

POLIPÉPTIDOS XTEN CON CÓDIGO DE BARRAS Y COMPOSICIONES DE ESTOS, Y MÉTODOS PARA PREPARAR Y USAR LOS MISMOS

LISTADO DE SECUENCIAS

La presente solicitud contiene un Listado de secuencias que se ha presentado electrónicamente en formato ASCII y se incorpora en la presente descripción como referencia en su totalidad. Dicha copia ASCII, creada el 6 de noviembre de 2020, tiene por nombre 20-1761-WO_Sequence_Listing_ST25.txt y tiene un tamaño de 1,494 bytes.

ANTECEDENTES

Un polipéptido puede producirse de una manera que dé como resultado una mezcla de polipéptidos. La mezcla de polipéptidos a menudo puede incluir el polipéptido de longitud completa, junto con variantes de tamaño (por ejemplo, truncamientos) de estos. La presencia de variantes que difieren en tamaño del producto de longitud completa deseado puede afectar el comportamiento biológico de una sustancia farmacológica polipeptídica, lo que afecta potencialmente la seguridad y/o eficacia de la sustancia farmacológica polipeptídica. Por ejemplo, los profármacos basados en proteínas para la terapia contra el cáncer pueden diseñarse con un mecanismo de activación dirigido al tumor. Más específicamente, la proteína terapéutica de longitud completa puede producirse y administrarse en forma de profármaco inactivo (no citotóxico), que se convierte en el fármaco activo mediante la eliminación preferencial de una porción del polipéptido profármaco en el lado biológico deseado (por ejemplo, el tumor). Las variantes de truncamiento de la construcción de longitud completa pueden perder secuencias protectoras y volverse citotóxicas (activas), lo que "contamina" así la composición de profármaco y produce una mezcla que tiene componentes que son involuntariamente activos fuera del sitio biológico previsto. En algunos casos, estas variantes de longitud más corta pueden presentar un mayor riesgo de inmunogenicidad, tener una toxicidad menos selectiva por las células tumorales o mostrar un perfil farmacocinético menos deseado (por ejemplo, que da como resultado una ventana terapéutica más estrecha) en comparación con la proteína de longitud completa, o tienen efectos nocivos no deseados en un receptor fuera del sitio previsto (por ejemplo en tejido sano). Como resultado, la detección y cuantificación de las variaciones estructurales de las proteínas pueden ser importantes para evaluar las propiedades biológicas (por ejemplo, seguridad clínica y eficacia farmacológica) de los productos bioterapéuticos y en el desarrollo de nuevos productos bioterapéuticos (por ejemplo, con mayor eficacia y efectos secundarios reducidos). Las técnicas y métodos existentes para identificar y cuantificar la cantidad de productos de truncamiento "contaminantes" pueden incluir uno o más inconvenientes, tales como ser de sensibilidad, facilidad, eficiencia o efectividad limitadas.

SUMARIO

En la presente descripción se describen polipéptidos que comprenden un polipéptido recombinante extendido (XTEN) que comprende una pluralidad de motivos de secuencia no superpuestos. En los polipéptidos XTEN de esta invención, la pluralidad de motivos de secuencia no superpuestos comprende: un conjunto de motivos de secuencia no superpuestos, en donde cada uno de dichos motivos de secuencia se repite al menos dos veces en el polipéptido XTEN; y también un motivo de secuencia no superpuesto único que aparece solo una vez dentro del polipéptido XTEN; en donde el polipéptido comprende además un primer fragmento de código de barras liberable del polipéptido tras la digestión por una proteasa. En dichas modalidades, el primer fragmento de código de barras es una porción del XTEN que incluye al menos parte del motivo de secuencia que aparece solo una vez dentro del XTEN y difiere en secuencia y peso molecular de todos los demás fragmentos peptídicos que pueden liberarse del polipéptido tras la digestión completa del polipéptido por la proteasa. Además, en las modalidades de XTEN de la invención proporcionadas en la presente descripción, el fragmento de código de barras no incluye el aminoácido N-terminal o el aminoácido C-terminal del polipéptido. Como se describe adicionalmente en la presente descripción, los polipéptidos XTEN de esta invención se caracterizan por comprender al menos 150 aminoácidos, más específicamente 150-3 000 aminoácidos de longitud. Los aminoácidos que comprenden los polipéptidos XTEN de la invención se caracterizan porque al menos 90 % de estos residuos son glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) o prolina (P), y el polipéptido XTEN comprende al menos cuatro de estos aminoácidos (G, A, S, T, E o P). Además, los polipéptidos XTEN que se proporcionan en la presente descripción comprenden motivos de secuencia no superpuestos que tienen una longitud de secuencia de 9 a 14 aminoácidos y dentro de cada uno de dichos motivos no superpuestos, la secuencia de aminoácidos G, A, S, T, E o P es sustancialmente aleatoria con respecto a cualquier otro motivo de secuencia no superpuesto que comprende el polipéptido XTEN.

En algunas modalidades, el fragmento de código de barras no incluye un ácido glutámico que está inmediatamente adyacente a otro ácido glutámico en el XTEN. En algunas modalidades, el fragmento de código de barras tiene un ácido glutámico en su extremo C-terminal. En algunas modalidades, el fragmento de código de barras tiene un aminoácido N-terminal inmediatamente precedido por un residuo de ácido glutámico. En algunas modalidades, el residuo de ácido glutámico que precede al aminoácido N-terminal no está inmediatamente adyacente a otro residuo de ácido glutámico. En algunas modalidades, el fragmento de código de barras no incluye un residuo de ácido glutámico en una posición distinta del extremo C-terminal del fragmento de código de barras, a menos que el ácido glutámico esté seguido inmediatamente por una prolina. En algunas modalidades, el fragmento de código de barras se ubica de 10 aminoácidos a 150 aminoácidos del extremo N-terminal del polipéptido o el extremo C-terminal del polipéptido.

En algunas modalidades, los motivos de secuencia del conjunto de motivos de secuencia no superpuestos se identifican en la presente descripción mediante las SEQ ID NO: 182-203 y

1715-1722. En algunas modalidades, los motivos de secuencia del conjunto de motivos de secuencia no superpuestos se identifican en la presente descripción mediante las SEQ ID NO: 186-189. En algunas modalidades, el conjunto de motivos de secuencia no superpuestos comprende al menos dos, al menos tres o los cuatro motivos de secuencia de las SEQ ID NO: 186-189.

En modalidades específicas, los polipéptidos proporcionados en la presente descripción comprenden un polipéptido XTEN como se describe en la presente descripción en donde el fragmento de código de barras no incluye el aminoácido N-terminal o el aminoácido C-terminal del polipéptido; no incluye un ácido glutámico que está inmediatamente adyacente a otro ácido glutámico en el XTEN; tiene un ácido glutámico en su extremo C-terminal; tiene un aminoácido N-terminal que está inmediatamente precedido por un residuo de ácido glutámico; y se ubica de 10 a 125 aminoácidos del extremo N-terminal del polipéptido o del extremo C-terminal del polipéptido.

En algunas de estas modalidades específicas, el residuo de ácido glutámico que precede al aminoácido N-terminal no está inmediatamente adyacente a otro residuo de ácido glutámico. En algunas de estas modalidades específicas, el fragmento de código de barras no incluye un residuo de ácido glutámico en una posición distinta del extremo C-terminal del fragmento de código de barras, a menos que el ácido glutámico esté seguido inmediatamente por una prolina.

En algunas modalidades, los polipéptidos XTEN proporcionados en la presente descripción comprenden una pluralidad de motivos de secuencia no superpuestos, en donde cada uno de dicho motivo de secuencia se repite al menos dos veces en el polipéptido XTEN y tiene una longitud de entre 9 y 14 aminoácidos. En algunas modalidades, los motivos de secuencia del conjunto de motivos de secuencia no superpuestos se identifican en la presente descripción mediante las SEQ ID NO: 182-203 y 1715-1722. En algunas modalidades, los motivos de secuencia del conjunto de motivos de secuencia no superpuestos se identifican en la presente descripción mediante las SEQ ID NO: 186-189. En algunas modalidades, el conjunto de motivos de secuencia no superpuestos comprende al menos dos, al menos tres o los cuatro motivos de secuencia de las SEQ ID NO: 186-189. En algunas modalidades, al menos 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de los residuos de aminoácidos del polipéptido XTEN son una combinación de glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) o prolina (P), en donde el polipéptido XTEN comprende al menos cuatro de estos aminoácidos (G, A, S, T, E o P). En algunas modalidades, el XTEN tiene una longitud de 150 a 3 000 aminoácidos. En algunas modalidades, el XTEN tiene una longitud de 150 a 1 000 aminoácidos. En algunas modalidades, el polipéptido puede escindirse mediante una proteasa que escinde en el lado C-terminal de los residuos de ácido glutámico que no van seguidos de prolina. En ciertas modalidades, la proteasa es una proteasa Glu-C.

En algunas modalidades de los polipéptidos XTEN proporcionados en la presente descripción, el fragmento de código de barras se ubica dentro de 200, dentro de 150, dentro

de 100 o dentro de 50 aminoácidos del extremo N-terminal del polipéptido. En algunas modalidades, el fragmento de código de barras se ubica entre 10 y 200, entre 30 y 200, entre 40 y 150, o entre 50 y 100 aminoácidos del extremo N-terminal de la proteína. En algunas modalidades, el fragmento de código de barras se ubica dentro de 200, dentro de 150, dentro de 100 o dentro de 50 aminoácidos del extremo C-terminal del polipéptido. En algunas modalidades, el fragmento de código de barras se ubica entre 10 y 200, entre 30 y 200, entre 40 y 150 o entre 50 y 100 aminoácidos del extremo C-terminal de la proteína. En algunas modalidades, el fragmento de código de barras tiene una longitud de al menos 4 aminoácidos. En algunas modalidades, el fragmento de código de barras tiene una longitud de entre 4 y 20, entre 5 y 15, entre 6 y 12 o entre 7 y 10 aminoácidos. En algunas modalidades, el fragmento de código de barras se identifica en la presente descripción mediante las SEQ ID NO: 8020-8030 (BAR001-BAR011).

En algunas modalidades, el polipéptido comprende además un segundo fragmento de código de barras en donde el segundo fragmento de código de barras es una porción de XTEN y difiere en secuencia y peso molecular de todos los demás fragmentos peptídicos que pueden liberarse del polipéptido tras la digestión completa del polipéptido por proteasa. En algunas modalidades, el polipéptido comprende además un tercer fragmento de código de barras en donde el tercer fragmento de código de barras es una porción de XTEN y difiere en secuencia y peso molecular de todos los demás fragmentos peptídicos que pueden liberarse del polipéptido tras la digestión completa del polipéptido por proteasa.

En algunas modalidades, el XTEN tiene al menos 90 %, al menos 92 %, al menos 95 %, al menos 98 %, al menos 99 % o 100 % de identidad de secuencia con una secuencia identificada en la presente descripción mediante las SEQ ID NO: 8001-8019. En algunas modalidades, el XTEN tiene una longitud de al menos 200, al menos 250, al menos 300, al menos 350, al menos 400, al menos 450 o al menos 500 aminoácidos.

En algunas modalidades, el polipéptido comprende además un polipéptido biológicamente activo unido al polipéptido XTEN (BPXTEN). En algunas modalidades, el polipéptido XTEN está unido al polipéptido biológicamente activo en el extremo amino o carboxilo del XTEN. En cualquier configuración, el fragmento de código de barras se ubica dentro de una región del XTEN que se extiende, según se mide desde el extremo amino o carboxilo unido al polipéptido biológicamente activo, entre 5 % y 50 %, entre 7 % y 40 %, o entre 10 % y 30 % de la longitud del XTEN.

En algunas modalidades, el polipéptido BPXTEN comprende además uno o más fragmentos de referencia liberables del polipéptido tras la digestión por la proteasa, en donde el uno o más fragmentos de referencia comprenden cada uno una porción del polipéptido biológicamente activo. En algunas modalidades, el uno o más fragmentos de referencia es un único fragmento de referencia que difiere en secuencia y peso molecular de todos los demás fragmentos peptídicos que pueden liberarse del polipéptido tras la digestión del polipéptido por la proteasa. En algunas modalidades, dicho fragmento de referencia comprende un péptido cuya presencia en una mezcla de polipéptidos indica su

existencia o integridad (es decir, que la proteína no se ha degradado o escindido proteolíticamente).

En algunas modalidades, el polipéptido BPXTEN comprende además un primer segmento de liberación (RS1) ubicado entre el XTEN y el polipéptido biológicamente activo. En algunas modalidades, el RS1 comprende una secuencia de aminoácidos que tiene al menos 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una secuencia identificada en la presente descripción mediante una cualquiera de las secuencias en las Tablas 4a-4h. En algunas modalidades, el polipéptido biológicamente activo se identifica en la presente descripción mediante una cualquiera o una combinación de las secuencias en las Tablas 4a-4h y 8a-8b.

En algunas modalidades, el polipéptido BPXTEN tiene ventajosamente una semivida terminal que es al menos dos veces más larga en comparación con el polipéptido biológicamente activo no unido a ningún XTEN.

En algunas modalidades, el polipéptido BPXTEN es ventajosamente menos inmunogénico en comparación con el polipéptido biológicamente activo no unido a ningún XTEN, en donde la inmunogenicidad puede determinarse por medición de la producción de anticuerpos IgG que se unen selectivamente al polipéptido biológicamente activo después de la administración de dosis comparables a un ser humano o animal.

En algunas modalidades, el polipéptido BPXTEN exhibe un factor de peso molecular aparente en condiciones fisiológicas mayor que aproximadamente 6.

En algunas modalidades, el polipéptido BPXTEN comprende además un segundo polipéptido XTEN en donde el segundo polipéptido XTEN comprende una secuencia de aminoácidos que tiene las mismas características expuestas anteriormente y a lo largo de esta descripción para el primer componente XTEN de estas modalidades del BPXTEN, y en donde el primer polipéptido XTEN se ubica en el extremo N-terminal del polipéptido biológicamente activo y el segundo polipéptido XTEN se ubica en el extremo C-terminal del polipéptido biológicamente activo. En algunas modalidades, el segundo polipéptido XTEN comprende una secuencia de aminoácidos que difiere de la secuencia de aminoácidos del primer XTEN que comprende estas modalidades del BPXTEN. En ciertas modalidades, la secuencia de aminoácidos del segundo polipéptido XTEN es más larga que la secuencia de aminoácidos del primer polipéptido XTEN.

En algunas modalidades, el polipéptido BPXTEN comprende además un segundo segmento de liberación (RS2) ubicado entre el polipéptido biológicamente activo y el segundo polipéptido XTEN. En algunas modalidades, el RS1 del primer polipéptido XTEN y el RS2 del segundo polipéptido XTEN son idénticos en secuencia. En algunas modalidades, el RS1 del primer polipéptido XTEN y el RS2 del segundo polipéptido XTEN son cada uno un sustrato para la escisión por múltiples proteasas en uno, dos o tres sitios de escisión dentro de cada secuencia de segmento de liberación.

En algunas de estas modalidades, el polipéptido BPXTEN comprende un fragmento de código de barras adicional que es una porción del segundo polipéptido XTEN y difiere en

secuencia y peso molecular de todos los demás fragmentos peptídicos que pueden liberarse del polipéptido tras la digestión completa del polipéptido por proteasa. En algunas de estas modalidades, el fragmento de código de barras adicional no incluye el aminoácido C-terminal del polipéptido. En algunas de estas modalidades, el fragmento de código de barras adicional comprende un residuo de ácido glutámico en su extremo C-terminal. En algunas de estas modalidades, el fragmento de código de barras adicional del segundo polipéptido XTEN se ubica dentro de 200, dentro de 150, dentro de 100 o dentro de 50 aminoácidos del extremo C-terminal del segundo componente XTEN del polipéptido BPXTEN. En algunas de estas modalidades, el fragmento de código de barras adicional del segundo polipéptido XTEN se ubica en una ubicación que está entre 10 y 200, entre 30 y 200, entre 40 y 150, o entre 50 y 100 aminoácidos desde el extremo C-terminal del segundo componente XTEN del polipéptido BPXTEN. En algunas de estas modalidades, el fragmento de código de barras adicional tiene una longitud de entre 4 y 20, entre 5 y 15, entre 6 y 12 o entre 7 y 10 aminoácidos. En algunas de estas modalidades, el fragmento de código de barras adicional se identifica en la presente descripción mediante las SEQ ID NO: 8020-8030 (BAR001-BAR011).

En algunas modalidades, el segundo polipéptido XTEN comprende además un conjunto de fragmentos de código de barras que incluye el fragmento de código de barras adicional y al menos un fragmento de código de barras adicional, en donde cada fragmento de código de barras del conjunto de fragmentos de código de barras difiere en secuencia y peso molecular de todos los demás fragmentos peptídicos que pueden liberarse del polipéptido BPXTEN tras la digestión completa del polipéptido por la proteasa. En algunas modalidades, el segundo polipéptido XTEN se identifica mediante las SEQ ID NO: 8001-8019. En algunas modalidades, el fragmento de código de barras adicional no incluye un residuo de ácido glutámico que está inmediatamente adyacente a otro residuo de ácido glutámico en el polipéptido.

En algunas modalidades, al menos 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de los residuos de aminoácidos del segundo polipéptido XTEN son una combinación de glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) y prolina (P), en donde el polipéptido XTEN comprende al menos cuatro de estos aminoácidos (G, A, S, T, E o P). En algunas modalidades, la suma del número total de aminoácidos en el primer polipéptido XTEN y el número total de aminoácidos en el segundo polipéptido XTEN es al menos 300, al menos 350, al menos 400, al menos 500, al menos 600, al menos 700 o al menos 800 aminoácidos. En algunas modalidades, el segundo polipéptido XTEN comprende una pluralidad de motivos de secuencia no superpuestos, en donde cada uno de dichos motivos de secuencia se repite al menos dos veces en la secuencia del segundo polipéptido XTEN y tiene una longitud de entre 9 y 14 aminoácidos.

En algunas modalidades, para el segundo polipéptido XTEN, los motivos de secuencia de la pluralidad de motivos de secuencia no superpuestos se identifican en la presente descripción mediante las SEQ ID NO: 182-203 y 1715-1722. En algunas modalidades, los

motivos de secuencia de la pluralidad de motivos de secuencia no superpuestos se identifican en la presente descripción mediante las SEQ ID NO: 186-189. En algunas modalidades, para el segundo polipéptido XTEN, la pluralidad de motivos de secuencia no superpuestos comprende al menos dos, al menos tres o los cuatro de los siguientes motivos: SEQ ID NOs: 186-189. En algunas modalidades, el segundo polipéptido XTEN tiene una longitud de 150 a 3 000 aminoácidos. En algunas modalidades, el segundo polipéptido XTEN tiene una longitud de 150 a 1 000 aminoácidos. En algunas modalidades, el segundo polipéptido XTEN tiene al menos 90 %, al menos 92 %, al menos 95 %, al menos 98 %, al menos 99 % o 100 % de identidad de secuencia con la secuencia identificada en la presente descripción mediante las SEQ ID NO: 8001-8019. En algunas modalidades, el segundo polipéptido XTEN tiene una longitud de al menos 200, al menos 250, al menos 300, al menos 350, al menos 400, al menos 450 o al menos 500 aminoácidos.

En modalidades particulares, los polipéptidos BPXTEN proporcionados en la presente descripción comprenden un primer polipéptido XTEN que comprende una primera secuencia RS proximal pero que no comprende el extremo C-terminal del polipéptido, unido covalentemente a un primer y segundo polipéptido biológicamente activo unido covalentemente en tándem, en donde un segundo polipéptido XTEN está unido covalentemente al extremo C-terminal de los polipéptidos biológicamente activos unidos en tándem, en donde el segundo polipéptido XTEN comprende una segunda secuencia RS próxima pero que no comprende el extremo N-terminal del segundo polipéptido XTEN, en donde la primera y segunda secuencias RS pueden ser iguales o diferentes. En modalidades particulares, el segundo polipéptido XTEN comprende una secuencia de aminoácidos que es más larga que la secuencia de aminoácidos del primer polipéptido XTEN. En ciertas modalidades, la primera o la segunda proteína biológicamente activa o ambas comprenden una proteína de unión específica, en ciertas modalidades en donde la proteína de unión específica se une específicamente a un antígeno o agonista expresado en un sitio biológico deseado. En modalidades particulares, el sitio biológico deseado es un tumor y el antígeno es un antígeno específico de tumor. En modalidades particulares, el primer y segundo polipéptido biológicamente activo son diferentes, que incluyen pero sin limitarse a tener diferentes afinidades de unión específicas.

En la presente descripción se describe además un ácido nucleico que comprende un polinucleótido que codifica un polipéptido tal como cualquier polipéptido XTEN o BPXTEN descrito en la presente descripción o el complemento inverso de dicho polinucleótido.

En la presente descripción también se describe un vector de expresión que comprende cualquier secuencia polinucleotídica tal como se describe en la presente descripción y una secuencia reguladora unida operativamente a la secuencia polinucleotídica que regula la expresión u otra actividad biológica de dicho polinucleótido.

En la presente descripción se describe una célula huésped que comprende el vector de expresión como se describe en la presente descripción. En algunas modalidades, la célula

huésped es una procariota. En algunas de estas modalidades, la célula huésped es *E. coli*. En algunas modalidades alternativas, la célula huésped es una célula de mamífero.

En la presente descripción se describe adicionalmente una composición farmacéutica que comprende un polipéptido como se describe en la presente descripción y uno o más excipientes farmacéuticamente aceptables. En algunas modalidades, la composición farmacéutica se formula para su administración a un animal y, en particular, a un ser humano, en donde dicha administración puede ser por cualquier vía de administración terapéuticamente efectiva. Las composiciones farmacéuticas descritas en la presente descripción pueden prepararse y usarse en cualquier formulación conocida en la técnica y particularmente adaptada a la vía, sitio y efecto previsto de la administración en el ser humano o animal.

En la presente descripción se describe el uso de un polipéptido como se describe en la presente descripción y, particularmente, un polipéptido BPXTEN en la preparación de un medicamento para el tratamiento de una enfermedad, trastorno o afección en un ser humano o animal. En algunas modalidades, la enfermedad, trastorno o afección puede ser cáncer.

En la presente descripción se describe un método para tratar una enfermedad en un ser humano o animal humano o animal como se describió anteriormente y en toda esta descripción, el método comprende administrar al ser humano o animal que lo necesita una o más dosis terapéuticamente efectivas de una composición farmacéutica. En algunas modalidades, la composición farmacéutica se administra al ser humano o animal como una o más dosis terapéuticamente efectivas administradas en un cronograma clínicamente adecuado diariamente, semanalmente, mensualmente o anualmente y en una dosis clínicamente apropiada.

En la presente descripción se describe una mezcla que comprende una pluralidad de polipéptidos, particularmente polipéptidos XTEN y BPXTEN como se describe en la presente descripción de longitud variable, la mezcla comprende:

un primer conjunto de polipéptidos, en donde cada polipéptido del primer conjunto de polipéptidos comprende un fragmento de código de barras que puede liberarse del polipéptido por digestión con una proteasa y tiene una secuencia y peso molecular que difieren de la secuencia y el peso molecular de todos los demás fragmentos que pueden liberarse del primer conjunto de polipéptidos; y

un segundo conjunto de polipéptidos que carecen del fragmento de código de barras del primer conjunto de polipéptidos;

en donde tanto el primer conjunto de polipéptidos como el segundo conjunto de polipéptidos comprenden cada uno un fragmento de referencia que es común al primer conjunto de polipéptidos y al segundo conjunto de polipéptidos y que se produce por digestión con la proteasa; y

en donde la relación del primer conjunto de polipéptidos respecto a los polipéptidos que comprenden el fragmento de referencia es mayor que 0,7.

En algunas modalidades, la relación del primer conjunto de polipéptidos respecto a los polipéptidos que comprenden el fragmento de referencia es mayor que 0,8, 0,9, 0,95 o 0,98. En algunas modalidades, el fragmento de referencia aparece no más de una vez en cada polipéptido del primer conjunto de polipéptidos y el segundo conjunto de polipéptidos. En algunas modalidades, la proteasa es una proteasa que escinde en el lado C-terminal de los residuos de ácido glutámico. En algunas modalidades, la liberación del código de barras de polipéptidos que comprenden el primer conjunto de polipéptidos se ve facilitada por pepsina, elastasa, termolisina o proteasas Glu-C. En algunas modalidades, la proteasa Glu-C facilita la liberación del código de barras. En algunas modalidades, la proteasa no es tripsina. En algunas modalidades, los polipéptidos de longitudes variables comprenden polipéptidos que comprenden al menos un polipéptido XTEN como se expone en la presente descripción.

En algunas modalidades, el primer conjunto de polipéptidos comprende un polipéptido de longitud completa, en donde el fragmento de código de barras es una porción del polipéptido de longitud completa. En algunas modalidades, el polipéptido de longitud completa es cualquier polipéptido descrito en la presente descripción y, particularmente, los polipéptidos XTEN y BPXTEN. En algunas modalidades, el fragmento de código de barras no comprende ni el aminoácido N-terminal ni el aminoácido C-terminal del polipéptido de longitud completa. En algunas modalidades, la mezcla de polipéptidos de longitudes variables difieren entre sí debido al truncamiento N-terminal, el truncamiento C-terminal o el truncamiento tanto N- como C-terminal de un polipéptido de longitud completa.

En la presente descripción se describe un método para evaluar, en una mezcla que comprende polipéptidos de longitud variable, y en particular polipéptidos XTEN y BPXTEN como se describe en la presente descripción, una cantidad relativa de un primer conjunto de polipéptidos en la mezcla respecto a un segundo conjunto de polipéptidos en la mezcla, en donde cada polipéptido del primer conjunto de polipéptidos comparte un fragmento de código de barras que aparece una y solo una vez en el polipéptido y cada polipéptido del segundo conjunto de polipéptidos carece del fragmento de código de barras que comparten los polipéptidos del primer conjunto, en donde los polipéptidos individuales tanto del primer conjunto de polipéptidos como del segundo conjunto de polipéptidos comprenden cada uno un fragmento de referencia, el método comprende:

poner en contacto la mezcla con una proteasa para producir una pluralidad de fragmentos proteolíticos que son el resultado de la escisión del primer conjunto de polipéptidos y el segundo conjunto de polipéptidos, en donde la pluralidad de fragmentos proteolíticos comprende una pluralidad de fragmentos de referencia y una pluralidad de fragmentos de código de barras; y

determinar una relación de la cantidad de fragmentos de código de barras respecto a la cantidad de fragmentos de referencia, para evaluar de esta manera las cantidades relativas del primer conjunto de polipéptidos respecto al segundo conjunto de polipéptidos.

En algunas modalidades, el fragmento de referencia aparece no más de una vez en cada polipéptido del primer conjunto de polipéptidos y el segundo conjunto de polipéptidos.

En algunas modalidades, la proteasa escinde los polipéptidos de longitud variable en el lado C-terminal de los residuos de ácido glutámico que no van seguidos de un residuo de prolina.

En algunas modalidades, la proteasa es una proteasa Glu-C. En algunas modalidades, la proteasa no es tripsina. En algunas modalidades, la determinación de una relación de la cantidad de fragmentos de código de barras respecto a la cantidad de fragmentos de referencia comprende cuantificar los fragmentos de código de barras y los fragmentos de referencia de la mezcla después de poner en contacto la mezcla de polipéptidos con la proteasa. En algunas modalidades, los fragmentos de código de barras y los fragmentos de referencia se identifican en base a sus respectivas masas. En algunas modalidades, los fragmentos de código de barras y los fragmentos de referencia se identifican mediante espectrometría de masas. En algunas modalidades, los fragmentos de código de barras y los fragmentos de referencia se identifican mediante cromatografía líquida-espectrometría de masas (LC-MS). En algunas modalidades, la determinación de una relación de los fragmentos de código de barras respecto a los fragmentos de referencia comprende el marcaje isobárico o el marcaje con isótopos estables. En algunas modalidades, la determinación de una relación de los fragmentos de código de barras respecto a los fragmentos de referencia comprende enriquecer la mezcla con uno o ambos de un fragmento de referencia marcado con isótopo y un fragmento de código de barras marcado con isótopo.

En algunas de estas modalidades, los polipéptidos de longitud variable comprenden un polipéptido de longitud completa y fragmentos truncados de este. En algunas de estas modalidades, la mezcla de polipéptidos de longitudes variables difiere entre sí debido al truncamiento N-terminal, el truncamiento C-terminal o el truncamiento tanto N- como C-terminal de un polipéptido de longitud completa. En algunas de estas modalidades, la relación de la cantidad de fragmentos de código de barras respecto a fragmentos de referencia es mayor que 0,5, 0,6, 0,7, 0,8, 0,9, 0,95, 0,98 o 0,99.

En la presente descripción se describe una mezcla que comprende una pluralidad de polipéptidos de longitud variable, la mezcla comprende un primer conjunto de polipéptidos, en donde cada polipéptido del primer conjunto de polipéptidos comprende un fragmento de código de barras que puede liberarse del polipéptido por digestión con una proteasa y que tiene una secuencia y peso molecular que difieren de la secuencia y el peso molecular de todos los demás fragmentos que pueden liberarse del primer conjunto de polipéptidos. Dicha modalidad también incluye un segundo conjunto de polipéptidos que carecen del fragmento de código de barras del primer conjunto de polipéptidos, en donde tanto el primer conjunto de polipéptidos como el segundo conjunto de polipéptidos comprenden cada uno un fragmento de referencia que es común al primer conjunto de polipéptidos y al segundo conjunto de polipéptidos y liberable por digestión con la proteasa. En dicha modalidad, el número de fragmentos de referencia cuantificados en la mezcla de

polipéptidos después de la digestión con proteasa es igual a la suma del número del primer y segundo conjunto de polipéptidos en la mezcla, y el número de fragmentos de código de barras cuantificados en la mezcla de polipéptidos después de la digestión con proteasa es igual al número del primer conjunto de polipéptidos en la mezcla;. En dicha modalidad, el primer conjunto de polipéptidos comprende un fragmento de referencia, la relación del primer conjunto de polipéptidos respecto a los polipéptidos en la mezcla que comprenden el fragmento de referencia es mayor que 0,7.

En algunas modalidades, la mezcla tiene una relación del primer conjunto de polipéptidos respecto a los polipéptidos que comprenden el fragmento de referencia mayor que 0,8, 0,9 o 0,95.

En una modalidad particular, el fragmento de referencia aparece no más de una vez en cada polipéptido del primer conjunto de polipéptidos y el segundo conjunto de polipéptidos. En modalidades alternativas, el fragmento de referencia aparece dos veces en cada polipéptido del primer conjunto de polipéptidos y del segundo conjunto de polipéptidos.

En algunas modalidades, el primer conjunto de polipéptidos comprende un polipéptido de longitud completa, en donde el fragmento de código de barras es una porción del polipéptido de longitud completa.

En algunas modalidades, el polipéptido de longitud completa incluye los polipéptidos descritos en la presente descripción.

En una modalidad particular, el fragmento de código de barras de la mezcla no comprende el aminoácido N-terminal y el aminoácido C-terminal del polipéptido de longitud completa.

En algunas modalidades, la mezcla contiene polipéptidos de longitudes variables que difieren entre sí debido al truncamiento N-terminal, el truncamiento C-terminal o el truncamiento tanto N-terminal como C-terminal de un polipéptido de longitud completa.

En algunas modalidades, el fragmento de referencia aparece no más de una vez en cada polipéptido del primer conjunto de polipéptidos y el segundo conjunto de polipéptidos. En una modalidad alternativa, el número de fragmentos de referencia en el primer conjunto de polipéptidos puede diferir del número de fragmentos de referencia en el segundo conjunto de polipéptidos, pero el número de estos en cada polipéptido de cada conjunto debe ser el mismo.

En una modalidad particular, cada uno de los fragmentos de referencia en los polipéptidos de la mezcla tiene una secuencia y peso molecular que difieren de la secuencia y el peso molecular de todos los demás fragmentos.

En la presente descripción se describe una mezcla que comprende una pluralidad de polipéptidos de longitud variable, la mezcla comprende un primer conjunto de polipéptidos, en donde cada polipéptido del primer conjunto de polipéptidos comprende

un fragmento de código de barras que puede liberarse del polipéptido por digestión con una proteasa y tiene una secuencia y peso molecular que difieren de la secuencia y el peso molecular de todos los demás fragmentos que pueden liberarse del primer conjunto de polipéptidos. La mezcla comprende además un segundo conjunto de polipéptidos que

carecen del fragmento de código de barras del primer conjunto de polipéptidos, en donde tanto el primer conjunto de polipéptidos como el segundo conjunto de polipéptidos comprenden cada uno un fragmento de referencia que es común al primer conjunto de polipéptidos y al segundo conjunto de polipéptidos y liberable por digestión con la proteasa. La relación del primer conjunto de polipéptidos respecto a los polipéptidos en la mezcla tiene la fórmula:

$$\frac{[\text{polipéptidos que contienen código de barras}]}{[(\text{polipéptidos que contienen péptido de referencia}) \times N]}$$

donde N es el número de apariciones del péptido de referencia que se libera de cada polipéptido en la mezcla, y en donde cuando el primer conjunto de polipéptidos comprende un fragmento de referencia, la relación del primer conjunto de polipéptidos respecto a los polipéptidos en la mezcla que comprenden el fragmento de referencia es mayor que 0,7.

En una modalidad particular, la relación del primer conjunto de polipéptidos respecto a los polipéptidos que comprenden el fragmento de referencia es mayor que 0,8, 0,9 o 0,95.

En algunas modalidades, el fragmento de referencia aparece no más de una vez en cada polipéptido del primer conjunto de polipéptidos y el segundo conjunto de polipéptidos.

En algunas modalidades, el fragmento de referencia aparece dos veces en cada polipéptido del primer conjunto de polipéptidos y del segundo conjunto de polipéptidos.

En una modalidad particular, el primer conjunto de polipéptidos comprende un polipéptido de longitud completa, en donde el fragmento de código de barras es una porción del polipéptido de longitud completa.

En algunas modalidades, el polipéptido de longitud completa incluye los polipéptidos descritos en la presente descripción. En una modalidad particular, el fragmento de código de barras no comprende el aminoácido N-terminal y el aminoácido C-terminal del polipéptido de longitud completa.

En algunas modalidades, la mezcla de polipéptidos de diferentes longitudes difieren entre sí debido al truncamiento N-terminal, el truncamiento C-terminal o el truncamiento tanto N-terminal como C-terminal de un polipéptido de longitud completa.

En algunas modalidades, el fragmento de referencia aparece no más de una vez en cada polipéptido del primer conjunto de polipéptidos y el segundo conjunto de polipéptidos. En modalidades adicionales, el número de fragmentos de referencia en el primer conjunto de polipéptidos puede diferir del número de fragmentos de referencia en el segundo conjunto de polipéptidos, pero el número de estos en cada polipéptido de cada conjunto debe ser el mismo. En algunas modalidades, los fragmentos de referencia en los polipéptidos de la mezcla tienen una secuencia y peso molecular que difieren de la secuencia y el peso molecular de todos los demás fragmentos.

En la presente descripción se describe un método para detectar la integridad de la secuencia de polipéptidos que comprenden el primer conjunto de polipéptidos en la mezcla descrita en la presente descripción, el método comprende las etapas de digerir la mezcla de

polipéptidos con una proteasa que libera el fragmento de código de barras y el fragmento de referencia del primer conjunto de polipéptidos y libera el fragmento de referencia del segundo conjunto de polipéptidos, y determinar una relación de los fragmentos de código de barras del primer conjunto de polipéptidos respecto a los fragmentos de referencia del primer y segundo conjunto de polipéptidos. En una modalidad particular, la integridad de la secuencia de los polipéptidos del primer conjunto de polipéptidos se detecta mediante una comparación de la relación de los fragmentos con la relación esperada de los fragmentos en base al número de fragmentos de código de barras y fragmentos de referencia en los polipéptidos que comprenden el primer y segundo conjuntos de polipéptidos.

Los métodos contemplados en la presente descripción se prestan fácilmente al análisis cualitativo y cuantitativo de los polipéptidos que contienen los códigos de barras y/o fragmentos de referencia, por ejemplo mediante el uso de LC/MS. En una modalidad particular, la LC/MS es cuantitativa y detecta una cantidad isotópicamente distinguible de fragmentos de códigos de barras, fragmentos de referencia o ambos. En dichos métodos ilustrativos, la mezcla de polipéptidos se enriquece con una cantidad conocida de un "material estándar" para facilitar dicho análisis. Por ejemplo, un material estándar de este tipo es uno que comprende una versión marcada isotópicamente de dicha mezcla de una pluralidad de polipéptidos de longitud variable a analizar. Este estándar marcado isotópicamente puede añadirse a la mezcla como una secuencia completa antes de la digestión por dicha proteasa. Alternativamente, la proteasa digiere la muestra de prueba de la mezcla de polipéptidos de longitud variable y el material estándar marcado isotópicamente en reacciones separadas y el material estándar marcado isotópicamente digerido con proteasa se añade a la muestra de prueba antes del análisis por LC/MS. Los métodos de la presente invención comprenden además cuantificar la cantidad de fragmentos de código de barras, fragmentos de referencia o ambos de la muestra de prueba en comparación con la cuantificación de las cantidades detectadas isotópicamente distinguibles de fragmentos de código de barras, fragmentos de referencia o ambos.

A los expertos en la técnica se les ocurrirán variaciones y modificaciones de estas modalidades después de revisar esta descripción. Las características y aspectos anteriores pueden implementarse, en cualquier combinación y subcombinación (que incluyen múltiples combinaciones y subcombinaciones dependientes), con una o más características descritas en la presente descripción. Las diversas características descritas o ilustradas anteriormente, que incluyen los componentes de estas, pueden combinarse o integrarse en otras modalidades. Por otra parte, ciertas funciones pueden omitirse o no implementarse.

INCORPORACIÓN COMO REFERENCIA

Todas las publicaciones, patentes y solicitudes de patente mencionadas en esta descripción se incorporan en la presente descripción como referencia en la misma medida que si cada publicación, patente, o solicitud de patente individual se indicara específica e individualmente como incorporada como referencia.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Varias características de esta descripción se exponen con particularidad en las reivindicaciones adjuntas. Se podrá obtener una mejor comprensión de las características y ventajas de la presente descripción como referencia a la siguiente descripción detallada que expone modalidades ilustrativas, en las que se utilizan los principios de la invención, y los dibujos acompañantes de los cuales:

La Figura 1 representa una mezcla de polipéptidos acopladores de linfocitos T activados por proteasa XTENilados ("XPAT") que tienen longitudes variables de polipéptidos XTEN. El XPAT de longitud completa (arriba) comprende un polipéptido XTEN de 288 aminoácidos de longitud en el extremo N-terminal y un polipéptido XTEN de 864 aminoácidos de longitud en el extremo C-terminal. Pueden producirse varios truncamientos en el XPAT en uno o ambos polipéptidos XTEN N-terminal y C-terminal, por ejemplo, durante la fermentación, purificación u otras etapas en la preparación del producto. Si bien los productos que tienen truncamientos limitados (truncamientos cerca de una porción del polipéptido XTEN distal del acoplador de linfocitos T activado por proteasa unido a este) pueden funcionar de una manera similar a la construcción de longitud completa, truncamientos intensos (truncamientos más cercanos a una porción del polipéptido XTEN próximo del acoplador de linfocitos T activado por proteasa unido a este) pueden poseer propiedades farmacológicas significativamente diferentes de sus contrapartes de longitud completa. La presencia de truncamientos plantea un desafío para cuantificar las variantes farmacológicamente eficaces e ineficaces en un producto XPAT. Como se ilustra en la Figura 1 mediante el uso de XPAT de longitud completa, cada polipéptido XTEN tiene un extremo proximal y un extremo distal, en donde el extremo proximal se ubica, con relación al extremo distal, más cerca del polipéptido biológicamente activo (por ejemplo, acoplador de linfocitos T, citocina, anticuerpo monoclonal (mAb), fragmento de anticuerpo u otra proteína que esté XTENilada). En dependencia de la orientación del enlace, los extremos proximal o distal del polipéptido XTEN pueden corresponder al extremo N-terminal o al extremo C-terminal del polipéptido XTEN.

