proliferación) B1 (Captación de LDL acetiladas) C2 (Activación de neutrófilos) E4 (Liberación de citocinas)
ThrRS1 <sup>N4</sup>	N-terminal	229	A7 (Proliferación) B1 (Captación de LDL acetiladas) C2 (Activación de neutrófilos) D1 (Modulación de receptores de tipo Toll) E12 (Liberación de citocinas) G1 (Diferenciación celular)
ThrRS1 <sup>C1</sup>	N-terminal	70	B1 (Captación de LDL acetiladas) C2 (Activación de neutrófilos) G1, G2, G3, G4 (Diferenciación celular)
ThrRS1 <sup>C1</sup>	C-terminal	131	A7 (Proliferación) C2 (Activación de neutrófilos) A8 (Proliferación) A7 (Proliferación)
ThrRS1 <sup>C5</sup>	C-terminal	95	C2 (Activación de neutrófilos) D2 (Modulación de receptores de tipo Toll) E1, E8 (Liberación de citocinas)

5 Se llega a la conclusión de que ThrRS1<sup>N1</sup>, ThrRS1<sup>N2</sup>, ThrRS1<sup>N3</sup>, ThrRS1<sup>N4</sup>, ThrRS1<sup>C1</sup> y ThrRS1<sup>C5</sup> parecen ser reguladores principales de la proliferación celular, la diferenciación, la liberación de citocinas, la captación de LDL acetiladas, la adhesión celular y la quimiotaxis y son posiblemente reguladores de los receptores de tipo Toll y la activación de neutrófilos.

10 Los datos de exploración fenotípica demuestran que los polipéptidos de AARS ThrRS1<sup>N1</sup>, (aminoácidos 1-322), y ThrRS1<sup>N3</sup> (aminoácidos 1-328+1 aa) definen los límites de un primer nuevo dominio de proteína que es altamente activo en una amplia serie de ensayos de exploración fenotípica.

15 Por consiguiente, se llega a la conclusión de que los polipéptidos de AARS que comprenden los aminoácidos 1-328 de treonil ARNt sintetasa definen los límites aproximados (es decir, en aproximadamente +/- 5 aminoácidos) de un primer nuevo dominio de proteína AARS altamente activo, que es i) altamente activo funcionalmente, ii) puede fabricarse y producirse fácilmente en *E. coli* e iii) muestra características de estabilidad de proteínas y agregación favorables. Los expertos en la materia apreciarán que cualquier polipéptido de AARS que comprenda tan pocos como 322 aminoácidos de la treonil ARNt sintetasa, hasta tan grandes como polipéptidos que comprenden los primeros 328 aminoácidos de la treonil ARNt sintetasa, representan equivalentes funcionales de los polipéptidos de AARS específicos descritos.

20 Los datos de exploración fenotípica también demuestran que los polipéptidos de AARS ThrRS1<sup>N2</sup>, (aminoácidos 1-146), definen los límites de un segundo nuevo dominio de proteína que es altamente activo en una amplia serie de ensayos de exploración fenotípica. Por consiguiente, se llega a la conclusión de que los polipéptidos de AARS que comprenden los aminoácidos 1-146 de treonil ARNt sintetasa definen los límites aproximados (es decir, en aproximadamente +/- 5 aminoácidos) de un segundo nuevo dominio de proteína AARS altamente activo, que es i) altamente activo funcionalmente, ii) puede fabricarse y producirse fácilmente en *E. coli* e iii) muestra características de estabilidad de proteínas y agregación favorables.

25 Además, los datos de exploración fenotípica también demuestran que los polipéptidos de AARS ThrRS1<sup>C1</sup>, (aminoácidos 307-723), y ThrRS1<sup>C5</sup> (aminoácidos 294-723) definen los límites de un tercer nuevo dominio de proteína que es altamente activo en una amplia serie de ensayos de exploración fenotípica.

30 Por consiguiente, se llega a la conclusión de que los polipéptidos de AARS que comprenden los aminoácidos 294-723 de treonil ARNt sintetasa definen los límites aproximados (es decir, en aproximadamente +/- 5 aminoácidos) de un tercer nuevo dominio de proteína AARS altamente activo, que es i) altamente activo funcionalmente, ii) puede fabricarse y producirse fácilmente en *E. coli* e iii) muestra características de estabilidad de proteínas y agregación favorables. Los expertos en la materia apreciarán que cualquier polipéptido de AARS que comprende tan pocos como los aminoácidos 307-723 de la treonil ARNt sintetasa, hasta tan grandes como polipéptidos que comprenden los aminoácidos 294-723 de la treonil ARNt sintetasa, representan equivalentes funcionales de los polipéptidos de

AARS específicos descritos.  
LISTADO DE SECUENCIAS

- 5 <110> aTyr Pharma Inc.  
Pangu Biopharma Limited  
Greene, Leslie Ann  
Chiang, Kyle P.  
Hong, Fei  
Vasserot, Alain P.
- 10 Lo, Wing-Sze  
Watkins, Jeffry D.  
Mendlein, John D.  
Quinn, Cheryl L.
- 15 <120> DESCUBRIMIENTO INNOVADOR DE COMPOSICIONES TERAPÉUTICAS, DIAGNÓSTICAS Y DE ANTICUERPOS RELACIONADAS CON FRAGMENTOS DE PROTEÍNA TREONIL ARNt SINTETASAS
- <130> 120161.483PC
- 20 <140> PCT  
<141> 27-04-2011
- <150> US 61/328.369  
<151> 27-04-2010
- 25 <150> US 61/328.371  
<151> 27-04-2010
- <150> US 61/328.373  
<151> 27-04-2010
- 30 <160> 126
- <170> FastSEQ para Windows Versión 4.0
- 35 <210> 1  
<211> 77  
<212> ADN  
<213> Secuencia artificial
- 40 <220>  
<223> Oligonucleótido
- <400> 1
- 45 **aggaggtaaa acatatgcat catcatcatc atcacggtaa gcctatccct aaccctttgc 60**  
**tcggtctcga ttctacg 77**
- <210> 2  
<211> 12  
<212> ADN  
<213> Secuencia artificial
- <220>  
<223> Oligonucleótido
- 55 <400> 2  
taatgactcg ag 12
- <210> 3  
<211> 17  
<212> ADN  
<213> Secuencia artificial
- <220>



ES 2 638 311 T3

<212> ADN  
 <213> Secuencia artificial

5  
 <220>  
 <223> Oligonucleótido

<400> 9  
 ctcgagtaat ga 12

10  
 <210> 10  
 <211> 12  
 <212> ADN  
 <213> Secuencia artificial

15  
 <220>  
 <223> Oligonucleótido

<400> 10  
 catatgggat cc 12

20  
 <210> 11  
 <211> 72  
 <212> ADN  
 <213> Secuencia artificial

25  
 <220>  
 <223> Oligonucleótido

30  
 <400> 11  
  
**ctcgagggta agcctatccc taacctctc ctcggtctcg attctacgca ccaccaccac 60**  
**caccactaat ga . 72**

35  
 <210> 12  
 <211> 322  
 <212> PRT  
 <213> *Homo sapiens*

<400> 12

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

210 215 220  
 Thr Leu Leu Ala Met Phe Lys Tyr Asn Lys Phe Lys Cys Arg Ile Leu  
 225 230 235 240  
 Asn Glu Lys Val Asn Thr Pro Thr Thr Thr Val Tyr Arg Cys Gly Pro  
 245 250 255  
 Leu Ile Asp Leu Cys Arg Gly Pro His Val Arg His Thr Gly Lys Ile  
 260 265 270  
 Lys Ala Leu Lys Ile His Lys Asn Ser Ser Thr Tyr Trp Glu Gly Lys  
 275 280 285  
 Ala Asp Met Glu Thr Leu Gln Arg Ile Tyr Gly Ile Ser Phe Pro Asp  
 290 295 300  
 Pro Lys Met Leu Lys Glu Trp Glu Lys Phe Gln Glu Glu Ala Lys Asn  
 305 310 315 320  
 Arg Asp

<210> 13  
 <211> 966  
 <212> ADN  
 <213> *Homo sapiens*

5

<400> 13

```

atgtttgagg agaaggccag cagtccttca ggaagatgg gagggcagga gaagccgatt 60
ggtgctgggtg aagagaagca aaaggaagga ggcaaaaaga agaacaaaga aggatctgga 120
gatggaggtc gagctgagtt gaatccttgg cctgaatata tttacacacg tcttgagatg 180
tataatatac taaaagcaga acatgattcc attctggcag aaaaggcaga aaaagatagc 240
aagccaatta aagtcacttt gcctgatggt aaacagggtg atgcggaatc ttggaaaact 300
acaccatatac aaattgcctg tggatttagt caaggcctgg ccgacaacac cgttattgct 360
aaagtaaata atggtgtgtg ggacctggac cgccctctgg aagaagattg taccttggag 420
cttctcaagt ttgaggatga ggaagctcag gcagtgtatt ggcactctag tgctcacata 480
atgggtgaag ccatggaaag agtctatggt ggatgtttat gctacgggtcc gccaatagaa 540
aatggattct attatgacat gtacctcgaa gaaggggggtg tgtctagcaa tgatttctct 600
tctctggagg ctttgtgtaa gaaaatcatt aaagaaaaac aagcttttga aagactggaa 660
gttaagaaag aaactttact ggcaatgttt aagtacaaca agttcaaata cgggatattg 720
aatgaaaag tgaatactcc aactaccaca gtctatagat gtggcccttt gatagatctc 780
tgccggggtc ctcatgtag acacacgggc aaaattaagg ctttaaaaat acacaaaaat 840
tcctccacgt actgggaagg caaagcagat atggagactc tccagagaat ttatggcatt 900
tcattcccag atcctaaaat gttgaaagag tgggagaagt tccaagagga agctaaaaac 960
cgagat
    
```

<210> 14  
 <211> 146  
 <212> PRT  
 <213> *Homo sapiens*

5

<400> 14

```

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

Glu Asp  
 145

10

<210> 15  
 <211> 438  
 <212> ADN  
 <213> *Homo sapiens*

15

<400> 15

```

atgtttgagg agaaggccag cagtccttca ggaagatgg gaggcgagga gaagccgatt 60
gggtgctggtg aagagaagca aaaggaagga ggcaaaaaga agaacaaaga aggatctgga 120
gatggaggtc gagctgagtt gaatccttgg cctgaatata ttacacacg tcttgagatg 180
tataatatac taaaagcaga acatgattcc attctggcag aaaaggcaga aaaagatagc 240
aagccaatta aagtcacttt gcctgatggt aaacagggtg atgcggaatc ttggaaaact 300
acaccatata aaattgctg tggaattagt caaggcctgg cgcacaacac cgttattgct 360
aaagtaaata atgttgtgtg ggacctggac cgcctctctg aagaagattg taccttgagg 420
cttctcaagt ttgaggat                                     438
    
```

<210> 16  
 <211> 17  
 <212> PRT  
 <213> *Mus musculus*

5

<400> 16

```

Met Ser Gln Glu Lys Ala Ser Ser Pro Ser Gly Lys Met Asp Gly Glu
  1           5                10          15
Lys
    
```

10

<210> 17  
 <211> 45  
 <212> PRT  
 <213> *Mus musculus*

15

<400> 17

```

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

20

<210> 18  
 <211> 12  
 <212> PRT  
 <213> *Mus musculus*

25

<400> 18

```

Leu Lys Ala Glu His Asp Ser Ile Leu Ala Glu Lys
  1           5                10
    
```

30

<210> 19  
 <211> 74  
 <212> PRT  
 <213> *Mus musculus*

35

<400> 19

```

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

ES 2 638 311 T3

<210> 20  
 <211> 12  
 <212> PRT  
 <213> *Mus musculus*  
 5  
 <400> 20

Leu Lys Ala Glu His Asp Ser Ile Leu Ala Glu Lys  
 1 5 10

10 <210> 21  
 <211> 24  
 <212> PRT  
 <213> *Mus musculus*  
 15 <400> 21

Ala Ala Lys Asp Ser Lys Pro Ile Lys Val Thr Leu Pro Asp Gly Lys  
 1 5 10 15  
 Gln Val Asp Ala Glu Ser Trp Lys  
 20

20 <210> 22  
 <211> 22  
 <212> PRT  
 <213> *Mus musculus*  
 25 <400> 22

Thr Thr Pro Tyr Gln Ile Ala Cys Gly Ile Ser Gln Gly Leu Ala Asp  
 1 5 10 15  
 Asn Thr Val Val Ala Lys  
 20

30 <210> 23  
 <211> 58  
 <212> PRT  
 <213> *Mus musculus*  
 <400> 23

Leu Lys Ala Glu His Asp Ser Ile Leu Ala Glu Lys Ala Ala Lys Asp  
 1 5 10 15  
 Ser Lys Pro Ile Lys Val Thr Leu Pro Asp Gly Lys Gln Val Asp Ala  
 20 25 30  
 Glu Ser Trp Lys Thr Thr Pro Tyr Gln Ile Ala Cys Gly Ile Ser Gln  
 35 40 45  
 Gly Leu Ala Asp Asn Thr Val Val Ala Lys  
 50 55

40 <210> 24  
 <211> 329  
 <212> PRT  
 <213> *Homo sapiens*  
 <400> 24

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

<210> 25  
 <211> 990  
 <212> ADN  
 <213> *Homo sapiens*

5

<400> 25

```

atgtttgagg agaaggccag cagtccttca ggaagatgg gaggcgagga gaagccgatt 60
ggtgctggtg aagagaagca aaaggaagga ggcaaaaaga agaacaaaga aggatctgga 120
gatggaggtc gagctgagtt gaatccttgg cctgaatata tttacacacg tcttgagatg 180
tataatatac taaaagcaga acatgattcc attctggcag aaaaggcaga aaaagatagc 240
aagccaatta aagtcacttt gcctgatggt aaacaggttg atgcggaatc ttggaaaact 300
acaccatata aaattgcctg tggatttagt caaggcctgg ccgacaacac cgttattgct 360
aaagtaaata atgttgtgtg ggacctggac cgccctctgg aagaagattg taccttggag 420
cttctcaagt ttgaggatga ggaagctcag gcagtgtatt ggcactctag tgctcacata 480
atgggtgaag ccatggaaag agtctatggt ggatgtttat gctacgggtcc gccaatagaa 540
aatggattct attatgacat gtacctcgaa gaagggggtg tgtctagcaa tgatttctct 600
tctctggagg ctttgtgtaa gaaaatcatt aaagaaaaac aagcttttga aagactggaa 660
gttaagaaag aaactttact ggcaatgttt aagtacaaca agttcaaata ccggatattg 720
aatgaaaagc tgaatactcc aactaccaca gtctatagat gtggcccttt gatagatctc 780
tgccggggtc ctcattgtag acacacgggc aaaattaagg ctttaaaaat acacaaaaat 840
tctccacgt actgggaagg caaagcagat atggagactc tccagagaat ttatggcatt 900

tcattcccag atcctaaaat gttgaaagag tgggagaagt tccaagagga agctaaaaac 960
cgagatcata ggaaaattgg caggccatga 990

```

<210> 26  
 <211> 118  
 <212> PRT  
 <213> *Homo sapiens*

<400> 26

```

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

```

<210> 27  
 <211> 357  
 <212> ADN  
 <213> *Homo sapiens*

<400> 27