La Figura 2 representa una mezcla de polipéptidos XPAT que tienen longitudes variables de polipéptidos XTEN con código de barras. En el XPAT de longitud completa (arriba), el polipéptido XTEN N-terminal de 288 aminoácidos de longitud contiene tres secuencias de código de barras fusionadas de manera escindible, "NA", "NB" y "NC" (desde el extremo distal hasta el extremo proximal), y el polipéptido XTEN C-terminal de 864 aminoácidos de longitud contiene tres secuencias de código de barras fusionadas de manera escindible, "CC", "CB" y "CA" (desde el extremo proximal hasta el extremo distal). Cada código de barras se ubica para indicar una longitud farmacológicamente relevante del polipéptido XTEN correspondiente. Por ejemplo, los productos minoritarios de truncamiento N-terminal del XPAT, que carecen del código de barras "NA" pero que tienen los códigos de barras más proximales "NB" y "NC", pueden mostrar sustancialmente las mismas propiedades farmacológicas que la construcción de longitud completa. Por el contrario, los

principales productos de truncamiento N-terminal del XPAT, por ejemplo, que carece de los tres códigos de barras en el extremo N-terminal, pueden diferir claramente en la actividad farmacológica de la construcción de longitud completa. Se identifica una secuencia escindible proteolíticamente única a partir del polipéptido biológicamente activo (aquí, los scFv en tándem que comprenden la porción activa del acoplador de linfocitos T) del XPAT. Debido a su presencia en todas las variantes de longitud del XPAT (que incluye el XPAT de longitud completa, los truncamientos minoritarios y los truncamientos principales de este), la secuencia proteolíticamente escindible única puede usarse como referencia para cuantificar las cantidades de varios productos de truncamiento en relación con la cantidad total de la proteína biológicamente activa.

La Figura 3 ilustra un diseño potencial de un polipéptido XTEN con código de barras mediante la inserción de una secuencia generadora de código de barras en un polipéptido XTEN de propósito general (o común). El polipéptido XTEN de propósito general (o común) ilustrativo (arriba) comprende motivos de 12 mer no superpuestos en la secuencia "BCDABDCDABDCBDCDABDCB", en donde aparecen los motivos de secuencia "A", "B", "C" y "D" 3, 6, 5 y 7 veces, respectivamente. La digestión con proteasa Glu-C del polipéptido XTEN de propósito general ilustrativo (panel superior) no produce péptidos únicos excepto ambos extremos ("NT" y "CT"). La inserción de una secuencia generadora de código de barras, "X" (por ejemplo, un 12 mer único), en el polipéptido XTEN da como resultado una secuencia proteolíticamente escindible única (o secuencia de código de barras) que no aparece en ningún otro lugar del polipéptido XTEN. La secuencia generadora de código de barras, "X", puede ubicarse en donde el código de barras resultante marca una longitud farmacológicamente relevante del polipéptido XTEN. Por ejemplo, un polipéptido XTEN que carece de un código de barras puede diferir funcionalmente del polipéptido XTEN correspondiente con el código de barras. Un experto en la técnica entenderá que la secuencia generadora de código de barras ("X") puede ser la propia secuencia del código de barras. Alternativamente, la secuencia generadora de código de barras ("X") puede diferir de la secuencia de código de barras resultante. Por ejemplo, la secuencia de código de barras puede superponerse y, por lo tanto, contener parte del motivo de 12 mer anterior o siguiente.

Las Figuras 4A-4B ilustran la cuantificación del nivel de truncamiento para un polipéptido XTEN N-terminal. La Figura 4A demuestra que puede construirse un polipéptido XTEN con código de barras (panel inferior) mediante el reemplazo de un motivo de secuencia en un polipéptido XTEN de propósito general (panel superior) (por ejemplo, el tercer motivo de secuencia desde el extremo N-terminal, "D") con un motivo generador de código de barras, "X"; y, en este ejemplo, el motivo generador de código de barras ("X") es en sí mismo la secuencia de código de barras escindible proteolíticamente única. Como se muestra en el panel inferior de la Figura 4A, el código de barras se ubica en donde todas las formas de truncamiento intenso del polipéptido XTEN carecen del código de barras, y todas las formas de truncamiento limitado del polipéptido XTEN contienen el código de barras.

La Figura 4B ilustra la abundancia relativa de varios productos de escisión en dos mezclas diferentes de XPAT. En una de las mezclas, el código de barras está presente en el 99 % de las construcciones que contienen la proteína biológicamente activa. En la otra de las mezclas, el 13 % de las construcciones carecen de código de barras. Las Figuras 4A-4B ilustran el uso del polipéptido XTEN con código de barras para diferenciar entre dos mezclas de polipéptidos que tienen pesos moleculares promedio sustancialmente similares pero actividades farmacológicas claramente diferentes.

La Figura 5A ilustra la cromatografía de exclusión por tamaño (SEC) analítica de la proteína XPAT y la detección de la proteína de longitud completa y sus derivados truncados. La fracción de proteína sintética + truncamientos incluye fragmentos tan grandes como la proteína sintética intacta.

La Figura 5B ilustra la abundancia de péptidos de código de barras en preparaciones de XPAT detectadas por espectrometría de masas. Cada medición es el área XIC del código de barras en N SGPGSTPAE (SEQ ID NO. 8029) y el código de barras en C GSAPGTE (SEQ ID NO. 8023) normalizada respecto a un pico de 400 nM de su péptido sintético marcado con isótopo pesado correspondiente.

El expediente de patente o solicitud contiene al menos un dibujo ejecutado en color. La Oficina proporcionará copias de esta publicación de patente o solicitud de patente con dibujos en color previa solicitud y pago de la tasa necesaria.

Terminología

Como se usa en la presente descripción, los siguientes términos tienen los significados que se les atribuyen a menos que se especifique de cualquier otra manera.

Como se usa en la descripción y las reivindicaciones, las formas singulares "un", "una" y "el/la" incluyen referencias en plural a menos que el contexto lo indique claramente de cualquier otra manera. Por ejemplo, el término "una célula" incluye una pluralidad de células, que incluye sus mezclas.

Los términos "polipéptido", "péptido", y "proteína" se usan indistintamente en la presente descripción para referirse a polímeros de aminoácidos de cualquier longitud. El polímero puede ser lineal o ramificado, puede comprender aminoácidos modificados y puede interrumpirse por no aminoácidos. Los términos también incluyen un polímero de aminoácido que se ha modificado, por ejemplo, mediante formación de enlaces disulfuro, glucosilación, lipidación, acetilación, fosforilación, o cualquier otra manipulación, tal como conjugación con un componente de marcaje.

Como se usa en la presente descripción, el término "aminoácido" se refiere a aminoácidos ya sea naturales y/o no naturales o sintéticos, que incluyen pero no sin limitarse a glicina y tanto los isómeros ópticos D o L, como a los análogos de aminoácidos y peptidomiméticos. Se usan códigos estándar de una o tres letras para denominar aminoácidos.

Una "célula huésped" incluye una célula individual o un cultivo de células que puede ser o ha sido un receptor de los vectores humanos o animales. Las células huésped incluyen la progenie de una célula huésped única. La progenie no es necesariamente completamente

idéntica (en la morfología o en la genómica del complemento de ADN total) a la célula original debido a la variación de origen natural o la modificación genética.

Una proteína “quimérica” contiene al menos un polipéptido que comprende regiones en una posición diferente en la secuencia a la que aparece en la naturaleza. Las regiones pueden existir normalmente en proteínas separadas y se unen en el polipéptido de fusión; o pueden existir normalmente en la misma proteína pero se colocan en una nueva disposición en el polipéptido de fusión. Dichas proteínas pueden describirse como proteínas “conjugadas”, “unidas”, “fusionadas” o “de fusión”; estos términos se usan indistintamente en la presente descripción y se refieren a la unión de dos secuencias polipeptídicas más por cualquier medio, que incluye la conjugación química o los medios recombinantes. Una proteína quimérica puede crearse, por ejemplo, por síntesis química, o mediante la creación y traducción de un polinucleótido en el que las regiones peptídicas se codifican en la relación deseada.

Los términos "polinucleótidos", "ácidos nucleicos", "nucleótidos" y "oligonucleótidos" se usan indistintamente y se refieren a una forma polimérica de nucleótidos de cualquier longitud, ya sea desoxirribonucleótidos o ribonucleótidos, o análogos de estos. Los polinucleótidos pueden tener cualquier estructura tridimensional, y pueden realizar cualquier función, conocida o por descubrir o desarrollar. Un polinucleótido puede comprender nucleótidos modificados, tales como nucleótidos metilados y análogos de nucleótidos. Si están presentes, las modificaciones a la estructura de los nucleótidos pueden impartirse antes o después del ensamblaje del polímero. La secuencia de nucleótidos puede estar interrumpida por componentes no nucleotídicos. Un polinucleótido puede modificarse adicionalmente después de la polimerización, tal como mediante conjugación con un componente de marcaje.

El término “complemento de un polinucleótido” indica una molécula polinucleotídica que tiene una secuencia de bases complementaria y orientación inversa en comparación con una secuencia de referencia, en donde podría hibridarse con una secuencia de referencia con fidelidad completa.

Como se usa en la presente descripción, los polinucleótidos que tienen “homología” o que son “homólogos” son aquellos que se hibridan en condiciones rigurosas como se definen en la presente descripción y tienen al menos 70 %, preferentemente al menos 80 %, con mayor preferencia al menos 90 %, con mayor preferencia 95 %, con mayor preferencia 97 %, con mayor preferencia 98 % y aún con mayor preferencia 99 % de identidad de secuencia en comparación con esas secuencias.

Los términos “por ciento de identidad” y “% de identidad,” tal como se aplican a las secuencias polinucleotídicas, se refieren al porcentaje de coincidencias de residuos entre al menos dos secuencias polinucleotídicas alineadas mediante el uso de un algoritmo estandarizado. Un algoritmo de este tipo puede insertar, de manera estandarizada y reproducible, interrupciones en las secuencias que se comparan para optimizar la alineación entre dos secuencias, y por lo tanto lograr una comparación más significativa de las dos

secuencias. El por ciento de identidad puede medirse respecto a la longitud de una secuencia polinucleotídica definida completa, por ejemplo, como se define mediante un número de SEQ ID particular, o puede medirse respecto a una longitud más corta, por ejemplo, respecto a la longitud de un fragmento tomado de una secuencia polinucleotídica definida más larga, por ejemplo, un fragmento de al menos 45, al menos 60, al menos 90, al menos 120, al menos 150, al menos 210 o al menos 450 residuos contiguos. Tales longitudes son solo ilustrativas, y se entiende que cualquier longitud de fragmento soportada por las secuencias mostradas en la presente descripción, en las tablas, figuras o el listado de secuencias, puede usarse para describir una longitud respecto a la cual puede medirse el porcentaje de identidad.

“Por ciento (%) de identidad de secuencia de aminoácidos”, con respecto a las secuencias polipeptídicas identificadas en la presente descripción, se define como el porcentaje de residuos de aminoácidos en una secuencia de consulta que son idénticos a los residuos de aminoácidos de una segunda secuencia polipeptídica de referencia o una porción de esta, después de alinear las secuencias e introducir interrupciones, si fuera necesario, para lograr el por ciento máximo de identidad de secuencia, y sin considerar ninguna sustitución conservadora como parte de la identidad de secuencia. La alineación para los fines de determinar el por ciento de identidad de secuencia de aminoácidos puede lograrse de varias maneras que están dentro de la técnica, por ejemplo, mediante el uso de software informático disponible públicamente tal como BLAST, BLAST-2, ALIGN o software Megalign (DNASTAR). Los expertos en la técnica pueden determinar los parámetros adecuados para medir la alineación, que incluyen cualquier algoritmo necesario para lograr una alineación máxima a lo largo de toda la longitud de las secuencias que se comparan. El por ciento de identidad puede medirse respecto a la longitud de una secuencia polipeptídica definida completa, por ejemplo, como se define mediante un número de SEQ ID particular, o puede medirse respecto a una longitud más corta, por ejemplo, respecto a la longitud de un fragmento tomado de una secuencia polipeptídica definida más larga, por ejemplo, un fragmento de al menos 15, al menos 20, al menos 30, al menos 40, al menos 50, al menos 70 o al menos 150 residuos contiguos. Tales longitudes son solo ilustrativas, y se entiende que cualquier longitud de fragmento soportada por las secuencias mostradas en la presente descripción, en las tablas, figuras o el listado de secuencias, puede usarse para describir una longitud respecto a la cual puede medirse el porcentaje de identidad.

Como se usa en la presente descripción, "repetitividad" de una secuencia de aminoácidos del polipéptido XTEN se refiere a la repetitividad de 3 mer y puede medirse mediante programas o algoritmos informáticos o mediante otros medios conocidos en la técnica. La repetitividad de 3 mer de una secuencia de aminoácidos del polipéptido XTEN puede evaluarse mediante la determinación del número de apariciones de las secuencias de 3 mer superpuestas dentro del polipéptido. Por ejemplo, un polipéptido de 200 residuos de aminoácidos tiene 198 secuencias de 3 aminoácidos superpuestas (3 mer), pero el número de secuencias de 3-mer únicas depende de la cantidad de repetitividad dentro de la

secuencia. Se puede generar una puntuación (en lo sucesivo, "puntuación de subsecuencia") que refleja el grado de repetitividad de los 3 mer en la secuencia polipeptídica global. En el contexto de la presente invención, "puntuación de subsecuencia" significa la suma de las apariciones de cada marco único de 3 mer en una secuencia de 200 aminoácidos consecutivos del polipéptido dividida por el número absoluto de subsecuencias de 3 mer únicas dentro de la secuencia de 200 aminoácidos. En el Ejemplo 73 de la publicación de solicitud de patente Internacional núm. WO 2010/091122 A1, que se incorpora como referencia en su totalidad, se presentan ejemplos de dichas puntuaciones de subsecuencia derivadas de los primeros 200 aminoácidos de polipéptidos repetitivos y no repetitivos. En algunas modalidades, la presente invención proporciona polipéptidos BPXTEN, cada uno de los cuales comprende al menos un polipéptido XTEN en el que la secuencia de aminoácidos del polipéptido XTEN puede tener una puntuación de subsecuencia menor que 16, menor que 14, menor que 12 o, con mayor preferencia, menor que 10.

El término "secuencia de aminoácidos del polipéptido XTEN sustancialmente no repetitiva", como se usa en la presente descripción, se refiere a un polipéptido XTEN, en donde hay pocos o ningún caso de cuatro aminoácidos contiguos en la secuencia de aminoácidos del polipéptido XTEN que son tipos de aminoácidos idénticos y en donde la secuencia de aminoácidos del polipéptido XTEN tiene una puntuación de subsecuencia (definida en el párrafo anterior de la presente descripción) de 12, o 10 o menos, o que no hay un patrón en el orden, desde el extremo N- hasta el C-terminal, de los motivos de secuencia que constituyen la secuencia polipeptídica.

Como se expone en la presente descripción, el término "motivos de secuencia no superpuestos" incluye motivos de secuencia completamente no superpuestos, así como también motivos de secuencia no superpuestos solo parcialmente, siempre y cuando dichos motivos de secuencia no superpuestos parcialmente no se superpongan completamente.

Un "vector" es una molécula de ácido nucleico, preferentemente autorreplicativa en un huésped apropiado, que transfiere una molécula de ácido nucleico insertada hacia y/o entre células huésped. El término incluye vectores que funcionan principalmente para la inserción de ADN o ARN en una célula, replicación de vectores que funcionan principalmente para la replicación de ADN o ARN, y vectores de expresión que funcionan para la transcripción y/o traducción del ADN o ARN. También se incluyen vectores que proporcionan más de una de las funciones anteriores. Un "vector de expresión" es un polinucleótido que, cuando se introduce en una célula huésped adecuada, puede transcribirse y traducirse en polipéptido(s). Un "sistema de expresión" generalmente connota una célula huésped adecuada compuesta por un vector de expresión que puede funcionar para producir un producto de expresión deseado.

El término " $t_{1/2}$ " como se usa en la presente descripción significa la semivida terminal calculada como $\ln(2)/K_{el}$. K_{el} es la constante de la velocidad de eliminación terminal calculada mediante regresión lineal de la porción lineal terminal de la curva de concentración logarítmica frente al tiempo. La semivida típicamente se refiere al tiempo

requerido para que la mitad de la cantidad de una sustancia administrada depositada en un organismo vivo se metabolice o elimine mediante procesos biológicos normales. Los términos " $t_{1/2}$ ", "semivida terminal", "semivida de eliminación" y "semivida en circulación" se usan indistintamente en la presente descripción.

Los términos "antígeno", "antígeno objetivo" o "inmunógeno" se usan indistintamente en la presente descripción para referirse a la estructura o determinante de unión al que se une un fragmento de anticuerpo o un agente terapéutico basado en fragmento de anticuerpo o contra el que tiene especificidad.

El término "carga útil", como se usa en la presente descripción, se refiere a una secuencia proteica o peptídica que tiene actividad biológica o terapéutica; la contraparte del farmacóforo de moléculas pequeñas. Los ejemplos de cargas útiles incluyen, pero no se limitan a, citocinas, enzimas, hormonas y factores sanguíneos y de crecimiento. Las cargas útiles pueden comprender además restos fusionados genéticamente o conjugados químicamente, tales como agentes quimioterapéuticos, compuestos antivirales, toxinas o agentes de contraste. Estos restos conjugados pueden unirse al resto del polipéptido mediante un enlazador que puede ser escindible o no escindible.

Como se usa en la presente descripción, "tratamiento" o "tratar", "aliviar" y "mejorar" se usan indistintamente en la presente descripción y se refieren a un enfoque para obtener resultados beneficiosos o deseados que incluyen, pero sin limitarse a, un beneficio terapéutico y/o un beneficio profiláctico. Por "beneficio terapéutico" se entiende la erradicación o mejora del trastorno subyacente que se trata. Además, se logra un beneficio terapéutico con la erradicación o mejora de uno o más de los síntomas fisiológicos asociados con la afección patológica subyacente en donde se observa una mejora en el ser humano o animal, pese a que el ser humano o animal aún puede verse afectado con el trastorno subyacente. Para el beneficio profiláctico, las composiciones pueden administrarse a un ser humano o animal en riesgo de desarrollar una afección patológica particular, o a un ser humano o animal que informa uno o más de los síntomas fisiológicos de una enfermedad, aunque no se haya hecho un diagnóstico de esta enfermedad.

Un "efecto terapéutico" como se usa en la presente descripción, se refiere a un efecto fisiológico, que incluye, pero sin limitarse a, la cura, mitigación, mejora o prevención de la afección patológica en seres humanos u otros animales, o a mejorar de cualquier otra manera el bienestar físico o mental de seres humanos o animales, causado por la administración de un polipéptido de fusión de la invención diferente de la capacidad de inducir la producción de un anticuerpo contra un epítipo antigénico que la proteína biológicamente activa posee. La determinación de una cantidad terapéuticamente efectiva está bien dentro de la capacidad de los expertos en la técnica, especialmente a la luz de la descripción detallada proporcionada en la presente descripción.

Los términos "cantidad terapéuticamente efectiva" y "dosis terapéuticamente efectiva", como se usan en la presente descripción, se refieren a una cantidad de una proteína biológicamente activa, ya sea sola o como parte de una composición de proteína de fusión,

que puede tener cualquier efecto detectable y beneficioso en cualquier síntoma, aspecto, parámetro medido o características de un estado o afección patológica cuando se administra en una o dosis repetidas a un ser humano o animal. Tal efecto no necesariamente debe ser absoluto para ser beneficioso. La afección patológica puede referirse a un trastorno o una enfermedad.

El término “régimen de dosis terapéuticamente efectivo”, como se usa en la presente descripción, se refiere a un cronograma para múltiples dosis administradas consecutivamente de una proteína biológicamente activa, ya sea sola o como parte de una composición de proteína de fusión, en donde las dosis se administran en cantidades terapéuticamente efectivas para dar como resultado un efecto beneficioso sostenido en cualquier síntoma, aspecto, parámetro medido o característica de un estado o afección patológica.

Polipéptido de fusión

En la presente descripción se describen polipéptidos que comprenden uno o más polipéptidos recombinantes extendidos (XTEN o XTENS) (como se describe con más detalle a continuación) que pueden fusionarse o conjugarse de cualquier otra manera con otro polipéptido, particularmente un polipéptido biológicamente activo, en donde dichas modalidades se denominan BPXTEN en la presente descripción.

En algunas modalidades, el polipéptido comprende un primer polipéptido XTEN (tal como los que se describen a continuación en la sección "POLIPÉPTIDO RECOMBINANTE EXTENDIDO (XTEN)" o los descritos en cualquier otra parte en la presente descripción). En algunas modalidades, el polipéptido comprende además un segundo polipéptido XTEN (tal como los que se describen a continuación en la sección "POLIPÉPTIDO RECOMBINANTE EXTENDIDO (XTEN)" o los descritos en cualquier otra parte en la presente descripción). En algunas modalidades, el polipéptido comprende un polipéptido XTEN en o cerca de su extremo N-terminal (un "XTEN N-terminal"). En algunas modalidades, el polipéptido comprende un polipéptido XTEN en o cerca de su extremo C-terminal (un "XTEN C-terminal"). En algunas modalidades, el polipéptido comprende un polipéptido XTEN N-terminal y un polipéptido XTEN C-terminal. En algunas modalidades, el primer polipéptido XTEN es un polipéptido XTEN N-terminal y el segundo polipéptido XTEN es un polipéptido XTEN C-terminal.

El polipéptido puede comprender además un polipéptido biológicamente activo ("BP") unido al polipéptido XTEN, para formar así un polipéptido de fusión que contiene XTEN denominado polipéptido "BPXTEN" en la presente descripción.

El polipéptido XTEN puede comprender uno o más fragmentos de código de barras (como se describe con más detalle a continuación) liberables (configurados para liberarse) del polipéptido XTEN tras la digestión del polipéptido de fusión (o BPXTEN) por una proteasa. En algunas modalidades, cada fragmento de código de barras difiere en secuencia y peso molecular de todos los demás fragmentos peptídicos (que incluyen todos los demás

fragmentos de código de barras si están presentes) que pueden liberarse del polipéptido tras la digestión completa del polipéptido por la proteasa.

El polipéptido (de fusión) puede comprender uno o más fragmentos de referencia (como se describe con más detalle a continuación) liberables (configurados para liberarse) del polipéptido, por ejemplo, tras la digestión con proteasa que libera el(los) fragmento(s) de código de barras del polipéptido. En algunas modalidades, cada fragmento de referencia puede ser un único fragmento de referencia que difiere en secuencia y peso molecular de todos los demás fragmentos peptídicos que pueden liberarse del polipéptido tras la digestión del polipéptido por la proteasa.

Polipéptido recombinante extendido (XTEN)

Longitud de cadena y composición de aminoácidos

En algunas modalidades, el polipéptido XTEN comprende al menos 150 aminoácidos. En algunas modalidades, el polipéptido XTEN tiene una longitud de 150 a 3 000 aminoácidos, o una longitud de 150 a 1 000 aminoácidos, o una longitud de al menos 200, al menos 250, al menos 300, al menos 350, al menos 400, al menos 450 o al menos 500 aminoácidos. En algunas modalidades, al menos 90 % de los residuos de aminoácidos del polipéptido XTEN son glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) o prolina (P). En algunas modalidades, al menos 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de los residuos de aminoácidos del polipéptido XTEN se seleccionan de G, A, S, T, E o P. En algunas modalidades, el polipéptido XTEN comprende al menos 4 tipos diferentes de aminoácidos G, A, S, T, E o P. En algunas modalidades, el polipéptido XTEN se caracteriza porque comprende al menos 150 aminoácidos; al menos 90 % de los residuos de aminoácidos del polipéptido XTEN son G, A, S, T, E o P y comprende al menos 4 tipos diferentes de aminoácidos seleccionados de G, A, S, T, E y P que está sustancialmente aleatorizado con respecto a cualquier otro motivo de secuencia no superpuesto que comprende el polipéptido XTEN. En algunas modalidades, un polipéptido de fusión que contiene XTEN (por ejemplo, un polipéptido de fusión que comprende un polipéptido biológicamente activo conjugado con el mismo) comprende un primer polipéptido XTEN y un segundo polipéptido XTEN. En algunas modalidades, la suma del número total de aminoácidos en el primer XTEN y el número total de aminoácidos en el segundo polipéptido XTEN es al menos 300, al menos 350, al menos 400, al menos 500, al menos 600, al menos 700 o al menos 800 aminoácidos.

Motivo de secuencia no superpuesto

En algunas modalidades, los polipéptidos XTEN proporcionados en la presente descripción comprenden, o se forman a partir de, una pluralidad de motivos de secuencia no superpuestos. En algunas modalidades, al menos uno de los motivos de secuencia no superpuestos es recurrente (o se repite al menos dos veces en el XTEN), y en donde al menos otro de los motivos de secuencia no superpuestos no es recurrente (o se encuentra solo una vez dentro del XTEN). En algunas modalidades, la pluralidad de motivos de secuencia no superpuestos comprende un conjunto de motivos de secuencia no

superpuestos (recurrentes), en donde cada uno de dichos motivos de secuencia se repite al menos dos veces en el XTEN; y un motivo de secuencia no superpuesto (no recurrente) que aparece (o se encuentra) solo una vez dentro del XTEN. En algunas modalidades, cada motivo de secuencia no superpuesto tiene una longitud de 9 a 14 (o de 10 a 14, o de 11 a 13) aminoácidos. En algunas modalidades, cada motivo de secuencia no superpuesto tiene una longitud de 12 aminoácidos. En algunas modalidades, la pluralidad de motivos de secuencia no superpuestos comprende un conjunto de motivos de secuencia no superpuestos (recurrentes), en donde cada uno de dichos motivos de secuencia se repite al menos dos veces en el XTEN; y tiene una longitud de entre 9 y 14 aminoácidos. En algunas modalidades, el conjunto de motivos de secuencia no superpuestos (recurrentes) comprende motivos de secuencia de 12 mer identificados en la presente descripción mediante las SEQ ID NO: 182-203 y 1715-1722 en la Tabla 1. En algunas modalidades, el conjunto de motivos de secuencia no superpuestos (recurrentes) comprende motivos de secuencia de 12 mer identificados en la presente descripción mediante las SEQ ID NO: 186-189 en la Tabla 1. En algunas modalidades, el conjunto de motivos de secuencia no superpuestos (recurrentes) comprende al menos dos, al menos tres o los cuatro motivos de secuencia de 12 mer de las SEQ ID NO: 186-189 en la Tabla 1.

Tabla 1. Motivos de secuencia de 12 mer ilustrativos para la construcción de XTEN

Familia de motivos*	SEQ ID NO	Secuencia de aminoácidos
AD	182	GESPGGSSGSES
AD	183	GSEGSSGPGESE
AD	184	GSSESGSSEGGP
AD	185	GSGGEPSESGSS
AE, AM	186	GSPAGSPTSTEE
AE, AM, AQ	187	GSEPATSGSETP
AE, AM, AQ	188	GTSESATPESGP
AE, AM, AQ	189	GTSTEPSEGSAP
AF, AM	190	GSTSEPSGTAP
AF, AM	191	GTSTPESGSASP
AF, AM	192	GTSPSGESSTAP
AF, AM	193	GSTSSTAESPGP
AG, AM	194	GTPGSGTASSSP
AG, AM	195	GSSTPSGATGSP
AG, AM	196	GSSPSASTGTGP
AG, AM	197	GASPGTSSTGSP
AQ	198	GEPAGSPTSTSE
AQ	199	GTGEPSSTPASE
AQ	200	GSGPSTESAPTE
AQ	201	GSETPSGPSETA
AQ	202	GPSETSTSEPGA
AQ	203	GSPSEPTEGTSA
BC	1715	GSGASEPTSTEP
BC	1716	GSEPATSGTEPS
BC	1717	GTSEPSTSEPGA
BC	1718	GTSTEPSEPGSA
BD	1719	GSTAGSETSTEA
BD	1720	GSETATSGSETA
BD	1721	GTSESATSESGA

Familia de motivos*	SEQ ID NO	Secuencia de aminoácidos
BD	1722	GTSTEASEGSAS

*Indica secuencias de motivos individuales que, cuando se usan juntas en varias permutaciones, dan como resultado una "secuencia de familia"

Fragmento de código de barras

En algunas modalidades, los polipéptidos proporcionados en la presente descripción comprenden un fragmento de código de barras (por ejemplo, un primer, segundo o tercer fragmento de código de barras de un polipéptido XTEN) liberable del polipéptido tras la digestión por una proteasa. En algunas modalidades, el fragmento de código de barras es una porción del XTEN que incluye al menos parte del motivo de secuencia (no recurrente, no superpuesto) que aparece (o se encuentra) solo una vez dentro del XTEN; y difiere en secuencia y peso molecular de todos los demás fragmentos peptídicos que pueden liberarse del polipéptido tras la digestión completa del polipéptido por la proteasa. Un experto en la técnica entenderá que el término "fragmento de código de barras" (o "código de barras" o "secuencia de código de barras") puede referirse a la porción del XTEN identificado en la presente descripción por fusión de manera escindible dentro del polipéptido, o al fragmento peptídico resultante liberado del polipéptido.

En algunas modalidades, el fragmento de código de barras no incluye el aminoácido N-terminal o el aminoácido C-terminal del polipéptido XTEN. Como se describe con más detalle a continuación o se describe en cualquier parte de la presente descripción, en algunas modalidades, el fragmento de código de barras es liberable (configurado para liberarse) tras la digestión con Glu-C del polipéptido de fusión. En algunas modalidades, el fragmento de código de barras no incluye un ácido glutámico que está inmediatamente adyacente a otro ácido glutámico en el polipéptido XTEN. En algunas modalidades, el fragmento de código de barras tiene un ácido glutámico en su extremo C-terminal. Un experto en la técnica entenderá que el extremo C-terminal de un fragmento de código de barras puede referirse al "último" (o el más C-terminal) residuo de aminoácido dentro del fragmento de código de barras, cuando se fusiona de manera escindible dentro de un polipéptido XTEN, incluso si otros residuos de aminoácidos "que no son código de barras" se ubican en el extremo C-terminal del fragmento del código de barras dentro del mismo polipéptido XTEN. En algunas modalidades, el fragmento de código de barras tiene un aminoácido N-terminal inmediatamente precedido por un residuo de ácido glutámico. En algunas modalidades, el residuo de ácido glutámico que precede al aminoácido N-terminal no está inmediatamente adyacente a otro residuo de ácido glutámico. En algunas modalidades, el fragmento de código de barras no incluye un residuo de ácido glutámico en una posición distinta del extremo C-terminal del fragmento de código de barras, a menos que el ácido glutámico esté seguido inmediatamente por una prolina. En algunas modalidades, el fragmento de código de barras se ubica entre 10 y 150, o entre 10 y 125 aminoácidos del extremo N-terminal del polipéptido o el extremo C-terminal del

polipéptido. En algunas modalidades, el fragmento de código de barras se ubica dentro o en una ubicación de 300, 280, 260, 250, 240, 220, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 48, 40, 36, 30, 24, 20, 12 o 10 aminoácidos del extremo N-terminal del polipéptido, o en una ubicación en un intervalo entre cualquiera de los anteriores. En algunas modalidades, el fragmento de código de barras se ubica dentro de 200, dentro de 150, dentro de 100 o dentro de 50 aminoácidos del extremo N-terminal del polipéptido. En algunas modalidades, el fragmento de código de barras se ubica entre 10 y 200, entre 30 y 200, entre 40 y 150 o entre 50 y 100 aminoácidos del extremo N-terminal del polipéptido. En algunas modalidades, el fragmento de código de barras se ubica dentro de 300, 280, 260, 250, 240, 220, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 48, 40, 36, 30, 24, 20, 12 o 10 aminoácidos del extremo C-terminal del polipéptido, o en un intervalo entre cualquiera de los anteriores. En algunas modalidades, el fragmento de código de barras se ubica dentro de 200, dentro de 150, dentro de 100 o dentro de 50 aminoácidos del extremo C-terminal del polipéptido. En algunas modalidades, el fragmento de código de barras se ubica entre 10 y 200, entre 30 y 200, entre 40 y 150, o entre 50 y 100 aminoácidos del extremo C-terminal del polipéptido. En algunas modalidades, el fragmento de código de barras no incluye el aminoácido N-terminal o el aminoácido C-terminal del polipéptido; no incluye un ácido glutámico que está inmediatamente adyacente a otro ácido glutámico en el XTEN; tiene un ácido glutámico en su extremo C-terminal; tiene un aminoácido N-terminal que está inmediatamente precedido por un residuo de ácido glutámico; y (v) se ubica de 10 a 150, o de 10 a 125 aminoácidos del extremo N-terminal del polipéptido o del extremo C-terminal del polipéptido. En algunas modalidades, el residuo de ácido glutámico que precede al aminoácido N-terminal no está inmediatamente adyacente a otro residuo de ácido glutámico. En algunas modalidades, el fragmento de código de barras no incluye un residuo de ácido glutámico en una posición distinta del extremo C-terminal del fragmento de código de barras, a menos que el ácido glutámico esté seguido inmediatamente por una prolina. En algunas modalidades, para un polipéptido XTEN con código de barras fusionado con un polipéptido biológicamente activo, al menos un fragmento de código de barras (o al menos dos fragmentos de código de barras, o tres fragmentos de código de barras) contenido en el XTEN con código de barras se ubica al menos 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300 aminoácidos del polipéptido biológicamente activo. En algunas modalidades, el fragmento de código de barras tiene una longitud de al menos 4, al menos 5, al menos 6, al menos 7 o al menos 8 aminoácidos. En algunas modalidades, el fragmento de código de barras tiene una longitud de al menos 4 aminoácidos. En algunas modalidades, el fragmento de código de barras tiene una longitud de 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 aminoácidos, o en un intervalo entre cualquiera de los valores anteriores. En algunas modalidades, el fragmento de código de barras tiene una longitud de entre 4 y 20, entre 5 y 15, entre 6 y 12 o entre 7 y 10 aminoácidos. En algunas modalidades,

el fragmento de código de barras se selecciona de las SEQ ID NO: 8020-8030 (BAR001-BAR011) en la Tabla 2.

Tabla 2. Fragmentos de códigos de barras ilustrativos liberables en la digestión con Glu-C

Secuencia de aminoácidos		SEQ ID NO:
SPATSGSTPE	BAR001	8020
GSAPATSE	BAR002	8021
GSAPGTATE	BAR003	8022
GSAPGTE	BAR004	8023
PATSGPTE	BAR005	8024
SASPE	BAR006	8025
PATSGSTE	BAR007	8026
GSAPGTSAE	BAR008	8027
SATSGSE	BAR009	8028
SGPGSTPAE	BAR010	8029
SGSE	BAR011	8030

En algunas modalidades, un polipéptido XTEN con código de barras comprende solo un fragmento de código de barras. En algunas modalidades, un polipéptido XTEN con código de barras comprende un conjunto de fragmentos de código de barras, que comprende un primer fragmento de código de barras, tal como los descritos anteriormente o en cualquier otra parte en la presente descripción. En estas modalidades, cada miembro del conjunto de secuencias de códigos de barras puede distinguirse de todas las demás secuencias de códigos de barras sobre la base de la secuencia de aminoácidos o el peso molecular (en donde estos métodos para distinguir diferentes secuencias de códigos de barras estarán relacionados). En algunas modalidades, el conjunto de fragmentos de código de barras comprende un segundo fragmento de código de barras (o un fragmento de código de barras adicional), tal como los descritos anteriormente o en cualquier otra parte en la presente descripción. En algunas modalidades, el conjunto de fragmentos de código de barras comprende un tercer fragmento de código de barras, tal como los descritos anteriormente o en cualquier otra parte en la presente descripción. El conjunto de fragmentos de código de barras fusionados dentro de un polipéptido XTEN N-terminal puede denominarse conjunto de códigos de barras en N-terminal ("un conjunto N-terminal"). El conjunto de fragmentos de código de barras fusionados dentro de un polipéptido XTEN C-terminal puede denominarse conjunto de códigos de barras en C-terminal ("un conjunto C-terminal"). En algunas modalidades, el conjunto N-terminal comprende un primer fragmento de código de barras y un segundo fragmento de código de barras. En algunas modalidades, el conjunto N-terminal comprende además un tercer fragmento de código de barras. En algunas modalidades, el conjunto C-terminal comprende un primer fragmento de código de barras y un segundo fragmento de código de barras. En algunas modalidades, el conjunto C-terminal comprende además un tercer fragmento de código de barras. En algunas modalidades, el segundo fragmento de código de barras se ubica en el extremo N-terminal del primer fragmento de código de barras del mismo conjunto. En algunas modalidades, el segundo fragmento de código de barras se ubica en el extremo C-terminal del primer fragmento de

código de barras del mismo conjunto. En algunas modalidades, el tercer fragmento de código de barras se ubica en el extremo N-terminal del primer y segundo fragmentos de código de barras. En algunas modalidades, el tercer fragmento de código de barras se ubica en el extremo C-terminal del primer y segundo fragmentos de código de barras. En algunas modalidades, el tercer fragmento de código de barras se ubica entre el primer y el segundo fragmento de código de barras. En algunas modalidades, el polipéptido comprende un conjunto de fragmentos de código de barras que incluye un primer fragmento de código de barras, un (segundo) fragmento de código de barras adicional y al menos un fragmento de código de barras adicional, en donde cada fragmento de código de barras del conjunto de fragmentos de código de barras es una parte del segundo polipéptido XTEN y difiere en secuencia y peso molecular de todos los demás fragmentos peptídicos que pueden liberarse del polipéptido tras la digestión completa del polipéptido por la proteasa.

XTEN con código de barras ilustrativo

Las secuencias de aminoácidos de 13 XTEN con código de barras ilustrativos, que contienen un código de barras (por ejemplo, SEQ ID NO: 8002-8003, 8005-8009 y 8013), o dos códigos de barras (por ejemplo, SEQ ID NO: 8001, 8004, 8010 y 8012), o tres códigos de barras (por ejemplo, SEQ ID NO: 8011), se ilustran en la Tabla 3a. Entre estos 13 polipéptidos XTEN con código de barras ilustrativos, seis (SEQ ID NO: 8001-8003, 8008-8009 y 8011) pueden fusionarse con una proteína biológicamente activa en el extremo C-terminal de la proteína biológicamente activa, y siete (SEQ ID NO: 8004-8007, 8010 y 8012-8013) pueden fusionarse en el extremo N-terminal de la proteína biológicamente activa. En algunas modalidades, el polipéptido XTEN tiene al menos 90 %, al menos 92 %, al menos 95 %, al menos 98 %, al menos 99 % o 100 % de identidad de secuencia con una secuencia seleccionada de las SEQ ID NO: 8001-8019 en la Tabla 3a.