```

atgtttgagg agaaggccag cagtccttca ggaagatgg gaggcgagga gaagccgatt 60
ggtgctggtg aagagaagca aaaggaagga ggcaaaaaga agaacaaaga aggatctgga 120
gatggaggtc gagctgagtt gaatccttgg cctgaatata tttacacacg tcttgagatg 180
tataatatac taaaagcaga acatgattcc attctggcag aaaaggcaga aaaagatagc 240
aagccaatta aagtcacttt gcctgatggt aaacaggttg atgcggaatc ttggaaaact 300
acaccatata aaattgcctg tggatttagt tgtattggca ctctagtgtc cacataa 357

```

<210> 28  
 <211> 688  
 <212> PRT  
 <213> *Homo sapiens*

ES 2 638 311 T3

<400> 28

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

ES 2 638 311 T3

145					150					155				160	
Met	Gly	Glu	Ala	Met	Glu	Arg	Val	Tyr	Gly	Gly	Cys	Leu	Cys	Tyr	Gly
				165					170					175	
Pro	Pro	Ile	Glu	Asn	Gly	Phe	Tyr	Tyr	Asp	Met	Tyr	Leu	Glu	Glu	Gly
			180					185					190		
Gly	Val	Ser	Ser	Asn	Asp	Phe	Ser	Ser	Leu	Glu	Ala	Leu	Cys	Lys	Lys
		195					200					205			
Ile	Ile	Lys	Glu	Lys	Gln	Ala	Phe	Glu	Arg	Leu	Glu	Val	Lys	Lys	Glu
	210					215					220				
Thr	Leu	Leu	Ala	Met	Phe	Lys	Tyr	Asn	Lys	Phe	Lys	Cys	Arg	Ile	Leu
225					230					235				240	
Asn	Glu	Lys	Val	Asn	Thr	Pro	Thr	Thr	Thr	Val	Tyr	Arg	Cys	Gly	Pro
				245					250					255	
Leu	Ile	Asp	Leu	Cys	Arg	Gly	Pro	His	Val	Arg	His	Thr	Gly	Lys	Ile
			260					265				270			
Lys	Ala	Leu	Lys	Ile	His	Lys	Asn	Ser	Ser	Thr	Tyr	Trp	Glu	Gly	Lys
		275					280					285			
Ala	Asp	Met	Glu	Thr	Leu	Gln	Arg	Ile	Tyr	Gly	Ile	Ser	Phe	Pro	Asp
	290					295					300				
Pro	Lys	Met	Leu	Lys	Glu	Trp	Glu	Lys	Phe	Gln	Glu	Glu	Ala	Lys	Asn
305						310					315			320	
Arg	Asp	His	Arg	Lys	Ile	Gly	Arg	Asp	Gln	Glu	Leu	Tyr	Phe	Phe	His
				325					330					335	
Glu	Leu	Ser	Pro	Gly	Ser	Cys	Phe	Phe	Leu	Pro	Lys	Gly	Ala	Tyr	Ile
			340					345					350		
Tyr	Asn	Ala	Leu	Ile	Glu	Phe	Ile	Arg	Ser	Glu	Tyr	Arg	Lys	Arg	Gly
		355					360					365			
Phe	Gln	Glu	Val	Val	Thr	Pro	Asn	Ile	Phe	Asn	Ser	Arg	Leu	Trp	Met
	370					375					380				
Thr	Ser	Gly	His	Trp	Gln	His	Tyr	Ser	Glu	Asn	Met	Phe	Ser	Phe	Glu
385					390					395				400	
Val	Glu	Lys	Glu	Leu	Phe	Ala	Leu	Lys	Pro	Met	Asn	Cys	Pro	Gly	His
				405					410					415	
Cys	Leu	Met	Phe	Asp	His	Arg	Pro	Arg	Ser	Trp	Arg	Glu	Leu	Pro	Leu
			420					425				430			
Arg	Leu	Ala	Asp	Phe	Gly	Val	Leu	His	Arg	Asn	Glu	Leu	Ser	Gly	Ala
		435					440					445			
Leu	Thr	Gly	Leu	Thr	Arg	Val	Arg	Arg	Phe	Gln	Gln	Asp	Asp	Ala	His
	450					455					460				
Ile	Phe	Cys	Ala	Met	Glu	Gln	Ile	Glu	Asp	Glu	Ile	Lys	Gly	Cys	Leu
465					470					475				480	
Asp	Phe	Leu	Arg	Thr	Val	Tyr	Ser	Val	Phe	Gly	Phe	Ser	Phe	Lys	Leu
				485					490					495	
Asn	Leu	Ser	Thr	Arg	Pro	Glu	Lys	Phe	Leu	Gly	Asp	Ile	Glu	Val	Trp
			500					505					510		
Asp	Gln	Ala	Glu	Lys	Gln	Leu	Glu	Asn	Ser	Leu	Asn	Glu	Phe	Gly	Glu
		515					520					525			
Lys	Trp	Glu	Leu	Asn	Ser	Gly	Asp	Gly	Ala	Phe	Tyr	Gly	Pro	Lys	Ile
	530					535					540				
Asp	Ile	Gln	Ile	Lys	Asp	Ala	Ile	Gly	Arg	Tyr	His	Gln	Cys	Ala	Thr
545					550					555				560	
Ile	Gln	Leu	Asp	Phe	Gln	Leu	Pro	Ile	Arg	Phe	Asn	Leu	Thr	Tyr	Val
				565					570					575	
Arg	Pro	Phe	Trp	Leu	Ser	Pro	Arg	Gln	Val	Met	Val	Val	Pro	Val	Gly
			580					585					590		
Pro	Thr	Cys	Asp	Glu	Tyr	Ala	Gln	Lys	Val	Arg	Gln	Gln	Phe	His	Asp
		595					600					605			
Ala	Lys	Phe	Met	Ala	Asp	Ile	Asp	Leu	Asp	Pro	Gly	Cys	Thr	Leu	Asn
	610					615					620				
Lys	Lys	Ile	Arg	Asn	Ala	Gln	Leu	Ala	Gln	Tyr	Asn	Phe	Ile	Leu	Val
625					630					635				640	
Val	Gly	Glu	Lys	Glu	Lys	Ile	Ser	Gly	Thr	Val	Asn	Ile	Arg	Thr	Arg
				645					650					655	
Asp	Asn	Lys	Val	His	Gly	Glu	Arg	Thr	Ile	Ser	Glu	Thr	Ile	Glu	Arg
			660					665					670		
Leu	Gln	Gln	Leu	Lys	Glu	Phe	Arg	Ser	Lys	Gln	Ala	Glu	Glu	Glu	Phe
		675						680					685		



```

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

```

<210> 36  
 <211> 346  
 5 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad C-terminal de 6xHis

10 <400> 36

```

Met Gly Ser Phe Glu Glu Lys Ala Ser Ser Pro Ser Gly Lys Met Gly
 1      5      10      15
Gly Glu Glu Lys Pro Ile Gly Ala Gly Glu Glu Lys Gln Lys Glu Gly
 20      25      30
Gly Lys Lys Lys Asn Lys Glu Gly Ser Gly Asp Gly Gly Arg Ala Glu
 35      40      45

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

```

<210> 37

<211> 166

5 <212> PRT

<213> Secuencia artificial

<220>

10 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad N-terminal de 6xHis (ADN2.0)

<400> 37

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

115 120 125  
 Ile Ser Gln Gly Leu Ala Asp Asn Thr Val Ile Ala Lys Val Asn Asn  
 130 135 140  
 Val Val Trp Asp Leu Asp Arg Pro Leu Glu Glu Asp Cys Thr Leu Glu  
 145 150 155 160  
 Leu Leu Lys Phe Glu Asp  
 165

<210> 38  
 <211> 166  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad C-terminal de 6xHis (ADN2.0)

<400> 38

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

<210> 39  
 <211> 349  
 <212> PRT  
 <213> Secuencia artificial

<220>

<223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad N-terminal de 6xHis

<400> 39

5

```

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

```

<210> 40

10 <211> 349

<212> PRT

<213> Secuencia artificial

<220>

15 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad C-terminal de 6xHis

<400> 40

```

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

```

<210> 41

5 <211> 142

<212> PRT

<213> Secuencia artificial

<220>

10 <223> Polipéptido de treonil ARNt sintetasas con marcador de afinidad N-terminal de 6xHis

<400> 41

```

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

```

5 <210> 42  
 <211> 142  
 <212> PRT  
 <213> Secuencia artificial

10 <220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad C-terminal de 6xHis  
 <400> 42

```

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

```

15 <210> 43  
 <211> 170  
 <212> PRT  
 <213> Secuencia artificial

20 <220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad N-terminal de 6xHis (Genewiz)  
 <400> 43

```

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

```

<210> 44  
 <211> 170  
 5 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> Polipéptido de treonil ARNt sintetasas con marcador de afinidad C-terminal de 6xHis (Genewiz)

10 <400> 44

```

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

```

15 <210> 45  
 <211> 975  
 <212> ADN  
 <213> Secuencia artificial  
 <220>

ES 2 638 311 T3

<223> Polinucleótido de treonil ARNt sintetasa optimizado por codones

<400> 45

```

ggatccttcg aagaaaaagc gagcagccct agcggtaaaa tgggtggcga agagaaacca 60
attggtgctg gcgaagagaa acagaaagaa ggccggtaaaa aaaaaaacia agagggcagc 120
ggcgacggcg gtcgtgctga actgaatcct tggccggagt atatctatac acgcctggag 180
atgtataaca tcctgaaagc cgagcatgat tctattctgg cggagaaagc agagaaagac 240
tccaaaccga ttaaagtgac cctgccggat ggaaaacaag ttgatgccga gtcctggaaa 300
acgacaccgt atcagattgc ctgtggtatc agtcaaggtc tggccgacaa tacctgtgatt 360
gctaaagtga ataacgtggt gtgggactct gaccgtcctc tggaggaaga ttgtaccctg 420
gagctgctga aatttgagga tgaggaagca caggccgttt attggcatag cagcggccac 480
atcatgggag aagcaatgga acgtgtgtat ggtgggtgtc tgtgtatgg tccgccgatt 540
gaaaatggct tctattatga catgtatctg gaagagggtg gtgttagtag taacgacttt 600
tcgtcgctgg aagctctgtg taaaaaaatc atcaaagaga aacaagcctt tgaacgcctg 660
gaagtgaaaa aagaaaccct gctggcaatg ttcaaataca acaaattcaa atgccgtatc 720
ctgaacgaga aagtgaacac cccgaccaca actgtttatc gctgtggccc actgattgat 780
ctgtgccgtg gtccacacgt ccgtcatacc ggcaaatca aagccctgaa aatccacaaa 840
aacagctcga cgtattggga aggcaaagcg gatatggaga ctctgcagcg cttttatggc 900
atctcattcc cggacccaaa aatgctgaaa gagtgggaga aattccaaga agagggcгаа 960
aaccgtgacc tcgag 975

```

5

<210> 46

<211> 447

<212> ADN

10 <213> Secuencia artificial

<220>

<223> Polinucleótido de treonil ARNt sintetasa optimizado por codones (ADN2.0)

15 <400> 46

```

tttgaagaga aagccagctc tcctagcggg aagatgggtg gtgaagagaa gccgatcggg 60
gctggcgaag aaaagcagaa agaaggcggc aagaagaaaa acaaagaggg ttccgggtgac 120
ggcggctcgt cggaactgaa tccgtggccg gagtacattt acaccctct ggagatgtat 180
aacatcctga aagcagaaca cgacagcatc ctggcagaga aggcggagaa ggacagcaaa 240

ccgatcaaag ttactctgcc ggacggcaag caagtggatg cagagagctg gaaaaccacc 300
ccgtatcaga ttgctgtgtg tattagccaa ggcttggcgg ataacacggt gattgccaag 360
gtcaataatg tcgtttggga cctggatcgc ccaactggaag aggattgcac gctggagctg 420
ttgaaattcg aggattaatg actcgag 447

```

20

<210> 47

<211> 996

<212> ADN

25 <213> Secuencia artificial

<220>

<223> Polinucleótido de treonil ARNt sintetasa optimizado por codones

30 <400> 47

ES 2 638 311 T3

```

ttcgaagaga aagcaagctc tccaagcggg aagatgggcg gtgaagagaa gccgattggt 60
gctggcgaag agaagcaaaa agagggtggt aagaagaaga ataaagaggg cagcggcgat 120
ggtggccgtg cggaaactgaa cccgtggccg gagtatatct acaccctctt ggagatgtac 180
aacattctga aagcggaaaca tgacagcatt ctggcggaaa aggccgagaa agacagcaa 240
ccgatcaaaag ttaccctgcc ggatggttaag caggtggatg ccgaatcgtg gaaaacgacg 300
ccgtaccaga ttgcctgcgg catcagccaa ggtctggctg ataacacggt tatcgcgaag 360
gtcaacaatg tcgtgtggga cttggaccgc ccgttgggaag aggactgcac cctggagctg 420
ctgaaattcg aggacgaaga agcgcaggcg gtctattggc acagcagcgc acacatcatg 480
ggtgaagcaa tggagcgcgt ctacggtggt tgctgtggtt acggtccgcc gattgagaat 540
ggtttctact atgacatgta tctggaagaa ggcggcgttt ccagcaacga cttctctagc 600
ctggaggcgc tgtgtaagaa gattatcaaa gagaaacagg cgttcgagcg tttggagggtg 660
aagaaagaaa ctctgctggc gatgtttaag tacaacaagt ttaagtccg tattctgaac 720
gagaaagtta atacgcgcac caccaccgtt taccgttgtg gtccgttgat cgacgtgtg 780
cgtggtcctc acgtgcgcca taccggcaag atcaaggcac tgaaaattca caagaattcc 840
agcacctatt gggagggcaa ggccgatatg gaaacgctgc aacgcatcta tggatcagc 900
ttccagatc cgaaaatgct gaaagaatgg gagaaatttc aagaagaggc gaaaaatcgt 960
gatcatcgca aaattggccg tccgtaatga ctcgag 996

```

<210> 48

<211> 363

5 <212> ADN

<213> Secuencia artificial

<220>

<223> Polinucleótido de treonil ARNt sintetasa optimizado por codones

10

<400> 48

```

ggatcctttg aggaaaaagc gtcaagccct tctggaaaaa tgggaggcga agagaaaccg 60
atcggagccg gcgaagagaa acagaaagag gggggcaaaa aaaaaaacia agaaggctcg 120
ggcgacggag ggcgtgctga actgaacca tggccggagt atatctatac ccgcctggaa 180
atgtataaca ttctgaaagc ggaacacgac agcattctgg ccgaaaaagc ggagaaagat 240
tccaaaccga tcaaagtcac actgcctgat ggcaaacagg tggacgccga aagctggaaa 300
accacgcctt atcaaatcgc ttgtggcatt cgctgtattg gtactctggt tctgacgctc 360
gag 363

```

15 <210> 49

<211> 447

<212> ADN

<213> Secuencia artificial

20 <220>

<223> Polinucleótido de treonil ARNt sintetasa optimizado por codones (Genewiz)

<400> 49

```

ggatccttcg aagagaaagc ctcaagtccg agtggcaaaa tgggtggcga ggaaaaacca 60
atcggagccg gtgaagaaaa acagaaagag ggcggcaaaa aaaaaaacia agaaggcagc 120
ggcgacggag gtcgtgctga actgaatcct tggccggagt atatctatac ccgcctggag 180
atgtataaca ttctgaaagc cgagcatgat tccattctgg ccgaaaagc cgagaaagac 240
tcaaaaccga tcaaagtgac tctgccagat ggcaacaag ttgatgccga atcttggaaa 300
acaaccctt atcagatcgc ctgtggtatc agtcagggac tggccgacaa tactgttatt 360
gccaaagtca acaacgttgt gtgggatctg gaccgtcctc tgggaagaaga ttgtaccctg 420
gaactgctga aattcgagga tctcgag 447

```

25

<210> 50

<211> 417

30 <212> PRT

<213> *Homo sapiens*

<400> 50

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

<210> 51  
 <211> 1254  
 <212> ADN  
 <213> *Homo sapiens*

5

<400> 51

```

atggtgaaag agtggggagaa gttccaagag gaagctaaaa accgagatca taggaaaatt 60
ggcagggacc aagaactata tttctttcat gaactcagcc ctggaagttg cttttttctg 120
ccaaaaggag cctacattta taatgcactt attgaattca ttaggagcga atataggaaa 180
agaggattcc aggaggtagt caccocaaac atcttcaaca gccgactctg gatgacctcg 240
ggccactggc agcactacag cgagaacatg ttctcctttg aggtggagaa ggagctgttt 300
gccctgaaac ccatgaactg cccaggacac tgccttatgt ttgatcatcg gccaaaggtcc 360
tggcgagaac tgcctctgcg gctagctgat tttggggtac ttcataggaa cgagctgtct 420
ggagcactca caggactcac ccgggtacga agattccaac aggatgatgc tcacatattc 480
tgtgccatgg agcagattga agatgaaata aaaggttgtt tggattttct acgtacggta 540
tatagcgtat ttggattttc ttttaacta aacctttcta ctgcccgga aaaattcctt 600
ggagatatcg aagtatggga tcaagctgag aaacaacttg aaaacagtct gaatgaattt 660
ggtgaaaagt gggagttaa ctctggagat ggagctttct atggcccaa gattgacata 720
cagattaaag atgcgattgg gcggtaccac cagtgtgcaa ccatccagct ggatttccag 780
ttgccatca gatttaactt tacttatgta agccatgatg gtgatgataa gaaaaggcca 840
gtgattgttc atcgagccat cttgggatca gtggaaagaa tgattgctat cctcacagaa 900
aactatgggg gcaaatggcc cttttggctg tcccctcgcc aggtaatggt agttccagtg 960
ggaccaacct gtgatgaata tgccaaaag gtacgacaac aattccacga tgccaaattc 1020
atggcagaca ttgatctgga tccaggctgt acattgaata aaaagattcg aaatgcacag 1080
ttagcacagt ataacttcat tttagttgtt ggtgaaaaag agaaaatcag tggcactgtt 1140
aatatccgca caagagacaa taaggtccac ggggaacgca ccatttctga aactatcgag 1200
cggctacagc agctcaaaga gttccgcagc aaacaggcag aagaagaatt ttaa 1254

```

<210> 52  
 <211> 378  
 <212> PRT  
 <213> *Homo sapiens*

5

<400> 52

```

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

```

10

225					230					235				240	
Pro	Val	Ile	Val	His	Arg	Ala	Ile	Leu	Gly	Ser	Val	Glu	Arg	Met	Ile
				245					250					255	
Ala	Ile	Leu	Thr	Glu	Asn	Tyr	Gly	Gly	Lys	Trp	Pro	Phe	Trp	Leu	Ser
			260				265						270		
Pro	Arg	Gln	Val	Met	Val	Val	Pro	Val	Gly	Pro	Thr	Cys	Asp	Glu	Tyr
		275					280					285			
Ala	Gln	Lys	Val	Arg	Gln	Gln	Phe	His	Asp	Ala	Lys	Phe	Met	Ala	Asp
	290				295					300					
Ile	Asp	Leu	Asp	Pro	Gly	Cys	Thr	Leu	Asn	Lys	Lys	Ile	Arg	Asn	Ala
305					310				315						320
Gln	Leu	Ala	Gln	Tyr	Asn	Phe	Ile	Leu	Val	Val	Gly	Glu	Lys	Glu	Lys
				325				330						335	
Ile	Ser	Gly	Thr	Val	Asn	Ile	Arg	Thr	Arg	Asp	Asn	Lys	Val	His	Gly
			340				345						350		
Glu	Arg	Thr	Ile	Ser	Glu	Thr	Ile	Glu	Arg	Leu	Gln	Gln	Leu	Lys	Glu
		355					360					365			
Phe	Arg	Ser	Lys	Gln	Ala	Glu	Glu	Glu	Phe						
	370					375									

<210> 53  
 <211> 1137  
 <212> ADN  
 <213> *Homo sapiens*

5

<400> 53

```

ctgccaaaag gagcctacat ttataatgca cttattgaat tcattaggag cgaatatagg 60
aaaagaggat tccaggaggt agtcacccca aacatcttca acagccgact ctggatgacc 120
tcgggccact ggcagcacta cagcgagaac atgttctcct ttgaggtgga gaaggagctg 180
tttgccctga aacccatgaa ctgcccagga cactgcctta tgtttgatca tcggccaagg 240
tcctggcgag aactgcctct ggggctagct gattttgggg tacttcatag gaacgagctg 300
tctggagcac tcacaggact caccgggta cgaagattcc aacaggatga tgctcacata 360
ttctgtgcca tggagcagat tgaagatgaa ataaaaggtt gtttggattt tctacgtacg 420
gtatatagcg tatttggatt ttcttttaa ctaaaccttt ctactgccc ggaaaaattc 480
cttggagata tcgaagtatg ggatcaagct gagaacaac ttgaaaacag tctgaatgaa 540
tttggtgaaa agtgggagtt aaactctgga gatggagctt tctatggccc aaagattgac 600
atacagatta aagatgcgat tgggcggtac caccagtgtg caaccatcca gctggatttc 660
cagttgcca tcagatttaa tottacttat gtaagccatg atggtgatga taagaaaagg 720
ccagtgattg ttcctcgagc catcttggga tcagtggaaa gaatgattgc tctcctcaca 780
gaaaactatg ggggcaaatg gcccttttgg ctgtcccctc gccaggtaat ggtagttcca 840
gtgggaccaa cctgtgatga atatgcccaa aaggtacgac aacaattcca cgatgccaaa 900
ttcatggcag acattgatct ggatccaggc tgtacattga ataaaaagat tcgaaatgca 960
cagttagcac agtataactt cattttagtt gttggtgaaa aagagaaaat cagtggcact 1020
gttaatatcc gcacaagaga caataaggtc cacggggaac gcaccatttc tgaaactatc 1080
gagcggctac agcagctcaa agagttccgc agcaaacagg cagaagaaga attttaa 1137
    