Tabla 3a. XTEN con código de barras ilustrativos

SEQ ID NO.	Tipo de XTEN	# de código(s) de barras	Secuencia de aminoácidos	# Total de AA
8001	XTEN C-terminal	2	PGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGSEPATSG SETPGSPAGSPTSTEEGTSESATPE SGPGTSTEPSEGSAPGTSTEPSEGSAPG PGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT	864

SEQ ID NO.	Tipo de XTEN	# de código(s) de barras	Secuencia de aminoácidos	# Total de AA
			STEPSEGSAPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGSPAGS PTSTEEGTSESATPESGPGTSTEPSE GSAPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGS PAGSPTSTEEGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGftabTSESATPESGPGSE PATSGP TE SGSEPATSGSETPGSPAGSPTSTE EGTSTEPSE GSAPGTE STPSEGSAPG SEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGEPEA	
8002	XTEN C-terminal	1	PGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGSEPAT GSETPGSPAGSPTSTEEGTSESATPE SGPGTSTEPSEGSAPGTSTEPSEGSA PGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGSPAGS PTSTEEGTSESATPESGPGTSTEPSE GSAPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGS PAGSPTSTEEGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGPTES GSEPATSGSETPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEP SEGSAPGEPEA	864
8003	XTEN C-terminal	1	PGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGSEPAT GSETPGSPAGSPTSTEEGTSESATPE SGPGTSTEPSEGSAPGTSTEPSEGSA PGSPAGSPTSTEEGTSTEPSEGSAPG	864

SEQ ID NO.	Tipo de XTEN	# de código(s) de barras	Secuencia de aminoácidos	# Total de AA
			TSTEPSEGSAPGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGSPAGS PTSTEEGTSESATPESGPGTSTEPSE GSAPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGS PAGSPTSTEEGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGSETP GSEPATSGSETPGSPAGSPTSTEEGT STEPSE <u>GSAPGTE</u> STPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEP SEGSAPGEPEA	
8004	XTEN N-terminal	2	ASSPAGSPTSTESGTSESATPESGPG TETEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPE <u>SGPGSTPA</u> <u>ES</u> GSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPE SGPGE <u>SPATSGSTPE</u> GTSESATPESG PGSPAGSPTSTEEGSPAGSPTSTEEG TSTEPSEGSAPGTSESATPESGPGTS ESATPESGPGTSESATPESGPGSEPA TSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEG SAPGGSAP	288
8005	XTEN N-terminal	1	ASSPAGSPTSTESGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGTSTEPSE GSAPGSPAGSPTSTEEGTSESATPES GPGE <u>SPATSGSTPE</u> GTSESATPESGP GSPAGSPTSTEEGSPAGSPTSTEEGT STEPSEGSAPGTSESATPESGPGTSE SATPESGPGTSESATPESGPGSEPAT SGSETPGSEPATSGSETPGSPAGSPT STEEGTSTEPSEGSAPGTSTEPSEGS APGGSAP	288
8006	XTEN N-terminal	1	ASSPAGSPTSTESGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPE <u>SGPGSTPA</u> <u>ES</u> GSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPE	288

SEQ ID NO.	Tipo de XTEN	# de código(s) de barras	Secuencia de aminoácidos	# Total de AA
			SGPGEEPATSGSTPEGTSESATPESG PGSPAGSPTSTEEGSPAGSPTSTEEG TSTEPSEGSAPGTSESATPESGPGTS ESATPESGPGTSESATPESGPGSEPA TSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEG SAPGGSAP	
8007	XTEN N-terminal	1	ASSPAGSPTSTESGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPES <u>SGPGSTPA</u> <u>ESGSETPGTSESATPESGPGTSTEPS</u> EGSAPGSPAGSPTSTEEGTSESATPE SGPGSEPATSGSETPGTSESATPESG PGSPAGSPTSTEEGSPAGSPTSTEEG TSTEPSEGSAPGTSESATPESGPGTS ESATPESGPGTSESATPESGPGSEPA TSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEG SAPGGSAP	288
8008	XTEN C-terminal	1	PGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGSEPAT GSETPGSPAGSPTSTEEGTSESATPE SGPGTSTEPSEGSAPGTSTEPSEGSA PGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGSPAGS PTSTEEGTSESATPESGPGTSTEPSE GSAPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGS PAGSPTSTEEGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGSETP GSEPATSGSETPGSPAGSPTSTEEGT STEPSE <u>GSAPGTE</u> STPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEP SEGSAPG	864
8009	XTEN C-terminal	1	PGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGTSTEPSEGSAPGTSTE	576

SEQ ID NO.	Tipo de XTEN	# de código(s) de barras	Secuencia de aminoácidos	# Total de AA
			PSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTS TEEGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSPAG SPTSTEEGSPAGSPTSTEEGSPAGSP TSTEEGTSESATPESGPGTSTEPSEG SAPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGSP AGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEGTSTEPSEG SAPGTSESATPESGPGTSESATPESG PGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTS TEPSE <u>GSAPGTE</u> STPSEGSAPGSEPA TSGSETPGTSESATPESGPGTSTEPSE GSAPG	
8010	XTEN N-terminal	2	SAGSPGSPAGSPTSTEEGTSESATPE SGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPG TSESATPE <u>SGPGSTPAES</u> GSETPGSE PATSGSETPGSPAGSPTSTEEGTSES ATPESGPGTSTEPSEGSAPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEG SAPGTSTEPSEGSAPGTSESATPESG PGTSTEPSEGSAPGTSESATPESGPG <u>SEPATSGSTE</u> TPGTSTEPSEGSAPGT STEPSEGSAPGTSESATPESGPGTSE SATPESGPGSPAGSPTSTEEGTSESA TPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGS PAGSPTSTEEGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPGTSESATPES GPGSPAGSPTSTEEGSPAGSPTSTEE GSPAGSPTSTEEGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSE GSAPGSPAGSPTSTEEGTSESATPES GPGSEPATSGSETPGTSESATPESGP GSPAGSPTSTEEGSPAGSPTSTEEGT STEPSEGSAPGTSESATPESGPGTSE SATPESGPGTSESATPESGPGSEPAT SGSETPGSEPATSGSETPGSPAGSPT STEEGTSTEPSEGSAPGTSTEPSEGS APGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPS	1152

SEQ ID NO.	Tipo de XTEN	# de código(s) de barras	Secuencia de aminoácidos	# Total de AA
			EGSAPGTSESATPESGPGSEPATSGS ETPGSEPATSGSETPGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGTSES ATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEG SAPGTSTEPSEGSAPGTSESATPESG PGTESAS	
8011	XTEN C-terminal	3	SAGSPGSPAGSPTSTEEGTSESATPE SGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPG TSESATPESGPGSEPATSGSETPGSE PATSGSETPGSPAGSPTSTEEGTSES ATPESGPGTSTEPSEGSAPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEG SAPGTSTEPSEGSAPGTSESATPESG PGTSTEPSEGSAPGTSESATPESGPG SEPATSGSETPGTSTEPSEGSAPGTS TEPSEGSAPGTSESATPESGPGTSES ATPESGPGSPAGSPTSTEEGTSESAT PESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGTSTEPSEGS PGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGSP AGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGTSESATPESG PGSPAGSPTSTEEGSPAGSPTSTEEG SPAGSPTSTEEGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEG SAPGSPAGSPTSTEEGTSESATPESG PGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTS TEPSEGSAPGTSESATPESGPGTSES ATPESGPGTSESATPESGPGSEPATS GSETPGSEPATSGSETPGSPAGSPTS TEEGTSTEPSEGSAPGTSTEPSEGS PGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAPGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGSPAG SPTSTEEGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGSEPATSGSE TPGSE <u>PATSGSTE</u> TPGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGSPAGSPTSTEEGTS TEPSE <u>GSAPGTATE</u> SPEGSAPGTSES ATPESGPGTSTEPSE <u>GSAPG TSAES</u> ATPESGPGSEPATSGSETPGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPE SGPGTESAS	1152
8012	XTEN N-terminal	2	GSPAGSPTSTEEGTSESATPESGPGT STEPSEGSAPGSPAGSPTSTEEGTST	864

SEQ ID NO.	Tipo de XTEN	# de código(s) de barras	Secuencia de aminoácidos	# Total de AA
			EPSEGSAPGTSTEPSE <u>GSAPATSESA</u> TPESGPGSEPATSGSETPGSEPATSG SETPGSPAGSPTSTEEGTSE <u>SASPE</u> GPGTSTEPSEGSAPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGT STEPSEGSAPGTSESATPESGPGTST EPSEGSAPGTSESATPESGPGSEPAT SGSETPGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGSPAGS PTSTEEGTSESATPESGPGTSTEPSE GSAPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGS PAGSPTSTEEGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGSETP GSEPATSGSETPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEP SEGSAP	
8013	XTEN N-terminal	1	GSPAGSPTSTEEGTSESATPESGPGT STEPSEGSAPGSPAGSPTSTEEGTST EPSEGSAPGTSTEPSEGSAPGTSESA TPESGPGSE <u>SATSGSE</u> TPGSEPATSG SETPGSPAGSPTSTEEGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGT STEPSEGSAPGTSESATPESGPGTST EPSEGSAPGTSESATPESGPGSEPAT SGSETPGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGSPAGS PTSTEEGTSESATPESGPGTSTEPSE GSAPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETP	864

SEQ ID NO.	Tipo de XTEN	# de código(s) de barras	Secuencia de aminoácidos	# Total de AA
			GTSESATPESGPGTSTEPSEGSAPGS PAGSPTSTEEGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGSETP GSEPATSGSETPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEP SEGSAP	
8014	XTEN N-terminal (con etiqueta de His)	1	SPAGSPTSTESGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSESATPES <u>SGPGSTPAES</u> <u>GSET</u> PGTSESATPESGPGTSTEPSEG SAPGSPAGSPTSTEEGTSESATPESG PGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTS TEPSEGSAPGTSESATPESGPGTSES ATPESGPGTSESATPESGPGSEPATSG GSETPGSEPATSGSETPGSPAGSPTS TEEGTSTEPSEGSAPGTSTEPSEGSAP PGGSAP	292
8015	XTEN C-terminal	1	PGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTS TEEGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSPAG SPTSTEEGSPAGSPTSTEEGSPAGSP TSTEEGTSESATPESGPGTSTEPSEG SAPGTSESATPESGPGSEPATSGSETP PGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGSP AGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEGTSTEPSEG SAPGTSESATPESGPGTSESATPESG PGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTS TEPSE <u>GSAPGTE</u> STPSEGSAPGSEPA TSGSETPGTSESATPESGPGTSTEPSE GSAPGEPEA	582
8016	XTEN C-terminal	1	TPESGPGTSESATPESGPGSPAGSPT STEEGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGT STEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPTSTEEGS PAGSPTSTEEGSPAGSPTSTEEGTSE	576

SEQ ID NO.	Tipo de XTEN	# de código(s) de barras	Secuencia de aminoácidos	# Total de AA
			SATPESGPGTSTEPSEGSAPGTSESA TPESGPGSEPATSGSETPGTSESATP ESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGT SESATPESGPGSPAGSPTSTEEGSPA GSPTSTEEGTSTEPSEGSAPGTSESA TPESGPGTSESATPESGPGTSESATP ESGPGSEPATSGSETPGSE <u>SATSGSE</u> TPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGSEPATSGSETPGT SESA	
8017	XTEN C-terminal	1	GTSTEPSEGSAPGTSESATPESGPGS EPATSGSETPGSEPATSGSETPGSPA GSPTSTEEGTSESATPESGPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTS TEEGTSTEPSEGSAPGTSTEPSEGS APGTSESATPESGPGTSTEPSEGSAPG TSESATPESGPGSEPATSGSETPGTS TEPSEGSAPGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGSPAGSP TSTEEGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGSEPATSG GSETPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGS APGTSE <u>SASPE</u> SGPGSPAGSPTSTEEG SPAGSPTSTEEGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPE SGP	576
8018	XTEN C-terminal	1	GSETPGSPAGSPTSTEEGTSESATPE SGPGTSTEPSEGSAPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGSPAGS PTSTEEGTSESATPESGPGTSTEPSE GSAPGTSESATPESGPGSE <u>PATSGST</u> <u>ET</u> GTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGS	576

SEQ ID NO.	Tipo de XTEN	# de código(s) de barras	Secuencia de aminoácidos	# Total de AA
			PAGSPTSTEEGTSESATPESGPGSEP ATS	
8019	XTEN C-terminal	1	EGSAPGTSTEPSEGSAPGTSESATPE SGPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGTS ESATPESGPGSPAGSPTSTEEGTSES ATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSTEPSEG SAPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTS ESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSESATPE SGPGSPAGSPTSTEEGSPAGSPTSTE EGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSEPA TSGSETPGTSE <u>SASPESG</u> PGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPE SGPGSEPATSGSETPGTSESATPESG PGSPAGSPTSTEEGSPAGSPTSTEEG TSTEPSEGSAPGTSESATPESGPGTS ESAT	576

En algunas modalidades, puede obtenerse un polipéptido XTEN con código de barras al crear una o más mutaciones en un polipéptido XTEN de propósito general, tal como cualquiera de los enumerados en Tabla 3b, de acuerdo con uno o más de los siguientes criterios: minimizar el cambio de secuencia en el polipéptido XTEN, minimizar el cambio en la composición de aminoácidos en el polipéptido XTEN, mantener sustancialmente la carga neta del polipéptido XTEN, mantener sustancialmente (o mejorar) una baja inmunogenicidad del polipéptido XTEN, y mantener (o mejorar) sustancialmente las propiedades farmacocinéticas del polipéptido XTEN. En algunas modalidades, la secuencia de aminoácidos del polipéptido XTEN tiene al menos 90 %, al menos 92 %, al menos 95 %, al menos 98 %, al menos 99 % o 100 % de identidad de secuencia con cualquiera de las SEQ ID NO: 676-734 enumeradas en la Tabla 3b. En algunas modalidades, la secuencia XTEN, que tiene al menos 90 % (por ejemplo, al menos 92 %, al menos 95 %, al menos 98 % o al menos 99 %) pero menos de 100 % de identidad de secuencia con cualquiera de las SEQ ID NO: 676-734 enumeradas en la Tabla 3b, se obtiene por una o más mutaciones (por ejemplo, menos de 10, menos de 8, menos de 6, menos de 5, menos de 4, menos de 3, menos de 2 mutaciones) de la secuencia correspondiente de la Tabla 3b. En algunas modalidades, la una o más mutaciones comprenden la eliminación de un residuo de ácido glutámico, la inserción de un residuo de ácido glutámico, la sustitución de un residuo de ácido glutámico o la sustitución por un residuo de ácido glutámico, o cualquiera de sus combinaciones. En algunas modalidades, donde la secuencia de aminoácidos del

polipéptido XTEN difiere, pero tiene al menos 90 % (por ejemplo, al menos 92 %, al menos 95 %, al menos 98 % o al menos 99 %) de identidad de secuencia con cualquiera de las SEQ ID NO: 676-734 enumeradas en la Tabla 3b, al menos 80 %, al menos 90 %, al menos 95 %, al menos 97 % o aproximadamente 100 % de la diferencia entre la secuencia de aminoácidos del polipéptido XTEN y la secuencia correspondiente de la Tabla 3b involucra la eliminación de un residuo de ácido glutámico, la inserción de un residuo de ácido glutámico, la sustitución de un residuo de ácido glutámico o la sustitución por un residuo de ácido glutámico, o cualquiera de sus combinaciones. En algunas de tales modalidades, al menos 80 %, al menos 90 %, al menos 95 %, al menos 97 % o aproximadamente 100 % de la diferencia entre la secuencia de aminoácidos del polipéptido XTEN y la secuencia correspondiente de la Tabla 3b involucra una sustitución de un residuo de ácido glutámico, o una sustitución por un residuo de ácido glutámico, o ambos. El término "una sustitución de un primer aminoácido", como se usa en la presente descripción, se refiere al reemplazo del primer residuo de aminoácido por un segundo residuo de aminoácido, lo que da como resultado que el segundo residuo de aminoácido ocupe la posición de sustitución en la secuencia obtenida. Por ejemplo, "una sustitución de ácido glutámico" se refiere al reemplazo del residuo de ácido glutámico (E) por un residuo que no es ácido glutámico (por ejemplo, serina (S)). El término "una sustitución por un primer aminoácido", como se usa en la presente descripción, se refiere al reemplazo de un segundo residuo de aminoácido por el primer residuo de aminoácido, lo que da como resultado que el primer residuo de aminoácido ocupe la posición de sustitución en la secuencia obtenida. Por ejemplo, "una sustitución por ácido glutámico" se refiere a la sustitución de un residuo que no es ácido glutámico (por ejemplo, serina (S)) por un residuo de ácido glutámico.

Tabla 3b. XTEN de propósito general ilustrativos para la modificación a XTEN(s) con código de barras

SEQ ID NO	Nombre de XTEN	Secuencia de aminoácidos
676	AE144	GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP
677	AE144_1A	SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG
678	AE144_2A	TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG
679	AE144_2B	TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG
680	AE144_3A	SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG

SEQ ID NO	Nombre de XTEN	Secuencia de aminoácidos
		TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPG
681	AE144_3B	SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPG
682	AE144_4A	TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESAT PESGPGTSTEPSEGSAPG
683	AE144_4B	TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESAT PESGPGTSTEPSEGSAPG
684	AE144_5A	TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEG TSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEG
685	AE144_6B	TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESAT PESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPG
686	AE288_1	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP GTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPAT SGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP
687	AE288_2	GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEE GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP
688	AE576	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESA TPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGS PTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGS PTSTEEGTSESATPESGPGTSTEPSEGSAP
689	AE624	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTST

SEQ ID NO	Nombre de XTEN	Secuencia de aminoácidos
		SPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAG SPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAG SPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAG SPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPG
693	AE1152	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESA TPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGS PTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGS PTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEE GTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESA TPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAP
694	AE144A	STEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGT SESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPT STEEGSPAGSPTSTEEGS
695	AE144B	SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPG
696	AE180A	TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGSEPAT
697	AE216A	PESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP

SEQ ID NO	Nombre de XTEN	Secuencia de aminoácidos
		TSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSP TSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG TSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPG TSESAT
705	AE504A	EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPG TSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATS GSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPS
706	AE540A	TPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP GTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPAT SGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP
707	AE576A	TPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAP GTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPAT SGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGSEPATSGSETPGTSESA
708	AE612A	GSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPG TSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEG SPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG

SEQ ID NO	Nombre de XTEN	Secuencia de aminoácidos
		SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGSEPATSGSETPGTSESAT
709	AE648A	PESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG TSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESAT PESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEG TSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPG TSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATS GSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPG SEPATSGSETPGTSESAT
710	AE684A	EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPG TSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATS GSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPG TSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEG SPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGSEPATS
711	AE720A	TSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTE PSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTE EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSES ATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSES ATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSTE

SEQ ID NO	Nombre de XTEN	Secuencia de aminoácidos
712	AE756A	TSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSA PGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTE PSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTE EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSES ATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSES ATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSES
713	AE792A	EGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPG TSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPAT GSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPG TSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPAT GSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEG SPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGSEPATSGSETPGTSESATPESGPGTSTEPS
714	AE828A	PESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPAT GSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPG SEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESAT PESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPG TSESATPESGPGTSESATPESGPGTSESATPESGPGSEPAT GSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGSEPATSGSETPGTSESAT

SEQ ID NO	Nombre de XTEN	Secuencia de aminoácidos
715	AE869	GSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSP AGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTS TEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEG SAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSP AGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTS TEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEG SAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPE SGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGR
716	AE144_R1	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSESATPESGPGTESASR
717	AE288_R1	SAGSPTGPGSEPATSGSETPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSE SATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSE TPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAPASR
718	AE432_R1	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTESASR
719	AE576_R1	SAGSPTEEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTST EEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPA GSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPES GPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPA GSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAPASR

SEQ ID NO	Nombre de XTEN	Secuencia de aminoácidos
720	AE864_R1	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEG SPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESAT PESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSG SETPGTSESATPESGPGTESASR
721	AE712	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAG SPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSES ATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP PGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAG SPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAG SPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEAHHH
722	AE864_R2	GSPGAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEG SPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESAT PESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG

SEQ ID NO	Nombre de XTEN	Secuencia de aminoácidos
		SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTESASR
723	AE288_3	SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGTSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPG
724	AE284	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGTSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSE
725	AE292	SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGTSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGGSAP
726	AE864_2	AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEG SAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTS TEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTS TEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEG SAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTS TEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTS ESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTS TEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTS TEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTS ESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTS TEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTS ESATPESGPGTSTEPSEGA AEPEA
727	AE867	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESA TPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGS PTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGS

SEQ ID NO	Nombre de XTEN	Secuencia de aminoácidos
		PTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETP GTSESATPESGPGTSTEPSEGA AEPEA
728	AE867_2	SPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTST EEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTST EPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPA GSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTST EEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPA GSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPES GPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPA GSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAPG
729	AE868	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAG SPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSES ATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAG SPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPA SPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPA SPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAG SPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSET PGTSESATPESGPGTSTEPSEGA AEPEA
730	AE144_7A	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESA TPESGPGTSTEPSEGSAP
731	AE292	SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESAT

SEQ ID NO	Nombre de XTEN	Secuencia de aminoácidos
		PESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEG SPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESAT PESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGGSAP
732	AE293	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAG SPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSES ATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS PGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAG SPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAG SPTSTEEGTSESATPESGPGTSTEPSEGSAPGAAEPEA
733	AE300	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAG SPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSES ATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS PGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAG AAEPEA
734	AE584	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAG SPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSES ATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS PGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAG SPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAG SPTSTEEGTSESATPESGPGTSTEPSEGSAPGAAEPEA

En algunas modalidades, para construir la secuencia de un polipéptido XTEN con código de barras, se crean mutaciones de aminoácidos en polipéptidos XTEN de longitudes intermedias a las de la Tabla 3b, así como también polipéptidos XTEN de longitudes más largas que las de la Tabla 3b, tales como aquellos en los que uno o más motivos de 12 mer de la Tabla 1 se añaden al extremo N- o C-terminal de un XTEN de propósito general de la Tabla 3b.

Ejemplos adicionales de secuencias de aminoácidos de polipéptido XTEN de propósito general que pueden usarse de acuerdo con la presente descripción se describen en las publicaciones de patentes de Estados Unidos núms. 2010/0239554 A1, 2010/0323956 A1, 2011/0046060 A1, 2011/0046061 A1, 2011/0077199 A1, o 2011/0172146 A1 o

publicaciones de patente internacional núms. WO 2010091122 A1, WO 2010144502 A2, WO 2010144508 A1, WO 2011028228 A1, WO 2011028229 A1, WO 2011028344 A2, WO 2014/011819 A2 o WO 2015/023891, cuyas descripciones se incorporan expresamente como referencia en la presente descripción.

En algunas modalidades, un polipéptido XTEN con código de barras fusionado dentro de una cadena polipeptídica adyacente al extremo N-terminal de la cadena polipeptídica ("XTEN N-terminal") puede unirse a una etiqueta de His que comprende una pluralidad de residuos de poli(His), que incluyen de seis a ocho residuos de His en el extremo N-terminal para facilitar la purificación del polipéptido de fusión. En algunas modalidades, un polipéptido XTEN con código de barras fusionado dentro de una cadena polipeptídica en el extremo C-terminal de la cadena polipeptídica ("polipéptido XTEN C-terminal") puede comprender o unirse a la secuencia EPEA en el extremo C-terminal para facilitar la purificación del polipéptido de fusión. En alguna modalidad, el polipéptido de fusión comprende un polipéptido XTEN con código de barras N-terminal y un polipéptido XTEN con código de barras C-terminal, en donde el XTEN con código de barras N-terminal está unido a una etiqueta de His que comprende una pluralidad de residuos de poli(His), que incluyen de seis a ocho residuos de His en el extremo N-terminal; y en donde el polipéptido XTEN con código de barras C-terminal está unido a la secuencia EPEA en el extremo C-terminal, para facilitar de esta manera la purificación del polipéptido de fusión, por ejemplo, al menos en un 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o al menos 99 % de pureza mediante métodos cromatográficos conocidos en la técnica, que incluyen, pero sin limitarse a, cromatografía IMAC, matriz de afinidad C-tagXL y otros métodos similares, que incluyen, pero sin limitarse a, los descritos en la sección de EJEMPLOS a continuación.

Digestión con proteasa

Un fragmento de código de barras, como se describió anteriormente o en cualquier otra parte en la presente descripción, puede fusionarse de manera escindible dentro del polipéptido XTEN y liberarse (configurarse para liberarlo) del polipéptido XTEN tras la digestión del polipéptido por una proteasa. En algunas modalidades, la proteasa es una proteasa Glu-C. En algunas modalidades, la proteasa se escinde en el lado C-terminal de los residuos de ácido glutámico que no van seguidos por la prolina. Un experto en la técnica entenderá que un polipéptido XTEN con código de barras (un polipéptido XTEN que contiene fragmento(s) de código de barras en su interior) está diseñado para lograr una alta eficiencia, precisión y exactitud de la digestión con proteasa. Por ejemplo, un experto en la técnica entenderá que los residuos Glu-Glu (EE) adyacentes en una secuencia de XTEN pueden dar como resultado patrones de escisión variables tras la digestión con Glu-C. En consecuencia, cuando se usa la proteasa Glu-C para la liberación del código de barras, el polipéptido XTEN con código de barras o el(los) fragmento(s) de código de barras pueden estar sin ninguna secuencia Glu-Glu (EE). Un experto en la técnica también entenderá que una secuencia de dipéptido Glu-Pro (EP), si está presente en el polipéptido de fusión, puede

ser incapaz de escindirse por la proteasa Glu-C durante el proceso de liberación del código de barras.

Configuración estructural de BPXTEN

En algunas modalidades, una proteína de fusión BPXTEN comprende un solo polipéptido BP y un solo polipéptido XTEN. Dichas proteínas BPXTEN pueden tener al menos las siguientes permutaciones de configuraciones, cada una enumerada en una orientación N- a C-terminal: BP-XTEN; XTEN-BP; BP-S-XTEN; y XTEN-S-BP, en donde “S” es una secuencia espaciadora como se expone a continuación.

En algunas modalidades, la proteína BPXTEN comprende un polipéptido XTEN C-terminal y, opcionalmente, una secuencia espaciadora (S) entre el polipéptido XTEN y el polipéptido BP. Tal proteína BPXTEN puede representarse mediante la Fórmula I (representada de extremo N- a C-terminal):

$$(BP)-(S)_x-(XTEN) \quad (I),$$

en donde BP es una proteína biológicamente activa como se describe a continuación; S es una secuencia espaciadora que tiene entre 1 y aproximadamente 50 residuos de aminoácidos que puede incluir opcionalmente un segmento de liberación de BP (como se describe con más detalle a continuación); x es 0 o 1; y XTEN puede ser cualquier polipéptido XTEN descrito en la presente descripción.

En algunas modalidades, la proteína BPXTEN comprende un polipéptido XTEN N-terminal y, opcionalmente, una secuencia espaciadora (S) entre el polipéptido XTEN y la proteína BP. Tales proteínas BPXTEN pueden representarse mediante la Fórmula II (representada de extremo N- a C-terminal):

$$(XTEN)-(S)_x-(BP) \quad (II),$$

en donde BP es una proteína biológicamente activa como se describe a continuación; S es una secuencia espaciadora que tiene entre 1 y aproximadamente 50 residuos de aminoácidos que puede incluir opcionalmente un segmento de liberación de BP (como se describe con más detalle a continuación); x es 0 o 1; y XTEN puede ser cualquier polipéptido XTEN como se describe en la presente descripción.

En alguna modalidad, la proteína BPXTEN comprende un polipéptido XTEN N-terminal y un polipéptido XTEN C-terminal. Tales proteínas BPXTEN (por ejemplo, los XPAT en las Figuras 1-2) pueden representarse mediante la Fórmula III:

$$(XTEN)-(S)_y-(BP)-(S)_z-(XTEN) \quad (III)$$

en donde BP es una proteína biológicamente activa como se describe a continuación; S es una secuencia espaciadora que tiene entre 1 y aproximadamente 50 residuos de aminoácidos que puede incluir opcionalmente un segmento de liberación de BP (como se describe con más detalle a continuación); y es 0 o 1; z es 0 o 1; y XTEN puede ser cualquier polipéptido XTEN como se describe en la presente descripción.

Polipéptido biológicamente activo

Una proteína biológicamente activa (BP) que puede fusionarse con uno o más polipéptidos XTEN (como se describe en la presente descripción), particularmente los que se describen a

continuación, que comprende secuencias identificadas en la presente descripción en las Tablas 4a-4h y Tablas 6a-6f, junto con sus secuencias de aminoácidos y ácidos nucleicos correspondientes, se conocen bien en la técnica. Las descripciones y secuencias de estas BP están disponibles en bases de datos públicas tales como Chemical Abstracts Services Databases (por ejemplo, el Registro CAS), GenBank, The Universal Protein Resource (UniProt) y bases de datos proporcionadas por suscripción tales como GenSeq (por ejemplo, Derwent). Las secuencias polinucleotídicas que codifican las BP pueden ser secuencias polinucleotídicas de tipo salvaje que codifican una BP nativa (por ejemplo, ya sea de longitud completa o madura), o en algunos casos la secuencia puede ser una variante de una secuencia polinucleotídica de tipo salvaje (por ejemplo, un polinucleótido que codifica la proteína biológicamente activa de tipo salvaje), en donde la secuencia de nucleótidos del polinucleótido se ha optimizado, por ejemplo, para la expresión en una especie particular; o un polinucleótido que codifica una variante de la proteína de tipo salvaje, tal como un mutante dirigido al sitio o una variante alélica. Está dentro de la capacidad del experto en la técnica usar una secuencia de ADNc de consenso o de tipo salvaje o una variante con codones optimizados de una BP para crear construcciones de BPXTEN contempladas por la invención mediante el uso de métodos conocidos en la técnica y/o junto con la guía y métodos proporcionados en la presente descripción.

La BP para su inclusión en las proteínas BPXTEN descritas en la presente descripción (por ejemplo, un polipéptido de fusión que comprende al menos una BP y al menos un polipéptido XTEN) puede incluir cualquier proteína de interés o función biológica, terapéutica, profiláctica o de diagnóstico, o que sea útil para mediar una actividad biológica o prevenir o mejorar una enfermedad, trastorno o afecciones cuando se administra a un ser humano o animal. Particularmente ventajosas son las BP para las que se busca un aumento en un parámetro farmacocinético, una mayor solubilidad, una mayor estabilidad, un enmascaramiento de la actividad o alguna otra propiedad farmacéutica mejorada, o aquellas BP para las que el aumento de la semivida terminal mejoraría la eficacia, la seguridad o daría como resultado la reducción de la frecuencia de dosificación y/o la mejora del cumplimiento del paciente. Por lo tanto, las composiciones de proteína de fusión BPXTEN pueden prepararse con varios objetivos en mente, que incluyen la mejora de la eficacia terapéutica del compuesto bioactivo, por ejemplo, mediante el aumento de la exposición *in vivo* o el tiempo que la BPXTEN permanece dentro de la ventana terapéutica cuando se administra a un ser humano o animal en comparación con una BP no unida a un polipéptido XTEN.

Una BP puede ser una proteína nativa de longitud completa o puede ser un fragmento o una variante de secuencia de una proteína biológicamente activa que retiene al menos una porción de la actividad biológica de la proteína nativa.

En una modalidad, la BP incorporada en las composiciones humanas o animales puede ser un polipéptido recombinante con una secuencia correspondiente a una proteína que se encuentra en la naturaleza. En otra modalidad, la BP puede ser variantes de secuencia,

fragmentos, homólogos y miméticos de una secuencia natural que retienen al menos una porción de la actividad biológica de la BP nativa. En ejemplos no limitantes, una BP puede ser una secuencia que exhibe al menos aproximadamente 80 % de identidad de secuencia o, alternativamente, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una secuencia proteica seleccionada de las Tablas 4a-4h. En otros ejemplos no limitantes, una BP puede ser una secuencia biespecífica que comprende un primer dominio de unión y un segundo dominio de unión, en donde el primer dominio de unión, que tiene afinidad de unión específica por un marcador específico de tumor o un antígeno de una célula objetivo, exhibe al menos aproximadamente 80 % de identidad de secuencia, o alternativamente 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con las secuencias VL y VH pareadas de un anticuerpo anti-CD3 identificado en la Tabla 6f; y en donde el segundo dominio de unión, que tiene una afinidad de unión específica por una célula efectora, exhibe al menos aproximadamente 80 % de identidad de secuencia, o alternativamente 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con secuencias VL y VH pareadas de un anticuerpo anti-células objetivo identificado en la Tabla 6a. En una modalidad, una proteína de fusión BPXTEN puede comprender una sola proteína BP unida a un polipéptido XTEN. En otra modalidad, la proteína BPXTEN puede comprender una primera BP y una segunda molécula de la misma BP, lo que da como resultado una proteína de fusión que comprende las dos BP unidas a uno o más polipéptidos XTEN (por ejemplo, dos moléculas de glucagón o dos moléculas de hGH).

En general, la BP exhibe una especificidad de unión a un objetivo determinado (o a un número determinado de objetivos) u otra característica biológica deseada cuando se usa *in vivo* o cuando se utiliza en un ensayo *in vitro*. Por ejemplo, la BP puede ser un agonista, un receptor, un ligando, un antagonista, una enzima, un anticuerpo (por ejemplo, mono- o biespecífico) o una hormona. De particular interés son las BP que se usan o se conoce que son útiles para una enfermedad o trastorno en donde la BP nativa tiene una semivida terminal relativamente corta y para la cual una mejora de un parámetro farmacocinético (que opcionalmente podría liberarse de la proteína de fusión mediante la escisión de una secuencia espaciadora) permitiría una dosificación menos frecuente o un efecto farmacológico mejorado. También son de interés las BP que tienen una ventana terapéutica estrecha entre la dosis efectiva o concentración en sangre mínima (C_{mín}) y la dosis tolerada o concentración en sangre máxima (C_{máx}). En tales casos, la unión de la BP a una proteína de fusión que comprende una(s) secuencia(s) polipeptídica(s) seleccionada(s) de XTEN puede dar como resultado una mejora en estas propiedades, lo que las hace más útiles como agentes terapéuticos o preventivos en comparación con la BP no unida a uno o más polipéptidos XTEN.

Péptidos reguladores de glucosa

Las enfermedades o trastornos endocrinos y relacionados con la obesidad han alcanzado proporciones epidémicas en la mayoría de los países desarrollados y representan una carga importante y creciente para la atención de la salud en la mayoría de los países desarrollados, que incluyen una gran variedad de afecciones que afectan los órganos, los tejidos y el sistema circulatorio del cuerpo. De particular preocupación son las enfermedades y los trastornos endocrinos y relacionados con la obesidad, entre los que destaca la diabetes, una de las principales causas de muerte en los Estados Unidos.

La mayoría de los procesos metabólicos en la homeostasis de la glucosa y la respuesta de la insulina están regulados por múltiples péptidos y hormonas, y muchos de estos péptidos y hormonas, así como también sus análogos, han encontrado utilidad en el tratamiento de enfermedades y trastornos metabólicos. Muchos de estos péptidos tienden a ser altamente homólogos entre sí, incluso cuando poseen funciones biológicas opuestas. Los péptidos que aumentan la glucosa se ejemplifican con la hormona peptídica glucagón, mientras que los péptidos que reducen la glucosa incluyen la exendina 4, el péptido 1 similar al glucagón y la amilina. Sin embargo, el uso de péptidos y/u hormonas terapéuticas, incluso cuando se complementa con el uso de fármacos de molécula pequeña, ha tenido un éxito limitado en el tratamiento de dichas enfermedades y trastornos. En particular, la optimización de la dosis es importante para los fármacos y productos biológicos usados en el tratamiento de enfermedades metabólicas, especialmente aquellas con una ventana terapéutica estrecha. Las hormonas en general y los péptidos involucrados en la homeostasis de la glucosa a menudo tienen una ventana terapéutica estrecha. La ventana terapéutica estrecha, junto con el hecho de que tales hormonas y péptidos típicamente tienen una semivida corta, lo que requiere una dosificación frecuente para lograr un beneficio clínico, da como resultado dificultades en el tratamiento de tales pacientes. Si bien las modificaciones químicas de una proteína terapéutica, tales como la pegilación, pueden modificar su tasa de eliminación in vivo y la subsecuente semivida en suero, requiere etapas de fabricación adicionales y da como resultado un producto final heterogéneo. Además, se han informado efectos secundarios inaceptables por la administración crónica. Alternativamente, la modificación genética mediante la fusión de un dominio Fc con la proteína o el péptido terapéutico aumenta el tamaño de la proteína terapéutica, lo que reduce la tasa de eliminación a través del riñón y promueve el reciclaje de los lisosomas por el receptor FcRn. Desafortunadamente, el dominio Fc no se pliega con eficiencia durante la expresión recombinante y tiende a formar precipitados insolubles conocidos como cuerpos de inclusión. Estos cuerpos de inclusión deben solubilizarse y la proteína funcional debe renaturalizarse; un proceso lento, ineficiente y costoso.

Por lo tanto, un aspecto de la presente invención es la incorporación de péptidos involucrados en la homeostasis de la glucosa, la resistencia a la insulina y la obesidad (colectivamente, "péptidos reguladores de la glucosa") en proteínas de fusión BPXTEN para crear composiciones con utilidad en el tratamiento de trastornos, enfermedades y afecciones relacionadas con la glucosa, la insulina y la obesidad. Los péptidos reguladores

de la glucosa adecuados que pueden unirse a los polipéptidos XTEN descritos en la presente descripción para crear proteínas BPXTEN que incluyen todos los polipéptidos biológicamente activos, entre otros, que aumentan la secreción de insulina dependiente de glucosa por las células beta pancreáticas o potencian la acción de la insulina. Los péptidos reguladores de la glucosa también pueden incluir polipéptidos biológicamente activos que estimulan la transcripción del gen de la proinsulina en las células beta pancreáticas. Además, los péptidos reguladores de la glucosa también pueden incluir polipéptidos biológicamente activos que ralentizan el tiempo de vaciado gástrico y reducen la ingestión de alimentos. Los péptidos reguladores de la glucosa también pueden incluir polipéptidos biológicamente activos que inhiben la liberación de glucagón de las células alfa de los islotes de Langerhans. La Tabla 4a proporciona una lista no limitante de secuencias de péptidos reguladores de la glucosa que pueden incluirse en las proteínas de fusión BPXTEN de la invención. Los péptidos reguladores de la glucosa de las composiciones de BPXTEN de la invención descritas en la presente descripción pueden ser un péptido que exhibe al menos aproximadamente 80 % de identidad de secuencia (por ejemplo, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia) con una secuencia de aminoácidos seleccionada de la Tabla 4a.

Tabla 4a: Péptidos reguladores de glucosa

Nombre de la proteína (Sinónimo)	SEQ ID NO	Secuencia de aminoácidos
Adrenomedulina (ADM)	1	YRQSMNNFQGLRSFGCRFGTCTVQKLAHQI YQFTDKDKDNVAPRSKISPQGY
Amilina, rata	2	KCNTATCATQRLANFLVRSSNNLGPVLPPT NVGSNTY
Amilina humana	3	KCNTATCATQRLANFLVHSSNNFGAILSSTN VGSNTY
Calcitonina (hCT)	4	CGNLSTCMLGTYTQDFNKFHTFPQTAIGVG AP
Calcitonina, salmón	5	CSNLSTCVLGKLSQELHKLQTYPRNTGSGT P
Péptido relacionado con el gen de la calcitonina (h-CGRP α)	6	ACDTATCVTHRLAGLLSRSGGVVKNMVPT NVGSKAF
Péptido relacionado con el gen de la calcitonina (h-CGRP β)	7	ACNTATCVTHRLAGLLSRSGGMVKSNFVPT NVGSKAF
colecistocinina (CCK)	8	MNSGVCLCVLMAVLAAGALTQPVPADPA GSQLQRAEEAPRRQLRVSQRTDGESRAHLG ALLARYIQQARKAPSGRMSIVKNLQNLDP SHRISDRDYMGMDFGRRSAEEYEYPS
CCK-33	9	KAPSGRMSIVKNLQNLDP SHRISDRDYMGMDF
CCK-8	10	DYMGWMDF
Exendina 3	11	HSDGTFTSDLSKQMEEEAVRLFIEWLKNGG PSSGAPPS
Exendina 4	12	HGEGTFTSDLSKQMEEEAVR LFIEWLKNGGPSSGAPPS
FGF-19	13	MRSGCVVVHVWILAGLWLAVAGRPLAFSD

Nombre de la proteína (Sinónimo)	SEQ ID NO	Secuencia de aminoácidos
		AGPHVHYGWGDPIRLRHL YTSGPHGLSSCF LRIRADGVVDCARGQSAHSLLEIKAVALRT VAIKGVHSVRYLCMGADGKMQLLQYSEE DCAFEIEIRPDGYNVYRSEKHRLPVSLSSAK QRQLYKNRGFLPLSHFLPMLPMVPEEPEDL RGHLESDMFSSPLETDSMDPFGLVTGLEAV RSPSFEK
FGF-21	14	MDSDETGFEHSGLWVSVLAGLLLGACQAH PIPDSSPLLQFGGQVRQRYLYTDDAQQTEAH LEIREDGTVGGAADQSPESLLQLKALKPGVI QILGVKTSRFLCQRPDGALYGSLHFDPEACS FRELLLEDGYNVYQSEAHGLPLHLPGNKSP HRDPAPRGPARFLPLPGLPPALPEPPGILAPQ PPDVGSSDPLSMVGPSQGRSPSYAS
Gastrina	15	QLGPQGPPHLVADPSKKQGPWLEEEEEAYG WMDF
Gastrina-17	16	DPSKKQGPWLEEEEEAYGWMDF
Polipéptido inhibidor gástrico (GIP)	17	YAEFTFISDYSIAMDKIHQQDFVNWLLAQK GKKNDWKHNITQ
Grelina	18	GSSFLSPEHQRVQQRKESKKPPAKLQPR
Glucagón	19	HSQGTFTSDYSKYLDSRRAQDFVQWLMNT
Péptido 1 similar al glucagón (hGLP-1) (GLP-1; 1-37)	20	HDEFERHAEGTFTSDVSSTLEGQAALFIW LVKGRG
GLP-1 (7-36), humano	21	HAEGTFTSDVSSYLEGQAALFIWLVKGR
GLP-1 (7-37), humano	22	HAEGTFTSDVSSYLEGQAALFIWLVKGR G
GLP-1, rana	23	HAEGTYTNDVTEYLEEKAAKEFIEWLIKGG PKKIRYS
Péptido 2 similar al glucagón (GLP-2), humano	24	HADGSFSDEMNTILDNLAAARDFINWLIETKI TD
GLP-2, rana	25	HAEGTFTNDMTNYLEEKAAKEFVGWLIK G
IGF-1	26	GPETLCGAELVDALQFVCGDRGFYFNKPTG YGSSSRRAPQTGIVDECCFRSCDLRRLEMYC APLKPAKSA
IGF-2	27	AYRPSETLCGGELVDTLQFVCGDRGFYFSRP ASRVSRRSRGIVECCFRSCDLALLETYCAT PAKSE
Péptido INGAP (proteína asociada a la neogénesis de los islotes)	28	EESQKKLPSSRITCPQGSVAYGSYCYSLILIP QTWSNAELSCQMHFSGHLAFLSTGEITFVS SLVKNSLTAYQYIWIGLHDP SHGTL PNGSG WKWSSSNVLT FYNWERNPSIAADRGYCAV LSQKSGFQKWRDFNCENELPYICKFKV
Intermedina (AFP-6)	29	TQAQLLRVGCVLGTCQVQNLSHRLWQLMG PAGRQDSAPVDPSSPHSY
Leptina, humana	30	VPIQKVQDDTKTLIKTIVTRINDISHTQSVSS KQKVTGLDFIPGLHPILTL SKMDQTLAVYQQ ILTSMPSRNVIQISNDLENLRDLLHVLAFSKS CHLPWASGLETDSLGGVLEASGYSTEVA LSRLQGS LQDMLWQLDLSPGC
Neuromedina (U-8) porcina	31	YFLFRPRN
Neuromedina (U-9)	32	GYFLFRPRN
neuromedina (U25)	33	FRVDEEFQSPFASQSRGYFLFRPRN

Nombre de la proteína (Sinónimo)	SEQ ID NO	Secuencia de aminoácidos
humana)		
Neuromedina (U25) cerdo	34	FKVDEEFQGPIVSQNRRYFLFRPRN
Neuromedina S, humana	35	ILQRGSGTAAVDFTKKDHTATWGRPFFLFRP RN
Neuromedina U, rata	36	YKVNEYQGPVAPSGGFFLFRPRN
Oxintomodulina (OXM)	37	HSQGTFTSDYSKYLDSRRAQDFVQWLMNT KRNRRNNIA
Péptido YY (PYY)	38	YPIKPEAPGEDASPEELNRYYASLRHYLNLV TRQRY
Pramlintida	39	KCNTATCATNRLANFLVHSSNNFGPILPPTN VGSNTY-NH2
Urocortina (Ucn-1)	40	DNPSLSIDLTFHLLRTLLELARTQSQRERAEQ NRIIFDSV
Urocortina (Ucn-2)	41	IVLSLDVPIGLLQILLEQARARAAREQATTN ARILARVGHC
Urocortina (Ucn-3)	42	FTLSLDVPTNIMNLLFNIAKAKNLRAQAAA NAHLMAQI

“Adrenomedulina” o “ADM” significa la hormona peptídica de adrenomedulina humana y las variantes de especie y secuencia de esta que tienen al menos una porción de la actividad biológica de la ADM madura. La ADM se genera a partir de una preprohormona de 185 aminoácidos a través de la escisión enzimática y la amidación consecutivas, lo que da como resultado un péptido bioactivo de 52 aminoácidos con una semivida plasmática medida de 22 min. Las proteínas de fusión que contienen ADM de la invención pueden encontrar un uso particular en la diabetes para efectos estimulantes sobre la secreción de insulina de las células de los islotes para la regulación de la glucosa o en seres humanos o animales con hipotensión sostenida. Se ha informado de la infraestructura genómica completa de la AM humana (Ishimitsu y otros, 1994, Biochem. Biophys. Res. Commun 203:631-639), y se han clonado análogos de péptidos ADM, como se describe en la patente de Estados Unidos núm. 6,320,022.

“Amilina” significa la hormona peptídica humana denominada amilina, pramlintida y variaciones de especies de estas, como se describe en la patente de Estados Unidos núm. 5,234,906, que tiene al menos una porción de la actividad biológica de la amilina madura. La amilina es una hormona polipeptídica de 37 aminoácidos secretada junto con la insulina por las células beta pancreáticas en respuesta a la ingestión de nutrientes (Koda y otros, 1992, Lancet 339:1179-1180), y se ha informado que modula varias rutas clave del metabolismo de carbohidratos, que incluyen la incorporación de glucosa en glucógeno. Las proteínas de fusión que contienen amilina de la invención pueden complementar la acción de la insulina, que regula la tasa de desaparición de la glucosa de la circulación y su captación por los tejidos periféricos. Se han clonado análogos de amilina, como se describe en la patente de Estados Unidos núm. 5,686,411 y 7,271,238.

Se pueden crear miméticos de amilina que retienen la actividad biológica. Por ejemplo, la pramlintida tiene la secuencia KCNTATCATNRLANFLVHSSNNFGPILPPTNVGSNTY

(SEQ ID NO: 43), en donde los aminoácidos de la secuencia de amilina de rata se sustituyen por aminoácidos en la secuencia de amilina humana. En una modalidad, la invención contempla proteínas de fusión que comprenden miméticos de amilina de la secuencia KCNTATCATX₁RLANFLVHSSNNFGX₂ILX₂X₂TNVGSNTY (SEQ ID NO: 44) en donde X₁ es independientemente N o Q y X₂ es independientemente S, P o G. En una modalidad, el mimético de amilina incorporado en una BPXTEN puede tener la secuencia KCNTATCATNRLANFLVHSSNNFGGILGGTNVGSNTY (SEQ ID NO: 45). En otra modalidad, en donde el mimético de amilina se usa en el extremo C-terminal de BPXTEN, el mimético puede tener la secuencia KCNTATCATNRLANFLVHSSNNFGGILGGTNVGSNTY(NH₂) (SEQ ID NO: 46).

“Calcitonina” (CT) significa la proteína calcitonina humana y variantes de especie y secuencia de esta, que incluyen la calcitonina de salmón (“sCT”), que tiene al menos una porción de la actividad biológica de la CT madura. La CT es un péptido de 32 aminoácidos escindido de una prohormona más grande de la tiroides que parece funcionar en los sistemas nervioso y vascular, pero también se ha informado que es un potente mediador hormonal del reflejo de saciedad. (Revisado en Becker, JCEM, 89(4): 1512-1525 (2004) y Sexton, Current Medicinal Chemistry 6: 1067-1093 (1999)). Las proteínas de fusión que contienen calcitonina de la invención pueden encontrar un uso particular para el tratamiento de la osteoporosis y como terapia para la enfermedad ósea de Paget. Se han creado péptidos de calcitonina sintéticos, como se describe en las patentes de Estados Unidos núm. 5,175,146 y 5,364,840.

“Péptido relacionado con el gen de la calcitonina” o “CGRP” significa el péptido CGRP humano y las variantes de especie y secuencia de este que tienen al menos una porción de la actividad biológica del CGRP maduro, que es un miembro de la familia de péptidos de la calcitonina, que en seres humanos existe en dos formas, α -CGRP (un péptido de 37 aminoácidos) y β -CGRP. CGRP tiene una identidad de secuencia del 43-46 % con la amilina humana. Las proteínas de fusión que contienen CGRP de la invención pueden encontrar un uso particular para disminuir la morbilidad asociada con la diabetes, mejorar la hiperglucemia y la deficiencia de insulina, inhibir la infiltración de linfocitos en los islotes y proteger las células beta contra la destrucción autoinmunitaria. Los métodos para producir CGRP sintético y recombinante se describen en las patentes de Estados Unidos núm. 5,374,618.

“Colecistocinina” o “CCK” significa el péptido CCK humano y las variantes de especie y secuencia de este que tienen al menos una porción de la actividad biológica de la CCK madura. CCK-58 es la secuencia madura, mientras que la secuencia de aminoácidos de CCK-33 identificada por primera vez en seres humanos es la principal forma circulante del péptido. La familia CCK también incluye un fragmento C-terminal *in vivo* de 8-aminoácidos (“CCK-8”), donde la pentagastrina o CCK-5 es el péptido C-terminal CCK(29-33), y CCK-4 es el tetrapéptido C-terminal CCK(30-33). La CCK es una hormona peptídica del sistema gastrointestinal responsable de estimular la digestión de grasas y

proteínas. Las proteínas de fusión de la invención que contienen CCK-33 y CCK-8 pueden encontrar un uso particular para reducir el aumento de la glucosa circulante después de la ingestión de alimentos y potenciar el aumento de la insulina en circulación. Se han preparado análogos de CCK-8, como se describe en la patente de Estados Unidos núm. 5,631,230.

“Exendina 3” significa un péptido regulador de la glucosa aislado de *Heloderma horridum* y variantes de secuencia de este que tienen al menos una porción de la actividad biológica de la exendina 3 madura. La exendina 3 amida es un antagonista específico del receptor de la exendina que media un aumento del AMPc pancreático y la liberación de insulina y amilasa. Las proteínas de fusión que contienen exendina 3 de la invención pueden encontrar un uso particular en el tratamiento de la diabetes y los trastornos de resistencia a la insulina. La secuencia y los métodos para su ensayo se describen en la patente de Estados Unidos 5,424,86.

Exendina 4” significa un péptido regulador de la glucosa que se encuentra en la saliva del monstruo de Gila *Heloderma suspectum*, así como también variantes de especie y secuencia de este, e incluye la secuencia nativa de 39 aminoácidos HGEFTFTSDLSKQMEEEAVRLFIEYLNKNGGPSSGAPPPS (SEQ ID NO: 47) y secuencias homólogas y peptidomiméticos, y variantes de estos; secuencias naturales, tales como de primates y no naturales que tienen al menos una porción de la actividad biológica de la exendina 4 madura. La exendina 4 es una hormona polipeptídica incretina que disminuye la glucosa en sangre, promueve la secreción de insulina, ralentiza el vaciamiento gástrico y mejora la saciedad, lo que proporciona una marcada mejoría en la hiperglucemia posprandial. La Tabla 4b muestra las secuencias de una amplia variedad de especies, mientras que Tabla 4c muestra una lista de análogos de GLP-1 sintéticos; todos los cuales se contemplan para el uso en las proteínas BPXTEN descritas en la presente descripción.

Factor de crecimiento de fibroblastos 21, o "FGF-21" significa la proteína humana codificada por el gen FGF-21, o variantes de especie y secuencia de este que tienen al menos una porción de la actividad biológica del FGF-21 maduro. FGF-21 estimula la captación de glucosa en los adipocitos pero no en otros tipos de células; el efecto es aditivo a la actividad de la insulina. Las proteínas de fusión que contienen FGF-21 de la invención pueden encontrar un uso particular en el tratamiento de la diabetes, lo que incluye provocar el aumento del gasto de energía, la utilización de grasas y la excreción de lípidos. Se ha clonado FGF-21, como se describe en la patente de Estados Unidos núm. 6,716,626.

"Factor de crecimiento de fibroblastos 19" o "FGF-19" significa la proteína humana codificada por el gen de FGF-19, o variantes de especie y secuencia de este que tienen al menos una porción de la actividad biológica del FGF-19 maduro. FGF-19 es una proteína miembro de la familia del factor de crecimiento de fibroblastos (FGF). FGF-19 aumenta la expresión hepática del receptor de leptina, la tasa metabólica, estimula la captación de glucosa en los adipocitos y conduce a la pérdida de peso en un modelo de ratón obeso (Fu y otros, 2004, *Endocrinology* 145: 2504-2603) las proteínas de fusión que contienen FGF-19

de la invención pueden encontrar un uso particular para aumentar la tasa metabólica y revertir la diabetes dietética y deficiente en leptina. FGF-19 se ha clonado y expresado, como se describe en la solicitud de patente de Estados Unidos núm. 20020042367.

“Gastrina” significa el péptido de gastrina humana, versiones truncadas y variantes de especie y secuencia de este que tienen al menos una porción de la actividad biológica de la gastrina madura. La gastrina se encuentra principalmente en tres formas: gastrina 34 ("gastrina grande"); gastrina 17 ("gastrina pequeña"); y gastrina 14 ("minigastrina") y comparte homología de secuencia con CCK. Las proteínas de fusión que contienen gastrina de la invención pueden encontrar un uso particular en el tratamiento de la obesidad y la diabetes para la regulación de la glucosa. Se ha sintetizado gastrina, como se describe en la patente de Estados Unidos núm. 5,843,446.

"Grelina" significa una hormona humana que induce la saciedad, o variantes de especie y secuencia de esta, que incluyen la secuencia nativa procesada de 27 o 28 aminoácidos y secuencias homólogas. Los niveles de grelina aumentan antes de las comidas y disminuyen después de las comidas, y pueden dar como resultado un aumento de la ingestión de alimentos y un aumento de la masa grasa por una acción ejercida a nivel del hipotálamo. Las proteínas de fusión que contienen grelina de la invención pueden encontrar un uso particular como agonistas; por ejemplo, para estimular selectivamente la motilidad del tracto GI en el trastorno de la motilidad gastrointestinal, para acelerar el vaciado gástrico o para estimular la liberación de la hormona del crecimiento. Los análogos de grelina con sustituciones de secuencia o variantes truncadas, como se describe en la patente de Estados Unidos núm. 7,385,026, puede encontrar un uso particular como parejas de fusión con polipéptidos XTEN para el uso como antagonistas para mejorar la homeostasis de la glucosa, el tratamiento de la resistencia a la insulina y el tratamiento de la obesidad. Se ha informado sobre el aislamiento y la caracterización de la grelina (Kojima y otros, 1999, Nature. 402:656-660) y se han preparado análogos sintéticos mediante síntesis de péptidos, como se describe en la patente de Estados Unidos núm. 6,967,237.

“Glucagón” significa el péptido regulador de la glucosa del glucagón humano, o variantes de especie y secuencia de este, que incluyen la secuencia nativa de 29 aminoácidos y secuencias homólogas; naturales, tales como de primates, y variantes de secuencias no naturales que tienen al menos una porción de la actividad biológica del glucagón maduro. El término "glucagón", como se usa en la presente descripción, también incluye peptidomiméticos del glucagón. Las proteínas de fusión que contienen glucagón de la invención pueden encontrar un uso particular para aumentar los niveles de glucosa en sangre en individuos con reservas de glucógeno hepático existentes y mantener la homeostasis de la glucosa en la diabetes. Se ha clonado el glucagón, como se describe en la patente de Estados Unidos núm. 4,826,763.

“GLP-1” significa péptido 1 similar al glucagón humano y variantes de secuencia de este que tienen al menos una porción de la actividad biológica del GLP-1 maduro. El término "GLP-1" incluye GLP-1(1-37), GLP-1(7-37) y GLP-1(7-36)amida humano. El GLP-1

estimula la secreción de insulina, pero solo durante los períodos de hiperglucemia. La seguridad del GLP-1 en comparación con la insulina mejora con esta propiedad y por la observación de que la cantidad de insulina secretada es proporcional a la magnitud de la hiperglucemia. La semivida biológica de GLP-1(7-37)OH es de tan solo 3 a 5 minutos (patente de Estados Unidos núm. 5,118,666). Las proteínas de fusión que contienen GLP-1 de la invención pueden encontrar un uso particular en el tratamiento de la diabetes y los trastornos de resistencia a la insulina para la regulación de la glucosa. Se ha clonado GLP-1 y se han preparado derivados, como se describe en la patente de Estados Unidos núm. 5,118,666. La Tabla 4b muestra ejemplos no limitados de secuencias de GLP-1 de una amplia variedad de especies, mientras que la Tabla 4c muestra las secuencias de varios análogos de GLP-1 sintéticos; todos los cuales se contemplan para el uso en las composiciones de BPXTEN descritas en la presente descripción.

Tabla 4b: Homólogos de GLP-1 naturales representativos como candidatos a BP

Nombre del gen	SEQ ID NO	Secuencia de aminoácidos
GLP-1 {rana}	48	HAEGTYTNDVTEYLEEKAAKEFIEWLI KGKPKKIRYS
GLP-1a {Xenopus laevis}	49	HAEGTFTSDVTQQLDEKAAKEFIDWLI NGGPSKEIIS
GLP-1b {Xenopus laevis}	50	HAEGTYTNDVTEYLEEKAAKEFIIEWL IKGKPK
GLP-1c {Xenopus laevis}	51	HAEGTFTNDMTNYLEEKAAKEFVGWL IKGRPK
Polipéptido inhibidor gástrico {Mus musculus}	52	HAEGTFISDYSIAMDKIRQQDFVNWLL
Polipéptido insulinotrópico dependiente de glucosa {Equus caballus}	53	HAEGTFISDYSIAMDKIRQQDFVNWLL
Péptido similar al glucagón {Petromyzon marinus}	54	HADGTFTNDMTSYLDAKAARDFVSW LARSDKS
Péptido similar al glucagón {Anguila rostrata}	55	HAEGTYTSDVSSYLQDQAAKEFVSWL KTGR
Péptido similar al glucagón {Anguila anguila}	56	HAEGTYTSDVSSYLQDQAAKEFVSWL KTGR
Péptido similar al glucagón {Hydrolagus colliei}	57	HADGIYTSDVASLTDYLKSKRFVESLS NYNKRQNDRRM
Péptido similar al glucagón {Amia calva}	58	YADAPYISDVYSYLQDQVAKKWLKSG QDRRE
GLUC_ICTPU/38-65	59	HADGTYTSDVSSYLQEQAAKDFITWL KS
GLUCL_ANGRO/1-28	60	HAEGTYTSDVSSYLQDQAAKEFVSWL KT
GLUC_BOVIN/98-125	61	HAEGTFTSDVSSYLEGQAAKEFIAWL V K
GLUC1_LOPAM/91-118	62	HADGTFTSDVSSYLKDQAIKDFVDR LK A
GLUCL_HYDCO/1-28	63	HADGIYTSDVASLTDYLKSKRFVESLS

Nombre del gen	SEQ ID NO	Secuencia de aminoácidos
		N
GLUC_CAVPO/53-80	64	HSQGTFTSDYSKYLDSRRAQQFLKWL LN
GLUC_CHIBR/1-28	65	HSQGTFTSDYSKHLDSRYAQEFVQWL MN
GLUC1_LOPAM/53-80	66	HSEGTFSDYISKYLEDRKAQEFVRWL MN
GLUC_HYDCO/1-28	67	HTDGIFSSDYSKYLDNRRTKDFVQWLL S
GLUC_CALMI/1-28	68	HSEGTFSSDYSKYLDSRRAKDFVQWL MS
GIP_BOVIN/1-28	69	YAEGTFISDYSIAMDKIRQQDFVNWLL A
VIP_MELGA/89-116	70	HADGIFTTVYSHLLAKLAVKRYLHSLI R
PACA_CHICK/131-158	71	HIDGIFTDSYSRYRKQMAVKKYLA AV LG
VIP_CAVPO/45-72	72	HSDALFTDTYTRLRKQMAMKKYLSV LN
VIP_DIDMA/1-28	73	HSDAVFTDSYTRLLKQMAMRKYLDSI LN
EXE1_HELSU/1-28	74	HSDATFTAESKLLAKLALQKYLESIL G
SLIB_CAPHI/1-28	75	YADAIFTNSYRKVLGQLSARKLLQDIM N
SLIB_RAT/31-58	76	HADAIFTSSYRRILGQLYARKLLHEIM N
SLIB_MOUSE/31-58	77	HVDAIFTTNYRKLLSQLYARKVIQDIM N
PACA_HUMAN/83-110	78	VAHGILNEAYRKVLDQLSAGKHLQSL VA
PACA_SHEEP/83-110	79	VAHGILDKAYRKVLDQLSARRYLQTL MA
PACA_ONCNE/82-109	80	HADGMFNKAYRKALGQLSARKYLHS LMA
GLUC_BOVIN/146-173	81	HADGSFSDEMNTVLDSLATRDFINWLL Q
SECR_CANFA/1-27	82	HSDGTFTSELSRLRESARLQRLQLGLV
SECR_CHICK/1-27	83	HSDGLFTSEYSKMRGNAQVQKFIQNL M
EXE3_HELHO/48-75	84	HSDGTFTSDLSKQMEEEAVRLFIEWLK N

Tabla 4c: Análogos sintéticos representativos de GLP-1

SEQ ID NO	Secuencia de aminoácidos
85	HAEGTFTSDVSSYLEGQAAREFIAWLVKGRG
86	HAEGTFTSDVSSYLEGQAAKEFIAWLVRGRG
87	HAEGTFTSDVSSYLEGQAAKEFIAWLVKGKG
88	HAEGTFTSDVSSYLEGQAAREFIAWLVRGKG
89	HAEGTFTSDVSSYLEGQAAREFIAWLVRGKGR
90	HAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRK
91	HAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRRK
92	HAEGTFTSDVSSYLEGQAAREFIAWLVKGKG
93	HAEGTFTSDVSSYLEGQAAKEFIAWLVRGKG
94	HAEGTFTSDVSSYLEGQAAREFIAWLVKGRGRK
95	HAEGTFTSDVSSYLEGQAAKEFIAWLVRGRGRRK

SEQ ID NO	Secuencia de aminoácidos
96	HAEGTFTSDVSSYLEGQAAREFIAWLVRGKGRK
97	HAEGTFTSDVSSYLEGQAAREFIAWLVRGKGRRK
98	HGEGTFTSDVSSYLEGQAAREFIAWLVKGRG
99	HGEGTFTSDVSSYLEGQAAKEFIAWLVRGRG
100	HGEGTFTSDVSSYLEGQAAKEFIAWLVKGKG
101	HGEGTFTSDVSSYLEGQAAREFIAWLVRGKG
102	HGEGTFTSDVSSYLEGQAAREFIAWLVRGRGRK
103	HGEGTFTSDVSSYLEGQAAREFIAWLVRGRGRRK
104	HGEGTFTSDVSSYLEGQAAREFIAWLVKGKG
105	HGEGTFTSDVSSYLEGQAAKEFIAWLVRGKG
106	HGEGTFTSDVSSYLEGQAAREFIAWLVKGRGRK
107	HGEGTFTSDVSSYLEGQAAKEFIAWLVRGRGRRK
108	HGEGTFTSDVSSYLEGQAAREFIAWLVRGKGRK
109	HGEGTFTSDVSSYLEGQAAREFIAWLVRGKGRRK
110	HAEGTFTSDVSSYLEGQAAREFIAWLVRGRGK
111	HAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRK
112	HAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRRK
113	HAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREK
114	HAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFK
115	HAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPK
116	HAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEK
117	HAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEEK
118	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGK
119	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRK
120	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRRK
121	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREK
122	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFK
123	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPK
124	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEK
125	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEEK
126	DEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRK
127	DEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRRK
128	DEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREK
129	DEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFK
130	DEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPK
131	DEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEK
132	DEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEEK
133	EFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGK
134	EFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRK
135	EFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRRK
136	EFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREK
137	EFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFK
138	EFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPK
139	EFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEK
140	EFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEEK
141	FERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGK
142	FERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRK
143	FERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRRK
144	FERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREK
145	FERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFK
146	FERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPK
147	FERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEK
148	FERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEEK
149	ERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGK
150	ERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRK
151	ERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRRK

SEQ ID NO	Secuencia de aminoácidos
152	ERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREK
153	ERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFK
154	ERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPK
155	ERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEK
156	ERHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEEK
157	RHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGK
158	RHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRK
159	RHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRRK
160	RHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREK
161	RHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFK
162	RHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPK
163	RHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEK
164	RHAEGTFTSDVSSYLEGQAAREFIAWLVRGRGRREFPEEK
165	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVKGRGK
166	HDEFERHAEGTFTSDVSSYLEGQAAKEFIAWLVRGRGK
167	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGK GK
168	HAEGTFTSDVSSYLEGQAAREFIAWLVKGRGK
169	HAEGTFTSDVSSYLEGQAAKEFIAWLVRGRGK
170	HAEGTFTSDVSSYLEGQAAREFIAWLVRGK GK
171	HAEGTFTSDVSSYLEGQAAREFIAWLVRGRGK
172	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVKGRGRK
173	HDEFERHAEGTFTSDVSSYLEGQAAKEFIAWLVRGRGRK
174	HDEFERHAEGTFTSDVSSYLEGQAAREFIAWLVRGK GK
175	HAEGTFTSDVSSYLEGQAAREFIAWLVKGRGRK
176	HAEGTFTSDVSSYLEGQAAKEFIAWLVRGRGRK
177	HAEGTFTSDVSSYLEGQAAREFIAWLVRGK GK
178	HGEGTFTSDVSSYLEGQAAREFIAWLVKGRGK
179	HGEGTFTSDVSSYLEGQAAREFIAWLVRGK GK

Las secuencias nativas de GLP pueden describirse mediante varios motivos de secuencia, que se presentan a continuación. Las letras entre paréntesis representan aminoácidos aceptables en cada posición de la secuencia: {H V Y} {A G I S T V} {D E H Q} {A G} {I L M P S T V} {F L Y} {D I N S T} {A D E K N S T} {A D E N S T V} {L M V Y} {A N R S T Y} {E H I K N Q R S T} {A H I L M Q V Y} {L M R T} {A D E G K Q S} {A D E G K N Q S Y} {A E I K L M Q R} {A K Q R S V Y} {{A I L M Q S T V} {G K Q R} {D E K L Q R} {F H L V W Y} {I L V} {A D E G H I K N Q R S T} {A D E G N R S T W} {G I L V W} {A I K L M Q S V} {A D G I K N Q R S T} {G K R S Y} (SEQ ID NO: 9399). Además, los análogos sintéticos de GLP-1 pueden ser útiles como parejas de fusión de polipéptidos XTEN para crear proteína BPXTEN con actividad biológica útil en el tratamiento de trastornos relacionados con la glucosa.

“GLP-2” significa péptido 2 similar al glucagón humano y variantes de secuencia de este que tienen al menos una porción de la actividad biológica del GLP-2 maduro. Más particularmente, el GLP-2 es un péptido de 33 aminoácidos, cosecretado junto con GLP-1 por las células endocrinas intestinales en el intestino delgado y grueso.

"Factor de crecimiento similar a la insulina 1" o "IGF-1" significa la proteína IGF-1 humana y variantes de especie y secuencia de esta que tienen al menos una porción de la actividad biológica del IGF-1 maduro. El IGF-1 consiste en 70 aminoácidos y se produce principalmente en el hígado como hormona endocrina, así como también en los tejidos objetivo de forma paracrina/autocrina. Las proteínas de fusión que contienen IGF-1 de la

invención pueden encontrar un uso particular en el tratamiento de la diabetes y los trastornos de resistencia a la insulina para la regulación de la glucosa. El IGF-1 se ha clonado y expresado en *E. coli* y levadura, como se describe en la patente de Estados Unidos núm. 5,324,639.

"Factor de crecimiento 2 similar a la insulina" o "IGF-2" significa la proteína IGF-2 humana y variantes de especie y secuencia de esta que tienen al menos una porción de la actividad biológica del IGF-2 maduro. El IGF-2 se ha clonado, como se describe en Bell y otros, 1985, *Proc Natl Acad Sci U S A.* 82:6450-4.

"Proteína asociada a la neogénesis de los islotes" (INGAP), o "factor de crecimiento de células beta pancreáticas" significa el péptido INGAP humano y variantes de especie y secuencia de este que tienen al menos una porción de la actividad biológica de INGAP madura. Las proteínas de fusión que contienen INGAP de la invención pueden encontrar un uso particular en el tratamiento o prevención de la diabetes y trastornos de resistencia a la insulina. INGAP se ha clonado y expresado, como se describe en R Rafaeloff y otros, 1997, *J Clin Invest.* 99(9): 2100–2109.

"Intermedina" o "AFP-6" significa el péptido intermedina humano y las variantes de especie y secuencia de este que tienen al menos una porción de la actividad biológica de la intermedina madura. El tratamiento con intermedina conduce a la reducción de la presión arterial tanto en seres humanos o animales, normales como hipertensos, así como también a la supresión de la actividad de vaciado gástrico, y está implicado en la homeostasis de la glucosa. Las proteínas de fusión que contienen intermedina de la invención pueden encontrar un uso particular en el tratamiento de la diabetes, los trastornos de resistencia a la insulina y la obesidad. Se han clonado péptidos y variantes de intermedina, como se describe en la patente de Estados Unidos núm. 6,965,013.

"Leptina" significa la leptina de origen natural de cualquier especie, así como también las isoformas D biológicamente activas, o fragmentos y variantes de secuencia de las mismas. Las proteínas de fusión que contienen leptina de la invención pueden encontrar un uso particular en el tratamiento de la diabetes para la regulación de la glucosa, los trastornos de resistencia a la insulina y la obesidad. La leptina se ha clonado, como se describe en la patente de Estados Unidos núm. 7,112,659, y fragmentos y análogos de leptina en la patente de Estados Unidos núm. 5,521,283, la patente de Estados Unidos núm. 5,532,336, los documentos PCT/US96/22308 y PCT/US96/01471.

"Neuromedina" significa la familia de péptidos de neuromedina que incluye péptidos U y S de neuromedina, y variantes de secuencia de estos. En la familia de neuromedina U se incluyen varias variantes truncadas o empalmadas, por ejemplo, FLFHYSKTQKLKGKSNVVEELQSPFASQSRGYFLFRPRN (SEQ ID NO: 180). Un ejemplo de la familia de neuromedina S es neuromedina S humana con la secuencia ILQRGSGTAAVDFTKKDHTATWGRPFFLFRPRN (SEQ ID NO: 181), particularmente su forma amida. Las proteínas de fusión de neuromedina de la invención pueden encontrar un uso particular en el tratamiento de la obesidad, la diabetes, la reducción de la ingestión

de alimentos y otras afecciones y trastornos relacionados, como se describe en la presente descripción.

"Oxintomodulina" u "OXM" significa oxintomodulina humana y variantes de especie y secuencia de esta que tienen al menos una porción de la actividad biológica de la OXM madura. OXM es un péptido de 37 aminoácidos producido en el colon que contiene la secuencia de 29 aminoácidos del glucagón seguida de una extensión carboxiterminal de 8 aminoácidos. Las proteínas de fusión que contienen OXM de la invención pueden encontrar un uso particular en el tratamiento de la diabetes para la regulación de la glucosa, los trastornos de resistencia a la insulina, la obesidad y pueden usarse como un tratamiento para la pérdida de peso.

"PYY" significa el polipéptido YY del péptido humano y las variantes de especie y secuencia de este que tienen al menos una porción de la actividad biológica del PYY maduro. Las proteínas de fusión que contienen PYY de la invención pueden encontrar un uso particular en el tratamiento de la diabetes para la regulación de la glucosa, los trastornos de resistencia a la insulina y la obesidad. Se han preparado análogos de PYY, como se describe en la patente de Estados Unidos núm. 5,604,203, 5,574,010 y 7,166,575.

"Urocortina" significa una hormona peptídica de urocortina humana y variantes de secuencia de la misma que tienen al menos una porción de la actividad biológica de la urocortina madura. Hay tres urocortinas humanas: Ucn-1, Ucn-2 y Ucn-3. Se han descrito urocortinas y análogos adicionales en las patentes de Estados Unidos núm. 6,214,797. Las proteínas BPXTEN que comprenden la urocortina de la invención también pueden encontrar un uso particular en el tratamiento o prevención de afecciones asociadas con la estimulación de la liberación de ACTH, hipertensión debido a efectos vasodilatadores, inflamación mediada por otros medios que no sean la elevación de ACTH, hipertermia, trastorno del apetito, insuficiencia cardíaca congestiva, estrés, ansiedad y psoriasis. Las proteínas de fusión que contienen urocortina también pueden combinarse con un módulo de péptido natriurético, una familia de amilina y una familia de exendina, o un módulo de la familia GLP1 para proporcionar un beneficio cardiovascular mejorado, por ejemplo, en el tratamiento de CHF, para proporcionar un efecto de vasodilatación beneficioso.

Enfermedad metabólica y proteínas cardiovasculares

Las enfermedades metabólicas y cardiovasculares representan una carga sustancial para la atención de la salud en la mayoría de los países desarrollados, y las enfermedades cardiovasculares siguen siendo la principal causa de muerte y discapacidad en los Estados Unidos y la mayoría de los países europeos. Las enfermedades y trastornos metabólicos incluyen una gran variedad de afecciones que afectan los órganos, los tejidos y el sistema circulatorio del cuerpo

La dislipidemia es un caso frecuente entre diabéticos y seres humanos o animales con enfermedad cardiovascular; caracterizada típicamente por parámetros tales como triglicéridos plasmáticos elevados, colesterol HDL (lipoproteína de alta densidad) bajo, niveles normales a elevados de colesterol LDL (lipoproteína de baja densidad) y niveles

elevados de partículas LDL pequeñas y densas en la sangre. La dislipidemia y la hipertensión son los principales contribuyentes a una mayor incidencia de eventos coronarios, enfermedades renales y muertes entre seres humanos o animales con enfermedades metabólicas como diabetes y enfermedad cardiovascular.

La enfermedad cardiovascular puede manifestarse por muchos trastornos, síntomas y cambios en los parámetros clínicos que involucran el corazón, la vasculatura y los sistemas de órganos en todo el cuerpo, que incluyen aneurismas, angina, aterosclerosis, accidente cerebrovascular (ictus), enfermedad cerebrovascular, insuficiencia cardíaca congestiva, enfermedad de las arterias coronarias, infarto de miocardio, gasto cardíaco reducido y enfermedad vascular periférica, hipertensión, hipotensión, marcadores sanguíneos (por ejemplo, proteína C reactiva, BNP y enzimas tales como CPK, LDH, SGPT, SGOT), entre otros.

La mayoría de los procesos metabólicos y muchos parámetros cardiovasculares están regulados por múltiples péptidos y hormonas (“proteínas metabólicas”), y muchos de estos péptidos y hormonas, así como también sus análogos, han encontrado utilidad en el tratamiento de dichas enfermedades y trastornos. Sin embargo, el uso de péptidos y/u hormonas terapéuticas, incluso cuando se complementa con el uso de fármacos de molécula pequeña, ha tenido un éxito limitado en el tratamiento de dichas enfermedades y trastornos. En particular, la optimización de la dosis es importante para los fármacos y productos biológicos usados en el tratamiento de enfermedades metabólicas, especialmente aquellas con una ventana terapéutica estrecha. Las hormonas en general y los péptidos involucrados en la homeostasis de la glucosa a menudo tienen una ventana terapéutica estrecha. La ventana terapéutica estrecha, junto con el hecho de que tales hormonas y péptidos típicamente tienen una semivida corta que requiere una dosificación frecuente para lograr un beneficio clínico, da como resultado dificultades en el tratamiento de tales pacientes. Por lo tanto, sigue existiendo la necesidad de agentes terapéuticos con mayor eficacia y seguridad en el tratamiento de enfermedades metabólicas.

Por lo tanto, un aspecto de la presente invención es la incorporación de proteínas metabólicas biológicamente activas e involucradas o usadas en el tratamiento de enfermedades y trastornos metabólicos y cardiovasculares en proteínas de fusión BPXTEN para crear composiciones con utilidad en el tratamiento de tales trastornos, enfermedades y afecciones relacionadas. Las proteínas metabólicas pueden incluir cualquier proteína de interés o función biológica, terapéutica o profiláctica que sea útil para prevenir, tratar, mediar o mejorar una enfermedad, trastorno o afección metabólica o cardiovascular. La Tabla 4d proporciona una lista no limitante de tales secuencias de BP metabólicas que se incluyen en las proteínas de fusión BPXTEN de la invención. Las proteínas metabólicas de las composiciones de BPXTEN de la invención pueden ser una proteína que exhibe al menos aproximadamente 80 % de identidad de secuencia, o alternativamente un 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una proteína de la lista de la Tabla 4d.

96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una secuencia proteica seleccionada de la Tabla 4d.

Tabla 4d: Proteínas biológicamente activas para trastornos metabólicos y cardiología

Nombre de la proteína (Sinónimo)	Secuencia	SEQ ID NO.
Anti-CD3	Ver las patentes de Estados Unidos núm. 5,885,573 y 6,491,916	
IL-1ra, longitud completa humana	MEICRGLRSHLITLLLFLFHSETICRPSGRKSSK MQAFRIWDVNQKTFYLRNNQLVAGYLQGPN VNLEEKIDVVPIEPHALFLGIHGGKMCLSCVK SGDETRLQLEAVNITDLSENRKQDKRFAFIRS DSGPTTSFESAACPGWFLCTAMEADQPVS LT NMPDEGVMVTKFYFQEDE	1723
IL-1ra, Perro	METCRCPLSYLISFLLFLPHSETACRLGKRPCR MQAFRIWDVNQKTFYLRNNQLVAGYLQGSN TKLEEKLDVVPVEPHAVFLGIHGGKLC LACV KSGDETRLQLEAVNITDLSKNKDQDKRFTFIL SDSGPTTSFESAACPGWFLCTALEADRPVSLT NRPEEAMMVTKFYFQKE	1724
IL-1ra, Conejo	MRPSRSTRRHLSLLLFLFHSETACRPSGKRPC RMQAFRIWDVNQKTFYLRNNQLVAGYLQGP NAKLEERIDVVPLEPQLLFLGIQRGKLC LSCV KSGDKMKLHLEAVNITDLGKNKEQDKRFTFI RSNSGPTTTFESASCPGWFLCTALEADQPVS L TNTPDDSIVVTKFYFQED	1725
IL-1ra, Rata	MEICRGPYSHLISLLLILLFRSESAGHIPAGKRP CKMQAFRIWDTNQKTFYLRNNQLIAGYLQGP NTKLEEKIDMVPIDFRNVFLGIHGGKLC LSCV KSGDDTKLQLEEVNITDLNKNKEEDKRFTFIR SETGPTTSFESLACPGWFLCTTLEADHPVSLT NTPKEPCTVTKFYFQED	1726
IL-1ra, ratón	MEICWGPYSHLISLLLILLFHSEAACRPSGKRP CKMQAFRIWDTNQKTFYLRNNQLIAGYLQGP NIKLEEKIDMVPIDLHSVFLGIHGGKLC LSCAK SGDDIKLQLEEVNITDLSKNKEEDKRFTFIRSE KGPTTSFESAACPGWFLCTTLEADRPVSLTNT PEEPLIVTKFYFQEDQ	1727
Anakinra	MRPSGRKSSKMQAFRIWDVNQKTFYLRNNQ LVAGYLQGPNVNLEEKIDVVPIEPHALFLGIH GGKMCLSCVKSGDETRLQLEAVNITDLSENR KQDKRFAFIRSDSGPTTSFESAACPGWFLCTA MEADQPVS LTNMPDEGVMVTKFYFQEDE	1728
péptido α -natriurético (ANP)	SLRRSSCFGGRMDRIGAQSGLGCNSFRY	1729
péptido β -natriurético humano (BNP humano)	SPKMVQGS GGFGRKMDRISSSSGLGCKVLRR H	1730
Cerebro natriurético péptido, Rata; (BNP Rata)	NSKMAHSSSCFGQKIDRIGAVSRLGCDGLRLF	1731
Péptido natriurético tipo C (CNP, porcino)	GLSKGCFGLKLDRIGSMSGLGC	1732
Factor de	PALPEDGGSGAFPPGHFKDPKRLYCKNGGFFL	1733

Nombre de la proteína (Sinónimo)	Secuencia	SEQ ID NO.
crecimiento de fibroblastos (FGF-2) 2	RIHPDGRVDGVREKSDPHIKLQLQAEERGVS IKGVCANRYLAMKEDGRLLASKCVTDECFFF ERLESNNYNTYRS RKYTSWYVAL KRTGQYKLGS KTGPGQKAIL FLPMSEAKS	
Receptor de TNF (TNFR)	LPAQVAFTPYAPEPGSTCRLREYYDQTAQMC CSKCSPGQHAKVFCTKTSDTVCDSCEDSTYT QLWNWVPECLSCGSRSSDQVETQACTREQN RICTCRPGWYCALSKQEGCRLCAPLRKCRPGF GVARPGTETSDVVCKPCAPGTFSNTTSSTDIC RPHQICNVVAIPGNASMDAVCTSTSPTRSMAP GAVHLPQPVSTRSQHTQPTPEPSTAPSTSFLLP MGSPPPAEGSTGD	1734

“Anti-CD3” significa el anticuerpo monoclonal contra la proteína de superficie de linfocitos T CD3, variantes de especie y secuencia, y fragmentos de este, que incluye OKT3 (también llamado muromonab) y anticuerpo monoclonal anti-CD3 humanizado (hOKT31(Ala-Ala)) (Herold y otros, 2002, New England Journal of Medicine 346:1692-1698) Las proteínas de fusión que contienen anti-CD3 de la invención pueden encontrar un uso particular para retrasar la aparición de diabetes tipo 1, que incluye el uso del anti-CD3 como efector terapéutico así como también un resto de direccionamiento para una segunda BP terapéutica en la composición de BPXTEN. Las secuencias para la región variable y la creación de anti-CD3 se han descrito en las patentes de Estados Unidos núm. 5,885,573 y 6,491,916.

“IL-1ra” significa la proteína antagonista del receptor de IL-1 humana y las variantes de especie y secuencia de esta, que incluyen la variante de secuencia anakinra (Kineret®), que tiene al menos una porción de la actividad biológica de la IL-1ra madura. Anakinra es una IL-1ra humana recombinante no glucosilada y se diferencia de la IL-1ra humana endógena por la adición de una metionina N-terminal. Una versión comercializada de anakinra se comercializa como Kineret®. Se une al receptor de IL-1 con la misma avidez que la IL-1ra y la IL-1b nativas, pero no da como resultado la activación del receptor (transducción de señales), un efecto atribuido a la presencia de un solo motivo de unión al receptor en IL-1ra frente a dos de tales motivos en IL-1 α e IL-1 β . Anakinra tiene 153 aminoácidos y 17,3 kD de tamaño, y tiene una semivida informada de aproximadamente 4-6 horas.

Se ha informado un aumento de la producción de IL-1 en pacientes con diversas enfermedades infecciosas microbianas y una variedad de otras enfermedades. Las proteínas de fusión que contienen IL-1ra de la invención pueden encontrar un uso particular en el tratamiento de cualquiera de las enfermedades y trastornos anteriores. La IL-1ra se ha clonado, como se describe en las patentes de Estados Unidos núm. 5,075,222 y 6,858,409.