```

10

<210> 54  
 <211> 292  
 <212> PRT  
 <213> *Homo sapiens*

15

<400> 54

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

<210> 55  
 <211> 879  
 <212> ADN  
 <213> *Homo sapiens*

<400> 55

ctgCGGctag ctgatttttg ggtacttcat aggaacgagc tgtctggagc actcacagga 60  
 ctcaccCGgg tacgaagatt ccaacaggat gatgctcaca tattctgtgc catggagcag 120  
 attgaagatg aaataaaagg ttgtttggat tttctacgta cggtatatag cgtatttggg 180  
 ttttctttta aactaaacct ttctactcgc ccggaaaaat tccttggaga tctcgaagta 240  
 tgggatcaag ctgagaaaca acttgaaaac agtctgaatg aatttggatg aaagtgggag 300  
 ttaaaactctg gagatggagc tttctatggc ccaaagattg acatacagat taaagatgcg 360  
 attgggCGgt accaccagtg tgcaaccatc cagctggatt tccagttgcc catcagattt 420  
 aatcttactt atgtaagcca tgatggtgat gataagaaaa ggccagtgat tgttcacgca 480  
 gccatcttgg gatcagtgga aagaatgatt gctatcctca cagaaaacta tgggggcaaa 540  
 tggccctttt ggctgtcccc tcgccaggta atggtagttc cagtgggacc aacctgtgat 600  
 gaatatgccc aaaaggtagc acaacaattc cagcatgcca aattcatggc agacattgat 660  
 ctggatccag gctgtacatt gaataaaaag attcgaatg cacagttagc acagtataac 720  
 ttcatcttag ttgttggatg aaaagagaaa atcagtggca ctgttaatat ccgcacaaga 780  
 gacaataagg tccacgggga acgcaccatt tctgaaacta tcgagcggct acagcagctc 840  
 aaagagttcc gcagcaaca gccagaaga gaattttaa 879

<210> 56  
 <211> 344  
 <212> PRT  
 <213> *Homo sapiens*

5

<400> 56

```

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

10

<210> 57  
 <211> 1035  
 <212> ADN  
 <213> *Homo sapiens*

15

<400> 57

ES 2 638 311 T3

```

agccgactct ggatgacctc gggccactgg cagcactaca gcgagaacat gttctccttt 60
gaggtggaga aggagctggt tgccttgaaa cccatgaact gcccaggaca ctgccttatg 120
tttgatcatc ggccaaggtc ctggcgagaa ctgcctctgc ggctagctga ttttggggta 180
cttcatagga acgagctgtc tggagcactc acaggactca cccgggtacg aagattccaa 240
caggatgatg ctccacatatt ctgtgccatg gagcagattg aagatgaaat aaaaggttgt 300
ttggattttc tacgtacggg atatagcgta tttggatttt cttttaaact aaacctttct 360
actcgccögg aaaaattcct tggagatata gaagtatggg atcaagctga gaaacaactt 420
gaaaacagtc tgaatgaatt tggtgaaaag tgggagttaa actctggaga tggagctttc 480
tatggcccaa agattgacat acagattaaa gatgögatg ggöggtagca ccagtgtgca 540
accatccagc tggatttcca gttgöccatc agatttaatc ttacttatgt aagccatgat 600
ggtgatgata agaaaaggcc agtgattggt catcgagcca tcttgggatc agtggaaaga 660
atgattgcta tcctcacaga aaactatggg ggcaaatggc ccttttggct gtcccctcgc 720
caggtaatgg tagttccagt gggaccaacc tgtgatgaat atgccccaaa ggtacgacaa 780
caattccagc atgöcaaatt catggcagac attgatctgg atccaggctg tacattgaat 840
aaaaagattc gaaatgcaca gttagcacag tataacttca ttttagttgt tggtgaaaaa 900
gagaaaatca gtggcactgt taatatccgc acaagagaca ataaggcca cggggaacgc 960
accatttctg aaactatcga göggctacag cagctcaaag agttccgag caaacaggca 1020
gaagaagaat tttaa 1035

```

<210> 58  
 <211> 430  
 <212> PRT  
 <213> *Homo sapiens*  
  
 <400> 58

5

Leu Gln Arg Ile Tyr Gly Ile Ser Phe Pro Asp Pro Lys Met Leu Lys  
 1 5 10 15  
 Glu Trp Glu Lys Phe Gln Glu Glu Ala Lys Asn Arg Asp His Arg Lys  
 20 25 30  
 Ile Gly Arg Asp Gln Glu Leu Tyr Phe Phe His Glu Leu Ser Pro Gly  
 35 40 45  
 Ser Cys Phe Phe Leu Pro Lys Gly Ala Tyr Ile Tyr Asn Ala Leu Ile  
 50 55 60  
 Glu Phe Ile Arg Ser Glu Tyr Arg Lys Arg Gly Phe Gln Glu Val Val  
 65 70 75 80  
 Thr Pro Asn Ile Phe Asn Ser Arg Leu Trp Met Thr Ser Gly His Trp  
 85 90 95  
 Gln His Tyr Ser Glu Asn Met Phe Ser Phe Glu Val Glu Lys Glu Leu  
 100 105 110  
 Phe Ala Leu Lys Pro Met Asn Cys Pro Gly His Cys Leu Met Phe Asp  
 115 120 125  
 His Arg Pro Arg Ser Trp Arg Glu Leu Pro Leu Arg Leu Ala Asp Phe  
 130 135 140  
 Gly Val Leu His Arg Asn Glu Leu Ser Gly Ala Leu Thr Gly Leu Thr  
 145 150 155 160  
 Arg Val Arg Arg Phe Gln Gln Asp Asp Ala His Ile Phe Cys Ala Met  
 165 170 175  
 Glu Gln Ile Glu Asp Glu Ile Lys Gly Cys Leu Asp Phe Leu Arg Thr  
 180 185 190  
 Val Tyr Ser Val Phe Gly Phe Ser Phe Lys Leu Asn Leu Ser Thr Arg  
 195 200 205  
 Pro Glu Lys Phe Leu Gly Asp Ile Glu Val Trp Asp Gln Ala Glu Lys  
 210 215 220  
 Gln Leu Glu Asn Ser Leu Asn Glu Phe Gly Glu Lys Trp Glu Leu Asn  
 225 230 235 240  
 Ser Gly Asp Gly Ala Phe Tyr Gly Pro Lys Ile Asp Ile Gln Ile Lys  
 245 250 255  
 Asp Ala Ile Gly Arg Tyr His Gln Cys Ala Thr Ile Gln Leu Asp Phe  
 260 265 270  
 Gln Leu Pro Ile Arg Phe Asn Leu Thr Tyr Val Ser His Asp Gly Asp  
 275 280 285  
 Asp Lys Lys Arg Pro Val Ile Val His Arg Ala Ile Leu Gly Ser Val  
 290 295 300  
 Glu Arg Met Ile Ala Ile Leu Thr Glu Asn Tyr Gly Gly Lys Trp Pro  
 305 310 315 320  
 Phe Trp Leu Ser Pro Arg Gln Val Met Val Val Pro Val Gly Pro Thr  
 325 330 335  
 Cys Asp Glu Tyr Ala Gln Lys Val Arg Gln Gln Phe His Asp Ala Lys  
 340 345 350  
 Phe Met Ala Asp Ile Asp Leu Asp Pro Gly Cys Thr Leu Asn Lys Lys  
 355 360 365  
 Ile Arg Asn Ala Gln Leu Ala Gln Tyr Asn Phe Ile Leu Val Val Gly  
 370 375 380  
 Glu Lys Glu Lys Ile Ser Gly Thr Val Asn Ile Arg Thr Arg Asp Asn  
 385 390 395 400  
 Lys Val His Gly Glu Arg Thr Ile Ser Glu Thr Ile Glu Arg Leu Gln  
 405 410 415  
 Gln Leu Lys Glu Phe Arg Ser Lys Gln Ala Glu Glu Glu Phe  
 420 425 430

<210> 59  
 <211> 1293  
 <212> ADN  
 <213> *Homo sapiens*

5

<400> 59

ES 2 638 311 T3

```

ctccagagaa tttatggcat ttcattocca gatcctaana tgttgaaaga gtgggagaag 60
ttccaagagg aagctaaaaa ccgagatcat aggaaaattg gcagggacca agaactatat 120
ttctttcatg aactcagccc tgggaagttgc ttttttctgc caaaaggagc ctacatttat 180
aatgcactta ttgaattcat taggagcgaa tataggaaaa gaggattcca ggaggtagtc 240
accccaaaaca tcttcaacag ccgactctgg atgacctcgg gccactggca gcactacagc 300
gagaacatgt tctcctttga ggtggagaag gagctgttt ccctgaaacc catgaactgc 360
ccaggacact gccttatggt tgatcatcgg ccaaggctct ggcgagaact gcctctgagg 420
ctagctgatt ttgggtact tcataggaac gagctgtctg gagcactcac aggactcacc 480
cgggtacgaa gattccaaca ggatgatgct cacatattct gtgcatgga gcagattgaa 540
gatgaaataa aaggttgttt ggattttcta cgtacggtat atagcgtatt tggattttct 600
tttaactaa acctttctac tcgcccggaa aaattccttg gagatatcga agtatgggat 660
caagctgaga aacaacttga aaacagtctg aatgaatttg gtgaaaagtg ggagttaaac 720
tctggagatg gagctttcta tggoccaaag attgacatac agattaaaga tgcgattggg 780
cggtagacc agtgtgcaac catccagctg gatttcagtg tggccatcag atttaactct 840
acttatgtaa gccatgatgg tgatgataag aaaaggccag tgattgttca tgcgagccatc 900
ttgggatcag tggaaagaat gattgctatc ctcacagaaa actatggggg caaatggccc 960
ttttggctgt cccctcgcca ggtaatggta gttccagttg gaccaacctg tgatgaatat 1020
gcccaaaagg tacgacaaca attccacgat gccaaattca tggcagacat tgatctggat 1080
ccaggctgta cattgaataa aaagattcga aatgcacagt tagcacagta taacttcatt 1140
ttagttgttg gtgaaaaaga gaaaatcagt ggcactgtta atatccgcac aagagacaat 1200
aaggtccacg gggaaacgac catttctgaa actatcgagc ggctacagca gctcaaagag 1260
ttccgcagca aacaggcgaga agaagaattt taa 1293

```

<210> 60  
 <211> 20  
 <212> PRT  
 <213> *Mus musculus*

5

<400> 60

```

    Asp Gln Glu Leu Tyr Phe Phe His Glu Leu Ser Pro Gly Ser Cys Phe
    1           5           10           15
    Phe Leu Pro Lys
    20

```

10

<210> 61  
 <211> 13  
 <212> PRT  
 <213> *Mus musculus*

15

<400> 61

```

    Gly Ala Tyr Ile Tyr Asn Thr Leu Met Glu Phe Ile Arg
    1           5           10

```

20

<210> 62  
 <211> 81  
 <212> PRT  
 <213> *Mus musculus*

25

<400> 62

ES 2 638 311 T3

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

5 <210> 63  
 <211> 12  
 <212> PRT  
 <213> *Mus musculus*

<400> 63

10 Asn Glu Leu Ser Gly Ala Leu Thr Gly Leu Thr Arg  
 1 5 10

15 <210> 64  
 <211> 3  
 <212> PRT  
 <213> *Mus musculus*

<400> 64

Val Arg Arg  
 1

20 <210> 65  
 <211> 20  
 <212> PRT  
 <213> *Mus musculus*

<400> 65

Phe Gln Gln Asp Asp Ala His Ile Phe Cys Ala Met Glu Gln Ile Glu  
 1 5 10 15  
 Asp Glu Ile Lys  
 20

30 <210> 66  
 <211> 149  
 <212> PRT  
 <213> *Mus musculus*

35 <400> 66

ES 2 638 311 T3

```

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

```

5 <210> 67  
 <211> 13  
 <212> PRT  
 <213> *Mus musculus*

<400> 67

```

10      Gly Ala Tyr Ile Tyr Asn Thr Leu Met Glu Phe Ile Arg
          1          5          10

```

15 <210> 68  
 <211> 224  
 <212> PRT  
 <213> *Mus musculus*

<400> 68

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

<210> 69  
 <211> 25  
 5 <212> PRT  
 <213> *Mus musculus*

<400> 69

Pro Val Ile Val His Arg Ala Ile Leu Gly Ser Val Glu Arg Met Ile  
 1 5 10 15  
 Ala Ile Leu Thr Glu Asn Tyr Gly Gly  
 20 25

10  
 <210> 70  
 <211> 262  
 <212> PRT  
 15 <213> *Mus musculus*

<400> 70

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

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

5 <210> 71  
 <211> 12  
 <212> PRT  
 <213> *Mus musculus*

<400> 71

10 Asn Glu Leu Ser Gly Ala Leu Thr Gly Leu Thr Arg  
 1 5 10

15 <210> 72  
 <211> 165  
 <212> PRT  
 <213> *Mus musculus*

<400> 72



ES 2 638 311 T3

```

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

5 <210> 75  
 <211> 12  
 <212> PRT  
 <213> *Mus musculus*

<400> 75

```

Asn Glu Leu Ser Gly Ala Leu Thr Gly Leu Thr Arg
10           5           10
    
```

15 <210> 76  
 <211> 145  
 <212> PRT  
 <213> *Mus musculus*

<400> 76

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

<210> 77  
 <211> 12  
 <212> PRT  
 <213> *Mus musculus*

5

<400> 77

Met Ile Ala Ile Leu Thr Glu Asn Tyr Gly Gly Lys  
 1 5 10

10

<210> 78  
 <211> 169  
 <212> PRT  
 <213> *Mus musculus*

15

<400> 78

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

20

ES 2 638 311 T3

<210> 79  
 <211> 10  
 <212> PRT  
 <213> *Mus musculus*

5 <400> 79

Ile Tyr Gly Ile Ser Phe Pro Asp Pro Lys  
 1 5 10

10 <210> 80  
 <211> 61  
 <212> PRT  
 <213> *Mus musculus*

15 <400> 80

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

20 <210> 81  
 <211> 14  
 <212> PRT  
 <213> *Mus musculus*

25 <400> 81

Gly Phe Gln Glu Val Val Thr Pro Asn Ile Phe Asn Ser Arg  
 1 5 10

30 <210> 82  
 <211> 218  
 <212> PRT  
 <213> *Mus musculus*

<400> 82

```

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

```

5 <210> 83  
 <211> 12  
 <212> PRT  
 <213> *Mus musculus*

10 <400> 83

```

Met Ile Ala Ile Leu Thr Glu Asn Tyr Gly Gly Lys
 1          5          10

```

15 <210> 84  
 <211> 315  
 <212> PRT  
 <213> *Mus musculus*

<400> 84

```

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

```

<210> 85  
 <211> 124  
 <212> PRT  
 <213> *Homo sapiens*

5

<400> 85

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

<210> 86  
 <211> 375  
 5 <212> ADN  
 <213> *Homo sapiens*

<400> 86

atgattgcta tcctcacaga aaactatggg ggcaaattggc ccttttggct gtcccctcgc 60  
 caggtaatgg tagttccagt gggaccaacc tgtgatgaat atgcccacaaa ggtacgacaa 120  
 caattccacg atgccaat catggcagac attgatctgg atccaggctg tacattgaat 180  
 aaaaagattc gaaatgcaca gttagcacag tataacttca ttttagttgt tggtgaaaaa 240  
 gagaaaatca gtggcactgt taatatccgc acaagagaca ataaggtcca cggggaacgc 300  
 accatttctg aaactatcga gcggctacag cagctcaaag agttccgcag caaacaggca 360  
 10 gaagaagaat ttttaa 375

<210> 87  
 <211> 696  
 <212> PRT  
 15 <213> *Homo sapiens*

<400> 87

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

Ala Leu Cys Lys Lys Ile Ile Lys Glu Lys Gln Ala Phe Glu Arg Leu  
 180 185 190  
 Glu Val Lys Lys Glu Thr Leu Leu Ala Met Phe Lys Tyr Asn Lys Phe  
 195 200 205  
 Lys Cys Arg Ile Leu Asn Glu Lys Val Asn Thr Pro Thr Thr Thr Val  
 210 215 220  
 Tyr Arg Cys Gly Pro Leu Ile Asp Leu Cys Arg Gly Pro His Val Arg  
 225 230 235 240  
 His Thr Gly Lys Ile Lys Ala Leu Lys Ile His Lys Asn Ser Ser Thr  
 245 250 255  
 Tyr Trp Glu Gly Lys Ala Asp Met Glu Thr Leu Gln Arg Ile Tyr Gly  
 260 265 270  
 Ile Ser Phe Pro Asp Pro Lys Met Leu Lys Glu Trp Glu Lys Phe Gln  
 275 280 285  
 Glu Glu Ala Lys Asn Arg Asp His Arg Lys Ile Gly Arg Asp Gln Glu  
 290 295 300  
 Leu Tyr Phe Phe His Glu Leu Ser Pro Gly Ser Cys Phe Phe Leu Pro  
 305 310 315 320  
 Lys Gly Ala Tyr Ile Tyr Asn Ala Leu Ile Glu Phe Ile Arg Ser Glu  
 325 330 335  
 Tyr Arg Lys Arg Gly Phe Gln Glu Val Val Thr Pro Asn Ile Phe Asn  
 340 345 350  
 Ser Arg Leu Trp Met Thr Ser Gly His Trp Gln His Tyr Ser Glu Asn  
 355 360 365  
 Met Phe Ser Phe Glu Val Glu Lys Glu Leu Phe Ala Leu Lys Pro Met  
 370 375 380  
 Asn Cys Pro Gly His Cys Leu Met Phe Asp His Arg Pro Arg Ser Trp  
 385 390 395 400  
 Arg Glu Leu Pro Leu Arg Leu Ala Asp Phe Gly Val Leu His Arg Asn  
 405 410 415  
 Glu Leu Ser Gly Ala Leu Thr Gly Leu Thr Arg Val Arg Arg Phe Gln  
 420 425 430  
 Gln Asp Asp Ala His Ile Phe Cys Ala Met Glu Gln Ile Glu Asp Glu  
 435 440 445  
 Ile Lys Gly Cys Leu Asp Phe Leu Arg Thr Val Tyr Ser Val Phe Gly  
 450 455 460  
 Phe Ser Phe Lys Leu Asn Leu Ser Thr Arg Pro Glu Lys Phe Leu Gly  
 465 470 475 480  
 Asp Ile Glu Val Trp Asp Gln Ala Glu Lys Gln Leu Glu Asn Ser Leu  
 485 490 495  
 Asn Glu Phe Gly Glu Lys Trp Glu Leu Asn Ser Gly Asp Gly Ala Phe  
 500 505 510  
 Tyr Gly Pro Lys Ile Asp Ile Gln Ile Lys Asp Ala Ile Gly Arg Tyr  
 515 520 525  
 His Gln Cys Ala Thr Ile Gln Leu Asp Phe Gln Leu Pro Ile Arg Phe  
 530 535 540  
 Asn Leu Thr Tyr Val Ser His Asp Gly Asp Asp Lys Lys Arg Pro Val  
 545 550 555 560  
 Ile Val His Arg Ala Ile Leu Gly Ser Val Glu Arg Met Ile Ala Ile  
 565 570 575  
 Leu Thr Glu Asn Tyr Gly Gly Lys Trp Pro Phe Trp Leu Ser Pro Arg  
 580 585 590  
 Gln Val Met Val Val Pro Val Gly Pro Thr Cys Asp Glu Tyr Ala Gln  
 595 600 605  
 Lys Val Arg Gln Gln Phe His Asp Ala Lys Phe Met Ala Asp Ile Asp  
 610 615 620  
 Leu Asp Pro Gly Cys Thr Leu Asn Lys Lys Ile Arg Asn Ala Gln Leu  
 625 630 635 640  
 Ala Gln Tyr Asn Phe Ile Leu Val Val Gly Glu Lys Glu Lys Ile Ser  
 645 650 655  
 Gly Thr Val Asn Ile Arg Thr Arg Asp Asn Lys Val His Gly Glu Arg  
 660 665 670  
 Thr Ile Ser Glu Thr Ile Glu Arg Leu Gln Gln Leu Lys Glu Phe Arg  
 675 680 685  
 Ser Lys Gln Ala Glu Glu Glu Phe

<210> 88  
 <211> 2091  
 <212> ADN  
 <213> *Homo sapiens*

5

<400> 88