“Péptidos natriuréticos” significa péptido natriurético auricular (ANP), péptido natriurético cerebral (BNP o péptido natriurético tipo B) y péptido natriurético tipo C (CNP); de especies tanto humanas como no humanas y variantes de secuencia de estos que tienen al menos una porción de la actividad biológica de los péptidos natriuréticos homólogos

maduros. Las secuencias de formas útiles de péptidos natriuréticos se describen en la publicación de patente de Estados Unidos 20010027181. Los ejemplos de ANP incluyen ANP humano (Kangawa y otros, 1984, BBRC 118:131) o el de varias especies, que incluyen ANP de cerdo y rata (Kangawa y otros, 1984, BBRC 121:585). El análisis de la secuencia reveló que el preproBNP consiste en 134 residuos y se escinde en un ProBNP de 108 aminoácidos. La escisión de una secuencia de 32 aminoácidos del extremo C-terminal de ProBNP da como resultado el BNP humano (77-108), que es la forma fisiológicamente activa en circulación. El BNP humano de 32 aminoácidos involucra la formación de un enlace disulfuro (Sudoh y otros, 1989, BBRC 159:1420) y las patentes de Estados Unidos núm. 5,114,923, 5,674,710, 5,674,710 y 5,948,761. La BPXTEN que contiene una o más funciones natriuréticas puede ser útil en el tratamiento de la hipertensión, la inducción de la diuresis, la inducción de la natriuresis, la dilatación o relajación de los conductos vasculares, la unión a los receptores de péptidos natriuréticos (tales como NPR-A), la supresión de la secreción de aldosterona de la glándula suprarrenal, el tratamiento de enfermedades y trastornos cardiovasculares, reducción, detención o inversión de la remodelación cardíaca después de un evento cardíaco o como resultado de insuficiencia cardíaca congestiva, tratamiento de enfermedades y trastornos renales; tratamiento o prevención del ictus isquémico, y tratamiento del asma.

“Factor 2 de crecimiento de unión a heparina” o “FGF-2” significa la proteína FGF-2 humana, y variantes de especie y secuencia de esta que tienen al menos una porción de la actividad biológica de la contraparte madura. La FGF-2 se ha clonado, como se describe en Burgess, W. H. y Maciag, T., Ann. Rev. Biochem., 58:575-606 (1989); Coulier, F. y otros, 1994, Prog. Growth Factor Res. 5:1; y la publicación PCT WO 87/01728.

“Receptor de TNF” significa el receptor humano para TNF, y variantes de especie y secuencia de este que tienen al menos una porción de la actividad biológica receptora del TNFR maduro. Se ha determinado la estructura cristalina de rayos x del complejo formado por el dominio extracelular del receptor de TNF p55 humano y TNF β (Banner y otros, 1993 Cell 73:431, incorporado en la presente descripción como referencia).

Factores de coagulación

En la hemofilia, la coagulación de la sangre se ve alterada por la falta de ciertos factores de coagulación de la sangre en plasma. El factor IX humano (FIX) es un zimógeno de una serina proteasa que es un componente importante de la vía intrínseca de la cascada de la coagulación sanguínea. Las proteínas del factor VIIa (FVIIa) han encontrado utilidad para el tratamiento de episodios hemorrágicos en pacientes con hemofilia A o B con inhibidores de FVIII o FIX y en pacientes con hemofilia adquirida, así como también para la prevención de hemorragias en intervenciones quirúrgicas o procedimientos invasivos en pacientes con hemofilia A o B con inhibidores de FVIII o FIX. Por lo tanto, sigue existiendo la necesidad de composiciones de factor IX y factor VIIa con semivida extendida y retención de actividad cuando se administran como parte de un régimen preventivo y/o terapéutico para la hemofilia B, así como también formulaciones que

reduzcan los efectos secundarios y puedan administrarse tanto por vía intravenosa como subcutánea.

Los factores de coagulación para su inclusión en la BPXTEN de la invención pueden incluir proteínas de interés o función biológica, terapéutica o profiláctica que son útiles para prevenir, tratar, mediar o mejorar los trastornos, enfermedades o deficiencias de la coagulación sanguínea. Las proteínas de coagulación adecuadas incluyen polipéptidos biológicamente activos que están involucrados en la cascada de coagulación como sustratos, enzimas o cofactores.

La Tabla 4e proporciona una lista no limitante de secuencias de factores de coagulación que se incluyen en las proteínas de fusión BPXTEN de la invención. Los factores de coagulación para su inclusión en la BPXTEN de la invención pueden ser una proteína que exhibe al menos aproximadamente 80 % de identidad de secuencia o, alternativamente, un 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una secuencia proteica seleccionada de la Tabla 4e.

Tabla 4e: Secuencias de polipéptidos de factores de coagulación

Nombre de BPXTEN	SEQ ID NO:	Secuencia de aminoácidos
Precursor de FIX	1735	<u>MQRVNMIMAES</u> PGLITICLLGYLLSAECTVFLDHENA <u>NKILNRPKRY</u> NSGKLEEFVQGNLERECMEEKCSFEE AREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCK DDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFC KNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRV SVSQTSKLTRAETVFPDVDYVNSTEAEITLDNITQSTQ SFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGG SIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHT QKRNVIIRIIPHHYNAAINKYNHDIALLELDEPLVLNS YVTPICIAADKEYTNIFLKFGSGYVSGWGRVVFHKGRSA LVLQYLRVPLVDRATCLRSTKFTIYNMFCAGFHEG GRDSCQGDSSGPHVTEVEGTSFLTGIISWGEECAMK GKYGIYTKVSRVYNWIKETKLT
FIX Homo sapiens	1736	YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTE RTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYEC WCPFGFEGKNCELDVTCNIKNGRCEQFC KNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLT RAETVFPDVDYVNSTEAEITLDNITQSTQSFNDFTRV VGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWI VTAAHCVETGVKITVVAGEHNIEETEHTQKRNVIIRI IPHHYNAAINKYNHDIALLELDEPLVLNSYVTPICIA DKEYTNIFLKFGSGYVSGWGRVVFHKGRSALVLQYLR VPLVDRATCLRSTKFTIYNMFCAGFHEGGRDSCQG DSGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTK VSRVYNWIKETKLT
Secuencia 4 de la patente US 20080214462	1737	MQRVNMIMAES <u>PGLITICLLGYLLSAECTVFLDHENA</u> <u>NKILNRPKRY</u> NSGKLEEFVQGNLERECMEEKCSFEE AREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCK DDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFC KNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRV SVSQTSKLTRAEA VFPDVDYVNSTEAEITLDNITQST QSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCG GSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHT

Nombre de BPXTEN	SEQ ID NO:	Secuencia de aminoácidos
		EQKRN VIRIIPHHNFNAAINTYNHDIALLELDEPLVNS YVTPIC IADKEYTNIFLKFGSGYVSGWGRVFHKGRSA LVLQYLRVPLVDRATCLRSTKFTIYNNMFCAGFHEG GRDSCQGDSSGGPHVTEVEGTSFLTGIISWGEECAMK GKYGIYTKVSR YVNWIKEKTKLT
Secuencia 6 de la patente US 20080214462	1738	MQRVNMIMAESPLITICLLGYLLSAECTVFLDHENA NKILNRPKRYNSGKLEEFVQGNLERECMEEKCSFEE AREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCK DDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFC KNSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRV SVSQT SKLTRAEAVFPDVDYVNSTEAE TILDNITQST QSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCG GSIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHT EQKRN VIRIIPHHNFNAAINTYNHDIALLELDEPLVNS YVTPIC IADKEYTNIFLKFGSGYVSGWGRVFHKGRSA LVLQYLRVPLVDRATCLRSTKFTIFNNMFCAGFHEG GRDSCQGDSSGGPHVTEVEGTSFLTGIISWGEECAMK GKYGIYTKVSR YVNWIKEKTKLT
Secuencia 8 de la patente US 20080214462	1739	MQRVNMIMAESPLITICLLGYLLSAECTVFLDHENA NKILNRPKRYNSGKLEEFVQGNLERECMEEKCSFEE AREVFENTERTTEFWKQYVDGDQCESNPCLNGGSCK DDINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFC KNSADNKVVCSCTEG YRLAENQKSCEPAVPFPCGRVSVSQT SKLTRAEAVFP DVDYVNSTEAE TILDNITQSTQSFNDFTRVVGGEDAK PGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCV ETGVKITVVAGEHNIEETEHTEQKRN VIRIIPHHNFNA AINTYNHDIALLELDEPLVLNSYVTPIC IADKEYTNIF LKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDAT CLRSTKFTIFNNMFCAGFHEGGRDSCQGDSSGGPHVT EVEGTSFLTGI VSWGEGCAMKGKYGIYTKVSR YVNW IKEKTKLT
Secuencia 2 de la patente US 7125841	1740	MQRVNMIMAESPLITICLLGYLLSAECTVFLDHENA NKILNRPKRYNSGKLEEFVQGNLERECMEEKCSFEEP REVFENTEKITEFWKQYVDGDQCESNPCLNGGSCKD DINSYECWCPFGFEGKNCELDVTCNIKNGRCEQFCK NSADNKVVCSCTEGYRLAENQKSCEPAVPFPCGRVS VSQT SKLTRAEAVFPDVDYVNPTEAE TILDNITQGTQ SFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGG SIVNEKWIVTAAHCVETGVKITVVAGEHNIEETEHT EQKRN VIRAIIPHHNYNAAINKYNHDIALLELDEPLVL NSYVTPIC IADKEYTNIFLKFGSGYVSGWARVFHKGR SALVLQYLRVPLVDRATCLRSTKFTIYNNMFCAGFH EGGRDSCQGDSSGGPHVTEVEGTSFLTGIISWGEECA MKGKYGIYTKVSR YVNWIKEKTKLT
Secuencia 1 de la patente US 20080167219	1741	YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTE RTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYEC WCPFGFEGKNCELDATCN IKNGRCEQFCKNSADNKV VCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQT SKLT RAETVFPDVDYVNSTEAE TILDNITQSTQSFNDFTRV VGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWI VTAAHCVETGVKITVVAGEHNIEETEHTEQKRN VIRII PHHNYNAAINKYNHDIALLELDEPLVLNSYVTPIC I ADKEYTNIFLKFGSGYVSGWGRVFHKGRSALVLQYLR VPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSCQG DSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTK VSR YVNWIKEKTKLT

Nombre de BPXTEN	SEQ ID NO:	Secuencia de aminoácidos
Secuencia 2 de la patente US 20080167219	1742	YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTE RTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYEC WCPFGFEGKNCELDVTCTNIKNGRCEQFCKNSADNKV VCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLT RAETVFPDVDYVNSTEAEITLDNITQSTQSFNDFTRV VGGEDAKPGQFPWQVVLNGKVDADFCCGSIVNEKWI VTAAHCVETGVKITVVAGEHNIEETEHETEQKRNVIIRI PHHNYNAAINKYNHDIALLELDAPLVLNSYVTPICIA DKEYTNIFLKFGSGYVSGWGRVFHKGGRSALVLQYLR VPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSCQG DSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGITYK VSRYVNWIKEKTKLT
Secuencia 3 de la patente US 20080167219	1743	YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTE RTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYEC WCPFGFEGKNCELDATCTNIKNGRCEQFCKNSADNKV VCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLT RAETVFPDVDYVNSTEAEITLDNITQSTQSFNDFTRV VGGEDAKPGQFPWQVVLNGKVDADFCCGSIVNEKWI VTAAHCVETGVKITVVAGEHNIEETEHETEQKRNVIIRI PHHNYNAAINKYNHDIALLELDAPLVLNSYVTPICIA DKEYTNIFLKFGSGYVSGWGRVFHKGGRSALVLQYLR VPLVDRATCLRSTKFTIYNNMFCAGFHEGGRDSCQG DSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGITYK VSRYVNWIKEKTKLT
Secuencia 4 de la patente US 20080167219	1744	YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTE RTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYEC WCPFGFEGKNCELDATCTNIKNGRCEQFCKNSADNKV VCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLT RAETVFPDVDYVNSTEAEITLDNITQSTQSFNDFTRV VGGEDAKPGQFPWQVVLNGKVDADFCCGSIVNEKWI VTAAHCVETGVKITVVAGEHNIEETEHETEQKRNVIIRI PHHNYNAAINKYNHDIALLELDEPLVLNSYVTPICIA DKEYTNIFLKFGSGYVSGWGRVFHKGGRSALVLQYLR VPLVDRATCLASTKFTIYNNMFCAGFHEGGRDSCQG DSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGITYK VSRYVNWIKEKTKLT
Secuencia 5 de la patente US 20080167219	1745	YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTE RTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYEC WCPFGFEGKNCELDVTCTNIKNGRCEQFCKNSADNKV VCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLT RAETVFPDVDYVNSTEAEITLDNITQSTQSFNDFTRV VGGEDAKPGQFPWQVVLNGKVDADFCCGSIVNEKWI VTAAHCVETGVKITVVAGEHNIEETEHETEQKRNVIIRI PHHNYNAAINKYNHDIALLELDAPLVLNSYVTPICIA DKEYTNIFLKFGSGYVSGWGRVFHKGGRSALVLQYLR VPLVDRATCLASTKFTIYNNMFCAGFHEGGRDSCQG DSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGITYK VSRYVNWIKEKTKLT
Secuencia 6 de la patente US 20080167219	1746	YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTE RTTEFWKQYVDGDQCESNPCLNGGSCKDDINSYEC WCPFGFEGKNCELDATCTNIKNGRCEQFCKNSADNKV VCSCTEGYRLAENQKSCEPAVPFPCGRVSVSQTSKLT RAETVFPDVDYVNSTEAEITLDNITQSTQSFNDFTRV VGGEDAKPGQFPWQVVLNGKVDADFCCGSIVNEKWI VTAAHCVETGVKITVVAGEHNIEETEHETEQKRNVIIRI PHHNYNAAINKYNHDIALLELDAPLVLNSYVTPICIA DKEYTNIFLKFGSGYVSGWGRVFHKGGRSALVLQYLR

Nombre de BPXTEN	SEQ ID NO:	Secuencia de aminoácidos
		VPLVDRATCLASTKFTIYNNMFCAGFHEGGRDSCQG DSGGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTK VSRVYNWIKETKLT
Factor VII/ VIIa	1747	ANAFLEELRPGSLERECKEEQCSFEEAREIFKDAERT KLFWISYSDGDQCASSPCQNGGSKDQLQSYICFCLP AFEGRNCETHKDDQLICVNENGGCEQYCSDHTGTKR SRCHEGYSLADGVSCTPTVEYPCGKIPILEKRNAS KPQGRIVGGKVC PKGECPWQVLLL VNGAQLCGGTLI NTIWVVSAAHCFDKIKNWRNLI AVLGEHDLSEHDGD EQSRRVAQVIIPSTYVPGTTNHDIALLR LHQPVVLT D HVVPLCLPERTFSERTLAFVRFSLVSGWGQLLDRGAT ALELMVLNVPRLMTQDCLQQRKVGDSPNITEYMFC AGYSDGSKDSCKGDSGGPHATHYRGTWYLTGIVSW GQGCATVGHFVYTRVSQYIEWLQKLMRSEPRPGV LLRAPFP

El "Factor IX" ("FIX") incluye la proteína del Factor IX humano y variantes de especie y secuencia de esta que tienen al menos una porción de la actividad biológica receptora del Factor IX maduro. En algunas modalidades, el péptido FIX es un análogo estructural o un peptidomimético de cualquiera de los péptidos FIX descritos en la presente descripción, que incluyen las secuencias de la Tabla 4e. En algunas modalidades, el péptido FIX es un análogo estructural o un peptidomimético de cualquiera de los péptidos FIX descritos en la presente descripción, que incluyen las secuencias de la Tabla 4e. En un ejemplo específico de la presente invención, el FIX es FIX humano. En otra modalidad, el FIX es una secuencia polipeptídica de la Tabla 4e. El factor IX maduro es una proteína de cadena sencilla de 415 residuos de aminoácidos que contiene aproximadamente un 17 % de carbohidratos en peso (Schmidt 2003, Trends Cardiovasc Med, 13: 39).

En algunos casos, el factor de coagulación es el Factor IX, una variante de secuencia del Factor IX, o un resto del Factor IX, tal como las secuencias ilustrativas de la Tabla 4e, así como también cualquier proteína o polipéptido sustancialmente homólogo al mismo cuyas propiedades biológicas den como resultado la actividad del Factor IX.

"Factor VII" (FVII) significa la proteína humana, y las variantes de especie y secuencia de esta que tienen al menos una porción de la actividad biológica del Factor VII activado. El Factor VII y el FVIIa humano recombinante se han introducido para su uso en hemorragias incontrolables en pacientes con hemofilia (con deficiencia de Factor VIII o IX) que han desarrollado inhibidores contra el factor de coagulación de reemplazo. El Factor VIIa humano recombinante tiene utilidad en el tratamiento de la hemorragia incontrolable en pacientes con hemofilia (con deficiencia de Factor VIII o IX), que incluyen aquellos que han desarrollado inhibidores contra el factor de coagulación de reemplazo. En algunas modalidades, el péptido FVII es la forma activada (FVIIa), un análogo estructural o un peptidomimético de cualquiera de los péptidos FVII descritos en la presente descripción, que incluyen las secuencias de la Tabla 4e. Los factores VII y VIIa se han clonado, como se

describe en la patente de Estados Unidos núm. 6,806,063 y la solicitud de patente de Estados Unidos núm. 20080261886.

Proteínas de la hormona del crecimiento

"Hormona del crecimiento" o "GH" significa la proteína de la hormona del crecimiento humana y las variantes de especie y secuencia de esta, e incluye, pero no se limita a, la secuencia humana de 191 aminoácidos de cadena sencilla de la GH. La invención contempla la inclusión en la BPXTEN de cualquier secuencia homóloga de GH, fragmentos de secuencia que sean naturales, tales como los de primates, mamíferos (que incluyen los animales domésticos) y variantes de secuencia no naturales que retienen al menos una porción de la actividad biológica o la función biológica de GH y/o que son útiles para prevenir, tratar, mediar o mejorar una enfermedad, deficiencia, trastorno o afección relacionada con la GH. Las secuencias de GH de no mamíferos están bien descritas en la bibliografía. Por ejemplo, se puede encontrar una alineación de secuencias de GH de pescado en *Genetics and Molecular Biology* 2003 26 p.295-300. Además, las secuencias nativas homólogas a la GH humana pueden encontrarse mediante técnicas estándar de búsqueda de homología, tales como NCBI BLAST.

En una modalidad, la GH incorporada en las composiciones humanas o animales puede ser un polipéptido recombinante con una secuencia correspondiente a una proteína que se encuentra en la naturaleza. En otra modalidad, la GH puede ser una variante de secuencia, un fragmento, un homólogo o un mimético de una secuencia natural que retiene al menos una porción de la actividad biológica de la GH nativa. La Tabla 4f proporciona una lista no limitante de secuencias de GH de una amplia variedad de especies de mamíferos que se incluyen en las proteínas de fusión BPXTEN de la invención. Cualquiera de estas secuencias de GH o derivados homólogos contruidos mediante mezcla de mutaciones individuales entre especies o familias puede ser útil para las proteínas de fusión de esta invención. La GH que puede incorporarse en una proteína de fusión BPXTEN puede incluir una proteína que exhibe al menos aproximadamente 80 % de identidad de secuencia o, alternativamente, un 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una proteína seleccionada de la Tabla 4f.

Tabla 4f: Secuencias de aminoácidos de la hormona del crecimiento de especies animales

Especie de GH	Secuencia de aminoácidos	SEQ ID NO.
Hombre	FPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPK EQKYSFLQNPQTSLCFSESIPTPSNREETQQKSNLELLRI SLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKD LEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDD ALLKNYGLLYCFRKDMDKVETFLRIVQCRSVEGSCGF	1750
Cerdo	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPE GQRYSIQNAQAACFSETIPAPTGKDEAQQRSDVELLR FLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEKDKD LEEGIQALMRELEDGSPRAGQILKQTYDKFDTNLRSD ALLKNYGLLSCFKKDLHKAETYLRV MKCRRFVESSCAF	1751

Especie de GH	Secuencia de aminoácidos	SEQ ID NO.
Alpaca	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERTYIPE GQRYSIQNAQA AFCFSETIPAPT GKDEAQQRSDVELLR FSLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEK LKD LEEGIQALMRELEDGSPRAGQILRQTYDKFDTNLRSD ALLKNYGLLSCFKKDLHKAETYL RV MKCRRFVESSCAF	1752
Camello	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERTYIPE GQRYSIQNAQA AFCFSETIPAPT GKDEAQQRSDVELLR FSLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEK LKD LEEGIQALMRELEDGSPRAGQILRQTYDKFDTNLRSD ALLKNYGLLSCFKKDLHKAETYL RV MKCRRFVESSCAF	1753
Caballo	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPE GQRYSIQNAQA AFCFSETIPAPT GKDEAQQRSDMELLR FSLLLIQSWLGPVQLLSRVFTNSLVFGTSDRVYEK LRD LEEGIQALMRELEDGSPRAGQILKQTYDKFDTNLRSD ALLKNYGLLSCFKKDLHKAETYL RV MKCRRFVESSCAF	1754
Elefante	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPE GQRYSIQNAQA AFCFSETIPAPT GKDEAQQRSDVELLR FSLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEK LKD LEEGIQALMRELEDGSPRPGQVLKQTYDKFDTNMRSD DALLKNYGLLSCFKKDLHKAETYL RV MKCRRFVESSCAF	1755
Zorro rojo	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPE GQRYSIQNAQA AFCFSETIPAPT GKDEAQQRSDVELLR FSLVLIQSWLGPLQFLSRVFTNSLVFGTSDRVYEK LKD LEEGIQALMRELEDGSPRAGQILKQTYDKFDTNLRSD ALLKNYGLLSCFKKDLHKAETYL RV MKCRRFVESSCAF	1756
Perro	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPE GQRYSIQNAQA AFCFSETIPAPT GKDEAQQRSDVELLR FSLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEK LKD LEEGIQALMRELEDGSPRAGQILKQTYDKFDTNLRSD ALLKNYGLLSCFKKDLHKAETYL RV MKCRRFVESSCAF	1757
Gato	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPE GQRYSIQNAQA AFCFSETIPAPT GKDEAQQRSDVELLR FSLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEK LKD LEEGIQALMRELEDGSPRGGQILKQTYDKFDTNLRSD ALLKNYGLLSCFKKDLHKAETYL RV MKCRRFVESSCAF	1758
Visón americano	FPAMPLSSLFANAVLRAQHLHQLAADTYKDFERAYIP EGQRYSIQNAQA AFCFSETIPAPT GKDEAQQRSDMELL RFSLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEK LK DLEEGIQALMRELEDGSPRAGPILKQTYDKFDTNLRSD DALLKNYGLLSCFKKDLHKAETYL RV MKCRRFVESSCAF	1759
Rorcual común	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPE GQRYSIQNAQA AFCFSETIPAPT GKDEAQQRSDVELLR FSLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEK LKD LEEGIQALMRELEDGSPRAGQILKQTYDKFDTNMRSD DALLKNYGLLSCFKKDLHKAETYL RV MKCRRFVESSCAF	1760
Delfín	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPE GQRYSIQNTQA AFCFSETIPAPT GKDEAQQRSDVELLR FSLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEK LKD	1761

Especie de GH	Secuencia de aminoácidos	SEQ ID NO.
	LEEGIQALMRELEDGSPRAGQILKQTYDKFDTNMRSDDALLKNYGLLSCFKKDLHKAETYLRLVMKCRRFVESSCAF	
Hipopótamo	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPEGQRYSIQNTQA AFCFSETIPAPT GKDEAQQRSDVELLRFSLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEKDKDLEEGIQALMRELEDGSPRAGQILKQTYDKFDTNMRSDDALLKNYGLLSCFKKDLHKAETYLRLVMKCRRFVESSCAF	1762
Conejo	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPEGQRYSIQNAQA AFCFSETIPAPT GKDEAQQRSDMELLRFSLLLIQSWLGPVQFLSRAFTNTLVFGTSDRVYEKDKDLEEGIQALMRELEDGSPRVGQLLKQTYDKFDTNLRGDDALLKNYGLLSCFKKDLHKAETYLRLVMKCRRFVESSCVF	1763
Rata	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPEGQRYSIQNAQA AFCFSETIPAPT GKDEAQQRTDMELLRFSLLLIQSWLGPVQFLSRIFTNSLMFGTSDRVYEKDKDLEEGIQALMQELEDGSPRIGQILKQTYDKFDANMRSDDALLKNYGLLSCFKKDLHKAETYLRLVMKCRRFAESSCAF	1764
Ratón	FPAMPLSSLFSNAV LRAQHLHQLAADTYKEFERAYIPEGQRYSIQNAQA AFCFSETIPAPT GKDEAQQRTDMELLRFSLLLIQSWLGPVQFLSRIFTNSLMFGTSDRVYEKDKDLEEGIQALMQELEDGSPRVGQILKQTYDKFDANMRSDDALLKNYGLLSCFKKDLHKAETYLRLVMKCRRFVESSCAF	1765
Hámster	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPEGQRYSIQNAQT AFCFSETIPAPT GKDEAQQRSDMELLRFSLLLIQSWLGPVQFLSRIFTNSLMFGTSDRVYEKDKDLEEGIQALMQELEDGSPRVGQILKQTYDKFDTNMRSDDALLKNYGLLSCFKKDLHKAETYLRLVMKCRRFVESSCAF	1766
Rata topo	FPAMPLSNLFANAVLRAQHLHQLAADTYKEFERAYIPEGQRYSIQNAQA AFCFSETIPAPT GKDEAQQRSDMELLRFSLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVFEKDKDLEEGIQALMRELEDGSLRAGQLLKQTYDKFDTNMRSDDALLKNYGLLSCFKKDLHKAETYLRLVMKCRRFVESSCAF	1767
Conejillo de indias	FPAMPLSSLFGNAV LRAQHLHQLAADTYKEFERTYIPEGQRYSIHNTQT AFCFSETIPAPTDKEEAQQRSDVELLHFSLLLIQSWLGPVQFLSRVFTNSLVFGTSDRVYEKDKDLEEGIQALMRELEDGTPRAGQILKQTYDKFDTNLRSDALLKNYGLLSCFRKDLHRTETYLRLVMKCRRFVESSCAF	1768
Buey	AFPAMSLSGLFANAVLRAQHLHQLAADTFKEFERTYIPEGQRYSIQNTQV AFCFSETIPAPT GKNEAQQKSDLELLRISLLLIQSWLGPLQFLSRVFTNSLVFGTSDRVYEKDKDLEEGILALMRELEDGTPRAGQILKQTYDKFDTNMRSDDALLKNYGLLSCFRKDLHKTETYLRLVMKCRRFGEASCAF	1769
Oveja/Cabra	AFPAMSLSGLFANAVLRAQHLHQLAADTFKEFERTYIPEGQRYSIQNTQV AFCFSETIPAPT GKNEAQQKSDLELLRISLLLIQSWLGPLQFLSRVFTNSLVFGTSDRVYEKDKDLEEGILALMRELEDVTPRAGQILKQTYDKFDTNMRSDDALLKNYGLLSCFRKDLHKTETYLRLVMKCRRFGEASCAF	1770

Especie de GH	Secuencia de aminoácidos	SEQ ID NO.
Ciervo rojo	FPAMSLSGLFANAVLRAQHLHQLAADTFKEFERTYIPE GQRYSIQNTQVAFCFSETIPAPTGKNEAQQKSDLELLRI SLLLIQSWLGPLQFLSRVFTNSLVFGTSDRVYEKLKDL EEGILALMRELEDGTPRAGQILKQTYDKFDTNMRSDD ALLKNYGLLSCFRKDLHKTETETYL RV MKCRRFGEASCAF	1771
Jirafa	AFPAMSLSGLFANAVLRAQHLHQLAADTFKEFERTYIP EGQRYSIQNTQVAFCFSETIPAPTGKNEAQQKSDLELL RISLLLIQSWLGPLQFLSRVFSNSLVFGTSDRVYEKLKDL LEEGILALMRELEDGTPRAGQILKQTYDKFDTNMRSDD DALLKNYGLLSCFRKDLHKTETETYL RV MKCRRFGEASCAF	1772
Chevrotain a-1	FPAMSLSGLFANAVLRVQHLHQLAADTFKEFERTYIPE GQRYSIQNTQVAFCFSETIPAPTGKNEAQQKSDLELLRI SLLLIQSWLGPLQFLSRVFTNSLVFGTSDRVYEKLKDL EEGILALMRELEDGPPRAGQILKQTYDKFDTNMRSDD ALLKNYGLLSCFRKDLHKTETETYL RV MKCRRFGEASCAF	1773
Loris lento	FPAMPLSSLFANAVLRAQHLHQLAADTYKEFERAYIPE GQRYSIQNAQA AFCFSETIPAPTGKDEAQQRSDMELLR FSLLIQSWLGPVQLLSRVFTNSLVLTGSDRVYEKLKDL LEEGIQALMRELEDGSPRVGQILKQTYDKFDTNLRSD ALLKNYGLLSCFKKDLHKAETYL RV MKCRRFVESSCAF	1774
Tití	FPTIPLSRLLDNAMLRAHRLHQLAFD TYQEFEEAYIPK EQKYSFLQNPQTSLCFSESIPTASKKETQQKSNLELLR MSLLLIQSWFEPVQFLRSVFANSLLYGVSDSDVY EYLK DLEEGIQ TLMGRLEDGSPRTGEIFMQTYRKFDVNSQN NDALLKNYGLLYCFRKDMDKVETFLRI VQCR- SVEGSCGF	1775
Zarigüeya de cola negra	FPAMPLSSLFANAVLRAQHLHQLVADTYKEFERTYIPE AQRHSIQSTQTAF CFSETIPAPTGKDEAQQRSDVELLR SLLLIQSWLSPVQFLSRVFTNSLVFGTSDRVYEKL RDL EEGIQALMQELEDGSSRGGLVLKTTYDKFDTNLRSD ALLKNYGLLSCFKKDLHKAETYL RV MKCRRFVESSCAF	1776
Mono (rhesus)	FPTIPLSRLFDNAMLRAHRLHQLAFD TYQEFEEAYIPK EQKYSFLQNPQTSLCFSESIPTPSNREETQQKSNLELLRI SLLLIQSWLEPVQFLRSVFANSLVYGT SYSDVYDLLKDL LEEGIQ TLMGRLEDGSSRTGQIFKQTYSKFDTN SHNND ALLKNYGLLYCFRKDMDKIETFLRI VQCR-SVEGSCGF	1777

Citocinas

La BP puede ser una citocina o una o más citocinas. Las citocinas se refieren a proteínas (por ejemplo, quimiocinas, interferones, linfocinas, interleucinas y factores de necrosis tumoral) liberadas por las células que pueden afectar el comportamiento celular. Las citocinas pueden producirse en una amplia variedad de células, que incluye células inmunitarias tales como macrófagos, linfocitos B, linfocitos T y mastocitos, así como también células endoteliales, fibroblastos y diversas células estromales. Una citocina dada puede producirse en más de un tipo de célula. Las citocinas pueden estar involucradas en la producción de efectos inmunomoduladores sistémicos o locales.

Ciertas citocinas pueden funcionar como citocinas proinflamatorias. Las citocinas proinflamatorias se refieren a las citocinas involucradas en la inducción o amplificación de una reacción inflamatoria. Las citocinas proinflamatorias pueden trabajar con varias células del sistema inmunitario, tales como los neutrófilos y los leucocitos, para generar una respuesta inmunitaria. Ciertas citocinas pueden funcionar como citocinas antiinflamatorias. Las citocinas antiinflamatorias se refieren a las citocinas involucradas en la reducción de una reacción inflamatoria. Las citocinas antiinflamatorias, en algunos casos, pueden regular una respuesta de citocinas proinflamatorias. Algunas citocinas pueden funcionar como citocinas tanto proinflamatorias como antiinflamatorias.

Las citocinas incluidas en las composiciones de la invención pueden tener utilidad en el tratamiento de diversas categorías terapéuticas o de enfermedades, que incluyen, pero sin limitarse a, cáncer, artritis reumatoide, esclerosis múltiple, miastenia grave, lupus eritematoso sistémico, enfermedad de Alzheimer, esquizofrenia, infecciones virales (por ejemplo, hepatitis C crónica, SIDA), asma alérgica, procesos neurodegenerativos de la retina, trastornos metabólicos, resistencia a la insulina y miocardiopatía diabética. Las citocinas pueden ser especialmente útiles en el tratamiento de afecciones inflamatorias y autoinmunitarias.

Los ejemplos de citocinas que son regulables por los sistemas y composiciones de la presente descripción incluyen, pero no se limitan a, linfocinas, monocinas y hormonas polipeptídicas tradicionales excepto la hormona del crecimiento humano. Entre las citocinas se incluyen la hormona paratiroidea; tiroxina; insulina; proinsulina; relaxina; prorrelaxina; hormonas glicoproteicas tales como hormona estimulante del folículo (FSH), hormona estimulante de la tiroides (TSH) y hormona luteinizante (LH); factor de crecimiento hepático; factor de crecimiento de fibroblastos; prolactina; lactógeno placentario; factor alfa de necrosis tumoral; sustancia inhibidora de Müller; péptido asociado a gonadotropina de ratón; inhibina; activina; factor de crecimiento vascular endotelial; integrina; trombopoyetina (TPO); factores de crecimiento nervioso tales como NGF-alfa; factor de crecimiento plaquetario; factores de crecimiento transformante (TGF) tales como TGF-alfa, TGF-beta, TGF-beta1, TGF-beta2 y TGF-beta3; factor de crecimiento similar a la insulina I y II; eritropoyetina (EPO); Flt-3L; factor de células madre (SCF); factores osteoinductivos; interferones (IFN) tales como IFN- α , IFN- β , IFN- γ ; factores estimulantes de colonias (CSF) tales como CSF de macrófagos (M-CSF); CSF de granulocitos y macrófagos (GM-CSF); CSF de granulocitos (G-CSF); factor estimulante de macrófagos (MSP); interleucinas (IL) tales como IL-1, IL-1a, IL-1b, IL-1RA, IL-18, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-12b, IL-13, IL-14, IL-15, IL-16, IL-17, IL-20; un factor de necrosis tumoral tal como CD154, LT-beta, TNF-alfa, TNF-beta, 4-1BBL, APRIL, CD70, CD153, CD178, GITRL, LIGHT, OX40L, TALL-1, TRAIL, TWEAK, TRANCE; y otros factores polipeptídicos que incluyen LIF, oncostatina M (OSM) y ligando kit (KL). Los receptores de citocinas se refieren a las proteínas

receptoras que se unen a las citocinas. Los receptores de citocinas pueden estar unidos a la membrana y ser solubles.

El polinucleótido objetivo puede codificar una citocina. Los ejemplos no limitantes de citocinas incluyen 4-1BBL, activina β A, activina β B, activina β C, activina β E, artemina (ARTN), BAFF/BLyS/TNFSF138, BMP10, BMP15, BMP2, BMP3, BMP4, BMP5, BMP6, BMP7, BMP8a, BMP8b, proteína morfogénica ósea 1 (BMP1), CCL1/TCA3, CCL11, CCL12/MCP-5, CCL13/MCP-4, CCL14, CCL15, CCL16, CCL17/TARC, CCL18, CCL19, CCL2/MCP-1, CCL20, CCL21, CCL22/MDC, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, CCL3, CCL3L3, CCL4, CCL4L1/LAG-1, CCL5, CCL6, CCL7, CCL8, CCL9, CD153/CD30L/TNFSF8, CD40L/CD154/TNFSF5, CD40LG, CD70, CD70/CD27L/TNFSF7, CLCF1, c-MPL/CD110/TPOR, CNTF, CX3CL1, CXCL1, CXCL10, CXCL11, CXCL12, CXCL13, CXCL14, CXCL15, CXCL16, CXCL17, CXCL2/MIP-2, CXCL3, CXCL4, CXCL5, CXCL6, CXCL7/Ppbp, CXCL9, EDA-A1, FAM19A1, FAM19A2, FAM19A3, FAM19A4, FAM19A5, ligando Fas/FASLG/CD95L/CD178, GDF10, GDF11, GDF15, GDF2, GDF3, GDF4, GDF5, GDF6, GDF7, GDF8, GDF9, factor neurotrófico derivado de la línea de células gliales (GDNF), factor de diferenciación del crecimiento 1 (GDF1), IFNA1, IFNA10, IFNA13, IFNA14, IFNA2, IFNA4, IFNA5/IFNaG, IFNA7, IFNA8, IFNB1, IFNE, IFNG, IFNZ, IFN ω /IFNW1, IL11, IL18, IL18BP, IL1A, IL1B, IL1F10, IL1F3/IL1RA, IL1F5, IL1F6, IL1F7, IL1F8, IL1F9, IL1RL2, IL31, IL33, IL6, IL8/CXCL8, inhibina A, inhibina B, leptina, LIF, LTA/TNFB/TNFSF1, LTB/TNFC, neurturina (NRTN), OSM, OX-40L/TNFSF4/CD252, persefina (PSPN), RANKL/OPGL/TNFSF11(CD254), TL1A/TNFSF15, TNFA, TNF-alfa/TNFA, TNFSF10/TRAIL/APO-2L(CD253), TNFSF12, TNFSF13, TNFSF14/LIGHT/CD258, XCL1 y XCL2. En algunas modalidades, el gen objetivo codifica un inhibidor del punto de regulación inmunitario. Los ejemplos no limitantes de tales inhibidores de puntos de regulación inmunitarios incluyen PD-1, CTLA-4, LAG3, TIM-3, A2AR, B7-H3, B7-H4, BTLA, IDO, KIR y VISTA. En algunas modalidades, el gen objetivo codifica una cadena alfa, beta, gamma y/o delta del receptor de linfocitos T (TCR).

En algunos casos, la citocina puede ser una quimiocina. La quimiocina puede seleccionarse de un grupo que incluye, pero sin limitarse a, ARM CX2, BCA-1 / CXCL13, CCL11, CCL12/MCP-5, CCL13/MCP-4, CCL15/MIP-5/MIP-1 delta, CCL16 / HCC-4 / NCC4, CCL17/TARC, CCL18 / PARC / MIP-4, CCL19/MIP-3b, CCL2/MCP-1, CCL20/MIP-3 alfa/MIP3A, CCL21/6Ckine, CCL22/MDC, CCL23 / MIP 3, CCL24/Eotaxina 2/MPIF-2, CCL25/TECK, CCL26/Eotaxina 3, CCL27/CTACK, CCL28, CCL3/Mip1a, CCL4 / MIP1B, CCL4L1/LAG-1, CCL5/RANTES, CCL6/C10, CCL8/MCP-2, CCL9, CML5, CXCL1, CXCL10 / Crg-2, CXCL12 / SDF-1 beta, CXCL14/BRAK, CXCL15/Lungkina, CXCL16 / SR-PSOX, CXCL17, CXCL2/MIP-2, CXCL3 / GRO gamma, CXCL4 / PF4, CXCL5, CXCL6 / GCP-2, CXCL9 / MIG, FAM19A1, FAM19A2, FAM19A3, FAM19A4

/ TAFA4, FAM19A5, Fractalkina/CX3CL1, I-309/CCL1/TCA-3, IL-8/CXCL8, MCP-3/CCL7, NAP-2 / PPBP / CXCL7, XCL2 y Armo IL10.

La Tabla 4g proporciona una lista no limitante de tales secuencias de BP que se incluyen en las proteínas de fusión BPXTEN de la invención. Las proteínas metabólicas de las composiciones de BPXTEN de la invención pueden ser una proteína que exhibe al menos aproximadamente 80 % de identidad de secuencia, o alternativamente un 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una secuencia proteica seleccionada de la Tabla 4g.