```

atgtttgagg agaaggccag cagtccttca ggaagatgg gaggcgagga gaagccgttg 60
aatccttggc ctgaatatat ttacacacgt cttgagatgt ataataact aaaagcagaa 120
catgattcca ttctggcaga aaaggcagaa aaagatagca agccaattaa agtcactttg 180
cctgatggta aacaggttga tgcggaatct tggaaaacta caccatatca aattgcctgt 240
ggaattagtc aaggcctggc cgacaacacc gttattgcta aagtaaataa tgttgtgtgg 300
gacctggacc gccctctgga agaagattgt accttgagc ttctcaagtt tgaggatgag 360
gaagctcagg cagtgtattg gcaactctag gctcacataa tgggtgaagc catggaaaga 420
gtctatggtg gatgtttatg ctacgggtcc ccaatagaaa atggattcta ttatgacatg 480
tacctcgaag aagggggtgt gtctagcaat gatttctctt ctctggaggc tttgtgtaag 540
aaaatcatta aagaaaaaca agcttttgaa agactggaag ttaagaaaga aactttactg 600
gcaatgttta agtacaacaa gttcaaatgc cggatattga atgaaaaggt gaatactcca 660
actaccacag tctatagatg tggccctttg atagatctct gccgggggtcc tcatgttaga 720
cacacgggca aaattaaggc tttaaaaata cacaaaaatt cctccacgta ctgggaaggc 780
aaagcagata tggagactct ccagagaatt tatggcattt cattcccaga tcoataaatg 840
ttgaaagagt gggagaagtt ccagaggaa gctaaaaacc gagatcatag gaaaattggc 900
agggaccaag aactatattt ctttcatgaa ctccagcctg gaagttgctt tttctgcca 960
aaaggagcct acatttataa tgcacttatt gaattcatta ggagcgaata taggaaaaga 1020
ggattccagg aggtagtcac cccaaacatc ttcaacagcc gactctggat gacctcgggc 1080
cactggcagc actacagcga gaacatgttc tcctttgagg tggagaagga gctgtttgcc 1140
ctgaaaccca tgaactgcc aggacactgc cttatgtttg atcatcggcc aaggtcctgg 1200
cgagaactgc ctctgcccgt agctgatttt ggggtacttc ataggaacga gctgtctgga 1260
gcactcacag gactcaccgg ggtacgaaga ttccaacagg atgatgctca catattctgt 1320
gccatggagc agattgaaga tgaataaaaa ggttgtttgg attttctacg tacggtatat 1380
agcgtatttg gattttcttt taaactaaac ctttctactc gcccgaaaaa attccttggg 1440
gatatcgaag tatgggatca agctgagaaa caacttgaaa acagtctgaa tgaatttggg 1500
gaaaagtggg agttaaactc tggagatgga gctttctatg gcccaaagat tgacatacag 1560
attaagatg cgattgggcg gtaccaccag tgtgcaacca tccagctgga tttccagttg 1620
cccatcagat ttaatcttac ttatgtaagc catgatggtg atgataagaa aaggccagtg 1680
attgttcacg gagccatctt gggatcagtg gaaagaatga ttgctatcct cacagaaaac 1740
tatgggggca aatggccctt ttggctgtcc cctcgccagg taatggtagt tccagtgagg 1800
ccaacctgtg atgaatatgc ccaaaaggta cgacaacaat tccacgatgc caaattcatg 1860
gcagacattg atctggatcc aggctgtaca ttgaataaaa agattcgaaa tgcacagtta 1920
gcacagtata acttcatttt agttgttggg gaaaaagaga aatcagtggt cactgttaat 1980
atccgcacaa gagacaataa ggtccacggg gaacgcacca tttctgaaac tatcgagcgg 2040
ctacagcagc tcaaagagtt ccgcagcaaa caggcagaag aagaatttta a 2091
    
```

10 <210> 89  
 <211> 563  
 <212> PRT  
 <213> *Homo sapiens*

15 <400> 89

ES 2 638 311 T3

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

Lys Ala Leu Lys Ile His Lys Asn Ser Ser Thr Tyr Trp Glu Gly Lys  
 115 120 125  
 Ala Asp Met Glu Thr Leu Gln Arg Ile Tyr Gly Ile Ser Phe Pro Asp  
 130 135 140  
 Pro Lys Met Leu Lys Glu Trp Glu Lys Phe Gln Glu Glu Ala Lys Asn  
 145 150 155 160  
 Arg Asp His Arg Lys Ile Gly Arg Asp Gln Glu Leu Tyr Phe Phe His  
 165 170 175  
 Glu Leu Ser Pro Gly Ser Cys Phe Phe Leu Pro Lys Gly Ala Tyr Ile  
 180 185 190  
 Tyr Asn Ala Leu Ile Glu Phe Ile Arg Ser Glu Tyr Arg Lys Arg Gly  
 195 200 205  
 Phe Gln Glu Val Val Thr Pro Asn Ile Phe Asn Ser Arg Leu Trp Met  
 210 215 220  
 Thr Ser Gly His Trp Gln His Tyr Ser Glu Asn Met Phe Ser Phe Glu  
 225 230 235 240  
 Val Glu Lys Glu Leu Phe Ala Leu Lys Pro Met Asn Cys Pro Gly His  
 245 250 255  
 Cys Leu Met Phe Asp His Arg Pro Arg Ser Trp Arg Glu Leu Pro Leu  
 260 265 270  
 Arg Leu Ala Asp Phe Gly Val Leu His Arg Asn Glu Leu Ser Gly Ala  
 275 280 285  
 Leu Thr Gly Leu Thr Arg Val Arg Arg Phe Gln Gln Asp Asp Ala His  
 290 295 300  
 Ile Phe Cys Ala Met Glu Gln Ile Glu Asp Glu Ile Lys Gly Cys Leu  
 305 310 315 320  
 Asp Phe Leu Arg Thr Val Tyr Ser Val Phe Gly Phe Ser Phe Lys Leu  
 325 330 335  
 Asn Leu Ser Thr Arg Pro Glu Lys Phe Leu Gly Asp Ile Glu Val Trp  
 340 345 350  
 Asp Gln Ala Glu Lys Gln Leu Glu Asn Ser Leu Asn Glu Phe Gly Glu  
 355 360 365  
 Lys Trp Glu Leu Asn Ser Gly Asp Gly Ala Phe Tyr Gly Pro Lys Ile  
 370 375 380  
 Asp Ile Gln Ile Lys Asp Ala Ile Gly Arg Tyr His Gln Cys Ala Thr  
 385 390 395 400  
 Ile Gln Leu Asp Phe Gln Leu Pro Ile Arg Phe Asn Leu Thr Tyr Val  
 405 410 415  
 Ser His Asp Gly Asp Asp Lys Lys Arg Pro Val Ile Val His Arg Ala  
 420 425 430  
 Ile Leu Gly Ser Val Glu Arg Met Ile Ala Ile Leu Thr Glu Asn Tyr  
 435 440 445  
 Gly Gly Lys Trp Pro Phe Trp Leu Ser Pro Arg Gln Val Met Val Val  
 450 455 460  
 Pro Val Gly Pro Thr Cys Asp Glu Tyr Ala Gln Lys Val Arg Gln Gln  
 465 470 475 480  
 Phe His Asp Ala Lys Phe Met Ala Asp Ile Asp Leu Asp Pro Gly Cys  
 485 490 495  
 Thr Leu Asn Lys Lys Ile Arg Asn Ala Gln Leu Ala Gln Tyr Asn Phe  
 500 505 510  
 Ile Leu Val Val Gly Glu Lys Glu Lys Ile Ser Gly Thr Val Asn Ile  
 515 520 525  
 Arg Thr Arg Asp Asn Lys Val His Gly Glu Arg Thr Ile Ser Glu Thr  
 530 535 540  
 Ile Glu Arg Leu Gln Gln Leu Lys Glu Phe Arg Ser Lys Gln Ala Glu  
 545 550 555 560  
 Glu Glu Phe

<210> 90  
 <211> 1692  
 <212> ADN  
 <213> *Homo sapiens*

<400> 90

```

atgggtgaag ccatggaaag agtctatggt ggatgtttat gctacgggtcc gccaatagaa 60
aatggattct attatgacat gtacctcgaa gaaggggggtg tgtctagcaa tgatttctct 120
tctctggagg ctttctgtaa gaaaatcatt aaagaaaaac aagcttttga aagactggaa 180
gttaagaaag aaactttact ggcaatgttt aagtacaaca agttcaaagt ccggatattg 240
aatgaaaagg tgaatactcc aactaccaca gtctatagat gtggcccttt gatagatctc 300
tgccgggggc ctcatgtag acacacgggc aaaattaagg ctttaaaaat acacaaaaat 360
tcctccacgt actgggaagg caaagcagat atgggagactc tccagagaat ttatggcatt 420
tcattcccag atcctaaaat gttgaaagag tgggagaagt tccaagagga agctaaaaac 480
cgagatcata ggaaaattgg cagggaccaaa gaactatatt tctttcatga actcagccct 540
ggaagttgct ttttctgccc aaaaggagcc tacatttata atgcacttat tgaattcatt 600
aggactcgaat ataggaagaagg aggattccag gaggtagtca ccccaaacat ctctttgag 660
cgactctgga tgacctcggg ccactggcag cactacagcg agaactggt ctctttgag 720
gtggagaagg agctgtttgc cctgaaaccc atgaaactgcc caggacactg ccttatgttt 780
gatcatcggc caaggtcctg gcgagaactg cctctgccc tagctgattt tggggtactt 840
cataggaacg agctgtctgg agcactcaca ggactcaccg ggtacgaag attccaacag 900
gatgatgctc acatattctg tgccatggag cagattgaag atgaaataaa aggttgtttg 960
gattttctac gtacggtata tagcgtattt ggattttctt ttaaactaaa ctttctact 1020
cgcccggaaa aattccttgg agatatcgaa gtatgggatc aagctgagaa acaacttgaa 1080
aacagtctga atgaatttgg tgaaaagtgg gagttaaact ctggagatgg agctttctat 1140
ggcccaaaga ttgacataca gattaagat gcgattgggc ggtaccacca gtgtgcaacc 1200
atccagctgg atttccagtt gcccatcaga tttaacttta cttatgtaag ccatgatggt 1260
gatgataaga aaaggccagt gattgttcat cgagccatct tgggatcagt ggaaagaatg 1320
attgctatcc tcacagaaaa ctatgggggc aatggccct tttggctgtc cctctgccag 1380
gtaatggtag ttccagtggg accaacctgt gatgaatatg cccaaaaggt acgacaacaa 1440
ttccacgatg ccaaattcat ggcagacatt gatctggatc caggctgtac attgaataaa 1500
aagattcgaa atgcacagtt agcacagtat aacttcattt tagttgttgg tgaaaaagag 1560
aaaatcagtg gcactgttaa tatccgcaca agagacaata aggtccacgg ggaacgcacc 1620
atctctgaaa ctatcgagcg gctacagcag ctcaaagagt tccgcagcaa acaggcagaa 1680
gaagaattht aa 1692
    
```

5 <210> 91  
 <211> 50  
 <212> ADN  
 <213> *Homo sapiens*

10 <400> 91  
 ccgagatcat aggaaaattg gcaggccatg atggtgatga taagaaaagg 50

<210> 92

15 <400> 92  
 000

<210> 93  
 <211> 50  
 <212> ADN  
 <213> *Homo sapiens*

<400> 93  
 gaagatggga ggcgaggaga agccgttgaa tccttgccct gaatatattt 50

25 <210> 94  
 <211> 16  
 <212> PRT  
 <213> *Homo sapiens*

30 <400> 94

Lys Met Gly Gly Glu Glu Lys Pro Leu Asn Pro Trp Pro Glu Tyr Ile  
 1 . 5 10 15

## ES 2 638 311 T3

<210> 95  
<211> 50  
<212> ADN  
<213> *Homo sapiens*

5

<400> 95  
catatcaaat tgcctgtgga attagtgta ttggactct agtgctcaca 50

10

<210> 96

<400> 96  
000

15

<210> 97  
<211> 437  
<212> PRT  
<213> Secuencia artificial

20

<220>  
<223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad N-terminal de 6xHis

<400> 97

```

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu
 1      5      10
Gly Leu Asp Ser Thr Leu Lys Glu Trp Glu Lys Phe Gln Glu Glu Ala
      20      25      30
Lys Asn Arg Asp His Arg Lys Ile Gly Arg Asp Gln Glu Leu Tyr Phe
      35      40      45
Phe His Glu Leu Ser Pro Gly Ser Cys Phe Phe Leu Pro Lys Gly Ala
      50      55      60
Tyr Ile Tyr Asn Ala Leu Ile Glu Phe Ile Arg Ser Glu Tyr Arg Lys
65      70      75      80
Arg Gly Phe Gln Glu Val Val Thr Pro Asn Ile Phe Asn Ser Arg Leu
      85      90      95
Trp Met Thr Ser Gly His Trp Gln His Tyr Ser Glu Asn Met Phe Ser
      100      105      110
Phe Glu Val Glu Lys Glu Leu Phe Ala Leu Lys Pro Met Asn Cys Pro
      115      120      125
Gly His Cys Leu Met Phe Asp His Arg Pro Arg Ser Trp Arg Glu Leu
      130      135      140
Pro Leu Arg Leu Ala Asp Phe Gly Val Leu His Arg Asn Glu Leu Ser
145      150      155      160
Gly Ala Leu Thr Gly Leu Thr Arg Val Arg Arg Phe Gln Gln Asp Asp
      165      170      175
Ala His Ile Phe Cys Ala Met Glu Gln Ile Glu Asp Glu Ile Lys Gly
      180      185      190
Cys Leu Asp Phe Leu Arg Thr Val Tyr Ser Val Phe Gly Phe Ser Phe
      195      200      205
Lys Leu Asn Leu Ser Thr Arg Pro Glu Lys Phe Leu Gly Asp Ile Glu
      210      215      220
Val Trp Asp Gln Ala Glu Lys Gln Leu Glu Asn Ser Leu Asn Glu Phe
225      230      235      240
Gly Glu Lys Trp Glu Leu Asn Ser Gly Asp Gly Ala Phe Tyr Gly Pro
      245      250      255
Lys Ile Asp Ile Gln Ile Lys Asp Ala Ile Gly Arg Tyr His Gln Cys
      260      265      270
Ala Thr Ile Gln Leu Asp Phe Gln Leu Pro Ile Arg Phe Asn Leu Thr
      275      280      285
Tyr Val Ser His Asp Gly Asp Lys Lys Arg Pro Val Ile Val His
      290      295      300
Arg Ala Ile Leu Gly Ser Val Glu Arg Met Ile Ala Ile Leu Thr Glu
305      310      315      320
Asn Tyr Gly Gly Lys Trp Pro Phe Trp Leu Ser Pro Arg Gln Val Met
      325      330      335
Val Val Pro Val Gly Pro Thr Cys Asp Glu Tyr Ala Gln Lys Val Arg
      340      345      350
Gln Gln Phe His Asp Ala Lys Phe Met Ala Asp Ile Asp Leu Asp Pro
      355      360      365
Gly Cys Thr Leu Asn Lys Lys Ile Arg Asn Ala Gln Leu Ala Gln Tyr
      370      375      380
Asn Phe Ile Leu Val Val Gly Glu Lys Glu Lys Ile Ser Gly Thr Val
385      390      395      400
Asn Ile Arg Thr Arg Asp Asn Lys Val His Gly Glu Arg Thr Ile Ser
      405      410      415
Glu Thr Ile Glu Arg Leu Gln Gln Leu Lys Glu Phe Arg Ser Lys Gln
      420      425      430
Ala Glu Glu Glu Phe
      435

```

<210> 98  
 <211> 437  
 <212> PRT  
 <213> Secuencia artificial

5

&lt;220&gt;

&lt;223&gt; Polipéptido de treonil ARNt sintetasa con marcador de afinidad C-terminal de 6xHis

&lt;400&gt; 98

5