Tabla 4g: Citocinas para la conjugación

Nombre de la proteína (Sinónimo)	Secuencia de aminoácidos	SEQ ID NO.
Anti-CD3	Ver las patentes de Estados Unidos núm. 5,885,573 y 6,491,916	
IL-1ra, longitud completa humana	MEICRGLRSHLITLLLFLFHSETICRPSGRKSSKMQAF RIWDVNQKTFYLRNNQLVAGYLQGPVNLEEKIDV VPIEPHALFLGIHGGKMCLSCVKSGDETRLQLEAVNI TDLSEN RKQDKRFAFIRSDSGPTTSFESAACPGWFLC TAMEADQPVSLTNMPDEGVMVTKFYFQEDE	270
IL-1ra, Perro	METCRCPLSYLISFLLFLPHSETACRLGKRPCRMAF RIWDVNQKTFYLRNNQLVAGYLQGSNTKLEEKLDV VPVEPHAVFLGIHGGKLCCLACVKSGDETRLQLEAVN ITDLSKNKDQDKRFTFILSDSGPTTSFESAACPGWFL CTALEADRPVSLTNRPEEAMMVTKFYFQKE	271
IL-1ra, Conejo	MRPSRSTRRHLSLILLFLFHSETACRPSGKRPCRMAF FRIWDVNQKTFYLRNNQLVAGYLQGPNAKLEERID VVPLEPQLLFLGIQRGKLCCLSCVKSGDKMKLHLEAV NITDLGKNKEQDKRFTFIRSNSGPTTTTFESASCPGWFL LCTALEADQPVSLTNTPDSDIVVTKFYFQED	272
IL-1ra, Rata	MEICRGPYSHLISLILLIFRSESAGHIPAGKRPCMKQ AFRIWDTNQKTFYLRNNQLIAGYLQGPNTKLEEKID MVPIDFRNVFLGIHGGKLCCLSCVKSGDDTKLQLEEV NITDLNKNKEEDKRFTFIRSETGPTTSFESLACPGWFL CTTLEADHPVSLTNTPKPCTVTKFYFQED	273
IL-1ra, ratón	MEICWGPYSHLISLILLIFHSEAACRPSGKRPCMKQ AFRIWDTNQKTFYLRNNQLIAGYLQGPNIKLEEKID MVPIDLHSVFLGIHGGKLCCLSCAKSGDDIKLQLEEVN ITDLSKNKEEDKRFTFIRSEKGPTTSFESAACPGWFLC TTLEADRPVSLTNTPEEPLIVTKFYFQEDQ	274
Anakinra	MRPSGRKSSKMQAFRIWDVNQKTFYLRNNQLVAGY LQGPVNLEEKIDVPIEPHALFLGIHGGKMCLSCVK SGDETRLQLEAVNITDLSENRKQDKRFAFIRSDSGPT TSFESAACPGWFLCTAMEADQPVSLTNMPDEGVMV TKFYFQEDE	275
IL-10	MHSSALLCCLVLLTGVRASPGGTQSENSCTHFPGN LPNMLRDLRDAFSRVKTFFQMKDQLDNLLKESLLE DFKGYLGCQALSEMIQFYLEEVMPPQAENQDPDIKAH VNSLGENLKTLLRLRLRRCHRFPCENKSKAVEQVKN AFNKLQEKGIYKAMSEFDIFINYIEAYMTMKIRN	276

“IL-1ra” significa la proteína antagonista del receptor de IL-1 humana y las variantes de especie y secuencia de esta, que incluyen la variante de secuencia anakinra (Kineret®), que

tiene al menos una porción de la actividad biológica de la IL-1ra madura. La IL-1ra humana es una glucoproteína madura de 152 residuos de aminoácidos. Las proteínas de fusión que contienen IL-1ra de la invención pueden encontrar un uso particular en el tratamiento de cualquiera de las enfermedades y trastornos anteriores. La IL-1ra se ha clonado, como se describe en las patentes de Estados Unidos núm. 5,075,222 y 6,858,409.

En algunos casos, la BP puede ser IL-10. La IL-10 puede ser una citocina antiinflamatoria efectiva que reprime la producción de citocinas y quimiocinas proinflamatorias. La IL-10 puede ser útil para el tratamiento de enfermedades autoinmunitarias y enfermedades inflamatorias tales como artritis reumatoide, esclerosis múltiple, miastenia gravis, lupus eritematoso sistémico, Alzheimer, esquizofrenia, asma alérgica, procesos neurodegenerativos retinianos y diabetes.

En algunos casos, la IL-10 puede modificarse para mejorar la estabilidad y disminuir la degradación prolítica. La modificación puede ser una o más sustituciones de enlaces amida. En algunos casos, pueden sustituirse uno o más enlaces amida dentro de la cadena principal de IL-10 para lograr los efectos mencionados anteriormente. El uno o más enlaces amida ($-\text{CONH}-$) en IL-10 pueden reemplazarse con un enlace que es un isótero de un enlace amida, tal como $-\text{CH}_2\text{NH}-$, $-\text{CH}_2\text{S}-$, $-\text{CH}_2\text{CH}_2-$, $-\text{CH}=\text{CH}-$ (cis y trans), $-\text{COCH}_2-$, $-\text{CH}(\text{OH})\text{CH}_2-$ o $-\text{CH}_2\text{SO}-$. Además, los enlaces amida en IL-10 también pueden reemplazarse por un enlace pseudopeptídico isótero reducido. Ver Couder y otros (1993) *Int. J. Peptide Protein Res.* 41:181-184, que se incorpora de este modo como referencia en su totalidad.

El uno o más aminoácidos ácidos, que incluyen ácido aspártico, ácido glutámico, ácido homoglutámico, tirosina, sulfonamidas de alquilo, arilo, arilalquilo y heteroarilo de ácido 2,4-diaminopropiónico, ornitina o lisina y alquilaminoácidos sustituidos con tetrazol; y residuos de amida de cadena lateral tales como asparagina, glutamina y derivados sustituidos con alquilo o aromáticos de asparagina o glutamina; así como también aminoácidos que contienen hidroxilo, que incluyen serina, treonina, homoserina, ácido 2,3-diaminopropiónico y derivados sustituidos con alquilo o aromáticos de serina o treonina.

El uno o más aminoácidos hidrófobos en la IL-10, tales como alanina, leucina, isoleucina, valina, norleucina, ácido (S)-2-aminobutírico, (S)-ciclohexilalanina u otros alfa-aminoácidos simples, pueden sustituirse por aminoácidos que incluyen, pero sin limitarse a, una cadena lateral alifática de carbonos C1-C10 que incluyen sustituciones de alquilo, alquenilo o alquinilo de cadena ramificada, cíclica y lineal

En algunos casos, el uno o más aminoácidos hidrófobos en la IL-10, pueden sustituirse tal como la sustitución de aminoácidos hidrófobos sustituidos con aromáticos que incluyen fenilalanina, triptófano, tirosina, sulfotirosina, bifenilalanina, 1-naftilalanina, 2-naftilalanina, 2-benzotienilalanina, 3-benzotienilalanina, histidina, que incluyen formas sustituidas con amino, alquilamino, dialquilamino, aza, halógeno (flúor, cloro, bromo o yodo) o alcoxi (de C₁-C₄) de los aminoácidos aromáticos enumerados anteriormente, ejemplos ilustrativos de los cuales son: 2-, 3- o 4-aminofenilalanina, 2-, 3- o 4-clorofenilalanina, 2-, 3- o 4-

metilfenilalanina, 2-, 3- o 4-metoxifenilalanina, 5-amino-, 5-cloro-, 5-metil- o 5-metoxitriptófano, 2', 3'- o 4'-amino-, 2', 3'- o 4'-cloro-, 2, 3 o 4-bifenilalanina, 2', 3'- o 4'-metil-, 2-, 3- o 4-bifenilalanina y 2- o 3-piridilalanina;

El uno o más aminoácidos hidrófobos en la IL-10 tales como fenilalanina, triptófano, tirosina, sulfotirosina, bifenilalanina, 1-naftilalanina, 2-naftilalanina, 2-benzotienilalanina, 3-benzotienilalanina, histidina, que incluyen amino, alquilamino, dialquilamino, aza, halogenados (flúor, cloro, bromo o yodo) o alcox pueden sustituirse por aminoácidos aromáticos que incluyen: 2-, 3- o 4-aminofenilalanina, 2-, 3- o 4-clorofenilalanina, 2-, 3- o 4- metilfenilalanina, 2-, 3- o 4-metoxifenilalanina, 5-amino-, 5-cloro-, 5-metil- o 5-metoxitriptófano, 2', 3'- o 4'-amino-, 2', 3'- o 4'-cloro-, 2, 3 o 4-bifenilalanina, 2', 3'- o 4'-metil-, 2-, 3- o 4-bifenilalanina y 2- o 3-piridilalanina

Los aminoácidos que comprenden cadenas laterales básicas, que incluyen arginina, lisina, histidina, ornitina, ácido 2,3-diaminopropiónico, homoarginina, que incluyen derivados sustituidos con alquilo, alquenilo o arilo de los aminoácidos anteriores, pueden estar sustituidos. Los ejemplos son N-épsilon-isopropil-lisina, 3-(4-tetrahidropiridil)-glicina, 3-(4-tetrahidropiridil)-alanina, N,N-gamma, gamma'-dietil-homoarginina, alfa-metil-arginina, ácido alfa-metil-2,3-diaminopropiónico, alfa-metil-histidina y alfa-metil-ornitina donde el grupo alquilo ocupa la posición pro-R del carbono alfa. La IL-10 modificada puede comprender amidas formadas a partir de cualquier combinación de alquilo, aromático, heteroaromático, ornitina o ácido 2,3-diaminopropiónico, ácidos carboxílicos o cualquiera de los muchos derivados activados bien conocidos, tales como cloruros de ácido, ésteres activos, azolidas y derivados relacionados, lisina y ornitina.

En algunos casos, los compuestos de IL-10 pueden comprender uno o más L-aminoácidos de origen natural, L-aminoácidos sintéticos y/o D-enantiómeros de un aminoácido. El polipéptido IL-10 puede comprender uno o más de los siguientes aminoácidos: ácido ω -aminodecanoico, ácido ω -aminotetradecanoico, ciclohexilalanina, ácido α,γ -diaminobutírico, ácido α,β -diaminopropiónico, ácido δ -amino valérico, t-butilalanina, t-butilglicina, N-metilisoleucina, fenilglicina, ciclohexilalanina, norleucina, naftilalanina, ornitina, citrulina, 4-clorofenilalanina, 2-fluorofenilalanina, piridilalanina 3-benzotienilalanina, hidroxiprolina, β -alanina, ácido o-aminobenzoico, ácido m-aminobenzoico, ácido p-aminobenzoico, ácido m-aminometilbenzoico, ácido 2,3-diaminopropiónico, ácido α -aminoisobutírico, N-metilglicina (sarcosina), 3-fluorofenilalanina, 4-fluorofenilalanina, penicilamina, ácido 1,2,3,4-tetrahidroisoquinolina-3-carboxílico, β -2-tienilalanina, sulfóxido de metionina, homoarginina, N-acetil lisina, ácido 2,4-diamino butírico, rho-aminofenilalanina, N-metilvalina, homocisteína, homoserina, ácido ϵ -amino hexanoico, ácido ω -aminohexanoico, ácido ω -aminoheptanoico, ácido ω -aminooctanoico y ácido 2,3-diaminobutírico.

La IL-10 puede comprender un residuo de cisteína o una cisteína que puede actuar como enlazador con otro péptido a través de un enlace disulfuro o para proporcionar la ciclación del polipéptido IL-10. Los métodos para introducir una cisteína o un análogo de cisteína se

conocen en la técnica; ver, por ejemplo, la patente de Estados Unidos núm. 8,067,532. Un polipéptido IL-10 puede ciclarse. Otros medios de ciclación incluyen la introducción de un enlazador de oxima o un enlazador de lantionina; ver, por ejemplo, la patente de Estados Unidos núm. 8,044,175. Puede usarse y/o introducirse cualquier combinación de aminoácidos (o restos no aminoacídicos) que puedan formar un enlace ciclante. Puede generarse un enlace ciclante con cualquier combinación de aminoácidos (o con un aminoácido y $-(CH_2)_nCO-$ o $-(CH_2)_nC_6H_4-CO-$) con grupos funcionales que permiten la introducción de un puente. Algunos ejemplos son disulfuros, miméticos de disulfuro tales como el puente $-(CH_2)_n$ -carba, tioacetal, puentes tioéter (cistationina o lantionina) y puentes que contienen ésteres y éteres.

La IL-10 puede sustituirse con un N-alquilo, arilo o reticulación de la cadena principal para construir lactamas y otras estructuras cíclicas, derivados de hidroximetilo C-terminal, derivados o-modificados, derivados modificados en el N-terminal que incluyen amidas sustituidas tales como alquilamidas e hidrazidas. En algunos casos, un polipéptido IL-10 es un análogo retroinverso.

La IL-10 puede ser IL-10 puede ser proteína nativa, fragmento de péptido IL-10 o péptido modificado, que tiene al menos una porción de la actividad biológica de la IL-10 nativa. La IL-10 puede modificarse para mejorar la captación intracelular. Una de tales modificaciones puede ser la unión de un dominio de transducción de proteínas. El dominio de transducción de proteínas puede unirse al extremo C-terminal de la IL-10. Alternativamente, el dominio de transducción de proteínas puede unirse al extremo N-terminal de la IL-10. El dominio de transducción de proteínas puede unirse a IL-10 a través de un enlace covalente. El dominio de transducción de proteínas puede elegirse de cualquiera de las secuencias enumeradas en la Tabla 4h.

Tabla 4h. Dominios de transducción de proteínas ilustrativos

SEQ ID NO	Secuencia de aminoácidos
277	YGRKKRRQRRR;
278	RRQRRTSKLMKR
279	GWTLNSAGYLLGKINLKALAALAKKIL
280	KALAWREAKLAKALAKALAKHLAKALAKALKCEA
281	RQIKIWFQNRRMKWKK
282	YGRKKRRQRRR
283	RKKRRQRRR
284	YGRKKRRQRRR
285	RKKRRQRR
286	YARAAARQARA
287	THRLPRRRRRR
288	GGRRARRRRRR

Las BP de las composiciones humanas o animales no se limitan a polipéptidos nativos de longitud completa, sino que también incluyen versiones recombinantes, así como también variantes biológica y/o farmacológicamente activas o fragmentos de estos. Por ejemplo, el trabajador experto apreciará que pueden realizarse varias sustituciones de aminoácidos en

la BP para crear variantes sin apartarse del espíritu de la invención con respecto a la actividad biológica o las propiedades farmacológicas de la BP. Ejemplos de sustituciones conservadoras de aminoácidos en secuencias polipeptídicas se muestran en la Tabla 5. Sin embargo, en modalidades de BPXTEN en las que la identidad de secuencia de la BP es inferior al 100 % en comparación con una secuencia específica descrita en la presente descripción, la invención contempla la sustitución de cualquiera de los otros 19 L-aminoácidos naturales por un residuo de aminoácido dado de la BP dada, que puede estar en cualquier posición dentro de la secuencia de la BP, que incluye los residuos de aminoácidos adyacentes. Si alguna sustitución particular da como resultado un cambio no deseado en la actividad biológica, entonces puede emplearse un aminoácido alternativo y evaluar la construcción mediante los métodos descritos en la presente descripción, o mediante el uso de cualquiera de las técnicas y pautas para mutaciones conservativas y no conservativas expuestas, por ejemplo, en la patente de Estados Unidos núm. 5,364,934, cuyo contenido se incorpora por referencia en su totalidad, o mediante el uso de métodos generalmente conocidos por los expertos en la técnica. Además, las variantes también pueden incluir, por ejemplo, polipéptidos en donde se añaden o eliminan uno o más residuos de aminoácidos en el extremo N- o C-terminal de la secuencia de aminoácidos nativa de longitud completa de una BP que retiene al menos una porción de la actividad biológica del péptido nativo.

Tabla 5: Sustituciones de aminoácidos conservativas ilustrativas

Residuo original	Sustituciones ilustrativas
Ala (A)	val; leu; ile
Arg (R)	lys; gln; asn
Asn (N)	gln; his; lys; arg
Asp (D)	glu
Cys (C)	ser
Gln (Q)	asn
Glu (E)	asp
Gly (G)	pro
His (H)	asn; gln; lys; arg
Ile (I)	leu; val; met; ala; phe; norleucina
Leu (L)	norleucina; ile; val; met; ala; phe
Lys (K)	arg; gln; asn
Met (M)	leu; phe; ile
Phe (F)	leu; val; ile; ala
Pro (P)	gly
Ser (S)	thr
Thr (T)	ser
Trp (W)	tyr
Tyr (Y)	trp; phe; thr; ser
Val (V)	ile; leu; met; phe; ala; norleucina

En algunas modalidades, una BP incorporada en un polipéptido BPXTEN puede tener una secuencia que exhibe al menos aproximadamente 80 % de identidad de secuencia con una secuencia de las Tablas 4a-4h, alternativamente al menos aproximadamente 81 %, aproximadamente 82 %, aproximadamente 83 %, aproximadamente 84 %, aproximadamente 85 %, aproximadamente 86 %, aproximadamente 87 %, aproximadamente 88 %, aproximadamente 89 %, aproximadamente 90 %, aproximadamente 91 %, aproximadamente 92 %, aproximadamente 93 %, aproximadamente 94 %, aproximadamente 95 %, aproximadamente 96 %, aproximadamente 97 %, aproximadamente 98 %, aproximadamente 99 %, o aproximadamente 100 % de identidad de secuencia con una secuencia de las Tablas 4a-4h.

aproximadamente 85 %, aproximadamente 86 %, aproximadamente 87 %, aproximadamente 88 %, aproximadamente 89 %, aproximadamente 90 %, o aproximadamente 91 %, o aproximadamente 92 %, o aproximadamente 93 %, o aproximadamente 94 %, o aproximadamente 95 %, o aproximadamente 96 %, o aproximadamente 97 %, o aproximadamente 98 %, o aproximadamente 99 %, o 100 % de identidad de secuencia en comparación con una secuencia de las Tablas 4a-4h. En algunas modalidades, una BP incorporada en una BPXTEN puede ser una secuencia biespecífica que comprende un primer dominio de unión y un segundo dominio de unión, en donde el primer dominio de unión, que tiene afinidad de unión específica por un marcador específico de tumor o un antígeno de una célula objetivo, exhibe al menos aproximadamente 80 % de identidad de secuencia, o alternativamente 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con secuencias VL y VH pareadas de un anticuerpo anti-CD3 seleccionado de la Tabla 6f; y en donde el segundo dominio de unión, que tiene una afinidad de unión específica por una célula efectora, exhibe al menos aproximadamente 80 % de identidad de secuencia, o alternativamente 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con secuencias VL y VH pareadas de un anticuerpo anti-célula objetivo seleccionado de la Tabla 6a. La actividad de la BP de las modalidades anteriores puede evaluarse mediante el uso de ensayos o parámetros medidos o determinados como se describe en la presente descripción, y aquellas secuencias que retienen al menos aproximadamente 40 %, aproximadamente 50 %, aproximadamente 55 %, aproximadamente 60 %, o aproximadamente 70 %, o aproximadamente 80 %, o aproximadamente 90 %, o aproximadamente 95 % o más actividad en comparación con la secuencia de BP nativa correspondiente se consideraría adecuada para su inclusión en la BPXTEN humana o animal. La BP que se encuentra que retiene un nivel adecuado de actividad puede unirse a uno o más polipéptidos XTEN descritos anteriormente o en cualquier otra parte en la presente descripción. En una modalidad, una BP que retiene un nivel adecuado de actividad puede unirse a uno o más polipéptidos XTEN, que tienen al menos aproximadamente 80 % de identidad de secuencia (por ejemplo, al menos aproximadamente 81 %, al menos aproximadamente 82 %, al menos aproximadamente 83 %, al menos aproximadamente 84 %, al menos aproximadamente 85 %, al menos aproximadamente 86 %, al menos aproximadamente 87 %, al menos aproximadamente 88 %, al menos aproximadamente 89 %, al menos aproximadamente 90 %, al menos aproximadamente 91 %, al menos aproximadamente 92 %, al menos aproximadamente 93 %, al menos aproximadamente 94 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 % o 100 % de identidad de secuencia) con una secuencia de las Tablas 3a-3b, lo que da como resultado una proteína de fusión quimérica.

Acopladores de linfocitos T

Fórmulas de configuración estructural adicionales de BPXTEN se relacionan con acopladores de linfocitos T activados por proteasa xtenilados (“XPAT” o “XPATs”), en donde la BP es un anticuerpo biespecífico (por ejemplo, un acoplador de linfocitos T biespecífico). En algunas modalidades, la composición de XPAT comprende una primera porción que comprende un primer dominio de unión y un segundo dominio de unión, una segunda porción que comprende el segmento de liberación y una tercera porción que comprende el resto voluminoso de XTEN. En algunas modalidades, la composición de XPAT tiene la configuración de Fórmula Ia (representada de extremo N-terminal a C-terminal):

(primera porción)-(segunda porción)-(tercera porción) **(Ia)**

en donde la primera porción es una biespecífica que comprende dos scFv en donde el primer dominio de unión tiene afinidad de unión específica por un marcador específico de tumor o un antígeno de una célula objetivo y el segundo dominio de unión tiene afinidad de unión específica por una célula efectora; la segunda porción comprende un segmento de liberación (RS) que puede escindirse por una proteasa de mamífero (como se describe con más detalle a continuación, la proteasa puede ser específica de un tumor o de un antígeno, de esta manera la activación); y la tercera porción es un resto voluminoso. En la modalidad anterior, los dominios de unión de la primera porción pueden estar en el orden (VL-VH)₁-(VL-VH)₂, en donde "1" y "2" representan el primer y segundo dominios de unión, respectivamente, o (VL-VH)₁-(VH-VL)₂, o (VH-VL)₁-(VL-VH)₂, o (VH-VL)₁-(VH-VL)₂, en donde los dominios de unión pareados se unen por un enlazador polipeptídico (como se describe con más detalle a continuación). En una modalidad, las alternativas para la primera porción VL y VH se identifican en las Tablas 6a-6f; las alternativas para RS se identifican en las secuencias expuestas en las Tablas 8a-8b (como se describe con más detalle a continuación); y las alternativas para el resto voluminoso se identifican en la presente descripción mediante: XTEN; dominio de unión a albúmina; albúmina; dominio de unión a IgG; polipéptidos que consisten en prolina, serina y alanina; ácido graso; dominio Fc; polietilenglicol (PEG), PLGA; e hidroxietilalmidón. Cuando se desee, el resto voluminoso es un XTEN que tiene al menos aproximadamente 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una secuencia identificada por las secuencias expuestas en las Tablas 3a-3b. En las modalidades anteriores, la composición es una proteína de fusión recombinante. En otra modalidad, las porciones se unen por conjugación química.

En otra modalidad, la composición XPAT tiene la configuración de Fórmula IIa (representada de extremo N-terminal a C-terminal):

(tercera porción)-(segunda porción)-(primera porción) **(IIa)**

en donde la primera porción es una biespecífica que comprende dos scFv en donde el primer dominio de unión tiene afinidad de unión específica por un marcador específico de tumor o un antígeno de una célula objetivo y el segundo dominio de unión tiene afinidad de unión específica por una célula efectora; la segunda porción comprende un segmento de

liberación (RS) que puede escindirse por una proteasa de mamífero; y la tercera porción es un resto voluminoso. En la modalidad anterior, los dominios de unión de la primera porción pueden estar en el orden (VL-VH)1-(VL-VH)2, en donde "1" y "2" representan el primer y segundo dominios de unión, respectivamente, o (VL-VH)1-(VH-VL)2, o (VH-VL)1-(VL-VH)2, o (VH-VL)1-(VH-VL)2, en donde los dominios de unión pareados se unen por un enlazador polipeptídico como se describe más adelante en la presente descripción. En una modalidad, las alternativas para la primera parte VL y VH se identifican en las Tablas 6a-6f; las alternativas para RS se identifican en las secuencias expuestas en las Tablas 8a-8b; y las alternativas para el resto voluminoso se identifican en la presente descripción mediante: XTEN; dominio de unión a albúmina; albúmina; dominio de unión a IgG; polipéptidos que consisten en prolina, serina y alanina; ácido graso; y dominio Fc. Cuando se desee, el resto voluminoso es un XTEN que tiene al menos aproximadamente 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una secuencia seleccionada del grupo de secuencias expuestas en las Tablas 3a-3b. En las modalidades anteriores, la composición es una proteína de fusión recombinante. En otra modalidad, las porciones se unen por conjugación química.

En otra modalidad, la composición de XPAT tiene la configuración de Fórmula IIIa (representada de extremo N-terminal a extremo C-terminal):

(quinta porción)-(cuarta porción)-(primera porción)-(segunda porción)-(tercera porción)

(IIIa)

en donde la primera porción es una biespecífica que comprende dos scFv en donde el primer dominio de unión tiene afinidad de unión específica por un marcador específico de tumor o un antígeno de una célula objetivo y el segundo dominio de unión tiene afinidad de unión específica por una célula efectora; la segunda porción comprende un segmento de liberación (RS) que puede escindirse por una proteasa de mamífero; la tercera porción es un resto voluminoso; la cuarta porción comprende un segmento de liberación (RS) que puede escindirse por una proteasa de mamífero que puede ser idéntica o diferente a la segunda porción; y la quinta porción es un resto voluminoso que puede ser idéntico o diferente a la tercera porción. En la modalidad anterior, los dominios de unión de la primera porción pueden estar en el orden (VL-VH)1-(VL-VH)2, en donde "1" y "2" representan el primer y segundo dominios de unión, respectivamente, o (VL-VH)1-(VH-VL)2, o (VH-VL)1-(VL-VH)2, o (VH-VL)1-(VH-VL)2, en donde los dominios de unión pareados se unen por un enlazador polipeptídico como se describe más adelante en la presente descripción. En las modalidades anteriores, las alternativas para RS se identifican en las secuencias expuestas en las Tablas 8a-8b. En las modalidades anteriores, las alternativas para el resto voluminoso se identifican en la presente descripción mediante: XTEN; dominio de unión a albúmina; albúmina; dominio de unión a IgG; polipéptidos que consisten en prolina, serina y alanina; ácido graso; y dominio Fc. Cuando se desee, el resto voluminoso es un XTEN que tiene al menos aproximadamente 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una secuencia seleccionada del grupo de secuencias

expuestas en las Tablas 3a-3b. En las modalidades anteriores, la composición es una proteína de fusión recombinante. En otra modalidad, las porciones se unen por conjugación química.

Las composiciones humanas o animales, basadas en su diseño y componentes específicos, proporcionan ventajosamente agentes terapéuticos biespecíficos que tienen más selectividad, mayor semivida y dan como resultado menos toxicidad y menos efectos secundarios una vez que se escinden por proteasas que se encuentran asociadas con los tejidos objetivo o tejidos afectados por una enfermedad, en donde las composiciones humanas o animales tienen un índice terapéutico mejorado en comparación con las composiciones de anticuerpos biespecíficos conocidas en la técnica. Tales composiciones son útiles en el tratamiento de ciertas enfermedades, que incluyen, pero no se limitan al cáncer, como se expone en la presente descripción. Sin limitarse a ninguna teoría mecanicista, el experto apreciará que las composiciones de la presente invención logran esta reducción en las interacciones inespecíficas mediante una combinación de mecanismos, que incluyen el impedimento estérico al ubicar los dominios de unión en las voluminosas moléculas de XTEN, en donde las características flexibles y no estructuradas de los polipéptidos XTEN, al unirse a la composición, pueden oscilar y moverse alrededor de los dominios de unión, lo que proporciona un bloqueo entre la composición y los tejidos o las células, así como también proporciona una reducción en la capacidad de la composición intacta para penetrar en una célula o tejido debido a la gran masa molecular (a la que contribuye tanto el peso molecular real del(de los) polipéptido(s) XTEN como debido al gran radio hidrodinámico de los polipéptidos XTEN no estructurados) en comparación con el tamaño de los dominios de unión individuales. Sin embargo, se diseñan composiciones en donde, cuando se encuentran en la proximidad de un tejido o célula objetivo que contiene o secreta una proteasa capaz de escindir el RS, o cuando se internalizan en una célula o tejido objetivo cuando un dominio de unión se ha unido al ligando, los dominios de unión biespecíficos se liberan del volumen del XTEN por la acción de la(s) proteasa(s), lo que elimina la barrera de impedimento estérico, y es más libre para ejercer su efecto farmacológico. Las composiciones humanas o animales encuentran uso en el tratamiento de una variedad de afecciones en las que se desea el suministro selectivo de una composición de anticuerpo biespecífico terapéutico a una célula, tejido u órgano. En una modalidad, el tejido objetivo es un cáncer, que puede ser una leucemia, un linfoma o un tumor de un órgano o sistema.

Dominios de unión

La descripción contempla el uso de dominios de unión de cadena sencilla, tales como, pero sin limitarse a, Fv, Fab, Fab', Fab'-SH, F(ab')₂, anticuerpos lineales, anticuerpos de un solo dominio, anticuerpos de camélidos de un solo dominio, moléculas de anticuerpos de cadena sencilla (scFv) y diacuerpos capaces de unirse a ligandos o receptores asociados con células efectoras y antígenos de tejidos o células enfermas que son cánceres, tumores u otros tejidos malignos. En algunas modalidades, el anticuerpo biespecífico comprende un primer

dominio de unión con especificidad de unión por un marcador de células objetivo y un segundo dominio de unión con especificidad de unión por un antígeno de célula efectora. En algunas modalidades, el primer y segundo dominios de unión pueden ser andamios que no son de anticuerpos tales como anticalinas, adnectinas, finómeros, afilinas, aficuerpos, centirinas, DARPins. En otras modalidades, el dominio de unión para el objetivo en la célula tumoral es un dominio variable de un receptor de linfocitos T que se ha modificado genéticamente para unirse al MHC que está cargado con un fragmento peptídico de una proteína sobreexpresada en las células tumorales. En algunas modalidades, las composiciones de XPAT se diseñan teniendo en cuenta la ubicación de la proteasa en el tejido objetivo, así como también la presencia de la misma proteasa en tejidos sanos no previstos como objetivos, así como también la presencia del ligando objetivo en tejido sano pero una mayor presencia del ligando en el tejido objetivo no sano, para proporcionar una amplia ventana terapéutica. Una “ventana terapéutica” se refiere a la mayor diferencia entre la dosis efectiva mínima y la dosis máxima tolerada para una composición terapéutica dada. Para ayudar a lograr una amplia ventana terapéutica, los dominios de unión de la primera porción de las composiciones están bloqueados por la proximidad del resto voluminoso (por ejemplo, un polipéptido XTEN), en donde la afinidad de unión de la composición intacta por uno o ambos ligandos se reduce en comparación con la composición que se ha escindido por una proteasa de mamífero, para liberar de esta manera la primera porción de los efectos de bloqueo del resto voluminoso.

Con respecto a los dominios de unión de cadena sencilla, como está bien establecido en la técnica, Fv es el fragmento de anticuerpo mínimo que contiene un sitio de unión y reconocimiento de antígeno completo, que consiste en un dímero de un dominio variable de cadena pesada (VH) y uno ligero (VL) en asociación no covalente. Dentro de cada cadena VH y VL hay tres regiones determinantes de la complementariedad (CDR) que interactúan para definir un sitio de unión a antígeno en la superficie del dímero VH-VL; las seis CDR de un dominio de unión confieren especificidad de unión a antígeno al anticuerpo o dominio de unión de cadena sencilla. En algunos casos, se crean scFv en los que cada uno tiene 3, 4 o 5 CHR dentro de cada dominio de unión. Las secuencias marco que flanquean las CDR tienen una estructura terciaria que se conserva esencialmente en las inmunoglobulinas nativas entre especies, y los residuos marco (FR) sirven para mantener las CDR en su orientación adecuada. Los dominios constantes no son necesarios para la función de unión, pero pueden ayudar a estabilizar la interacción VH-VL. En algunas modalidades, el dominio del sitio de unión del polipéptido puede ser un par de dominios VH-VL, VH-VH o VL-VL de la misma o de diferentes inmunoglobulinas, sin embargo, generalmente se prefiere producir dominios de unión de cadena sencilla mediante el uso de las respectivas cadenas VH y VL del anticuerpo original. El orden de los dominios VH y VL dentro de la cadena polipeptídica no limita la presente invención; el orden de los dominios dados puede invertirse generalmente sin ninguna pérdida de función, pero se entiende que los dominios VH y VL están dispuestos de modo que el sitio de unión a

antígeno pueda plegarse adecuadamente. Por lo tanto, los dominios de unión de cadena sencilla de las modalidades de scFv biespecíficos de las composiciones humanas o animales pueden estar en el orden (VL-VH)¹-(VL-VH)², en donde "1" y "2" representan el primer y segundo dominio de unión, respectivamente, o (VL-VH)¹-(VH-VL)², o (VH-VL)¹-(VL-VH)², o (VH-VL)¹-(VH-VL)², en donde los dominios de unión pareados se unen por un enlazador polipeptídico como se describe más adelante en la presente descripción.

La disposición de los dominios de unión en un anticuerpo de cadena sencilla biespecífico ilustrativo descrito en la presente descripción puede ser, por lo tanto, una en la que el primer dominio de unión esté ubicado en el extremo C-terminal del segundo dominio de unión. La disposición de las cadenas V puede ser VH (antígeno de superficie de célula objetivo)-VL(antígeno de superficie de célula objetivo)-VL(antígeno de célula efectora)-VH(antígeno de célula efectora), VH(antígeno de superficie de célula objetivo)-VL(antígeno de la célula objetivo) antígeno de superficie)-VH(antígeno de célula efectora)-VL(antígeno de célula efectora), VL(antígeno de superficie de célula objetivo)-VH(antígeno de superficie de célula objetivo)-VL(antígeno de célula efectora)-VH(antígeno de superficie de célula objetivo)-VL(antígeno de célula efectora)-VH(antígeno de célula efectora) o VL(antígeno de superficie de célula objetivo)-VH(antígeno de superficie de célula objetivo)-VH(antígeno de célula efectora)-VL(antígeno de célula efectora). Para una disposición, en la que el segundo dominio de unión se ubica en el extremo N-terminal del primer dominio de unión, son posibles los siguientes órdenes: VH (antígeno de célula efectora)-VL(antígeno de célula efectora)-VL(antígeno de superficie de célula objetivo)-VH(antígeno de superficie de célula objetivo), VH(antígeno de célula efectora)-VL(antígeno de célula efectora)-VH(antígeno de superficie de célula objetivo)-VL(antígeno de superficie de célula objetivo), VL(antígeno de célula efectora)-VH(antígeno de célula efectora)-VL(antígeno de superficie de célula objetivo)-VH(antígeno de superficie de célula objetivo) o VL(antígeno de célula efectora)-VH(antígeno de célula efectora)-VH (antígeno de superficie de célula objetivo)-VL(antígeno de superficie de célula objetivo). Como se usa en la presente descripción, "en el extremo N-terminal" o "en el extremo C-terminal" y sus variantes gramaticales indican una ubicación relativa dentro de la secuencia de aminoácidos primaria en lugar de una ubicación en el extremo N- o C-terminal absoluto del anticuerpo de cadena sencilla biespecífico. Por lo tanto, como ejemplo no limitante, un primer dominio de unión que está "ubicado en el extremo C-terminal del segundo dominio de unión" indica que la primera unión se ubica en el lado carboxilo del segundo dominio de unión dentro del anticuerpo de cadena sencilla biespecífico, y no excluye la posibilidad de que una secuencia adicional, por ejemplo, una etiqueta de His u otro compuesto, tal como un radioisótopo, se ubique en el extremo C-terminal del anticuerpo de cadena sencilla biespecífico.

En una modalidad, las composiciones de ensamblaje de polipéptidos quiméricos comprenden una primera porción que comprende un primer dominio de unión y un segundo dominio de unión en donde cada uno de dichos dominios de unión es un scFv y en donde cada scFv comprende un VL y un VH. En otra modalidad, las composiciones de ensamblaje

de polipéptidos quiméricos comprenden una primera porción que comprende un primer dominio de unión y un segundo dominio de unión en donde dichos dominios de unión están en una configuración de diacuerpo y en donde cada dominio comprende un dominio VL y un VH. En las modalidades anteriores, el primer dominio tiene especificidad de unión por un marcador específico de tumor o un antígeno de una célula objetivo y el segundo dominio de unión tiene especificidad de unión por un antígeno de célula efectora. En una modalidad de lo anterior, el antígeno de célula efectora se expresa en o dentro de una célula efectora. En una modalidad, el antígeno de célula efectora se expresa en un linfocito T, tal como una célula CD4+, CD8+ o asesina natural (NK). En otra modalidad, el antígeno de célula efectora se expresa en un linfocito B, un mastocito, una célula dendrítica o una célula mieloide. En una modalidad, el antígeno de célula efectora es CD3, el antígeno de conglomerado de diferenciación 3 de un linfocito T citotóxico. En algunas modalidades de lo anterior, el primer dominio de unión exhibe especificidad de unión por un marcador específico de tumor asociado con una célula tumoral. En una modalidad, el dominio de unión tiene afinidad de unión por un marcador específico de tumor en donde la célula tumoral puede incluir, sin limitación, células de tumor de células estromales, tumor de fibroblastos, tumor de miofibroblastos, tumor de células gliales, tumor de células epiteliales, tumor de células grasas, tumor de células inmunitarias, tumor de células vasculares y tumor de células de músculo liso. En una modalidad, el marcador específico de tumor o un antígeno de una célula objetivo puede ser alfa 4 integrina, Ang2, B7-H3, B7-H6, CEACAM5, cMET, CTLA4, FOLR1, EpCAM, CCR5, CD19, HER2, HER2 neu, HER3, HER4, HER1 (EGFR), PD-L1, PSMA, CEA, TROP-2, MUC1(mucina), MUC-2, MUC3, MUC4, MUC5AC, MUC5B, MUC7, MUC16 β hCG, Lewis-Y, CD20, CD33, CD38, CD30, CD56 (NCAM), CD133, gangliósido GD3; 9-O-Acetil-GD3, GM2, Globo H, fucosil GM1, GD2, anhidrasa carbónica IX, CD44v6, Nectina 4, Sonic Hedgehog (Shh), Wue-1, antígeno de células plasmáticas 1, proteoglicano de sulfato de condroitina de melanoma (MCSP), CCR8, antígeno epitelial 6-transmembrana de próstata (STEAP), mesotelina, antígeno A33, antígeno de células madre de próstata (PSCA), Ly-6, desmogleína 4, receptor de acetilcolina fetal (fnAChR), CD25, antígeno de cáncer 19-9 (CA19-9), antígeno de cáncer 125 (CA-125), receptor de sustancia inhibidora de Mueller tipo II (MISIIR), antígeno Tn sialilado (s TN), antígeno de activación de fibroblastos (FAP), endosialina (CD248), variante III del receptor del factor de crecimiento epidérmico (EGFRvIII), antígeno asociado a tumor L6 (TAL6), SAS, CD63, TAG72, antígeno de Thomsen-Friedenreich (antígeno TF), receptor del factor I de crecimiento similar a la insulina (IGF-IR), antígeno Cora, CD7, CD22, CD70, CD79a, CD79b, G250, MT-MMP, antígeno F19, CA19-9, CA-125, alfa-fetoproteína (AFP), VEGFR1, VEGFR2, DLK1, SP17, ROR1 y EphA2. En una modalidad, el primer dominio de unión que exhibe afinidad de unión a CD70 es su ligando natural, CD27 en lugar de un fragmento de anticuerpo. En otra modalidad, el primer dominio de unión que exhibe afinidad de unión por B7-H6 es su ligando natural Nkp30 en lugar de un fragmento de anticuerpo.

Las modalidades de scFv de las composiciones de XPAT de la invención comprenden un primer dominio de unión y un segundo dominio de unión en donde los dominios VL y VH se derivan de anticuerpos monoclonales con especificidad de unión por el marcador específico de tumor o un antígeno de una célula objetivo y antígenos de células efectoras, respectivamente. En otros casos, cada uno del primer y segundo dominios de unión comprende seis CDR derivadas de anticuerpos monoclonales con especificidad de unión por un marcador de células objetivo, tal como un marcador específico de tumor y antígenos de células efectoras, respectivamente. En otras modalidades, el primer y segundo dominios de unión de la primera porción de las composiciones humanas o animales pueden tener 3, 4 o 5 CHR dentro de cada dominio de unión. En otras modalidades, las modalidades de la invención comprenden un primer dominio de unión y un segundo dominio de unión en donde cada uno comprende una región CDR-H1, una región CDR-H2, una región CDR-H3, una región CDR-L1, una región CDR-L2 y una región CDR-H3, en donde cada una de dichas regiones se deriva de un anticuerpo monoclonal capaz de unirse al marcador específico de tumor o un antígeno de una célula objetivo y antígenos de células efectoras, respectivamente. En una modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos en donde el segundo dominio de unión comprende regiones VH y VL derivadas de un anticuerpo monoclonal capaz de unirse a CD3 humano. En otra modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos, en donde el segundo dominio de unión de scFv comprende regiones VH y VL en donde cada una de las regiones VH y VL exhibe al menos aproximadamente 90 %, o 91 %, o 92 %, o 93 %, o 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad o es idéntica a las secuencias VL y VH pareadas de un anticuerpo anti-CD3 expuesto en la Tabla 6a. En otro aspecto, las modalidades del segundo dominio de la invención comprenden una región CDR-H1, una región CDR-H2, una región CDR-H3, una región CDR-L1, una región CDR-L2 y una región CDR-H3, en donde cada una de dichas regiones se deriva de un anticuerpo monoclonal como se expone en la Tabla 6a. En las modalidades anteriores, los dominios VH y/o VL pueden configurarse como scFv, diacuerpos, un anticuerpo de dominio único o un anticuerpo de camélido de dominio único. En otras modalidades, los segundos dominios de las composiciones humanas o animales se derivan de un anticuerpo anti-CD3 como se expone en la Tabla 6a. En una modalidad de lo anterior, el segundo dominio de la composición humana o animal comprende las secuencias pareadas de la región VL y VH del anticuerpo anti-CD3 como se expone en la Tabla 6a. En otra modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos, en donde el segundo dominio de unión comprende regiones VH y VL en donde cada región VH y VL exhibe al menos aproximadamente 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad con o es idéntica a las secuencias VL y VH pareadas del anticuerpo anti-CD3 huUCHT1 de la Tabla 6a. En las modalidades anteriores, los dominios VH y/o VL pueden configurarse como scFv, una porción de un diacuerpo, un anticuerpo de dominio único o un anticuerpo de camélido de dominio único.