```

Met Leu Lys Glu Trp Glu Lys Phe Gln Glu Glu Ala Lys Asn Arg Asp
 1      5      10
His Arg Lys Ile Gly Arg Asp Gln Glu Leu Tyr Phe Phe His Glu Leu
 20      25      30
Ser Pro Gly Ser Cys Phe Phe Leu Pro Lys Gly Ala Tyr Ile Tyr Asn
 35      40      45
Ala Leu Ile Glu Phe Ile Arg Ser Glu Tyr Arg Lys Arg Gly Phe Gln
 50      55      60
Glu Val Val Thr Pro Asn Ile Phe Asn Ser Arg Leu Trp Met Thr Ser
 65      70      75      80
Gly His Trp Gln His Tyr Ser Glu Asn Met Phe Ser Phe Glu Val Glu
 85      90      95
Lys Glu Leu Phe Ala Leu Lys Pro Met Asn Cys Pro Gly His Cys Leu
 100     105
Met Phe Asp His Arg Pro Arg Ser Trp Arg Glu Leu Pro Leu Arg Leu
 115     120     125
Ala Asp Phe Gly Val Leu His Arg Asn Glu Leu Ser Gly Ala Leu Thr
 130     135     140
Gly Leu Thr Arg Val Arg Arg Phe Gln Gln Asp Asp Ala His Ile Phe
 145     150     155     160
Cys Ala Met Glu Gln Ile Glu Asp Glu Ile Lys Gly Cys Leu Asp Phe
 165     170     175
Leu Arg Thr Val Tyr Ser Val Phe Gly Phe Ser Phe Lys Leu Asn Leu
 180     185     190
Ser Thr Arg Pro Glu Lys Phe Leu Gly Asp Ile Glu Val Trp Asp Gln
 195     200     205
Ala Glu Lys Gln Leu Glu Asn Ser Leu Asn Glu Phe Gly Glu Lys Trp
 210     215     220
Glu Leu Asn Ser Gly Asp Gly Ala Phe Tyr Gly Pro Lys Ile Asp Ile
 225     230     235     240
Gln Ile Lys Asp Ala Ile Gly Arg Tyr His Gln Cys Ala Thr Ile Gln
 245     250     255
Leu Asp Phe Gln Leu Pro Ile Arg Phe Asn Leu Thr Tyr Val Ser His
 260     265     270
Asp Gly Asp Asp Lys Lys Arg Pro Val Ile Val His Arg Ala Ile Leu
 275     280     285
Gly Ser Val Glu Arg Met Ile Ala Ile Leu Thr Glu Asn Tyr Gly Gly
 290     295     300
Lys Trp Pro Phe Trp Leu Ser Pro Arg Gln Val Met Val Val Pro Val
 305     310     315     320
Gly Pro Thr Cys Asp Glu Tyr Ala Gln Lys Val Arg Gln Gln Phe His
 325     330     335
Asp Ala Lys Phe Met Ala Asp Ile Asp Leu Asp Pro Gly Cys Thr Leu
 340     345
Asn Lys Lys Ile Arg Asn Ala Gln Leu Ala Gln Tyr Asn Phe Ile Leu
 355     360     365
Val Val Gly Glu Lys Glu Lys Ile Ser Gly Thr Val Asn Ile Arg Thr
 370     375     380
Arg Asp Asn Lys Val His Gly Glu Arg Thr Ile Ser Glu Thr Ile Glu
 385     390     395     400
Arg Leu Gln Gln Leu Lys Glu Phe Arg Ser Lys Gln Ala Glu Glu Glu
 405     410     415
Phe Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr His
 420     425     430
His His His His
 435

```

<210> 99  
 <211> 403  
 <212> PRT  
 <213> Secuencia artificial

5

<220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad N-terminal de 6xHis

<400> 99

10

```

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

<210> 100  
 <211> 403  
 <212> PRT  
 <213> Secuencia artificial

5

<220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad C-terminal de 6xHis

<400> 100

10

```

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

```

ES 2 638 311 T3

<210> 101  
 <211> 313  
 <212> PRT  
 <213> Secuencia artificial

5

<220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad N-terminal de 6xHis

<400> 101

10

```

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

<210> 102  
 <211> 313  
 <212> PRT  
 <213> Secuencia artificial

15

<220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad C-terminal de 6xHis

<400> 102

20

ES 2 638 311 T3

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

<210> 103  
 <211> 369  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad N-terminal de 6xHis

<400> 103

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
 1 5 10 15

ES 2 638 311 T3

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

<210> 104  
 <211> 369  
 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad C-terminal de 6xHis

<400> 104

```

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

```

5 <210> 105  
 <211> 451  
 <212> PRT  
 <213> Secuencia artificial

10 <220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad N-terminal de 6xHis

<400> 105

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu  
1 5 10 15  
Gly Leu Asp Ser Thr Leu Gln Arg Ile Tyr Gly Ile Ser Phe Pro Asp  
20 25 30  
Pro Lys Met Leu Lys Glu Trp Glu Lys Phe Gln Glu Glu Ala Lys Asn  
35 40 45  
Arg Asp His Arg Lys Ile Gly Arg Asp Gln Glu Leu Tyr Phe Phe His  
50 55 60  
Glu Leu Ser Pro Gly Ser Cys Phe Phe Leu Pro Lys Gly Ala Tyr Ile  
65 70 75 80  
Tyr Asn Ala Leu Ile Glu Phe Ile Arg Ser Glu Tyr Arg Lys Arg Gly  
85 90 95  
Phe Gln Glu Val Val Thr Pro Asn Ile Phe Asn Ser Arg Leu Trp Met  
100 105 110  
Thr Ser Gly His Trp Gln His Tyr Ser Glu Asn Met Phe Ser Phe Glu  
115 120 125  
Val Glu Lys Glu Leu Phe Ala Leu Lys Pro Met Asn Cys Pro Gly His  
130 135 140  
Cys Leu Met Phe Asp His Arg Pro Arg Ser Trp Arg Glu Leu Pro Leu  
145 150 155 160  
Arg Leu Ala Asp Phe Gly Val Leu His Arg Asn Glu Leu Ser Gly Ala  
165 170 175  
Leu Thr Gly Leu Thr Arg Val Arg Arg Phe Gln Gln Asp Asp Ala His  
180 185 190  
Ile Phe Cys Ala Met Glu Gln Ile Glu Asp Glu Ile Lys Gly Cys Leu  
195 200 205  
Asp Phe Leu Arg Thr Val Tyr Ser Val Phe Gly Phe Ser Phe Lys Leu  
210 215 220  
Asn Leu Ser Thr Arg Pro Glu Lys Phe Leu Gly Asp Ile Glu Val Trp  
225 230 235 240  
Asp Gln Ala Glu Lys Gln Leu Glu Asn Ser Leu Asn Glu Phe Gly Glu  
245 250 255  
Lys Trp Glu Leu Asn Ser Gly Asp Gly Ala Phe Tyr Gly Pro Lys Ile  
260 265 270  
Asp Ile Gln Ile Lys Asp Ala Ile Gly Arg Tyr His Gln Cys Ala Thr  
275 280 285  
Ile Gln Leu Asp Phe Gln Leu Pro Ile Arg Phe Asn Leu Thr Tyr Val  
290 295 300  
Ser His Asp Gly Asp Asp Lys Lys Arg Pro Val Ile Val His Arg Ala  
305 310 315 320  
Ile Leu Gly Ser Val Glu Arg Met Ile Ala Ile Leu Thr Glu Asn Tyr  
325 330 335  
Gly Gly Lys Trp Pro Phe Trp Leu Ser Pro Arg Gln Val Met Val Val  
340 345 350  
Pro Val Gly Pro Thr Cys Asp Glu Tyr Ala Gln Lys Val Arg Gln Gln  
355 360 365  
Phe His Asp Ala Lys Phe Met Ala Asp Ile Asp Leu Asp Pro Gly Cys  
370 375 380  
Thr Leu Asn Lys Lys Ile Arg Asn Ala Gln Leu Ala Gln Tyr Asn Phe  
385 390 395 400  
Ile Leu Val Val Gly Glu Lys Glu Lys Ile Ser Gly Thr Val Asn Ile  
405 410 415  
Arg Thr Arg Asp Asn Lys Val His Gly Glu Arg Thr Ile Ser Glu Thr  
420 425 430  
Ile Glu Arg Leu Gln Gln Leu Lys Glu Phe Arg Ser Lys Gln Ala Glu  
435 440 445  
Glu Glu Phe  
450

<210> 106  
<211> 451

## ES 2 638 311 T3

<212> PRT

<213> Secuencia artificial

<220>

5 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad C-terminal de 6xHis

<400> 106

Met	Leu	Gln	Arg	Ile	Tyr	Gly	Ile	Ser	Phe	Pro	Asp	Pro	Lys	Met	Leu
1				5					10					15	
Lys	Glu	Trp	Glu	Lys	Phe	Gln	Glu	Glu	Ala	Lys	Asn	Arg	Asp	His	Arg
			20					25					30		
Lys	Ile	Gly	Arg	Asp	Gln	Glu	Leu	Tyr	Phe	Phe	His	Glu	Leu	Ser	Pro
		35					40					45			
Gly	Ser	Cys	Phe	Phe	Leu	Pro	Lys	Gly	Ala	Tyr	Ile	Tyr	Asn	Ala	Leu
	50					55				60					
Ile	Glu	Phe	Ile	Arg	Ser	Glu	Tyr	Arg	Lys	Arg	Gly	Phe	Gln	Glu	Val
65				70						75					80
Val	Thr	Pro	Asn	Ile	Phe	Asn	Ser	Arg	Leu	Trp	Met	Thr	Ser	Gly	His
			85						90					95	
Trp	Gln	His	Tyr	Ser	Glu	Asn	Met	Phe	Ser	Phe	Glu	Val	Glu	Lys	Glu
			100					105					110		
Leu	Phe	Ala	Leu	Lys	Pro	Met	Asn	Cys	Pro	Gly	His	Cys	Leu	Met	Phe
		115					120					125			
Asp	His	Arg	Pro	Arg	Ser	Trp	Arg	Glu	Leu	Pro	Leu	Arg	Leu	Ala	Asp
		130				135					140				
Phe	Gly	Val	Leu	His	Arg	Asn	Glu	Leu	Ser	Gly	Ala	Leu	Thr	Gly	Leu
145				150						155					160
Thr	Arg	Val	Arg	Arg	Phe	Gln	Gln	Asp	Asp	Ala	His	Ile	Phe	Cys	Ala
			165						170					175	
Met	Glu	Gln	Ile	Glu	Asp	Glu	Ile	Lys	Gly	Cys	Leu	Asp	Phe	Leu	Arg
			180					185					190		
Thr	Val	Tyr	Ser	Val	Phe	Gly	Phe	Ser	Phe	Lys	Leu	Asn	Leu	Ser	Thr
		195				200						205			
Arg	Pro	Glu	Lys	Phe	Leu	Gly	Asp	Ile	Glu	Val	Trp	Asp	Gln	Ala	Glu
	210					215					220				
Lys	Gln	Leu	Glu	Asn	Ser	Leu	Asn	Glu	Phe	Gly	Glu	Lys	Trp	Glu	Leu
225				230						235					240
Asn	Ser	Gly	Asp	Gly	Ala	Phe	Tyr	Gly	Pro	Lys	Ile	Asp	Ile	Gln	Ile
			245						250					255	
Lys	Asp	Ala	Ile	Gly	Arg	Tyr	His	Gln	Cys	Ala	Thr	Ile	Gln	Leu	Asp
			260					265					270		
Phe	Gln	Leu	Pro	Ile	Arg	Phe	Asn	Leu	Thr	Tyr	Val	Ser	His	Asp	Gly
		275					280					285			
Asp	Asp	Lys	Lys	Arg	Pro	Val	Ile	Val	His	Arg	Ala	Ile	Leu	Gly	Ser
	290					295					300				
Val	Glu	Arg	Met	Ile	Ala	Ile	Leu	Thr	Glu	Asn	Tyr	Gly	Gly	Lys	Trp
305				310							315				320
Pro	Phe	Trp	Leu	Ser	Pro	Arg	Gln	Val	Met	Val	Val	Pro	Val	Gly	Pro
			325						330					335	
Thr	Cys	Asp	Glu	Tyr	Ala	Gln	Lys	Val	Arg	Gln	Gln	Phe	His	Asp	Ala
			340					345					350		
Lys	Phe	Met	Ala	Asp	Ile	Asp	Leu	Asp	Pro	Gly	Cys	Thr	Leu	Asn	Lys
		355					360					365			
Lys	Ile	Arg	Asn	Ala	Gln	Leu	Ala	Gln	Tyr	Asn	Phe	Ile	Leu	Val	Val
	370					375					380				
Gly	Glu	Lys	Glu	Lys	Ile	Ser	Gly	Thr	Val	Asn	Ile	Arg	Thr	Arg	Asp
385				390						395					400
Asn	Lys	Val	His	Gly	Glu	Arg	Thr	Ile	Ser	Glu	Thr	Ile	Glu	Arg	Leu
			405						410					415	
Gln	Gln	Leu	Lys	Glu	Phe	Arg	Ser	Lys	Gln	Ala	Glu	Glu	Glu	Phe	Gly
			420					425					430		
Lys	Pro	Ile	Pro	Asn	Pro	Leu	Leu	Gly	Leu	Asp	Ser	Thr	His	His	His
		435					440					445			
His	His	His													
	450														

ES 2 638 311 T3

<211> 148  
 <212> PRT  
 <213> Secuencia artificial

5 <220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad N-terminal de 6xHis

<400> 107

```

Met His His His His His His Gly Lys Pro Ile Pro Asn Pro Leu Leu
 1          5          10          15

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

10

<210> 108  
 <211> 148  
 <212> PRT  
 <213> Secuencia artificial

15

<220>  
 <223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad C-terminal de 6xHis

<400> 108

20

```

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

<210> 109  
 <211> 1260  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> Polinucleótido de treonil ARNt sintetasa optimizado por codones

<400> 109

10

```

ctgaaagaat gggagaagtt ccaggaagag gccaaaaacc gcgatcatcg taagattggc 60
cgtgaccagg agctgtactt ttccacagag ctgagcccgg gtagctgctt tttcctgccg 120
aaaggcgcgt acatctataa cgctttgatc gaattcatcc gcagcagta ccgtaagcgt 180
ggttttcaag aggtcgtcac tccgaatate ttcaacagcc gtctgtggat gacttctggg 240
cactggcagc actacagcga gaatatgttt agcttcgagg ttgaaaaaga gctgtttggc 300
ctgaagccga tgaactgtcc gggccatttc ctgatgttog atcacccgcc gcgttcctgg 360
cgcgagctgc cgctgcgcct ggcggatttt ggtgtgttgc accgcaatga gctgagcggg 420
gcgctgaccg gtctgaccoc gtttcgcctt tttcagcaag atgacgcgca cattttctgc 480
gcgatggagc aaatcgagga tgagattaag ggttgtctgg acttctctgc taccgtgtat 540
tcggttttgc gctttagctt taaactgaac ttgtctacc gtccggaaaa gtttttggg 600
gacattgagg tctgggatca agcagagaag caactggaga acagcctgaa cgaattoggt 660
gaaaagtggg aactgaatag cggtagcggc gcgttctacg gtccctaagat cgacatocaa 720
atcaaagatg cgatcggccg ttatcaccag tgcgctacga tccagttgga cttccagttg 780
ccaatccgct tcaacctgac ctatgtgtcg catgacggcg acgacaagaa acgtccggtc 840
atcgttcacg gtgcaattct gggtagcgtc gagcgcgatg ttgcgattct gaccgagaat 900
tatggtggta aatggccggt ttggctgtcc cctcgccagg tgatgggggt tccggttggc 960
ccgacctgtg atgaatagc acagaaagtt cgtcaacaat ttcacgatgc aaagtttatg 1020
gcgacattg atctggatcc gggttgcacg ctgaataaga aattcgtaa tgcccagctg 1080
gcccaataca acttcatect ggtcgtgggc gagaaagaga agattagcgg cacggttaac 1140
attcgtacgc gtgacaataa ggtgcatggg gaacgtacca ttagcgaaac gattgaacgt 1200
ctgcagcaac tgaagaatt ccgcagcaaa caggcagaag aagaattcta atgactcgag 1260
    
```

<210> 110  
 <211> 1146  
 <212> ADN  
 <213> Secuencia artificial

15

<220>  
 <223> Polinucleótido de treonil ARNt sintetasa optimizado por codones

20

<400> 110

```

ggatccctgc cgaagggcgc ttatatctat aatgccctga tcgagtttat tcgctcggag 60
tatcgtaaac gtggctttca ggaagttggt acgcctaaca tttttaacag tcgcctgtgg 120
atgacatctg gtcactggca acactattct gagaacatgt tcagcttcca agtggagaaa 180
gagctgtttg ctctgaaacc gatgaattgt ccgggtcatt gtctgatggt tgatcaccgt 240
cctcgttctt ggcgtgaact gccactgcgt ctggctgact ttgggtgtgct gcaccgtaat 300
gaactgtctg gtgctctgac aggtctgaca cgtgtccgtc gtttccagca ggatgatgcc 360
cacatctttt gtgccatgga gcagatcgaa gatgagatca aaggttgccg ggattttctg 420
cgcacagtggt atagcgtggt cgggttttcc ttcaactga atctgagtac ccgtccagag 480
aaatttctgg gcgatatcga ggtttgggac caagccgaaa aacaactgga gaacagcctg 540
aacgaatttg gcgagaaatg ggaactgaac tcgggtgatg gtgcctttta tggccctaaa 600
atcgacatcc agattaaaga cgcattggc cgctatcatc agtgtgccac cattcagctg 660
gactttcagc tgcctattcg cttcaacctg acctatgta gccatgatgg agatgacaaa 720
aaacgtccgg tcatcgtaca ccgtgctatc ctgggttctg ttgagcgtat gattgccatt 780
ctgaccgaga attatggcgg aaaatggccg ttttggtgtt cacctcgtca agtgatgggt 840
gttccgggtg gaccaacttg tgatgagtat gctcaaaaag ttcgtcaaca atttcacgat 900
gccaaattca tggccgacat cgatctggac cctgggtgta ccctgaacaa aaaaatccgt 960
aacgcccagc tggcacagta taactttatc ctggctcgtg gcgaaaaaga aaaaatcagc 1020
ggcaccgtta atattcgtac tcgtgataac aaagttcatg gggagcgtac cattagtgag 1080
actatcgaac gtctgcaaca gctgaaagag ttccgctcta aacaggcaga agaggagttc 1140
ctcgag 1146
    
```

<210> 111  
 <211> 888  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> Polinucleótido de treonil ARNt sintetasa optimizado por codones

<400> 111

10

```

ctgcgcttgg ctgacttcgg cgtcctgcac cgcaacgaac tgagcgggtgc gctgaccggc 60
ctgaccgctg tgcgtcgctt ccaacaagat gatgcacaca tttctgcgc gatggagcag 120
atcgaggacg aatcaaggg ttgtctggac tttctgcgta cggtttacag cgtttttggt 180
ttcagcttta aactgaatct gagcacgcgt cctgagaagt tcctgggtga catcgaagtg 240

tgggatcagg cagaaaagca attggagaat agcctgaatg agttcggcga gaaatgggag 300
ttgaactctg gcgacggcgc gttctacggt ccgaagatcg atattcagat taaagacgcc 360
atcggtcggt accatcagtg cgctacgatt cagctggatt ttcagctgcc gattcgtttc 420
aatctgacct atgtttcca cgatggtgat gacaagaaac gcccggtcat cgtgcatcgc 480
gcgattctgg gtagcggtga gcgcatgatc gcgattttga ccgagaatta tggcggtaaa 540
tggccattct ggctgagccc gcgtcaagtt atggtggtcc cggttggtcc gacctgcgat 600
gagtatgcgc agaaggtccg tcaacagttc catgacgcca agtttatggc agatatacgc 660
ttggaccggy gttgtacct gaacaaaag atccgtaatg cgcagctggc ccaatacaac 720
ttatcctggy tggtaggtga aaaagagaag atttcggcga cggtgaacat tcgcaccgct 780
gacaacaaag tccacggcga acgtaccatc agcgaacta ttgaacgtct gcaacaactg 840
aaagagtttc gttctaaaca ggcagaagaa gagttttaat gactcgag 888
    
```

<210> 112  
 <211> 1044  
 <212> ADN  
 <213> Secuencia artificial

15

<220>  
 <223> Polinucleótido de treonil ARNt sintetasa optimizado por codones

<400> 112

20

```

ggatcctctc gtctgtggat gacttctggc cattggcagc attatagcga aaacatgttc 60
agcttcgagg ttgagaaaga gctgtttgcc ctgaaaccga tgaattgtcc gggtcactgt 120
ctgatgttcg atcacgctcc tcgttcttgg cgtgaaactgc ctctgcgcct ggctgatttt 180
ggtgttctgc accgcaatga actgtcaggt gctctgactg gtctgactcg tgttcgtcgt 240
ttccagcagg atgatgcca cattttctgt gccatggagc aaattgaaga tgagatcaaa 300
gggtgcctgg atttctgcg taccgtgtat agcgtattcg gcttctcctt taaactgaat 360
ctgtctacct gtccgaaaaa atttctgggg gacatcgagg tctgggatca agcggaaaaa 420
caactggaga acagcctgaa cgaattcggc gaaaaatggg aactgaactc cggagatggg 480
gccttttatg gtcctaaaaat tgatatccag atcaaagacg ccattgggcy ttatcatcag 540
tgtgccacca ttcagctgga ctttcaactg cctattcgtc tcaatctgac ctatgtgagc 600
catgatgggy acgacaaaaa acgtcctgtg attgtacacc gtgccattct gggttctgtg 660
gagcgtatga ttgccattct gaccgagaac tatgggggca aatggccttt ttggctgtct 720
cctcgtcagg ttatggttgt gccggttggg ccaacttgtg atgagtatgc tcagaaagtt 780
cgccagcaat tccacgatgc caaattcatg gccgatatcg acctggacce tggctgtacc 840
ctgaacaaaa aaatccgtaa tgcccagctg gcacagtata acttcatcct ggtagtgggc 900
gaaaaagaaa aaatccgtaa gaccgtgaat attcgtaccg gtgataataa agttcatggg 960
gagcgtacca tttcagaaac catcgagcgc ctgcagcaac tgaagaggtt tcgtagtaaa 1020
caggccgaag aggaattcct cgag 1044
    
```

25

<210> 113  
 <211> 1302  
 <212> ADN  
 <213> Secuencia artificial

30

<220>  
 <223> Polinucleótido de treonil ARNt sintetasa optimizado por codones

ES 2 638 311 T3

<400> 113

```

ctgcagcgca totacggcat ctcgttccca gaccctaaaa tgttgaaaga gtgggaaaag 60
ttccaagaag aggcgaaaaa ccgatgatcac cgcaagattg gtcgtgatca agagctgtat 120
ttcttccacg agctgtcccc gggttcgtgc ttcttccctgc cgaagggtgc gtacatctac 180
aatgctctga ttgagtttat tctgtagcgag tatcgtgaagc ggggcttcca ggaagttgtc 240
acgccgaata tctttaacag ccgctctgtgg atgaccagcg gtcattggca gcactacagc 300
gagaatatgt ttagctttga agttgaaaaa gagctgttcg cgctgaagcc gatgaattgt 360
ccaggtcact gcctgatggt cgaccaccgt ccgctgtctt ggcgtgaatt gccgctgcgc 420
ctggccgact tccgctgtgt gcacgtaac gaactgagcg gtgcgctgac cggctctgacc 480
cgtgtgcgcc gttttcagca ggatgatcgg catatctttt gcgcaatgga gcaaatcgag 540
gacgagatca agggttgttt ggacttcctg cgcacgggtg atagcgtttt tggttttagc 600
tttaaactga acctgtctac ccgctcggag aaattcctgg gcgacattga ggtctgggac 660
caggcagaaa agcaactgga gaacagcttg aatgagtttg gcgagaaatg ggagctgaat 720
agcggcgacg gtgccttcta cggccogaag attgatatcc agatcaaaga tgcgattggc 780
cgttatcacc agtgtgccac gatccagttg gactttcagc tgcgattcgc ctttaactctg 840

acctacgtga gccatgatgg cgatgataag aaacgtccgg tcattgtcca tctgtcgtatt 900
ctgggtagcg ttgaacgcat gatcgctatt ctgaccgaaa actacgggtg taagtggccg 960
ttctggttgt ccccgctca ggtgatggtg gttccggttg gtcctacgtg cgacgagtat 1020
gcgcaaaaag ttcgccaaca gttccaacgac gcaaagtta tggcagacat cgatctggac 1080
ccgggctgca ctttgaacaa gaagatccgt aatgcccaac tggcacaata caactttatt 1140
ctggtcgtgg gtgaaaaaga gaagattagc ggtactgtta acatccgtac ccgtgataac 1200
aaagtccacg gcgaacgcac catttccgaa acgattgagc gtctgcagca actgaaagaa 1260
ttccgcagca aacaagcgga agaagagttc taatgactcg ag 1302

```

5

<210> 114  
 <211> 381  
 <212> ADN  
 <213> Secuencia artificial

10

<220>  
 <223> Polinucleótido de treonil ARNt sintetasa optimizado por codones

15

<400> 114

```

ggatccatcg ctattctgac cgaaaattat ggcggcaaat ggcctttttg gctgtctcct 60
cgtcaagtca tggttgttcc ggttgggtccg acatgtgatg agtatgcca gaaagtccgt 120
cagcagtttc atgatgcaa attcatggcc gatatcgacc tggaccctgg ctgtacactg 180
aacaaaaaaa tccgtaacgc ccagctggct cagtataact tcatcctggt ggtgggtgag 240
aaagagaaaa tcagcggcac cgtcaacatt cgtaccogtg acaataaagt ccacggcgag 300
cgtacaattt cagagactat tgagcgcctg caacagctga aagaatttgc ctccaaacaa 360
gocgaagaag agttcctcga g

```

20

<210> 115  
 <211> 474  
 <212> PRT  
 <213> *Homo sapiens*

<400> 115

Glu 1 Glu Ala Gln Ala Val Tyr Trp His Ser Ser Ala His Ile Met Gly  
 5 10 15  
 Glu 20 Ala Met Glu Arg Val Tyr Gly Gly Cys Leu Cys Tyr Gly Pro Pro  
 25 30  
 Ile 35 Glu Asn Gly Phe Tyr Tyr Asp Met Tyr Leu Glu Glu Gly Gly Val  
 40 45  
 Ser 50 Ser Asn Asp Phe Ser Ser Leu Glu Ala Leu Cys Lys Lys Ile Ile  
 55 60  
 Lys 65 Glu Lys Gln Ala Phe Glu Arg Leu Glu Val Lys Lys Glu Thr Leu  
 70 75 80  
 Leu 85 Ala Met Phe Lys Tyr Asn Lys Phe Lys Cys Arg Ile Leu Asn Glu  
 90 95  
 Lys 100 Val Asn Thr Pro Thr Thr Thr Val Tyr Arg Cys Gly Pro Leu Ile  
 105 110  
 Asp 115 Leu Cys Arg Gly Pro His Val Arg His Thr Gly Lys Ile Lys Ala  
 120 125  
 Leu 130 Lys Ile His Lys Asn Ser Ser Thr Tyr Trp Glu Gly Lys Ala Asp  
 135 140  
 Met 145 Glu Thr Leu Gln Arg Ile Tyr Gly Ile Ser Phe Pro Asp Pro Lys  
 150 155 160  
 Met 165 Leu Lys Glu Trp Glu Lys Phe Gln Glu Glu Ala Lys Asn Arg Asp  
 170 175  
 His 180 Arg Lys Ile Gly Arg Asp Gln Glu Leu Tyr Phe Phe His Glu Leu  
 185 190  
 Ser 195 Pro Gly Ser Cys Phe Phe Leu Pro Lys Gly Ala Tyr Ile Tyr Asn  
 200 205  
 Ala 210 Leu Ile Glu Phe Ile Arg Ser Glu Tyr Arg Lys Arg Gly Phe Gln  
 215 220  
 Glu 225 Val Val Thr Pro Asn Ile Phe Asn Ser Arg Leu Trp Met Thr Ser  
 230 235 240  
 Gly 245 His Trp Gln His Tyr Ser Glu Asn Met Phe Ser Phe Glu Val Glu  
 250 255  
 Lys 260 Glu Leu Phe Ala Leu Lys Pro Met Asn Cys Pro Gly His Cys Leu  
 265 270  
 Met 275 Phe Asp His Arg Pro Arg Ser Trp Arg Glu Leu Pro Leu Arg Leu  
 280 285  
 Ala 290 Asp Phe Gly Val Leu His Arg Asn Glu Leu Ser Gly Ala Leu Thr  
 295 300  
 Gly 305 Leu Thr Arg Val Arg Arg Phe Gln Gln Asp Asp Ala His Ile Phe  
 310 315 320  
 Cys 325 Ala Met Glu Gln Ile Glu Asp Glu Ile Lys Gly Cys Leu Asp Phe  
 330 335  
 Leu 340 Arg Thr Val Tyr Ser Val Phe Gly Phe Ser Phe Lys Leu Asn Leu  
 345 350  
 Ser 355 Thr Arg Pro Glu Lys Phe Leu Gly Asp Ile Glu Val Trp Asp Gln  
 360 365  
 Ala 370 Glu Lys Gln Leu Glu Asn Ser Leu Asn Glu Phe Gly Glu Lys Trp  
 375 380  
 Glu 385 Leu Asn Ser Gly Asp Gly Ala Phe Tyr Gly Pro Lys Ile Asp Ile  
 390 395 400  
 Gln 405 Ile Lys Asp Ala Ile Gly Arg Tyr His Gln Cys Ala Thr Ile Gln  
 410 415  
 Leu 420 Asp Phe Gln Leu Pro Ile Arg Phe Asn Leu Thr Tyr Val Ser His  
 425 430  
 Asp 435 Gly Asp Asp Lys Lys Arg Pro Val Ile Val His Arg Ala Ile Leu  
 440 445  
 Gly 450 Ser Val Glu Arg Met Ile Ala Ile Leu Thr Glu Asn Tyr Gly Gly  
 455 460  
 Lys 465 Trp Pro Phe Trp Leu Ser Pro Arg Gln  
 470

ES 2 638 311 T3

<210> 116  
 <211> 1422  
 <212> ADN  
 <213> *Homo sapiens*

5 <400> 116

```

gaggaagctc aggcagtgta ttggcactct agtgcctcaca taatgggtga agccatggaa 60
agagtctatg gtggatgttt atgctacggg cgcaccaatag aaaatggatt ctattatgac 120
atgtacctcg aagaaggggg tgtgtctagc aatgatttct cttctctgga ggctttgtgt 180
aagaaaatca ttaaagaaaa acaagctttt gaaagactgg aagttaagaa agaaacttta 240
ctggcaatgt ttaagtacaa caagttcaaa tgccggatat tgaatgaaaa ggtgaatact 300
ccaactacca cagtctatag atgtggccct ttgatagatc tctgccgggg tctcatggtt 360
agacacacgg gcaaaattaa ggctttaaaa atacacaaaa attcctccac gtactgggaa 420
ggcaaagcag atatggagac tctccagaga atttatggca tttcattccc agatcctaaa 480
atgttgaaag agtgggagaa gttccaagag gaagctaaaa accgagatca taggaaaatt 540
ggcagggacc aagaactata tttctttcat gaactcagcc ctggaagttg cttttttctg 600
ccaaaaggag cctacattta taatgcactt attgaattca ttaggagcga atataggaaa 660
agaggattcc aggaggtagt caccocaaac atcttcaaca gccgactctg gatgacctcg 720
ggccactggc agcactacag cgagaacatg ttctcctttg aggtggagaa ggagctggtt 780
gccctgaaac ccatgaactg cccaggacac tgccttatgt ttgatcatcg gccaaagttc 840
tggcgagaac tgcctctgcg gctagctgat tttgggttac ttcataggaa cgagctgtct 900
ggagcactca caggactcac ccgggtacga agattccaac aggatgatgc tcacatattc 960
tgtgccatgg agcagattga agatgaaata aaaggttgtt tggattttct acgtacggta 1020
tatagcgtat ttggattttc ttttaacta aacctttcta ctgccocgga aaaattcctt 1080
ggagatatcg aagtatggga tcaagctgag aaacaacttg aaaacagtct gaatgaattt 1140
ggtgaaaagt gggagttaaa ctctggagat ggagctttct atggccaaa gattgacata 1200
cagattaag atgcgattgg gcggtaccac cagtgtgcaa ccatccagct ggatttccag 1260
ttgcccatca gatttaatct tacttatgta agccatgatg gtgatgataa gaaaaggcca 1320
gtgattgttc atcgagccat cttgggatca gtggaagaa tgattgctat cctcacagaa 1380
aactatgggg gcaaatggcc cttttggctg tcccctcgcc ag 1422
    
```

10 <210> 117  
 <211> 19  
 <212> PRT  
 <213> *Mus musculus*

15 <400> 117

```

    Glu Phe Glu Asp Ala Phe Pro Ala Asp Phe Ile Ala Glu Gly Ile Asp
     1           5           10           15
    Gln Thr Arg
    
```

<210> 118  
 <211> 20  
 <212> PRT  
 <213> *Mus musculus*

25 <400> 118

```

    Gly Trp Phe Tyr Thr Leu Leu Val Leu Ala Thr Ala Leu Phe Gly Gln
     1           5           10           15
    Pro Pro Phe Lys
     20
    
```

<210> 119  
 <211> 15  
 <212> PRT  
 <213> *Mus musculus*

30 <400> 119

Asn Val Ile Val Asn Gly Leu Ile Leu Ala Ser Asp Gly Gln Lys  
 1 5 10 15

<210> 120  
 <211> 438  
 5 <212> PRT  
 <213> *Mus musculus*  
 <400> 120

```

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

10

```

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

```

<210> 121  
 <211> 23  
 5 <212> PRT  
 <213> *Mus musculus*

<400> 121

```

Ala Pro Leu Lys Pro Tyr Pro Val Pro Thr Ser Asp Asn Ile Leu Ile
  1          5          10          15
Gln Glu Gln Thr Gln Leu Lys
          20

```

10

<210> 122  
 <211> 515  
 <212> PRT  
 15 <213> *Mus musculus*

<400> 122

ES 2 638 311 T3

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

```

Ile Thr Asp Arg Trp Val Leu Ser Phe Met Gln Ser Leu Leu Gly Phe
145          150          155
Phe Glu Thr Glu Met Ala Ala Tyr Arg Leu Tyr Thr Val Val Pro Arg
          165          170          175
Leu Val Lys Phe Val Asp Ile Leu Thr Asn Trp Tyr Val Arg Met Asn
          180          185          190
Arg Arg Arg Leu Lys Gly Glu Ser Gly Val Glu Asp Cys Val Met Ala
          195          200          205
Leu Glu Thr Leu Phe Ser Val Leu Leu Ser Leu Cys Arg Leu Met Ala
          210          215          220
Pro Tyr Thr Pro Phe Leu Thr Glu Leu Met Tyr Gln Asn Leu Lys Leu
225          230          235
Leu Ile Asp Pro Ala Ser Leu Arg Asp Lys Asp Thr Leu Ser Ile His
          245          250          255
Tyr Leu Met Leu Pro Arg Val Arg Glu Leu Ile Asp Lys Lys Thr
          260          265          270
Glu Asn Ala Val Ser Arg Met Gln Ser Val Ile Glu Leu Gly Arg Val
          275          280          285
Ile Arg Asp Arg Lys Thr Ile Pro Ile Lys Tyr Pro Leu Lys Glu Ile
          290          295          300
Val Val Ile His Gln Asp Pro Glu Ala Leu Glu Asp Ile Arg Ser Leu
305          310          315
Glu Lys Tyr Ile Ile Glu Glu Leu Asn Val Arg Lys Val Thr Leu Ser
          325          330          335
Thr Asp Lys Asn Lys Tyr Gly Ile Arg Leu Arg Ala Glu Pro Asp His
          340          345          350
Met Val Leu Gly Lys Arg Leu Lys Gly Ala Phe Lys Ala Val Met Met
          355          360          365
Ala Ile Lys Arg Leu Ser Asn Glu Glu Leu Glu Arg Phe Gln Lys Ser
          370          375          380
Gly Ser Ile Val Val Glu Gly His Glu Leu His Glu Glu Asp Ile Arg
385          390          395
Leu Met Tyr Thr Phe Asp Gln Ala Thr Gly Gly Thr Ala Gln Phe Glu
          405          410          415
Ala His Ser Asp Ala Gln Ala Leu Val Leu Leu Asp Val Thr Pro Asp
          420          425          430
Gln Ser Met Val Asp Glu Gly Met Ala Arg Glu Val Ile Asn Arg Ile
          435          440          445
Gln Lys Leu Arg Lys Lys Cys Asn Leu Val Pro Thr Asp Glu Ile Thr
          450          455          460
Val Tyr Tyr Asn Ala Lys Ser Glu Gly Arg Tyr Leu Asn Asn Val Ile
465          470          475
Glu Ser His Thr Asp Phe Val Phe Ala Thr Ile Lys Ala Pro Leu Lys
          485          490          495
Pro Tyr Pro Val Pro Thr Ser Asp Asn Ile Leu Ile Gln Glu Gln Thr
          500          505          510
Gln Leu Lys
          515