En otras modalidades, los scFv del primer dominio de la composición se derivan de un anticuerpo anti-células tumorales como se expone en la Tabla 6f. En otra modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos, en donde el primer dominio de unión comprende regiones VH y VL en donde cada región VH y VL exhibe al menos aproximadamente 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad o es idéntico a las secuencias VL y VH pareadas de un anticuerpo anti-células tumorales como se expone en la Tabla 6f. En una modalidad de lo anterior, el primer dominio de las composiciones mencionadas comprende las secuencias pareadas de las regiones VL y VH de un anticuerpo anti-células tumorales descrito en la presente descripción. En las modalidades anteriores, los dominios VH y/o VL pueden configurarse como scFv, una porción de un diacuerpo, un anticuerpo de dominio único o un anticuerpo de camélido de dominio único.

En otra modalidad, las composiciones de ensamblaje de polipéptidos quiméricos comprenden una primera porción que comprende un primer dominio de unión y un segundo dominio de unión en donde dichos dominios de unión están en una configuración de diacuerpo y cada uno de dichos dominios de unión comprende un dominio VL y un dominio VH. En una modalidad, las modalidades de diacuerpos de la invención comprenden un primer dominio de unión y un segundo dominio de unión en donde los dominios VL y VH se derivan de anticuerpos monoclonales con especificidad de unión por un marcador específico de tumor o un antígeno de una célula objetivo, y el antígeno de célula efectora, respectivamente. En otra modalidad, las modalidades de diacuerpos de la invención comprenden un primer dominio de unión y un segundo dominio de unión en donde cada uno comprende una región CDR-H1, una región CDR-H2, una región CDR-H3, una región CDR-L1, una región CDR-L2 y una región CDR-H3, en donde cada una de dichas regiones se deriva de un anticuerpo monoclonal capaz de unirse al marcador específico de tumor o al antígeno de la célula objetivo y al antígeno de célula efectora, respectivamente. Se prevé que las modalidades de diacuerpos de la invención comprendan un primer dominio de unión y un segundo dominio de unión en donde los dominios VL y VH se derivan de anticuerpos monoclonales con especificidad de unión por el marcador específico de tumor o antígeno de la célula objetivo, y el antígeno de célula efectora, respectivamente. En otro aspecto, las modalidades de diacuerpos de la invención comprenden un primer dominio de unión y un segundo dominio de unión en donde cada uno comprende una región CDR-H1, una región CDR-H2, una región CDR-H3, una región CDR-L1, una región CDR-L2 y una región CDR-H3, en donde cada una de dichas regiones se deriva de un anticuerpo monoclonal capaz de unirse al marcador específico de tumor o al antígeno de la célula objetivo, y al antígeno de célula efectora, respectivamente. En una modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos en donde el segundo dominio de unión del diacuerpo comprende las regiones VH y VL pareadas derivadas de un anticuerpo monoclonal capaz de unirse a CD3 humano. En otra modalidad, la invención proporciona una composición de ensamblaje de

polipéptidos quiméricos, en donde el segundo dominio de unión del diacuerpo comprende regiones VH y VL, en donde cada región VH y VL exhibe al menos aproximadamente 90 %, 91 %, 92 %, 93 % o 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad o es idéntica a las secuencias VL y VH pareadas de un anticuerpo anti-CD3 como se expone en la Tabla 6a. En otra modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos, en donde el segundo dominio de unión del diacuerpo comprende regiones VH y VL, en donde cada una de las regiones VH y VL exhibe al menos aproximadamente 90 %, 91 %, 92 %, 93 % o 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad o es idéntica a la secuencia VL y VH del anticuerpo huUCHT1 como se expone en la Tabla 6a. En otras modalidades, el segundo dominio de diacuerpo de la composición se deriva de un anticuerpo anti-CD3 descrito en la presente descripción. En otra modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos, en donde el primer dominio de unión del diacuerpo comprende regiones VH y VL, en donde cada región VH y VL exhibe al menos aproximadamente 90 %, 91 %, 92 %, 93 % o 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad o es idéntica a las secuencias VL y VH de un anticuerpo anti-célula tumoral como se expone en la Tabla 6f. En otras modalidades, el primer dominio de diacuerpo de la composición se deriva de un anticuerpo anti-célula tumoral descrito en la presente descripción.

Se conocen en la técnica anticuerpos monoclonales terapéuticos a partir de los cuales pueden derivarse los dominios VL y VH y CDR para las composiciones humanas o animales. Las secuencias de los anticuerpos anteriores pueden obtenerse de bases de datos, patentes o referencias bibliográficas disponibles públicamente. Además, los ejemplos no limitantes de anticuerpos monoclonales y secuencias VH y VL de anticuerpos anti-CD3 expuestos en la Tabla 6a y ejemplos no limitantes de anticuerpos monoclonales y secuencias de VH y VL de marcadores de cáncer, células tumorales u objetivo expuestas en la Tabla 6f.

Dominios de unión anti-CD3

En algunas modalidades, la invención proporciona composiciones de ensamblaje de polipéptidos quiméricos que comprenden un dominio de unión de la primera porción con afinidad de unión a linfocitos T. En una modalidad, el dominio de unión de la segunda porción comprende VL y VH derivados de un anticuerpo monoclonal contra un antígeno del CD3. En otra modalidad, el dominio de unión comprende VL y VH derivados de un anticuerpo monoclonal contra CD3épsilon y CD3delta. Los anticuerpos monoclonales contra CD3 neu se conocen en la técnica. Los ejemplos no limitantes ilustrativos de secuencias VL y VH de anticuerpos monoclonales contra CD3 se exponen en la Tabla 6a. En una modalidad, la invención proporciona un ensamblaje de polipéptidos quiméricos que comprende un dominio de unión con afinidad de unión a CD3 que comprende secuencias VL y VH anti-CD3 expuestas en la Tabla 6a. En otra modalidad, la invención proporciona un ensamblaje de polipéptidos quiméricos que comprende un dominio de unión de la primera porción con afinidad de unión a CD3épsilon que comprende secuencias VL y VH

anti-CD3épsilon expuestas en la Tabla 6a. En otra modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos, en donde el segundo dominio de unión de scFv de la primera porción comprende regiones VH y VL en donde cada una de las regiones VH y VL exhibe al menos aproximadamente 90 %, o 91 %, o 92 % o 93 %, o 94 %, o 95 %, o 96 %, o 97 %, o 98 % o 99 % de identidad o es idéntica a las secuencias VL y VH pareadas del anticuerpo anti-CD3 huUCHT1 de la Tabla 6a. En otra modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos que comprende un dominio de unión con afinidad de unión a CD3 que comprende la región CDR-L1, la región CDR-L2, la región CDR-L3, la región CDR-H1, la región CDR-H2 y la región CDR-H3, en donde cada una se deriva de las respectivas secuencias VL y VH anti-CD3 expuestas en la Tabla 6a. En otra modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos que comprende un dominio de unión con afinidad de unión por CD3 que comprende la región CDR-L1, la región CDR-L2, la región CDR-L3, la región CDR-H1, la región CDR-H2 y la región CDR-H3, en donde las secuencias CDR son RASQDIRNYLN (SEQ ID NO: 8034), YTSRLES (SEQ ID NO: 8035),QQGNTLPWT (SEQ ID NO: 8036), GYSFTGYTMN (SEQ ID NO: 8037), LINPYKGVST (SEQ ID NO: 8038) y SGYYGDSDWYFDV (SEQ ID NO: 8039)

El complejo CD3 es un grupo de moléculas de superficie celular que se asocia con el receptor de antígeno de linfocitos T (TCR) y funciona en la expresión de TCR en la superficie celular y en la cascada de transducción de señalización que se origina cuando un péptido:ligando MHC se une al TCR. Típicamente, cuando un antígeno se une al receptor de linfocitos T, el CD3 envía señales a través de la membrana celular al citoplasma dentro del linfocito T. Esto provoca la activación de los linfocitos T que se dividen rápidamente para producir nuevos linfocitos T sensibilizados para atacar el antígeno particular al que se expusieron los TCR. El complejo CD3 está compuesto por la molécula CD3épsilon, junto con otros cuatro polipéptidos unidos a la membrana (CD3-gamma, delta, zeta y beta). En seres humanos, CD3-épsilon está codificado por el gen CD3E en el cromosoma 11. Los dominios intracelulares de cada una de las cadenas de CD3 contienen motivos de activación del inmunorreceptor basado en tirosina (ITAM) que sirven como punto de nucleación para la maquinaria de transducción de señales intracelulares tras el acoplamiento del receptor de linfocitos T.

Varias estrategias terapéuticas modulan la inmunidad de los linfocitos T al dirigirse a la señalización de TCR, particularmente los anticuerpos monoclonales (mAb) anti-CD3 humanos que se usan ampliamente en la clínica en regímenes inmunosupresores. El mAb OKT3 de ratón específico contra CD3 fue el primer mAb autorizado para su uso en seres humanos (Sgro, C. Side-effects of a monoclonal antibody, muromonab CD3/orthoclone OKT3: bibliographic review. Toxicology 105:23-29, 1995) y se usa ampliamente en la clínica como agente inmunosupresor en trasplantes (Chatenoud, Clin. Transplant 7:422-430, (1993); Chatenoud, Nat. Rev. Immunol. 3:123-132 (2003); Kumar, Transplant. Proc. 30:1351-1352 (1998)), diabetes tipo 1 y psoriasis. Es importante destacar que los mAb anti-

CD3 pueden inducir la señalización parcial de los linfocitos T y la anergia clonal (Smith, JA, Nonmitogenic Anti-CD3 Monoclonal Antibodies Deliver a Partial T Cell Receptor Signal and Induce Clonal Anergy J. Exp. Med. 185:1413-1422 (1997)). OKT3 se ha descrito en la bibliografía como un mitógeno de linfocitos T, así como también un potente citolítico de linfocitos T (Wong, JT. The mechanism of anti-CD3 monoclonal antibodies. Mediation of cytolysis by inter-T cell bridging. Transplantation 50: 683-689 (1990)). En particular, los estudios de Wong demostraron que al unir los linfocitos T CD3 y las células objetivo, se podía lograr la destrucción del objetivo y que ni la ADCC mediada por FcR ni la fijación del complemento eran necesarias para que el MAB bivalente anti-CD3 lisara las células objetivo.

OKT3 exhibe una actividad mitógena y de destrucción de linfocitos T de una manera dependiente del tiempo; después de la activación temprana de los linfocitos T que conducen a la liberación de citocinas, luego de la administración posterior OKT3 bloquea todas las funciones conocidas de los linfocitos T. Es debido a este bloqueo posterior de la función de los linfocitos T que OKT3 ha encontrado una aplicación tan amplia como inmunosupresor en los regímenes de terapia para la reducción o incluso la supresión del rechazo del tejido de aloinjerto. Otros anticuerpos específicos para la molécula CD3 se describen en Tunnacliffe, Int. Immunol. 1 (1989), 546-50, los documentos WO2005/118635 y WO2007/033230 describen anticuerpos monoclonales anti-CD3 épsilon humano, la patente de Estados Unidos 5,821,337 describe las secuencias VL y VH del Ab monoclonal anti-CD3 murino UCHT1 (muxCD3, Shalaby y otros, J. Exp. Med. 175, 217-225 (1992) y una variante humanizada de este anticuerpo (hu UCHT1), y la solicitud de patente de Estados Unidos 20120034228 describe dominios de unión capaces de unirse a un epítopo de la cadena CD3 épsilon de humano y primate que no es chimpancé.

Tabla 6a: Anticuerpos monoclonales anti-CD3 y secuencias

Nom-bre del clon	Nom-bre del anti-cuer-po	Ob-je-tivo	Secuen-cia de VH	SEQ ID NO: (longi-tud comple-ta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi-tud com-pleta)	SEQ ID NO: (CDR)
huOK T3		CD3	QVQLV QSGGG VVQPG RSLRLS CKASG <u>YTFTR</u> <u>YTMH</u> WVRQA PGKGL EWIGYI <u>NPSRG</u> <u>YTNYN</u> <u>OKVKD</u> RFTISR DNSKN TAFLQ	8040	8900, 8901, 8902	DIQMT QSPSSL SASVG DRVITIT <u>CSASSS</u> <u>VSYMN</u> WYQQT PGKAP KRWIY <u>DTSKL</u> <u>ASGVPS</u> RFSGSG SGTDY TFTISSL QPEDIA TYYCQ	8049	8908, 8909, 8910

Nom- bre del clon	Nom- bre del anti- cuer- po	Ob- je- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud comple- ta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			MDSLR PEDTG VYFCA <u>RYDD</u> <u>HYCLD</u> <u>YWGQG</u> TPVTVS S			<u>QWSSN</u> <u>PFTFGQ</u> GTKLQI TR		
huUC HT1		CD3	EVQLV ESGGG LVQPG GSLRLS CAAS <u>G</u> <u>YSFTG</u> <u>YTMN</u> WVRQA PGKGL EWVAL <u>INPYK</u> <u>GVSTY</u> NQKFK DRFTIS VDKSK NTAYL QMNSL RAEDT AVYYC AR <u>SGY</u> <u>YGDSD</u> <u>WYFDV</u> WGQGT LVTVSS	8041	8037, 8038, 8039	DIQMT QSPSSL SASVG DRVIT <u>CRASQ</u> <u>DIRNY</u> <u>LNWYQ</u> QKPGK APKLLI <u>YTSR</u> <u>LES</u> GVP SRFSGS GSGTD YTLTIS SLQPED FATYY <u>COQGN</u> <u>TLPWT</u> FGQGT KVEIK	8050	8034, 8035, 8036
hu12F 6		CD3	QVQLV QSGGG VVQPG RSLRLS CKAS <u>G</u> <u>YTFTS</u> <u>YTMH</u> WVRQA PGKGL EWIG <u>YI</u> <u>NPSSG</u> <u>YTKYN</u> <u>QKFKD</u> RFTISA DKSKS TAFLQ MDSLR PEDTG VYFCA <u>RWQD</u> <u>YDVYE</u>	8042	8903, 8904, 8905	DIQMT QSPSSL SASVG DRVIT <u>TCRASS</u> <u>SVSYM</u> <u>HWYQQ</u> TPGKAP KPWIY <u>ATSNL</u> <u>ASG</u> VPS RFSGSG SGTDY TLTISSL QPEDIA TYYC <u>Q</u> <u>QWSSN</u> <u>PPTFGQ</u> GTKLQI TR	8051	8911, 8912, 8913

Nom- bre del clon	Nom- bre del anti- cuer- po	Ob- je- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud comple- ta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			DYWGQ GTPVT VSS					
mOK T3		CD3	QVQLQ QSGAE LARPG ASVKM SCKAS <u>GYTFT</u> <u>RYTMH</u> WVKQR PGQGL EWIG <u>YI</u> <u>NPSRG</u> <u>YTNYN</u> <u>QKFKD</u> KATLT TDKSSS TAYMQ LSSLTS EDSAV YYCAR <u>YYDDH</u> <u>YCLDY</u> WGQGT TLTVSS	8043	8900, 8906, 8902	QIVLTQ SPAIMS ASPGEK VTMTC <u>SASSSV</u> <u>SYMNW</u> YQQKS GTSPKR WIY <u>DT</u> <u>SKLAS</u> GVPAH FRGSGS GTSYSL TISGME AEDAA TYYC <u>Q</u> <u>QWSSN</u> <u>PFTFGS</u> GTKLEI NR	8052	8908, 8909, 8910
MT10 3	blinat umo mab	CD3	DIKLQQ SGAEL ARPGA SVKMS CKTSG <u>YTFTR</u> <u>YTMH</u> WVKQR PGQGL EWIG <u>YI</u> <u>NPSRG</u> <u>YTNYN</u> <u>QKFKD</u> KATLT TDKSSS TAYMQ LSSLTS EDSAV YYCAR <u>YYDDH</u> <u>YCLDY</u> WGQGT TLTVSS	8044	8900, 8906, 8902	DIQLTQ SPAIMS ASPGEK VTMTC <u>RASSSV</u> <u>SYMNW</u> YQQKS GTSPKR WIY <u>DT</u> <u>SKVAS</u> GVPYR FSGSGS GTSYSL TISSME AEDAA TYYC <u>Q</u> <u>QWSSN</u> <u>PLTFG</u> AGTKL ELK	8053	8914, 8915, 8916
MT11 0	solito mab	CD3	DVQLV QSGAE VKKPG ASVKV	8045	8900, 8907, 8902	DIVLTQ SPATLS LSPGER ATL <u>SCR</u>	8054	8917, 8915, 8916

Nom- bre del clon	Nom- bre del anti- cuer- po	Ob- je- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud comple- ta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			SCKAS <u>GYTFT</u> <u>RYTMH</u> WVRQA PGQGL EWIG <u>YI</u> <u>NPSRG</u> <u>YTNYA</u> <u>DSVKG</u> RFTITT DKSTST AYMEL SSLRSE DTATY YCARY <u>YDDHY</u> <u>CLDYW</u> GQGTT VTVSS			<u>ASQSVS</u> <u>YMNW</u> YQQKP GKAPK RWIY <u>D</u> <u>TSKVA</u> <u>SGVPA</u> RFSGSG SGTDYS LTINSL EAEDA ATYYC <u>QQWSS</u> <u>NPLTF</u> GGGTK VEIK		
CD3.7		CD3	EVQLV ESGGG LVQPG GSLKLS CAASG FTFNK YAMN WVRQA PGKGL EWVAR IRSKYN NYATY YADSV KDRFTI SRDDS KNTAY LQMNN LKTED TAVYY CVRHG NFGNS YISYW AYWGQ GTLVT VSS	8046		QTVVT QEPSLT VSPGGT VTLTC GSSTGA VTSGY YPNWV QQKPG QAPRG LIGGTK FLAPGT PARFSG SLLGG KAALT LSGVQP EDEAE YYCAL WYSNR WVFGG GTKLT VL	8055	
CD3.8		CD3	EVQLV ESGGG LVQPG GSLRLS CAASG FTFNTY AMNW VRQAP GKGLE	8047		QAVVT QEPSLT VSPGGT VTLTC GSSTGA VTTSN YANWV QQKPG QAPRG	8056	

Nom- bre del clon	Nom- bre del anti- cuer- po	Ob- je- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud comple- ta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			WVGRI RSKYN NYATY YADSV KGRFTI SRDDS KNTLY LQMNS LRAED TAVYY CVRHG NFGNS YVSWF AYWGQ GTLVT VSS			LIGGTN KRAPG VPARFS GSLLG GKAAL TLSGA QPEDE AEYYC ALWYS NLWVF GGGTK LTVL		
CD3.9		CD3	EVQLL ESGGG LVQPG GSLKLS CAASG FTFNTY AMNW VRQAP GKGLE WVARI RSKYN NYATY YADSV KDRFTI SRDDS KNTAY LQMNN LKTED TAVYY CVRHG NFGNS YVSWF AYWGQ GTLVT VSS	773		ELVVT QEPSLT VSPGGT VTLTCR SSTGAV TTSNY ANWVQ QKPGQ APRGLI GGTNK RAPGTP ARFSGS LLGGK AALTLS GVQPE DEAEY YCALW YSNLW VFGGG TKLTVL	772	
CD3.1 0		CD3	EVKLL ESGGG LVQPK GSLKLS CAASG FTFNTY AMNW VRQAP GKGLE WVARI RSKYN NYATY	8048		QAVVT QESALT TSPGET VTLTCR SSTGAV TTSNY ANWVQ EKPDH LFTGLI GGTNK RAPGV PARFSG	8057	

Nom- bre del clon	Nom- bre del anti- cuer- po	Ob- je- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud comple- ta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			YADSV KDRFTI SRDDS QSILYL QMNNL KTEDT AMYYC VRHGN FGNSY VSWFA YWGQG TLVTVS S			SLIGDK AALTIT GAQTE DEAIYF CALWY SNLWV FGGGT KLTVL		

* las secuencias subrayadas, si están presentes, son CDR dentro del VL y VH

Fragmentos de unión a antígeno celular CD3

En otro aspecto, la descripción se refiere a fragmentos de unión a antígeno (AF2) que tienen afinidad de unión específica por un antígeno de célula efectora que puede incorporarse en cualquiera de las modalidades de composiciones humanas o animales descritas en la presente descripción. En algunos casos, el antígeno de célula efectora se expresa en la superficie de una célula efectora seleccionada de una célula plasmática, un linfocito T, un linfocito B, una célula asesina inducida por citocinas (célula CIK), un mastocito, una célula dendrítica, un linfocito T regulador (linfocito RegT), un linfocito T cooperador, una célula mieloide y una célula NK.

Varios AF2 que se unen a antígenos de células efectoras tienen una utilidad particular para el apareamiento con un fragmento de unión a antígeno con afinidad de unión por antígenos EGFR asociados con una célula o tejido enfermos en formatos de composición para efectuar la destrucción celular de la célula o tejido enfermos. La especificidad de unión puede determinarse mediante regiones determinantes de la complementariedad, o CDR, tales como CDR de cadena ligera o CDR de cadena pesada. En muchos casos, la especificidad de unión se determina mediante las CDR de cadena ligera y las CDR de cadena pesada. Una combinación dada de CDR de cadena pesada y CDR de cadena ligera proporciona un bolsillo de unión dado que confiere mayor afinidad y/o especificidad hacia un antígeno de célula efectora en comparación con otros antígenos de referencia. Las composiciones biespecíficas resultantes, que tienen un primer fragmento de unión a antígeno (AF1) contra EGFR, unido por un enlazador peptídico flexible corto a un segundo fragmento de unión a antígeno (AF2) con especificidad de unión para un antígeno de célula efectora son biespecíficas, donde cada fragmento de unión a antígeno tiene afinidad de unión específica para sus respectivos ligandos. El experto entenderá que en tales composiciones, un AF1 dirigido contra un EGFR de un tejido enfermo se usa en

combinación con un AF2 dirigido hacia un marcador de células efectoras para acercar una célula efectora a la célula de un tejido enfermo para efectuar la citólisis de la célula del tejido enfermo. Además, los AF1 y AF2 se incorporan en los polipéptidos específicamente diseñados que comprenden segmentos de liberación escindibles y XTEN para conferir características de profármaco en las composiciones que se activan por la liberación del AF1 y AF2 fusionados tras la escisión de los segmentos de liberación cuando están cerca del tejido enfermo que tiene proteasas que pueden escindir los segmentos de liberación en una o más ubicaciones.

En una modalidad, el AF2 de las composiciones humanas o animales tiene afinidad de unión por un antígeno de célula efectora expresado en la superficie de un linfocito T. En otra modalidad, el AF2 de las composiciones humanas o animales tiene afinidad de unión por CD3. En otra modalidad, el AF2 de las composiciones humanas o animales tiene afinidad de unión por un miembro del complejo CD3, que incluye en forma individual o combinada independientemente todas las subunidades de CD3 conocidas del complejo CD3; por ejemplo, CD3 épsilon, CD3 delta, CD3 gamma, CD3 zeta, CD3 alfa y CD3 beta. En otra modalidad, el AF2 tiene afinidad de unión por CD3 épsilon, CD3 delta, CD3 gamma, CD3 zeta, CD3 alfa o CD3 beta.

El origen de los fragmentos de unión a antígeno contemplados por la descripción puede derivarse de un anticuerpo de origen natural o fragmento de este, un anticuerpo de origen no natural o fragmento de este, un anticuerpo humanizado o fragmento de este, un anticuerpo sintético o fragmento de este, un anticuerpo híbrido o fragmento de este, o un anticuerpo modificado genéticamente o fragmento de este. Los métodos para generar un anticuerpo para un marcador objetivo dado se conocen bien en la técnica. Por ejemplo, los anticuerpos monoclonales pueden producirse mediante el uso del método de hibridomas descrito por primera vez por Kohler y otros, *Nature*, 256:495 (1975), o pueden producirse mediante métodos de ADN recombinante (patente de Estados Unidos núm. 4,816,567). La estructura de anticuerpos y fragmentos de estos, regiones variables de cadenas pesadas y ligeras de un anticuerpo (VH y VL), regiones variables de cadena única (scFv), regiones determinantes de la complementariedad (CDR), y anticuerpos de dominio (dAb) se comprenden bien. Los métodos para generar un polipéptido que tiene un fragmento de unión a antígeno deseado con afinidad de unión por un antígeno dado se conocen en la técnica.

El experto entenderá que el uso de la expresión “fragmento de unión a antígeno” para las modalidades de la composición descritas en la presente descripción pretende incluir porciones o fragmentos de anticuerpos que retienen la capacidad de unirse a los antígenos que son los ligandos del anticuerpo intacto correspondiente. En tales modalidades, el fragmento de unión a antígeno puede ser, pero no se limitan a, CDR y regiones marco de intervención, regiones variables o hipervariables de cadenas ligeras y/o pesadas de un anticuerpo (VL, VH), fragmentos variables (Fv), fragmentos Fab', fragmentos F(ab')₂, fragmentos Fab, anticuerpos de cadena única (scAb), anticuerpos de camélidos contra VHH,

fragmento variable de cadena única (scFv), anticuerpos lineales, un anticuerpo de dominio único, regiones determinantes de la complementariedad (CDR), anticuerpos de dominio (dAb), inmunoglobulinas de cadena pesada de dominio único de tipo BHH o BNAR, inmunoglobulinas de cadena ligera de dominio único, u otros polipéptidos conocidos en la técnica que contienen un fragmento de un anticuerpo que puede unirse a un antígeno. Los fragmentos de unión a antígeno que tienen CDR-H y CDR-L pueden configurarse en una orientación (CDR-H)-(CDR-L) o (CDR-H)-(CDR-L), de extremo N-terminal a extremo C-terminal. La VL y la VH de dos fragmentos de unión a antígeno también pueden configurarse en una configuración de diacuerpo de cadena única; es decir, la VL y la VH del AF1 y el AF2 configuradas con enlazadores de una longitud adecuada para permitir la disposición como un diacuerpo.

Varios AF2 de unión a CD3 de la descripción se han modificado específicamente para mejorar su estabilidad en las modalidades de polipéptidos que se describen en la presente descripción. La agregación proteica de los anticuerpos sigue siendo un problema importante en su capacidad de desarrollo y sigue siendo un área principal de atención en la producción de anticuerpos. La agregación de anticuerpos puede desencadenarse por el desplegamiento parcial de sus dominios, lo que conduce a la asociación monómero-monómero seguida de nucleación y crecimiento agregado. Aunque las propensiones a la agregación de anticuerpos y proteínas basadas en anticuerpos pueden verse afectadas por las condiciones experimentales externas, estas dependen en gran medida de las propiedades intrínsecas del anticuerpo según se determina por sus secuencias y estructuras. Aunque es bien sabido que las proteínas solo son marginalmente estables en sus estados plegados, frecuentemente se aprecia menos que la mayoría de las proteínas son inherentemente propensas a la agregación en sus estados desplegados o parcialmente desplegados, y los agregados resultantes pueden ser extremadamente estables y de larga duración. La reducción de la propensión a la agregación también ha demostrado estar acompañada por un aumento en el título de expresión, lo que demuestra que la reducción de la agregación de proteínas es beneficiosa durante todo el proceso de desarrollo y puede conducir a un camino más eficiente hacia los estudios clínicos. En el caso de las proteínas terapéuticas, los agregados son un factor de riesgo importante para las respuestas inmunitarias nocivas en los pacientes, y pueden formarse a través de una variedad de mecanismos. El control de la agregación puede mejorar la estabilidad de las proteínas, la capacidad de fabricación, las tasas de deserción, la seguridad, la formulación, los títulos, la inmunogenia y la solubilidad. Las propiedades intrínsecas de las proteínas, tales como el tamaño, la hidrofobicidad, la electrostática y la distribución de cargas, desempeñan funciones importantes en la solubilidad de las proteínas. Se ha demostrado que la baja solubilidad de las proteínas terapéuticas debido a la hidrofobicidad superficial dificulta el desarrollo de formulaciones y puede provocar una biodistribución deficiente, un comportamiento farmacocinético no deseado e inmunogenicidad in vivo. Disminuir la hidrofobicidad superficial general de los anticuerpos monoclonales candidatos también puede proporcionar beneficios y ahorros de

costos relacionados con los regímenes de purificación y dosificación. Los aminoácidos individuales pueden identificarse mediante análisis estructural como contribuyentes al potencial de agregación en un anticuerpo, y pueden localizarse en las CDR así como también en las regiones marco. En particular, puede predecirse que los residuos tienen un alto riesgo de causar problemas de hidrofobicidad en un anticuerpo determinado. En una modalidad, la presente descripción proporciona un AF2 que tiene la capacidad de unirse específicamente a CD3 en la que el AF2 tiene al menos una sustitución de aminoácidos de un aminoácido hidrófobo en una región marco con relación al anticuerpo original o fragmento de anticuerpo en donde el aminoácido hidrófobo se selecciona de isoleucina, leucina o metionina. En otra modalidad, el AF2 contra CD3 tiene al menos dos sustituciones de aminoácidos hidrófobos en una o más regiones marco en donde los aminoácidos hidrófobos se seleccionan de isoleucina, leucina o metionina.

Los cambios en la carga eléctrica neta de un polipéptido, particularmente con respecto a los anticuerpos o fragmentos de anticuerpos que comprenden modalidades particulares de la invención expuestas en la presente descripción, se tuvieron en cuenta en el diseño de las secuencias del AF2 de las modalidades descritas en la presente descripción, en donde se hicieron sustituciones de aminoácidos individuales con relación al anticuerpo original utilizado como punto de partida. Relevante para estas consideraciones de diseño es el punto isoelectrico (pI) del polipéptido, que es el pH en el que el anticuerpo o fragmento de anticuerpo no tiene carga eléctrica neta. Un anticuerpo o fragmento de anticuerpo típicamente tiene una carga positiva neta que tiende a correlacionarse con un aumento de la eliminación de la sangre y retención en tejidos, con una semivida generalmente más corta, mientras que una carga negativa neta da como resultado una menor captación por los tejidos y una semivida más larga. Es posible manipular esta carga a través de mutaciones en los residuos del marco. El punto isoelectrico de un polipéptido puede determinarse matemáticamente (por ejemplo, mediante cálculos) o experimentalmente en un ensayo *in vitro*. En algunas modalidades, los puntos isoelectricos del AF1 y el AF2 se diseñan para estar dentro de un intervalo particular entre sí, para promover así la estabilidad.

En una modalidad, la presente descripción proporciona un AF2 para su uso en cualquiera de las modalidades de polipéptidos que se describen en la presente descripción que comprenden CDR-L y CDR-H, en donde el AF2 (a) se une específicamente al grupo de diferenciación 3 (CD3) receptor de linfocitos T; y (b) comprende CDR-H1, CDR-H2 y CDR-H3 que tienen las secuencias de aminoácidos de las SEQ ID NO: 742, 743 y 744, respectivamente. En otra modalidad, la presente descripción proporciona un AF2 para su uso en cualquiera de las modalidades de polipéptidos que se describen en la presente descripción que comprenden CDR-L y CDR-H, en donde el AF2 (a) se une específicamente al grupo de diferenciación 3 (CD3) receptor de linfocitos T; (b) comprende CDR-H1, CDR-H2 y CDR-H3, que tienen secuencias de aminoácidos de las SEQ ID NO: 742, 743 y 744, respectivamente; y (c) comprende CDR-L en donde la CDR-L comprende una CDR-L1 que tiene una secuencia de aminoácidos de las SEQ ID NO: 735 o 736, una CDR-L2 que tiene

una secuencia de aminoácidos de la SEQ ID NO: 738 o 739, y una CDR-L3 que tiene una secuencia de aminoácidos de la SEQ ID NO: 740. En otra modalidad, las modalidades de AF2 del párrafo anterior comprenden además regiones marco de cadena ligera (FR-L) y regiones marco de cadena pesada (FR-H) en donde AF2 comprende una FR-L1 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 746, una FR-L2 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 747, una FR-L3 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de una cualquiera de las SEQ ID NO: 748-751, una FR-L4 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 754, una FR-H1 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 755 o SEQ ID NO: 756, una FR-H2 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 759, una FR-H3 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 760; y una FR-H4 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 764. En otra modalidad, el AF2 para su uso en cualquiera de las modalidades de polipéptidos que se describen en la presente descripción comprende regiones marco de cadena ligera (FR-L) y regiones marco de cadena pesada (FR-H) en donde el AF2 comprende una FR-L1 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 746, una FR-L2 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 747, una FR-L3 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 748, FR-L4 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 754, una FR-H1 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 755, una FR-H2 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de

secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 759, una FR-H3 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 760; y una FR-H4 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 764. En otra modalidad, el AF2 para su uso en cualquiera de las modalidades de polipéptidos que se describen en la presente descripción comprende regiones marco de cadena ligera (FR-L) y regiones marco de cadena pesada (FR-H) en donde el AF2 comprende una FR-L1 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 746, una FR-L2 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 747, una FR-L3 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 749, una FR-L4 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 754, una FR-H1 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 755, una FR-H2 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 759, una FR-H3 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 760; y una FR-H4 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 764. En otra modalidad, el AF2 de las modalidades de polipéptidos humanos o animales descritas en la presente descripción comprende regiones marco de cadena ligera (FR-L) y regiones marco de cadena pesada (FR-H) en donde AF2 comprende una FR-L1 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 746, una FR-L2 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 747, una FR-L3 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 750, una FR-L4 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de

la SEQ ID NO: 754, una FR-H1 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 755, una FR-H2 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 759, una FR-H3 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 760, y una FR-H4 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 764. En otra modalidad, el AF2 de las modalidades de polipéptidos humanos o animales descritas en la presente descripción comprende regiones marco de cadena ligera (FR-L) y regiones marco de cadena pesada (FR-H) en donde AF2 comprende una FR-L1 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 746, una FR-L2 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 747, una FR-L3 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 751, una FR-L4 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 754, una FR-H1 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 756, una FR-H2 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 759, una FR-H3 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 760, y una FR-H4 que exhibe al menos 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a la secuencia de aminoácidos de la SEQ ID NO: 764.

En otra modalidad, la presente descripción proporciona un AF2 para su uso en cualquiera de las modalidades de polipéptidos que se describen en la presente descripción en donde el AF2 comprende una secuencia de aminoácidos pesada variable (VH) que tiene al menos 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a una secuencia de aminoácidos de la SEQ ID NO:766 o SEQ ID NO:769. En otra modalidad, la presente descripción proporciona un AF2 para su uso en cualquiera de las modalidades de polipéptidos que se describen en la presente descripción en donde el AF2 comprende una secuencia de aminoácidos ligera variable (VL) que tiene al menos

90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a una secuencia de aminoácidos de una cualquiera de las SEQ ID NO: 765, 767, 768, 770 o 771. En otra modalidad, la presente descripción proporciona un AF2 para su uso en cualquiera de las modalidades de polipéptidos que se describen en la presente descripción en donde el AF2 comprende una secuencia de aminoácidos pesada variable (VH) que tiene al menos 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a una secuencia de aminoácidos de la SEQ ID NO: 766 o SEQ ID NO: 769 y una secuencia de aminoácidos ligera variable (VL) que tiene al menos 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a una secuencia de aminoácidos de una cualquiera de las SEQ ID NO: 765, 767, 768, 770 o 771.

En otra modalidad, la presente descripción proporciona un AF2 para su uso en cualquiera de las modalidades de polipéptidos descritas en la presente descripción en donde el AF2 comprende una secuencia de aminoácidos que tiene al menos 95 %, 96 %, 97 %, 98 %, 99 % de identidad de secuencia o es idéntica a una secuencia de aminoácidos de una cualquiera de las SEQ ID NO: 776-780.

En otro aspecto, la presente descripción proporciona fragmentos de unión a antígeno AF2 que se unen al complejo proteico CD3 que tienen estabilidad mejorada en comparación con los anticuerpos de unión a CD3 o fragmentos de unión a antígeno conocidos en la técnica. Adicionalmente, los fragmentos de unión a antígeno CD3 de la descripción se diseñan para conferir un mayor grado de estabilidad en las composiciones de fragmentos de unión a antígeno biespecíficos quiméricos en las que se integran, lo que conduce a una mejor expresión y recuperación de la proteína de fusión, mayor vida útil y estabilidad mejorada cuando se administran a un ser humano o animal. En un enfoque, el AF2 contra CD3 de la presente descripción se diseña para tener un mayor grado de estabilidad térmica en comparación con determinados anticuerpos de unión a CD3 y fragmentos de unión a antígeno que se conocen en la técnica. Como resultado, el AF2 contra CD3 que se utiliza como componentes de las composiciones de fragmentos de unión a antígeno biespecíficos quiméricos en las que se integran exhibe propiedades farmacéuticas favorables, que incluyen alta termoestabilidad y baja propensión a la agregación, lo que resulta en una mejor expresión y recuperación durante la fabricación y el almacenamiento, así como también promueve un largo tiempo de semivida en suero. Las propiedades biofísicas tales como la termoestabilidad frecuentemente están limitadas por los dominios variables de anticuerpos, que difieren en gran medida en sus propiedades intrínsecas. La alta estabilidad térmica frecuentemente se asocia con altos niveles de expresión y otras propiedades deseadas, que incluyen ser menos susceptibles a la agregación (Buchanan A, y otros, Engineering a therapeutic IgG molecule to address cysteinylolation, aggregation and enhance thermal stability and expression. MAbs 2013; 5:255). La estabilidad térmica se determina midiendo la “temperatura de fusión” (T_m), que se define como la temperatura a la que la mitad de las moléculas se desnaturalizan. La temperatura de fusión de cada heterodímero es

indicativa de su estabilidad térmica. Los ensayos *in vitro* para determinar la T_m se conocen en la técnica, que incluyen métodos descritos en los Ejemplos, a continuación. El punto de fusión del heterodímero puede medirse mediante el uso de técnicas tales como calorimetría diferencial de barrido (Chen y otros (2003) *Pharm Res* 20:1952-60; Ghirlando y otros (1999) *Immunol Lett* 68:47-52). Alternativamente, la estabilidad térmica del heterodímero puede medirse mediante el uso de dicróismo circular (Murray y otros (2002) *J. Chromatogr Sci* 40:343-9), o como se describe en los ejemplos, más adelante.

Las curvas de desnaturalización térmica de los fragmentos de unión a CD3 y los anticuerpos biespecíficos anti-CD3 que comprenden dicho fragmento de unión anti-CD3 y una unión de referencia de la presente descripción muestran que las construcciones de la presente descripción son más resistentes a la desnaturalización térmica que el fragmento de unión a antígeno que consiste en una secuencia que se muestra en la SEQ ID NO: 781 o un anticuerpo biespecífico de control en donde dicho fragmento de unión a antígeno biespecífico de control comprende la SEQ ID NO: 781 y un fragmento de unión a antígeno de referencia que se une a una modalidad de EGFR descrita en la presente descripción. En una modalidad, los polipéptidos de cualquiera de las modalidades de composiciones humanas o animales descritas en la presente descripción comprenden un AF2 anti-CD3 de las modalidades descritas en la presente descripción, en donde la T_m del AF2 es al menos 2 °C mayor, o al menos 3 °C mayor, o al menos 4 °C mayor, o al menos 5 °C mayor, o al menos 6 °C mayor, o al menos 7 °C mayor, o al menos 8 °C mayor, o al menos 9 °C mayor, o al menos 10 °C mayor que la T_m de un fragmento de unión a antígeno que consiste en una secuencia de la SEQ ID NO: 781, según se determina por un aumento en la temperatura de fusión en un ensayo *in vitro*.