```

<210> 123  
 <211> 495  
 5 <212> PRT  
 <213> Secuencia artificial

<220>  
 <223> Polipéptido de treonil ARNt sintetasas con marcador de afinidad N-terminal de 6xHis

10 <400> 123

ES 2 638 311 T3

Met	His	His	His	His	His	His	Gly	Lys	Pro	Ile	Pro	Asn	Pro	Leu	Leu
1				5					10					15	
Gly	Leu	Asp	Ser	Thr	Glu	Glu	Ala	Gln	Ala	Val	Tyr	Trp	His	Ser	Ser
		20						25					30		
Ala	His	Ile	Met	Gly	Glu	Ala	Met	Glu	Arg	Val	Tyr	Gly	Gly	Cys	Leu

		35					40					45				
Cys	Tyr	Gly	Pro	Pro	Ile	Glu	Asn	Gly	Phe	Tyr	Tyr	Asp	Met	Tyr	Leu	
	50					55					60					
Glu	Glu	Gly	Gly	Val	Ser	Ser	Asn	Asp	Phe	Ser	Ser	Leu	Glu	Ala	Leu	
65					70					75					80	
Cys	Lys	Lys	Ile	Ile	Lys	Glu	Lys	Gln	Ala	Phe	Glu	Arg	Leu	Glu	Val	
			85						90					95		
Lys	Lys	Glu	Thr	Leu	Leu	Ala	Met	Phe	Lys	Tyr	Asn	Lys	Phe	Lys	Cys	
			100					105					110			
Arg	Ile	Leu	Asn	Glu	Lys	Val	Asn	Thr	Pro	Thr	Thr	Thr	Val	Tyr	Arg	
		115					120						125			
Cys	Gly	Pro	Leu	Ile	Asp	Leu	Cys	Arg	Gly	Pro	His	Val	Arg	His	Thr	
	130					135					140					
Gly	Lys	Ile	Lys	Ala	Leu	Lys	Ile	His	Lys	Asn	Ser	Ser	Thr	Tyr	Trp	
145					150					155					160	
Glu	Gly	Lys	Ala	Asp	Met	Glu	Thr	Leu	Gln	Arg	Ile	Tyr	Gly	Ile	Ser	
				165					170					175		
Phe	Pro	Asp	Pro	Lys	Met	Leu	Lys	Glu	Trp	Glu	Lys	Phe	Gln	Glu	Glu	
			180					185					190			
Ala	Lys	Asn	Arg	Asp	His	Arg	Lys	Ile	Gly	Arg	Asp	Gln	Glu	Leu	Tyr	
		195					200					205				
Phe	Phe	His	Glu	Leu	Ser	Pro	Gly	Ser	Cys	Phe	Phe	Leu	Pro	Lys	Gly	
	210					215					220					
Ala	Tyr	Ile	Tyr	Asn	Ala	Leu	Ile	Glu	Phe	Ile	Arg	Ser	Glu	Tyr	Arg	
225				230						235					240	
Lys	Arg	Gly	Phe	Gln	Glu	Val	Val	Thr	Pro	Asn	Ile	Phe	Asn	Ser	Arg	
				245					250					255		
Leu	Trp	Met	Thr	Ser	Gly	His	Trp	Gln	His	Tyr	Ser	Glu	Asn	Met	Phe	
		260						265					270			
Ser	Phe	Glu	Val	Glu	Lys	Glu	Leu	Phe	Ala	Leu	Lys	Pro	Met	Asn	Cys	
		275					280					285				
Pro	Gly	His	Cys	Leu	Met	Phe	Asp	His	Arg	Pro	Arg	Ser	Trp	Arg	Glu	
	290					295					300					
Leu	Pro	Leu	Arg	Leu	Ala	Asp	Phe	Gly	Val	Leu	His	Arg	Asn	Glu	Leu	
305					310					315					320	
Ser	Gly	Ala	Leu	Thr	Gly	Leu	Thr	Arg	Val	Arg	Arg	Phe	Gln	Gln	Asp	
				325					330					335		
Asp	Ala	His	Ile	Phe	Cys	Ala	Met	Glu	Gln	Ile	Glu	Asp	Glu	Ile	Lys	
			340					345					350			
Gly	Cys	Leu	Asp	Phe	Leu	Arg	Thr	Val	Tyr	Ser	Val	Phe	Gly	Phe	Ser	
		355					360					365				
Phe	Lys	Leu	Asn	Leu	Ser	Thr	Arg	Pro	Glu	Lys	Phe	Leu	Gly	Asp	Ile	
		370				375					380					
Glu	Val	Trp	Asp	Gln	Ala	Glu	Lys	Gln	Leu	Glu	Asn	Ser	Leu	Asn	Glu	
385					390					395					400	
Phe	Gly	Glu	Lys	Trp	Glu	Leu	Asn	Ser	Gly	Asp	Gly	Ala	Phe	Tyr	Gly	
				405					410					415		
Pro	Lys	Ile	Asp	Ile	Gln	Ile	Lys	Asp	Ala	Ile	Gly	Arg	Tyr	His	Gln	
			420					425					430			
Cys	Ala	Thr	Ile	Gln	Leu	Asp	Phe	Gln	Leu	Pro	Ile	Arg	Phe	Asn	Leu	
		435					440					445				
Thr	Tyr	Val	Ser	His	Asp	Gly	Asp	Asp	Lys	Lys	Arg	Pro	Val	Ile	Val	
	450					455					460					
His	Arg	Ala	Ile	Leu	Gly	Ser	Val	Glu	Arg	Met	Ile	Ala	Ile	Leu	Thr	
465					470					475					480	
Glu	Asn	Tyr	Gly	Gly	Lys	Trp	Pro	Phe	Trp	Leu	Ser	Pro	Arg	Gln		
				485					490					495		

<210> 124  
 <211> 495  
 <212> PRT

## ES 2 638 311 T3

<213> Secuencia artificial

<220>

<223> Polipéptido de treonil ARNt sintetasa con marcador de afinidad C-terminal de 6xHis

5

<400> 124

Met	Glu	Glu	Ala	Gln	Ala	Val	Tyr	Trp	His	Ser	Ser	Ala	His	Ile	Met
1				5					10					15	
Gly	Glu	Ala	Met	Glu	Arg	Val	Tyr	Gly	Gly	Cys	Leu	Cys	Tyr	Gly	Pro
			20					25					30		
Pro	Ile	Glu	Asn	Gly	Phe	Tyr	Tyr	Asp	Met	Tyr	Leu	Glu	Glu	Gly	Gly
		35					40					45			
Val	Ser	Ser	Asn	Asp	Phe	Ser	Ser	Leu	Glu	Ala	Leu	Cys	Lys	Lys	Ile
	50					55					60				
Ile	Lys	Glu	Lys	Gln	Ala	Phe	Glu	Arg	Leu	Glu	Val	Lys	Lys	Glu	Thr
65				70						75					80
Leu	Leu	Ala	Met	Phe	Lys	Tyr	Asn	Lys	Phe	Lys	Cys	Arg	Ile	Leu	Asn
				85					90					95	
Glu	Lys	Val	Asn	Thr	Pro	Thr	Thr	Thr	Val	Tyr	Arg	Cys	Gly	Pro	Leu
			100					105					110		
Ile	Asp	Leu	Cys	Arg	Gly	Pro	His	Val	Arg	His	Thr	Gly	Lys	Ile	Lys
	115						120					125			
Ala	Leu	Lys	Ile	His	Lys	Asn	Ser	Ser	Thr	Tyr	Trp	Glu	Gly	Lys	Ala
	130					135					140				
Asp	Met	Glu	Thr	Leu	Gln	Arg	Ile	Tyr	Gly	Ile	Ser	Phe	Pro	Asp	Pro
145					150					155					160
Lys	Met	Leu	Lys	Glu	Trp	Glu	Lys	Phe	Gln	Glu	Glu	Ala	Lys	Asn	Arg
				165					170					175	
Asp	His	Arg	Lys	Ile	Gly	Arg	Asp	Gln	Glu	Leu	Tyr	Phe	Phe	His	Glu
			180					185					190		
Leu	Ser	Pro	Gly	Ser	Cys	Phe	Phe	Leu	Pro	Lys	Gly	Ala	Tyr	Ile	Tyr
		195					200					205			
Asn	Ala	Leu	Ile	Glu	Phe	Ile	Arg	Ser	Glu	Tyr	Arg	Lys	Arg	Gly	Phe
	210					215						220			
Gln	Glu	Val	Val	Thr	Pro	Asn	Ile	Phe	Asn	Ser	Arg	Leu	Trp	Met	Thr
225					230					235					240
Ser	Gly	His	Trp	Gln	His	Tyr	Ser	Glu	Asn	Met	Phe	Ser	Phe	Glu	Val
				245					250					255	
Glu	Lys	Glu	Leu	Phe	Ala	Leu	Lys	Pro	Met	Asn	Cys	Pro	Gly	His	Cys
			260					265					270		
Leu	Met	Phe	Asp	His	Arg	Pro	Arg	Ser	Trp	Arg	Glu	Leu	Pro	Leu	Arg
		275					280					285			
Leu	Ala	Asp	Phe	Gly	Val	Leu	His	Arg	Asn	Glu	Leu	Ser	Gly	Ala	Leu
		290				295					300				
Thr	Gly	Leu	Thr	Arg	Val	Arg	Arg	Phe	Gln	Gln	Asp	Asp	Ala	His	Ile
305					310					315					320
Phe	Cys	Ala	Met	Glu	Gln	Ile	Glu	Asp	Glu	Ile	Lys	Gly	Cys	Leu	Asp
				325					330					335	
Phe	Leu	Arg	Thr	Val	Tyr	Ser	Val	Phe	Gly	Phe	Ser	Phe	Lys	Leu	Asn
			340					345					350		
Leu	Ser	Thr	Arg	Pro	Glu	Lys	Phe	Leu	Gly	Asp	Ile	Glu	Val	Trp	Asp
		355					360					365			
Gln	Ala	Glu	Lys	Gln	Leu	Glu	Asn	Ser	Leu	Asn	Glu	Phe	Gly	Glu	Lys
	370					375					380				
Trp	Glu	Leu	Asn	Ser	Gly	Asp	Gly	Ala	Phe	Tyr	Gly	Pro	Lys	Ile	Asp
385					390					395					400
Ile	Gln	Ile	Lys	Asp	Ala	Ile	Gly	Arg	Tyr	His	Gln	Cys	Ala	Thr	Ile
				405					410					415	
Gln	Leu	Asp	Phe	Gln	Leu	Pro	Ile	Arg	Phe	Asn	Leu	Thr	Tyr	Val	Ser
			420					425					430		
His	Asp	Gly	Asp	Asp	Lys	Lys	Arg	Pro	Val	Ile	Val	His	Arg	Ala	Ile
		435					440					445			
Leu	Gly	Ser	Val	Glu	Arg	Met	Ile	Ala	Ile	Leu	Thr	Glu	Asn	Tyr	Gly
	450					455					460				
Gly	Lys	Trp	Pro	Phe	Trp	Leu	Ser	Pro	Arg	Gln	Gly	Lys	Pro	Ile	Pro
465					470					475					480
Asn	Pro	Leu	Leu	Gly	Leu	Asp	Ser	Thr	His						

485

490

495

ES 2 638 311 T3

<210> 125  
 <211> 1434  
 <212> ADN  
 <213> Secuencia artificial

5

<220>  
 <223> Polinucleótido de treonil ARNt sintetasa optimizado por codones

<400> 125

10

```

gaggaagcac aagcggata ctggcacagc agcgcgcaca tcatgggtga ggcgatggag 60
cgtgtgtacg gtggctgcct gtgctacggc ccgccgattg agaatggttt ctattatgat 120
atgtacctgg aagaggggtg cgtgtctagc aatgatttct ccagcttggg ggcgctgtgt 180
aagaagatta tcaaagagaa gcaggccttc gagcgcctgg aagttaagaa agagactctg 240
ctggcaatgt ttaagtacaa caagttaag tgtcgcattt tgaatgaaaa agtgaacacc 300
ccaaccacga cgggtgatcg ctgtggtccg ttgatcgact tgtgtcgtgg ccctcacgtc 360
cgtcacaccg gtaaaatcaa ggcactgaaa attcacaaga acagcagcac ctactgggag 420
ggcaaggcag acatggaaac tctgcaacgt atttacggca tctcgttocc ggatccgaaa 480
atgctgaaag aatgggagaa atttcaggaa gaggcgaaga atcgtgacca tcgcaagatt 540
ggccgtgacc aagaactgta cttcttccac gagctgtctc cgggttctctg cttcttctctg 600
cctaaggggtg catacattta caacgctctg attgagttta ttcgcagcga gtatcgtaaa 660
cgtggctttc aagaggttgt taccocgaac attttcaata gccgtctgtg gatgaccagc 720
ggtcattggc aacattatag cgagaacatg ttttcttttg aggtcgagaa agagctgttt 780
gccctgaaac caatgaactg cccgggtcac tgtctgatgt tcgatcaccg tccgcgcagc 840
tggcgtgaac tgccgttgcg tttggcggat ttcggcgtgt tgcaccgcaa cgaactgagc 900
ggtgcgctga ccggtctgac ccgcgtccgc cgctttcagc aggatgacgc gcatacttt 960
tgcgctatgg agcagatcga ggacgaaatc aaaggttgct tggatttctt gcgcaccgtg 1020
tatagcgttt ttggtttcag ctttaagctg aacctgagca cgcgtccgga aaagttcctg 1080
ggtgacatcg aagtttggga ccaggccgaa aaacaactgg agaactccct gaatgaattt 1140
ggcgaaaagt gggaaactgaa tagcgggtgat ggtgccttct acggcccgaa aatcgacatc 1200
cagatcaaag acgctatcgg tcgttatcac caatgcgcga cgattcagct ggattttcag 1260
ctgccgatcc gtttcaatct gacctatgtg agccatgacg gcgatgacaa gaaacgtccg 1320
gttattgtcc atcgtgcgat tctgggcagc gtcgaacgta tgattgcaat cctgacggag 1380
aattacgggtg gcaaatggcc gttctggtctg tccccgcgtc aataatgact cgag 1434
    
```

<210> 126  
 <211> 2067  
 <212> ADN  
 <213> *Homo sapiens*

15

<400> 126

atgtttgagg	agaaggccag	cagtcottca	gggaagatgg	gaggcgagga	gaagccgatt	60
ggtgctggtg	aagagaagca	aaaggaagga	ggcaaaaaga	agaacaaaga	aggatctgga	120
gatggaggtc	gagctgagtt	gaatecctgg	cctgaatata	tttacacacg	tcttgagatg	180
tataatatac	taaaagcaga	acatgattcc	attctggcag	aaaaggcaga	aaaagatagc	240
aagccaatta	aagtcacttt	gcctgatggt	aaacaggttg	atgcggaatc	ttggaaaact	300
acaccatatac	aaattgcctg	tggaattagt	caaggcctgg	ccgacaacac	cgttattgct	360
aaagtaaata	atggtgtgtg	ggacctggac	cgccctctgg	aagaagattg	taccttggag	420
cttctcaagt	ttgaggatga	ggaagctcag	gcagtgatt	ggcactctag	tgctcacata	480
atgggtgaag	ccatggaaag	agtctatggt	ggatgtttat	gctacgggtcc	gccaatagaa	540
aatggattct	attatgacat	gtacctcgaa	gaagggggtg	tgtctagcaa	tgatttctct	600
tctctggagg	ctttgtgtaa	gaaaatcatt	aaagaaaac	aagcttttga	aagactggaa	660
gtaagaaag	aaactttact	ggcaatgttt	aagtacaaca	agttcaaatg	ccggatattg	720
aatgaaaagg	tgaatactcc	aactaccaca	gtctatagat	gtggcccttt	gatagatctc	780
tgccgggctc	ctcatgttag	acacacgggc	aaaattaagg	ctttaaaaat	acacaaaaat	840
tcctccacgt	actgggaagg	caaagcagat	atggagactc	tccagagaat	ttatggcatt	900
tcattcccag	atcctaaaat	gttgaaagag	tgggagaagt	tccaagagga	agctaaaaac	960
cgagatcata	ggaaaattgg	cagggacca	gaactatatt	tctttcatga	actcagccct	1020
ggaagtgtct	tttttctgcc	aaaaggagcc	tacatttata	atgcacttat	tgaattcatt	1080
aggagcgaat	ataggaaaag	aggattccag	gaggtagtca	ccccaaacat	cttcaacagc	1140
cgactctgga	tgacctcggg	ccactggcag	cactacagcg	agaacatggt	ctcctttgag	1200
gtggagaagg	agctgtttgc	cctgaaaccc	atgaactgcc	caggacactg	ccttatgttt	1260
gatcatcggc	caaggtcctg	gogagaactg	cctctgcggc	tagctgattt	tggggtactt	1320
cataggaacg	agctgtctgg	agcactcaca	ggactcacc	gggtacgaag	attccaacag	1380
gatgatgctc	acatattctg	tgccatggag	cagattgaag	atgaaataaa	aggttgtttg	1440
gattttctac	gtacggtata	tagcgtat	ggattttctt	ttaaactaaa	cctttctact	1500
cgcccggaaa	aattccttgg	agatatcgaa	gtatgggatc	aagctgagaa	acaacttgaa	1560
aacagtctga	atgaatttgg	tgaaaagtgg	gagttaaact	ctggagatgg	agctttctat	1620
ggcccaaaga	ttgacataca	gattaaagat	gcgattgggc	ggtaccacca	gtgtgcaacc	1680
atccagctgg	atttccagtt	gcccatacaga	tttaactcta	cttatgtaag	gcccttttgg	1740
ctgtcccctc	gccaggtaat	ggtagttcca	gtgggaccaa	cctgtgatga	atatgcccaa	1800
aaggtacgac	aacaattcca	cgatgccaaa	ttcatggcag	acattgatct	ggatccaggc	1860
tgtacattga	ataaaaagat	tcgaaatgca	cagttagcac	agtataactt	catttttagtt	1920
gttggtgaaa	aagagaaaat	cagtggcact	gttaatatcc	gcacaagaga	caataaggctc	1980
cacggggaac	gcaccatttc	tgaaactatc	gagcggctac	agcagctcaa	agagttccgc	2040
agcaaacagg	cagaagaaga	attttaa				2067

## REIVINDICACIONES

1. Una composición terapéutica que comprende un polipéptido de aminoacil-ARNt sintetasa (AARS) de 100 a 450 aminoácidos de longitud y que comprende una secuencia de aminoácidos que es al menos un 95 %, un 98 % o un 100 % idéntica a las SEQ ID NO: 12, 14, 50 o 58, en donde el polipéptido tiene una actividad de señalización extracelular y tiene una solubilidad de al menos 5 mg/ml y en donde la composición tiene una pureza de al menos el 95 % basándose en la proteína y menos de 10 UE de endotoxina/mg de proteína.
2. Un polipéptido de aminoacil-ARNt sintetasa (AARS) aislado, de 100 a 450 aminoácidos de longitud y que comprende una secuencia de aminoácidos que es al menos un 98 % o un 100 % idéntica a las SEQ ID NO: 12, 14, 50 o 58, en donde el polipéptido tiene una actividad de señalización extracelular, opcionalmente, en donde el polipéptido de AARS está fusionado a un polipéptido heterólogo.
3. Una composición que comprende un polipéptido de aminoacil-ARNt sintetasa (AARS) aislado de la reivindicación 2, en donde el polipéptido tiene una solubilidad de al menos 5 mg/ml y en donde la composición tiene una pureza de al menos el 95 % basándose en la proteína y menos de aproximadamente 10 UE de endotoxina/mg de proteína, opcionalmente en donde dicho polipéptido consiste en una secuencia de aminoácidos expuesta en las SEQ ID NO: 12, 14, 50 o 58 o difiere respecto de una secuencia de aminoácidos expuesta en las SEQ ID NO: 12, 14, 50 o 58, por la sustitución, la delección y/o la adición de 1, 2 o 3 aminoácidos.
4. La composición de la reivindicación 3, en donde al menos un resto o un sustrato sólido están unidos covalentemente o no covalentemente a dicho polipéptido.
5. Un anticuerpo aislado que muestra especificidad de unión por un polipéptido de AARS aislado de la reivindicación 2, opcionalmente en donde la afinidad del anticuerpo por el polipéptido de AARS es 10X mayor o más que su afinidad por un polipéptido de AARS correspondiente de longitud completa.
6. Una composición terapéutica que comprende un anticuerpo aislado de la reivindicación 5, en donde el anticuerpo tiene una afinidad de al menos 1 nM por el polipéptido, opcionalmente en donde el anticuerpo antagoniza una actividad no canónica del polipéptido de AARS.
7. Un sistema que comprende un polipéptido de aminoacil-ARNt sintetasa (AARS) sustancialmente puro de la reivindicación 2 y un elemento seleccionado entre el grupo que consiste en
- (i) una población de células modificadas por ingeniería genética en las que al menos una célula comprende un polinucleótido que codifica dicho polipéptido de AARS, en donde las células son capaces de crecer en un medio libre de suero,
  - (ii) una célula que comprende un receptor de la superficie celular o una porción extracelular del mismo que se unen al polipéptido y una molécula de menos de 2000 Dalton, o un segundo polipéptido, que modula la unión o la interacción del polipéptido de AARS y el receptor extracelular,
  - (iii) una célula que comprende un receptor de la superficie celular o una porción extracelular del mismo que se une específicamente a un polipéptido de AARS, en donde la célula comprende una molécula indicadora que permite la detección de un cambio en los niveles o la actividad del receptor de la superficie celular o una porción extracelular del mismo y
  - (iv) una población de células modificadas por ingeniería genética en las que al menos una célula comprende un polinucleótido que codifica dicho polipéptido de AARS, al menos 10 litros de un medio de crecimiento libre de suero y un recipiente estéril.
8. Un polinucleótido de aminoacil-ARNt sintetasa (AARS) aislado que comprende
- (i) una secuencia de nucleótidos que codifica un polipéptido de AARS de la reivindicación 2; o
  - (ii) una secuencia de nucleótidos que hibrida específicamente en condiciones de alta rigurosidad con una secuencia de nucleótidos de la cláusula (i).
9. Un método para determinar la presencia o los niveles de un polipéptido de AARS en una muestra, que comprende poner en contacto la muestra con uno o más agentes de unión que se unen específicamente a un polipéptido de AARS de la reivindicación 2, detectar la presencia o la ausencia del agente de unión y de este modo determinar la presencia o los niveles del polipéptido de AARS.
10. Un método para determinar la presencia o los niveles de una secuencia de polinucleótido de una variante de corte y empalme de AARS en una muestra, que comprende poner en contacto la muestra con uno o más oligonucleótidos que hibridan específicamente con un polinucleótido de AARS de la cláusula (i) de la reivindicación 8, detectar la presencia o la ausencia de los oligonucleótidos en la muestra y de este modo determinar la presencia o los niveles de la secuencia de polinucleótido de la variante de corte y empalme de AARS.
11. Un método para identificar un compuesto que se une específicamente a un polipéptido de aminoacil-ARNt

sintetasa (AARS) de la reivindicación 2, que comprende

a) combinar el polipéptido de AARS con al menos un compuesto de ensayo en condiciones adecuadas y

b) detectar la unión del polipéptido de AARS al compuesto de ensayo, identificando de este modo un compuesto que se une específicamente al polipéptido de AARS.

5 12. Una composición farmacéutica, que comprende un polipéptido de AARS de la reivindicación 2, un anticuerpo de la reivindicación 5 o un polinucleótido de AARS de la reivindicación 8 y un vehículo farmacéuticamente aceptable.

10 13. Un método *in vitro* para modular una actividad celular de una célula o una proteína, que comprende poner en contacto la célula o la proteína con un polipéptido de AARS de la reivindicación 2, un anticuerpo de la reivindicación 5 o un polinucleótido de AARS de la reivindicación 8 o una composición farmacéutica de la reivindicación 12.

15 14. Una proteína aminoacil-ARNt sintetasa (AARS) de longitud completa modificada por ingeniería genética que comprende un sitio de proteólisis heterólogo para permitir la generación proteolítica de un polipéptido de AARS de la reivindicación 2.

20 15. Una composición terapéutica de las reivindicaciones 1 o 6, un polipéptido de AARS de la reivindicación 2 o un ácido nucleico que codifica el polipéptido de AARS, para su uso en el tratamiento de una enfermedad o de un trastorno mediados por la desregulación de la expresión, la actividad o la ubicación espaciotemporal de una ARNt sintetasa, opcionalmente en donde la enfermedad se selecciona entre cáncer, neuropatía, diabetes y trastornos inflamatorios.

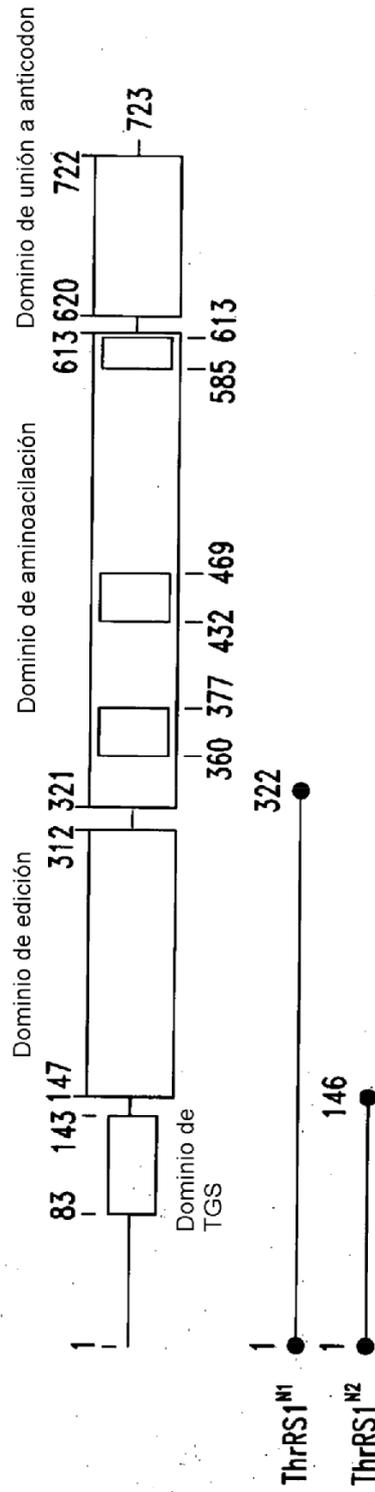


FIG. 1A

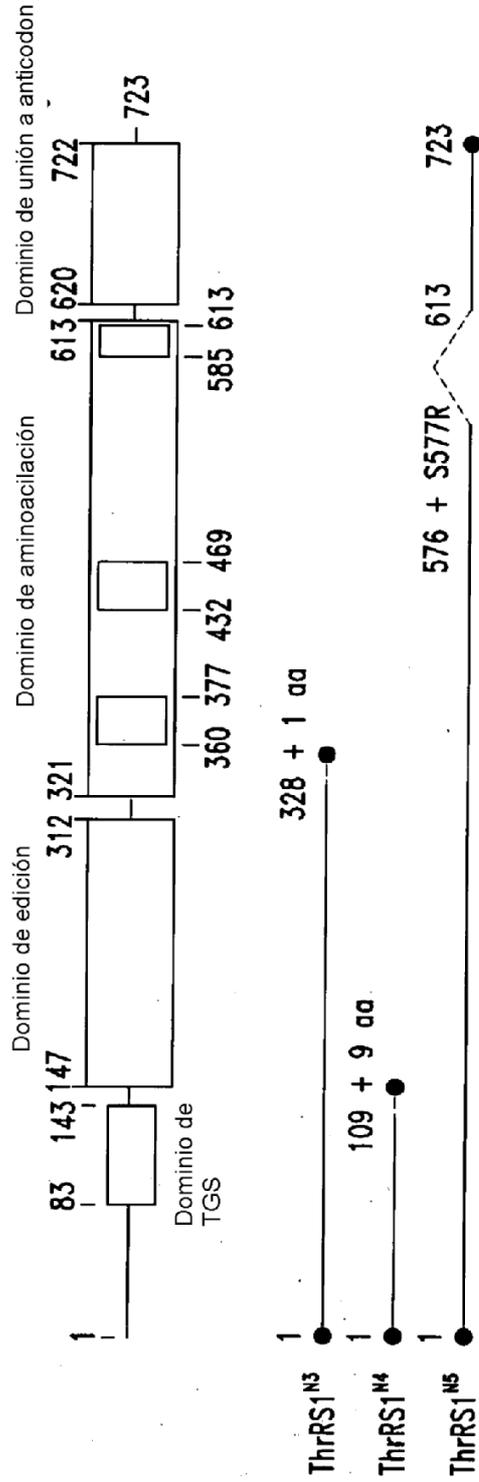


FIG. 1B

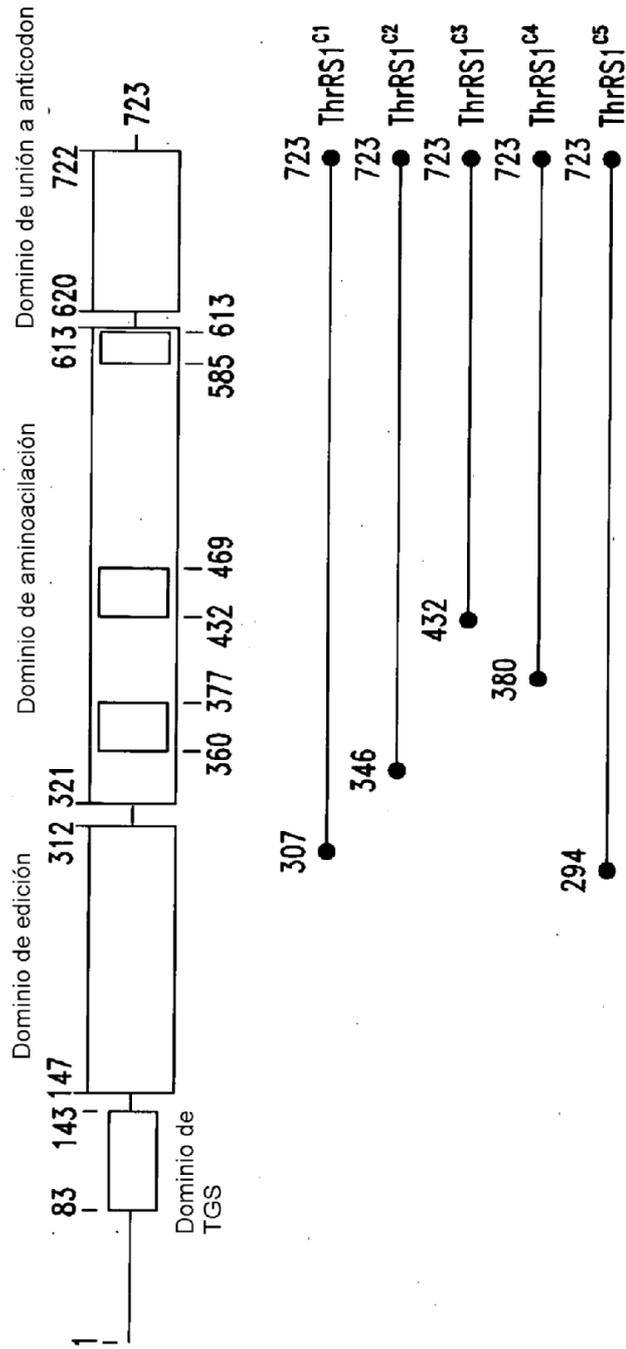


FIG. 2A

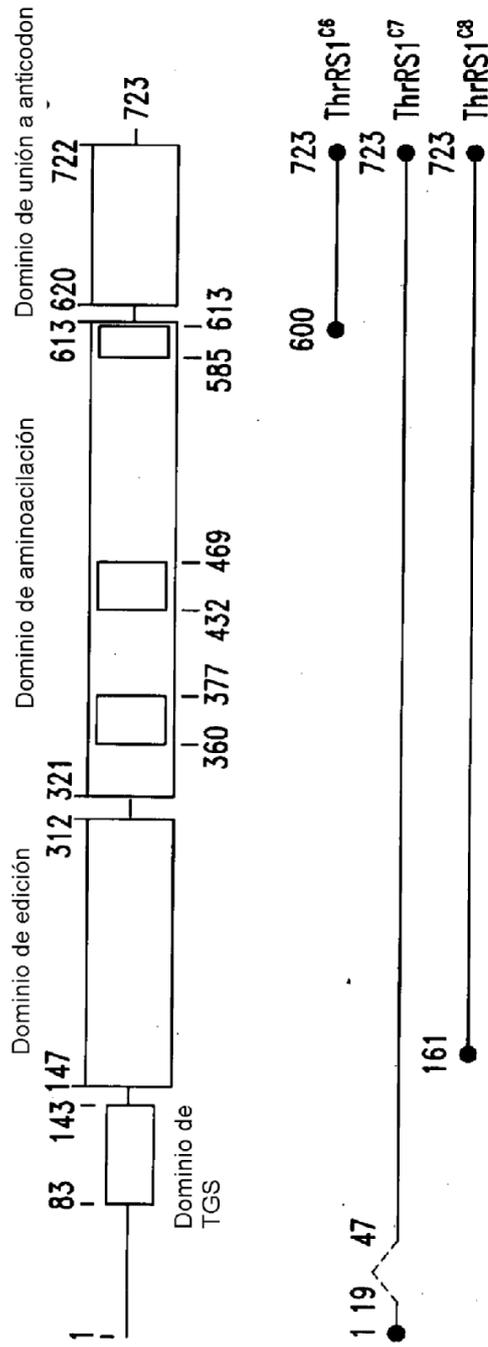
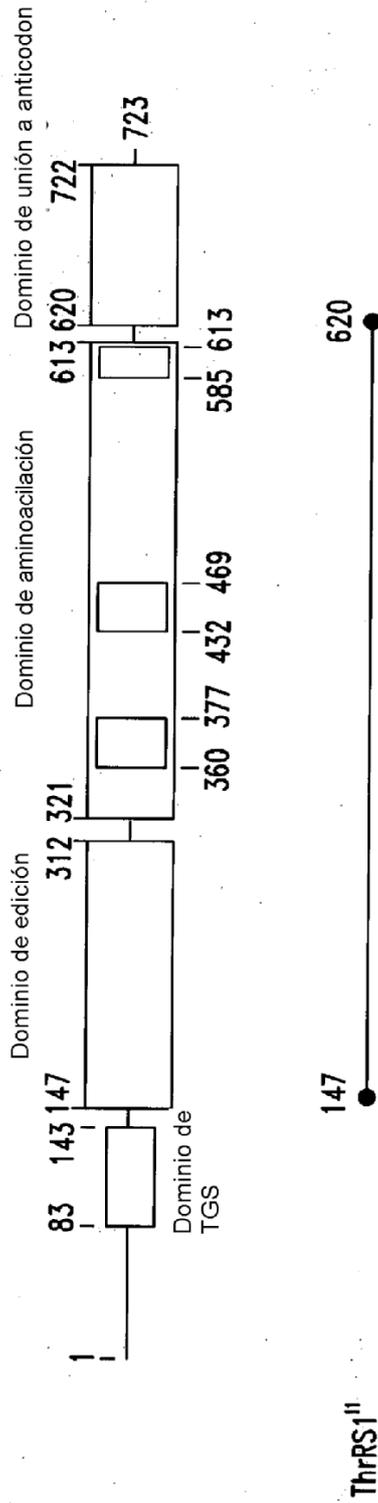


FIG. 2B



**FIG. 3**