En otra modalidad, los polipéptidos de cualquiera de las modalidades de composiciones humanas o animales descritas en la presente descripción comprenden un AF2 que se une específicamente a CD3 humano o de macaco cangrejero con una constante de disociación (K_d) constante entre aproximadamente 10 nM y aproximadamente 400 nM, o entre aproximadamente 50 nM y aproximadamente 350 nM, o entre aproximadamente 100 nM y 300 nM, según se determina en un ensayo de unión a antígeno *in vitro* que comprende un antígeno CD3 de macaco cangrejero o humano. En otra modalidad, los polipéptidos de cualquiera de las modalidades de composiciones humanas o animales descritas en la presente descripción comprenden un AF2 que se une específicamente a CD3 humano o de macaco cangrejero con una constante de disociación (K_d) más débil que aproximadamente 10 nM, o aproximadamente 50 nM, o aproximadamente 100 nM, o aproximadamente 150 nM, o aproximadamente 200 nM, o aproximadamente 250 nM, o aproximadamente 300 nM, o aproximadamente 350 nM, o más débil que aproximadamente 400 nM según se determina en un ensayo de unión a antígeno *in vitro*. Para mayor claridad, un fragmento de unión a antígeno con una K_d de 400 se une a su ligando más débilmente que uno con una K_d de 10 nM. En otra modalidad, los polipéptidos de cualquiera de las modalidades de composiciones humanas o animales descritas en la presente descripción comprenden un

AF2 que se une específicamente a CD3 humano o de macaco cangrejero con una afinidad de unión al menos 2 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces, o al menos 10 veces más débil que un fragmento de unión a antígeno que consiste en una secuencia de aminoácidos de la SEQ ID NO: 781, según se determina por las constantes de disociación respectivas (K_d) en un ensayo de unión a antígeno in vitro. En otra modalidad, la presente descripción proporciona polipéptidos biespecíficos que comprenden un AF2 que exhibe una afinidad de unión a CD3 que es al menos 2 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces, 10 veces, 20 veces, 50 veces, 100 veces, o al menos 1 000 veces, más débil con relación a la de las modalidades de EGFR AF1 descritas en la presente descripción que se incorporan en los polipéptidos humanos o animales, según se determina por las constantes de disociación respectivas (K_d) en un ensayo de unión a antígeno in vitro. La afinidad de unión de las composiciones humanas o animales por los ligandos objetivo puede analizarse mediante el uso de ensayos de unión o unión competitiva, tales como ensayos Biacore con receptores unidos a chip o proteínas de unión o ensayos ELISA, como se describe en la patente de Estados Unidos 5,534,617, ensayos descritos en los ejemplos en la presente descripción, ensayos de radioreceptor, u otros ensayos conocidos en la técnica. La constante de afinidad de unión puede determinarse después mediante el uso de métodos estándar, tales como análisis de Scatchard, como se describe por van Zoelen, y otros, Trends Pharmacol Sciences (1998) 19(12):487, u otros métodos conocidos en la técnica.

En un aspecto relacionado, la presente descripción proporciona AF2 que se unen a CD3 y se incorporan en composiciones polipeptídicas biespecíficas quiméricas que se diseñan para tener un punto isoeléctrico (pI) que confiere estabilidad mejorada en las composiciones de la descripción en comparación con composiciones correspondientes que comprenden anticuerpos de unión a CD3 o fragmentos de unión a antígeno conocidos en la técnica. En una modalidad, los polipéptidos de cualquiera de las modalidades de composiciones humanas o animales descritas en la presente descripción comprenden AF2 que se unen a CD3 en donde el AF2 exhibe un pI que se encuentra entre 6,0 y 6,6, incluidos. En otra modalidad, los polipéptidos de cualquiera de las modalidades de composiciones humanas o animales descritas en la presente descripción comprenden AF2 que se unen a CD3 en donde el AF2 exhibe un pI que es al menos 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, o 1,0 unidad de pH menor que el pI de un fragmento de unión a antígeno de referencia que consiste en una secuencia que se muestra en la SEQ ID NO: 781. En otra modalidad, los polipéptidos de cualquiera de las modalidades de composiciones humanas o animales descritas en la presente descripción comprenden un AF2 que se une a CD3 fusionado a un AF1 que se une a un antígeno de EGFR en donde el AF2 exhibe un pI que se encuentra dentro de al menos 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9, 1,0, 1,1, 1,2, 1,3, 1,4 o 1,5 unidades de pH del pI del AF1 que se une a un antígeno EGFR o un epítipo de este. En otra modalidad, los polipéptidos de cualquiera de las modalidades de composiciones humanas o animales descritas en la presente descripción comprenden un AF2 que se une a CD3, fusionado a un AF1 que se une a un antígeno EGFR en donde el AF2 exhibe un pI que se encuentra dentro

de al menos aproximadamente 0,1 a aproximadamente 1,5, o al menos aproximadamente 0,3 a aproximadamente 1,2, o al menos aproximadamente 0,5 a aproximadamente 1,0, o al menos aproximadamente 0,7 a aproximadamente 0,9 unidades de pH del pI del AF1. Se pretende específicamente que mediante tal diseño, en donde el pI de los dos fragmentos de unión a antígeno se encuentra dentro de tales intervalos, los fragmentos de unión a antígeno fusionados resultantes confieran un mayor grado de estabilidad en las composiciones de fragmentos de unión a antígeno, biespecíficas y quiméricas en las que se integran, lo que conduce a una mejor expresión y mejor recuperación de la proteína de fusión en forma soluble no agregada, aumento de la vida útil de las composiciones polipeptídicas biespecíficas quiméricas formuladas, y mayor estabilidad cuando la composición se administra a un ser humano o animal. Dicho de manera diferente, el hecho de que el pI del AF2 y el AF1 se encuentre dentro de un intervalo relativamente estrecho puede permitir la selección de un tampón u otra solución en la que tanto el AF2 como el AF1 sean estables, lo que promueve de esta manera la estabilidad general de la composición.

En ciertas modalidades, la VL y la VH de los fragmentos de unión a antígeno se fusionan a través de enlazadores relativamente largos, que consisten en 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 o 35 aminoácidos hidrófilos que, cuando se unen, tienen una característica flexible. En una modalidad, la VL y la VH de cualquiera de las modalidades de scFv descritas en la presente descripción se unen por enlazadores relativamente largos de aminoácidos hidrófilos que son GSGEGSEGGGEGGSEGGSGEGGEGEGSG (SEQ ID NO: 8058), TGS GEGSEGGGEGGSEGGSGEGGEGEGSGT (SEQ ID NO: 8059), GATPPETGAETESPGETTGGSAESEPPGEG (SEQ ID NO: 8060, o GSAAPTAGTTSPASPAPPTGGSSAAGSPST (SEQ ID NO: 8061). En otra modalidad, el AF1 y el AF2 se unen entre sí mediante un enlazador corto de aminoácidos hidrófilos que tienen 3, 4, 5, 6 o 7 aminoácidos. En una modalidad, las secuencias enlazadoras cortas son SGGGGS (SEQ ID NO: 8062), GGGGS (SEQ ID NO: 8063), GGSGGS (SEQ ID NO: 8064), GGS o GSP. En otra modalidad, la descripción proporciona composiciones que comprenden un diacuerpo de cadena sencilla en el cual después del plegado, el primer dominio (VL o VH) se aparea con el último dominio (VH o VL) para formar un scFv y los dos dominios en el medio se aparean para formar el otro scFv en el cual el primer y segundo dominios, así como también el tercer y último dominios, se fusionan entre sí mediante uno de los enlazadores cortos anteriores y el segundo y tercer dominios variables se fusionan mediante uno de los enlazadores relativamente largos anteriores. Como se apreciará por un experto en la técnica, la selección del enlazador corto y el enlazador relativamente largo es para evitar el apareamiento incorrecto de dominios variables adyacentes, y facilitar de esta manera la formación de la configuración de diacuerpo de cadena sencilla que comprende el VL y VH del primer fragmento de unión a antígeno y el segundo fragmento de unión a antígeno.

Tabla 6b. Secuencias ilustrativas de CDR contra CD3

Construcción	REGIÓN CDR	Secuencia de aminoácidos	SEQ ID NO:
3,23, 3,30, 3,31, 3,32	CDR-L1	RSSNGAVTSSNYAN	735
3,24	CDR-L1	RSSNGEVTTSNYAN	736
3,33, 3,9	CDR-L1	RSSTGAVTTSNYAN	737
3,23, 3,30, 3,31, 3,32, 3,9, 3,33	CDR-L2	GTNKRAP	738
3,24	CDR-L2	GTIKRAP	739
3,23, 3,24, 3,30, 3,31, 3,32	CDR-L3	ALWYPNLWVF	740
3,33, 3,9	CDR-L3	ALWYSNLWVF	741
3,23, 3,24, 3,30, 3,31, 3,32, 3,9	CDR-H1	GFTFNTYAMN	742
3,23, 3,24, 3,30, 3,31, 3,32, 3,9	CDR-H2	RIRSKYNNYATYYADSVKD	743
3,23, 3,24, 3,30, 3,31, 3,32	CDR-H3	HENFGNSYVSWFA H	744
3,9	CDR-H3	HGNFGNSYVSWFAY	745

Tabla 6c. Secuencias ilustrativas de FR contra CD3

Construcción	REGIÓN FR	Secuencia de aminoácidos	SEQ ID NO:
3,23, 3,24, 3,30, 3,31, 3,32, 3,9, 3,33	FR-L1	ELVVTQEPSLTVSPGGTVTLTC	746
3,23, 3,24, 3,30, 3,31, 3,32, 3,9, 3,33	FR-L2	WVQQKPGQAPRGLIG	747
3,23, 3,24	FR-L3	GTPARFSGSLLGGKAALTLSGVQPE DEAVYYC	748
3,30	FR-L3	GTPARFSGSSLGGKAALTLSGVQPED EAVYYC	749
3,31	FR-L3	GTPARFSGSLLGGSAAALTLSGVQPED EAVYYC	750
3,32	FR-L3	GTPARFSGSSLGGSAALTLSGVQPED EAVYYC	751
3,9	FR-L3	GTPARFSGSLLGGKAALTLSGVQPE DEAEYYC	752
3,33	FR-L3	GTPARFSGSSLGGSAALTLSGVQPED EAEYYC	753
3,23, 3,24, 3,30, 3,31, 3,32, 3,9	FR-L4	GGGTKLTVL	754
3,23, 3,24	FR-H1	EVQLLESGGGIVQPGGSLKLSCAAS	755
3,30, 3,31, 3,32	FR-H1	EVQL Q ESGGGIVQPGGSLKLSCAAS	756
3,33	FR-H1	EVQL Q ESGGGLVQPGGSLKLSCAAS	757
3,9	FR-H1	EVQLLESGGGLVQPGGSLKLSCAAS	758
3,23, 3,24, 3,30, 3,31, 3,32, 3,9, 3,33	FR-H2	WVRQAPGKGLEWVA	759
3,23, 3,24, 3,30, 3,31, 3,32	FR-H3	RFTISRDDSKNTVYLMNNLKTEDT AVYYCVR	760

Construcción	REGIÓN FR	Secuencia de aminoácidos	SEQ ID NO:
3,9	FR-H3	RFTISRDDSKNTAYLQMNNLKTEDT AVYYCVR	762
3,23, 3,24, 3,30, 3,31, 3,32, 3,9, 3,33	FR-H4	WGQGTLVTVSS	764

Tabla 6d: Secuencias VL & VH ilustrativas

Construcción	REGIÓN	Secuencia de aminoácidos	SEQ ID NO:
3,23	VL	ELVVTQEPSLTVSPGGTVTLTCRSSNG AVTSSNYANWVQQKPGQAPRGLIGGT NKRAPGTPARFSGSLLGGKAALTLSGV QPEDEAVYYCALWY P NLWVFGGGTK LTVL	765
3,23, 3,24	VH	EVQLLESGGGIVQPGGSLKLSCAASGF TFNTYAMNWVRQAPGKGLEWVARIRS KYNNYATYYADSVKDRFTISRDDSKN TVYLQMNNLKTEDTAVYYCVRHENF GNSYVSWFA H WGQGTLVTVSS	766
3,24	VL	ELVVTQEPSLTVSPGGTVTLTCRSSNG EVTTSNYANWVQQKPGQAPRGLIGGT I KRAPGTPARFSGSLLGGKAALTLSGV QPEDEAVYYCALWY P NLWVFGGGTK LTVL	767
3,30	VL	ELVVTQEPSLTVSPGGTVTLTCRSSNG AVTSSNYANWVQQKPGQAPRGLIGGT NKRAPGTPARFSGSSLLGGKAALTLSGV QPEDEAVYYCALWY P NLWVFGGGTK LTVL	768
3,30, 3,31, 3,32	VH	EVQL Q ESGGGIVQPGGSLKLSCAASGF TFNTYAMNWVRQAPGKGLEWVARIRS KYNNYATYYADSVKDRFTISRDDSKN TVYLQMNNLKTEDTAVYYCVRHENF GNSYVSWFA H WGQGTLVTVSS	769
3,31	VL	ELVVTQEPSLTVSPGGTVTLTCRSSNG AVTSSNYANWVQQKPGQAPRGLIGGT NKRAPGTPARFSGSLLGG S AALTLSGV QPEDEAVYYCALWY P NLWVFGGGTK LTVL	770
3,32	VL	ELVVTQEPSLTVSPGGTVTLTCRSSNG AVTSSNYANWVQQKPGQAPRGLIGGT NKRAPGTPARFSGSSLGG S AALTLSGV QPEDEAVYYCALWY P NLWVFGGGTK LTVL	771
3,9	VL	ELVVTQEPSLTVSPGGTVTLTCRS S TGA VTTSNYANWVQQKPGQAPRGLIGGTN KRAPGTPARFSGSLLGGKAALTLSGVQ PEDEAEYYCALWYSNLWVFGGGTKLT VL	772
3,9	VH	EVQLLESGGGLVQPGGSLKLSCAASGF TFNTYAMNWVRQAPGKGLEWVARIRS KYNNYATYYADSVKDRFTISRDDSKN TAYLQMNNLKTEDTAVYYCVR H GNF GNSYVSWFA Y WGQGTLVTVSS	773
3,33	VL	ELVVTQEPSLTVSPGGTVTLTCRS S TGA VTTSNYANWVQQKPGQAPRGLIGGTN KRAPGTPARFSGSSLGG S AALTLSGVQ	774

Construcción	REGIÓN	Secuencia de aminoácidos	SEQ ID NO:
		PEDEAEYYCALWYSNLWVFGGGTKLTVL	
3,33	VH	EVQLQESGGGLVQPGGSLKLSCAASGFTFNTYAMNWVRQAPGKGLEWVARIRSKYNNYATYYADSVKDRFTISRDDSKNTAYLQMNNLKTEDTAVYYCVRHGNFGNSYVSWFAYWGQGTLVTVSS	775

Tabla 6e: Secuencias de scFv ilustrativas

Construcción	Secuencia de aminoácidos	SEQ ID NO:
3,23	ELVVTQEPSLTVSPGGTVTLTCRSSNGAVTSSNYANWVQQKPGQAPRGLIGGTNKRAPGTPARFSGSLLGGKAA ALTLSGVQPEDEAVYYCALWYPNLWVFGGGTKLTVL LGATPPETGAETESPGETTGGSAESEPPGEGEVQLLES GGGIVQPGGSLKLSCAASGFTFNTYAMNWVRQAPG KGLEWVARIRSKYNNYATYYADSVKDRFTISRDDSK NTVYLMNNLKTEDTAVYYCVRHENFGNSYVSWFA HWGQGTLVTVSS	776
3,24	ELVVTQEPSLTVSPGGTVTLTCRSSNGEVTTSNYANWVQQKPGQAPRGLIGGTIKRAPGTPARFSGSLLGGKAA LTLSGVQPEDEAVYYCALWYPNLWVFGGGTKLTVL GATPPETGAETESPGETTGGSAESEPPGEGEVQLLESG GGIVQPGGSLKLSCAASGFTFNTYAMNWVRQAPGK GLEWVARIRSKYNNYATYYADSVKDRFTISRDDSKN TVYLMNNLKTEDTAVYYCVRHENFGNSYVSWFAH WGQGTLVTVSS	777
3,30	ELVVTQEPSLTVSPGGTVTLTCRSSNGAVTSSNYANWVQQKPGQAPRGLIGGTNKRAPGTPARFSGSSLGGKAA ALTLSGVQPEDEAVYYCALWYPNLWVFGGGTKLTVL LGATPPETGAETESPGETTGGSAESEPPGEGEVQLQES GGGIVQPGGSLKLSCAASGFTFNTYAMNWVRQAPG KGLEWVARIRSKYNNYATYYADSVKDRFTISRDDSK NTVYLMNNLKTEDTAVYYCVRHENFGNSYVSWFA HWGQGTLVTVSS	778
3,31	ELVVTQEPSLTVSPGGTVTLTCRSSNGAVTSSNYANWVQQKPGQAPRGLIGGTNKRAPGTPARFSGSLLGGSAA LTLSGVQPEDEAVYYCALWYPNLWVFGGGTKLTVL GATPPETGAETESPGETTGGSAESEPPGEGEVQLQES GGGIVQPGGSLKLSCAASGFTFNTYAMNWVRQAPG KGLEWVARIRSKYNNYATYYADSVKDRFTISRDDSK NTVYLMNNLKTEDTAVYYCVRHENFGNSYVSWFA HWGQGTLVTVSS	779
3,32	ELVVTQEPSLTVSPGGTVTLTCRSSNGAVTSSNYANWVQQKPGQAPRGLIGGTNKRAPGTPARFSGSSLGGSAA LTLSGVQPEDEAVYYCALWYPNLWVFGGGTKLTVL GATPPETGAETESPGETTGGSAESEPPGEGEVQLQES GGGIVQPGGSLKLSCAASGFTFNTYAMNWVRQAPG KGLEWVARIRSKYNNYATYYADSVKDRFTISRDDSK NTVYLMNNLKTEDTAVYYCVRHENFGNSYVSWFA HWGQGTLVTVSS	780
3,9	ELVVTQEPSLTVSPGGTVTLTCRSSTGAVTTSNYANWVQQKPGQAPRGLIGGTNKRAPGTPARFSGSLLGGKAA ALTLSGVQPEDEAEYYCALWYSNLWVFGGGTKLTVL LGATPPETGAETESPGETTGGSAESEPPGEGEVQLLES GGGLVQPGGSLKLSCAASGFTFNTYAMNWVRQAPG	781

Construcción	Secuencia de aminoácidos	SEQ ID NO:
	KGLEWVARIRSKYNNYATYYADSVKDRFTISRDDSK NTAYLQMNNLKTEDTAVYYCVRHGNFGNSYVSWFA YWGQGLVTVSS	
3,33	ELVVTQEPSLTVSPGGTVTLTCRSSTGAVTTSNYANW VQQKPGQAPRGLIGGTNKRAPGTPARFSGSSLGGSAA LTLSGVQPEDEAEYYCALWYSNLWVFGGGTKLTVL GATPPETGAETESPGETTGGSAESEPPGEGEVQLQESG GGLVQPGGSLKLSCAASGFTFNTYAMNWVRQAPGK GLEWVARIRSKYNNYATYYADSVKDRFTISRDDSKN TAYLQMNNLKTEDTAVYYCVRHGNFGNSYVSWFAY WGQGLVTVSS	782
4,11	QSVLTQPPSASGTPGQRVTISCSGSSSNIGSNYVYWY QQLPGTAPKLLIYRNNQRPSGVPDRFSGSKSGTSASL AISGLRSEDEADYYCAAWDDSLSGLWVFGGGTKLTV LGATPPETGAETESPGETTGGSAESEPPGEGQVQLQQ WGGGLVKPGGSLRLSCAASGFTFSSYSMNWVRQAP GKGLEWVSRINSDGSSTNYADSVKGRFTISRDNANT LYLQMNSLRAEDTAVYYCARELRWGNWGQGLVT VSS	783
4,12	QAGLTQPPSASGTPGQRVTLSCSGSYSNIGTYVYWY QQLPGTAPKLLIYSNDQRLSGVPDRFSGSKSGTSASL AISGLQSEDEAAYYCAAWDDSLNGWAFGGGTKLTV LGATPPETGAETESPGETTGGSAESEPPGEGQVQLQQ WGGGLVKPGGSLRLSCAASGFTFSSYSMNWVRQAP GKGLEWVSRINSDGSSTNYADSVKGRFTISRDNANT LYLQMNSLRAEDTAVYYCARELRWGNWGQGLVT VSS	784
4,13	QPGLTQPPSASGTPGQRVTLSCSGRSSNIGSYVYWY QHLPGMAPKLLIYRNSRRPSGVPDRFSGSKSGTSASL VISGLQSDDEADYYCAAWDDSLKSWVFGGGTKLTV LGATPPETGAETESPGETTGGSAESEPPGEGQVQLQQ WGGGLVKPGGSLRLSCAASGFTFSSYSMNWVRQAP GKGLEWVSRINSDGSSTNYADSVKGRFTISRDNANT LYLQMNSLRAEDTAVYYCARELRWGNWGQGLVT VSS	785
4,14	QSVLTQPPSASGTPGQRVTISCSGSSSNIGTNYVYWY QQFPGTAPKLLIYSNNQRPSGVPDRFSGSKSGTSGSLA ISGLQSEDEADYSCAAWDDSLNGWVFGGGTKLTVLG ATPPETGAETESPGETTGGSAESEPPGEGQVQLVQWG GGLVKPGGSLRLSCAASGFTFSSYSMNWVRQAPGKG LEWVSRINSDGSSTNYADSVKGRFTISRDNANTLYL QMNSLRAEDTAVYYCARELRWGNWGQGLVTVSS	786
4,15	QPGLTQPPSASGTPGQRVTISCSGSSSNIGSNYVYWY QQLPGTAPKLLIYRNNQRPSGVPDRLSGSKSGTSASL AISGLRSEDEADYYCAAWDDSLSGWVFGGGTKLTVL GATPPETGAETESPGETTGGSAESEPPGEGQVQLVQW GGGLVKPGGSLRLSCAASGFTFSSYSMNWVRQAPGK GLEWVSRINSDGSSTNYADSVKGRFTISRDNANTLY LQMNSLRAEDTAVYYCARELRWGNWGQGLVTVSS	787
4,16	QAVLTQPPSASGTPGQRVTISCSGSSSNIGSYVYWY QQVPGAAPKLLMRLNNQRPSGVPDRFSGAKSGTSAS LVISGLRSEDEADYYCAAWDDSLSGQWVFGGGTKLT VLGATPPETGAETESPGETTGGSAESEPPGEGQVQLQ QWGGGLVKPGGSLRLSCAASGFTFSSYSMNWVRQA PGKGLEWVSRINSDGSSTNYADSVKGRFTISRDNANT TLYLQMNSLRAEDTAVYYCARELRWGNWGQGLVT	788

Construcción	Secuencia de aminoácidos	SEQ ID NO:
	VSS	
4,17	QAGLTQPPSASGTPGQRVTISCSGSSSNIGSNYVYWY QQLPGTAPKLLIYRNNQRPSGVPDRFSGSKSGTSASL AISGLRSEDEADYYCATWDASLSGWVFGGGTKLTVL GATPPETGAETESPGETTGGSAESEPPGEGEVQLVQW GGGLVKPGGSLRLSCAASGFTFSSYSMNWVRQAPGK GLEWVSRINSDGSSTNYADSVKGRFTISRDNAKNTLY LQMNSLRAEDTAVYYCARELRWGNWQGQGLTVTVSS	789

Dominios de unión anti-EpCAM

En algunas modalidades, la invención proporciona composiciones de ensamblaje de polipéptidos quiméricos que comprenden un dominio de unión con afinidad de unión por el marcador EpCAM específico de tumor. En una modalidad, el dominio de unión comprende VL y VH derivados de un anticuerpo monoclonal contra EpCAM. Los anticuerpos monoclonales contra EpCAM se conocen en la técnica. Ejemplos ilustrativos no limitantes de anticuerpos monoclonales contra EpCAM y las secuencias VL y VH de estos se exponen en la Tabla 6f. En una modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos que comprende un dominio de unión con afinidad de unión por el marcador específico de tumor EpCAM que comprende secuencias VL y VH anti-EpCAM expuestas en las Tabla 6f. En otra modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos, en donde el primer dominio de unión de la primera porción comprende regiones VH y VL, en donde cada región VH y VL exhibe al menos aproximadamente 90 %, o 91 %, o 92 % o 93 %, o 94 %, o 95 %, o 96 %, o 97 %, o 98 %, o 99 % de identidad o es idéntica a las secuencias VL y VH pareadas del anticuerpo anti-EpCAM 4D5MUCB expuestas en la Tabla 6f. En otra modalidad, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos que comprende un dominio de unión con afinidad de unión por el marcador específico de tumor que comprende la región CDR-L1, la región CDR-L2, la región CDR-L3, la región CDR-H1, la región CDR-H2 y la región CDR-H3, en donde cada uno se deriva de las respectivas secuencias VL y VH expuestas en la Tabla 6f.

Tabla 6f. Secuencias y anticuerpos monoclonales anti-células objetivo

Nom-bre comercial	Nom-bre del anti-cuerpo	Objetivo	Secuencia de VH	SEQ ID NO: (longitud completa)	SEQ ID NO: (CDR)	Secuencia de VL	SEQ ID NO: (longitud completa)	SEQ ID NO: (CDR)
Tysabri TM	natalizumab	Alpha 4 Integrin	QVQLVQSGAEVKKPGAS	8065	8918, 8919, 8920	DIQMTQSPSSLSASVGDRV TITCKTS	8164	9178, 9179, 9180

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			VKVSC KASGF NIKDT YIHW VRQAP GQRLE WMGR IDPAN GYTK YDPK FOGR VTITA DTSAS TAYM ELSSL RSED TAVYY CARE GYYG NYGV YAMD YWGQ GTLVT VSS			<u>QDINKY</u> <u>MAWYQ</u> QTPGKA PRLLIH <u>Y</u> <u>TSALQP</u> GIPSRFS GSGSGR DYTFTIS SLQPEDI ATYYCL <u>QYDNL</u> <u>WTFGQG</u> TKVEIK		
REGN9 10	nesva cumab	Ang2	EVQLV ESGGG LVQPG GSLRL SCAAS <u>GFTFS</u> <u>SYDIH</u> WVRQ ATGK GLEW VSAIG <u>PAGD</u> <u>TYYP</u> <u>GSVK</u> GRFTI SRENA KNSLY LQMN SLRAG DTAV YYCA <u>RGLIT</u> <u>FGGLI</u> <u>APFDY</u> WGQG TLVTV	8066	8921, 8922, 8923	EIVLTQS PGTSLS PGERAT LSCRAS <u>QSVSST</u> <u>YLAWY</u> QQKPGQ APRLLIY <u>GASSRA</u> <u>TGIPDRF</u> SGSGSG TDFTLTI SRLEPED FAVYYC <u>QHYDNS</u> <u>QTFGQG</u> TKVEIK	8165	9181, 9182, 9183

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			SS					
hMFE2 3		CEA	QVKLE QSGAE VVKP GASV KLSCK <u>ASGEN</u> <u>IKDSY</u> MHWL RQGPG QRLE WIGWI <u>DPEN</u> <u>GDTE</u> YAPKF QGKA TFTTD TSANT AYLGL SSLRP EDTAV YYCN EG <u>TPT</u> <u>GPYY</u> <u>FDYW</u> GQCTL VTVSS	8067	8924, 8925, 8926	ENVLTQ SPSSMSA SVGDRV NIACSAS <u>SSVSYM</u> HWFQQK PGKSPK LWIYSTS <u>NLASGV</u> PSRFSGS GSGTDY SLTISSM QPEDAA TYYCQQ <u>RSSYPL</u> TFGGGT KLEIK	8166	9184, 9185, 9186
M5A (human ized T84.66)		CEA	EVQLV ESGGG LVQPG GSLRL SCAAS GFNIK <u>DTYM</u> <u>HWVR</u> QAPG KGLE WVAR <u>IDPAN</u> <u>GNSK</u> <u>YADS</u> <u>VKGR</u> FTISA DTSKN TAYLQ MNSL RAEDT AVYY CA <u>FG</u> <u>YYVS</u>	8068	8927, 8928, 8929	DIQLTQS PSSLSAS VGDRV ITC <u>RAG</u> <u>ESVDIF</u> <u>GVGFLH</u> WYQQKP GKAPKL LIY <u>RASN</u> <u>LESGVP</u> SRFSGSG SRTDFTL TISSLQP EDFATY YC <u>QQTN</u> <u>EDPYTF</u> GQGTVK EIK	8167	9187, 9188, 9189

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			DYAM AYWG QGTLV TVSS					
M5B (human ized T84.66)		CEA	EVQLV ESGGG LVQPG GSLRL SCAAS GFNIK DTYM HWVR QAPG KGLE WVAR IDPAN GNSK YVPK FQGR ATISA DTSKN TAYLQ MNSL RAEDT AVYY CAPE FG YYVS DYAM AYWG QGTLV TVSS	8069	8930, 8931, 8929	DIQLTQS PSSLSAS VGDRVT ITC RAG ESVDIF GVGFLH WYQQKP GKAPKL LIY RASN LESGVP SRFSGSG SRTDFTL TISSLQP EDFATY YC QQTN EDPYTF GQGTKV EIK	8167	9187, 9188, 9189
CEA- Cide	Labet uzuma b (MN- 14)	CEA CAM 5	EVQLV ESGGG VVQP GRSLR LSCSA SGFDF TTYW MSWV RQAPG KGLE WIGE I HPDSS TINYA PSLKD RFTISR DNAK NTLFL QMDS LRPED	8070	8932, 8933, 8934	DIQLTQS PSSLSAS VGDRVT ITC KAS QDVGTS VAWYQ QKPGKA PKLLIY WTSTRH TGVPSR FSGSGSG TDFTFTI SSLQPED IATYYC QQYSLY RSFGQG TKVEIK	8168	9190, 9191, 9192

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			TGVYF CAS <u>LY</u> <u>FGFP</u> <u>WFAY</u> WGQG TPVTV SS					
CEA- Scan	arcitu moma b	CEA CAM 5	EVKLV ESGGG LVQPG GSLRL SCATS <u>GFTFT</u> <u>DYYM</u> <u>NWVR</u> QPPGK ALEW <u>LGFIG</u> <u>NKAN</u> <u>GYTT</u> <u>EYSAS</u> VKGRF TISRD KSQSI LYLQ MNTL RAEDS <u>ATYY</u> <u>CTRD</u> <u>RGLRF</u> YFDY WGQG TTLTV SS	8071	8935, 8936, 8937	QTVLSQ SPAILS A SPGEKV TMT <u>CRA</u> <u>SSSVTYI</u> <u>HWYQQ</u> KPGSSPK SWIYAT <u>SNLASG</u> VPARFS GSGSGT SYSLTIS RVEAED AATYYC <u>QHWSS</u> <u>KPPTFG</u> GGTKLEI KR	8169	9193, 9194, 9195
MT110		CEA CAM 5	EVQLV ESGGG LVQPG RSLRL SCAAS GFTVS <u>SYWM</u> <u>HWVR</u> QAPG KGLE <u>WVGFI</u> <u>RNKA</u> <u>NGGT</u> <u>TEYA</u> <u>ASVK</u> <u>GRFTI</u>	8072	8938, 8939, 8940	QAVLTQ PASLSAS PGASAS LTCT <u>LR</u> <u>RGINVG</u> <u>AYSIW</u> YQQKPG SPPQYLL <u>RYKSDS</u> <u>DKQOGS</u> GVSSRFS ASKDAS ANAGIL LISGLQS EDEADY YCM <u>IW</u>	8170	9196, 9197, 9198

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			SRDDS KNTLY LQMN SLRAE DTAV YYCA <u>RDRG</u> <u>LRFYF</u> <u>DYWG</u> QGTTV TVSS			<u>HSGASA</u> <u>VFGGGT</u> KLTVL		
MT103	blinat umom ab	CD19	QVQL QQSG AELVR PGSSV KISCK ASGY AFSSY <u>WMN</u> WVKQ RPGQG LEWIG <u>QIWP</u> <u>GDGD</u> <u>TNYN</u> <u>GKFK</u> <u>GKAT</u> LTADE SSSTA YMQV SSLAS EDSAV YFCAR <u>RETTT</u> <u>VGRY</u> <u>YYAM</u> <u>DYWG</u> QGTTV TVSS	8073	8941, 8942, 8943	DIQLTQS PASLAV SLGQRA TISCK <u>KAS</u> <u>QSVDDYD</u> <u>GDSYLN</u> WYQQIP GQPPKL <u>LIYDASN</u> <u>LVSGIPP</u> RFSGSGS GTDFTL NIHPVEK VDAATY HC <u>QOQT</u> <u>EDPWTF</u> GGGTKL EIK	8171	9199, 9200, 9201
Arzerra	ofatu muma b	CD20	EVQLV ESGGG LVQPG RSLRL SCAAS GFTFN <u>DYAM</u> <u>HWVR</u> QAPG KGLE WVST <u>I</u>	8074	8944, 8945, 8946	EIVLTQS PATLSLS PGERAT LSC <u>RAS</u> <u>QSVSSY</u> <u>LAWYQ</u> QKPGQA PRLLIY <u>D</u> <u>ASNRAT</u> GIPARFS GSGSGT	8172	9202, 9203, 9204

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>SWNS</u> <u>GSIGY</u> <u>ADSV</u> <u>KGRFT</u> ISRDN AKKSL YLQM NSLRA EDTAL YYCA <u>KDIQY</u> <u>GNY Y</u> <u>YGMD</u> <u>VWGQ</u> GTTVT VSS			DFTLTIS SLEPEDF AVYYCQ <u>QRSNWP</u> <u>ITFGQGT</u> RLEIK		
Bexxar TM	tositu moma b	CD20	QAYL QQSG AELVR PGASV KMSC KASG YTFTS <u>YNMH</u> WVKQ TPRQG LEWIG <u>AIYPG</u> <u>NGDT</u> <u>SYNQ</u> <u>KEKG</u> KATLT VDKSS STAY MQLSS LTSED SAVYF CARV <u>VYYS</u> <u>NSYW</u> <u>YEDV</u> WGTG TTVTV SG	8075	8947, 8948, 8949	QIVLSQS PAILSAS PGEKVT MTCRAS <u>SSVSYM</u> <u>HWYQQ</u> KPGSSPK PWIYAP <u>SNLASG</u> VPARFS GSGSGT SYSLTIS RVEAED AATYYC <u>QQWSE</u> <u>NPPTFG</u> AGTKLE LK	8173	9205, 9206, 9207
GAZY VA	Obinu tuzum ab	CD20	QVQL VQSG AEVK KPGSS VKVSC KASG	8076	8950, 8951, 8952	DIVMTQ TPLSLPV TPGEPAS ISCRSSK <u>SLLHSN</u> <u>GITYLY</u>	8174	9208, 9209, 9210

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			YAFSY <u>SWIN</u> WVRQ APGQ GLEW MGR <u>IF</u> <u>PGDG</u> <u>DTDY</u> <u>NGKF</u> <u>KGRV</u> TITAD KSTST AYME LSSLR SEDTA VYYC ARN <u>V</u> <u>FDGY</u> <u>WLVY</u> WGQG TLVTV SS			WYLQKP GQSPQL LIY <u>QMS</u> <u>NLV</u> SGV PDRFSGS GSGTDF TLKISRV EAEDVG VYYC <u>AO</u> <u>NLELPY</u> <u>TFGGGT</u> KVEIK		
	Ocreli zumab / 2H7 v16	CD20	EVQLV ESGGG LVQPG GSLRL SCAAS <u>GYTF</u> <u>TSYN</u> MHWV RQAPG KGLE WVGA <u>IYPGN</u> <u>GDT</u> S YNQK FKGRF TISVD KSKNT LYLQ MNSL RAEDT AVYY CAR <u>V</u> <u>VYYS</u> <u>NSYW</u> <u>YFDV</u> WGQG TLVTV SS	8077	8953, 8954, 8949	DIQMTQ SPSSLSA SVGDRV TITC <u>RAS</u> <u>SSVSYM</u> <u>HWYQQ</u> KPGKAP KPLIY <u>AP</u> <u>SNLASG</u> VPSRFSG SGSGTD FTLTISS LQPEDF ATYYC <u>Q</u> <u>QWSFNP</u> <u>PTFGQG</u> TKVEIK	8175	8911, 9206, 9207

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
Rituxan ™	rituxi mab	CD20	QVQL QQPG AELVK PGASV KMSC KAS <u>G</u> <u>YTFTS</u> <u>YNMH</u> WVKQ TPGRG LEWIG <u>AIYPG</u> <u>NGDT</u> SYNQ KFKG KATLT ADKSS STAY MQLSS LTSED SAVY YC <u>ARS</u> <u>TYYG</u> <u>GDWY</u> <u>ENVW</u> GAGTT VTVSA	8078	8953, 8954, 8955	QIVLSQS PAILSAS PGEKVT MTCRAS <u>SSVSYIH</u> WFQQKP GSSPKP WIY <u>ATS</u> NLASGV PVRFSGS GSGTSY SLTISRV EAEDAA TYYC <u>QQ</u> <u>WTSNPP</u> <u>TFGGGT</u> KLEIK	8176	9211, 9212, 9213
Zevalin ™	ibritu moma b tieuxe tan	CD20	QAYL QQSG AELVR PGASV KMSC KAS <u>G</u> <u>YTFTS</u> <u>YNMH</u> WVKQ TPRQG LEWIG <u>AIYPG</u> <u>NGDT</u> <u>SYNQ</u> <u>KFKG</u> KATLT VDKSS STAY MQLSS LTSED SAVYF CAR <u>V</u> <u>VYYS</u>	8079	8953, 8956, 8949	QIVLSQS PAILSAS PGEKVT MTC <u>RAS</u> <u>SSVSYM</u> <u>HWYQQ</u> KPGSSPK PWIY <u>AP</u> <u>SNLASG</u> VPARFS GSGSGT SYSLTIS RVEAED AATYYC <u>QQWSE</u> <u>NPPTFG</u> AGTKLE LK	8173	8911, 9206, 9207

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>NSYW</u> <u>YFDV</u> WGTG TTVTV SA					
Mylotar g	Gemt uzuma b (hP67. 6)	CD33	QLVQS GAEV KKPGS SVKVS CKAS <u>GYTIT</u> <u>DSNIH</u> WVRQ APGQS LEWIG <u>YIYPY</u> <u>NGGT</u> <u>DYNQ</u> <u>KFKN</u> RATLT VDNPT NTAY MELSS LRSED TDFYY CVNG <u>NPWL</u> <u>AYWG</u> QGTLV TVSS	8080	8957, 8958, 8959	DIQLTQS PSTLSAS VGDRVT ITC <u>RASE</u> <u>SLDNYG</u> <u>IRFLTW</u> FQQKPG KAPKLL MYA <u>AASN</u> <u>QSG</u> SGVP SRFSGSG SGTEFTL TISSLQP DDFATY YC <u>QOTK</u> <u>EVPWSF</u> GQGTKV EVK	8177	9214, 9215, 9216
Daratu mumab		CD38	EVQLL ESGGG LVQPG GSLRL SCAVS <u>GFTFN</u> <u>SFAMS</u> WVRQ APGK GLEW VSA <u>IS</u> <u>GSGG</u> <u>GTYY</u> ADSV KGRFT ISRDN SKNTL YLQM NSLRA EDTAV	8081	8960, 8961, 8962	EIVLTQS PATLSLS PGERAT LSCRAS <u>QSVSSY</u> LAWYQ QKPGQA PRLLIY <u>D</u> <u>ASN</u> RAT GIPARFS GSGSGT DFTLTIS SLEPEDF AVYYC <u>Q</u> <u>QRSNWP</u> <u>PTFGQG</u> TKVEIK	8178	9217, 9218, 9219

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			YFCA <u>KDKIL</u> <u>WFGE</u> <u>PVFDY</u> WGQG TLVTV SS					
	1F6	CD70	QIQLV QSGPE VKKP GETVK ISCKA <u>SGYTF</u> <u>TNYG</u> <u>MNWV</u> KQAP GKGL KWMG <u>WINT</u> <u>YTGE</u> <u>PTYA</u> <u>DAFK</u> <u>GRFAF</u> SLETS ASTAY LQINN LKNED TATYF CARD <u>YGDY</u> <u>GMDY</u> WGQG TSVTV SS	8082	8963, 8964, 8965	DIVLTQS PASLAV SLGQRA TISC <u>RAS</u> <u>KSVSTS</u> <u>GYSEMH</u> WYQQKP GQPPKL LIY <u>LASN</u> <u>LESGVP</u> ARFSGS GSGTDF TLNIHPV EEEDAA TY YC <u>QHRS</u> <u>EVPWTF</u> GGGTKL EIK	8179	9220, 9221, 9222
	2F2	CD70	QVQL QQSGT ELMTP GASVT MSCK <u>TSGYT</u> <u>FSTY</u> <u>WIEW</u> VKQRP GHGLE WIGE <u>I</u> <u>LGPSG</u> <u>YTDY</u> <u>NEKE</u> <u>KAKA</u> TFTAD	8083	8966, 8967, 8968	DIVLTQS PASLTVS LGQKTTI <u>SCRASK</u> <u>SVSTSG</u> <u>YSFMH</u> WYQLKP GQSPKL LIY <u>LASD</u> <u>LPSGVP</u> ARFSGS GSGTDF TLKIHPV EEEDAA TY YC <u>QHRS</u>	8180	9220, 9223, 9224

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			TSSNT AYMQ LSSLA SEDSA VYYC <u>ARWD</u> <u>RLYA</u> <u>MDYW</u> GGGTS VTVSS			<u>EIPYTFG</u> GGTKLEI T		
	2H5	CD70	QVQL VESGG GVVQ PGRSL RLSCA ASGFT <u>FSSYI</u> <u>MHWV</u> RQAPG KGLE WVAV <u>ISYDG</u> <u>RNKY</u> <u>YADS</u> <u>VKGR</u> FTISR DNSK NTLYL QMNS LRAED TAVY YCAR <u>DTDG</u> <u>YDFD</u> <u>YWGQ</u> GTLVT VSS	8084	8969, 8970, 8971	EIVLTQS PATLSLS PGERAT <u>LSCRAS</u> <u>QSVSSY</u> <u>LAWYQ</u> QKPGQA PRLLIYD <u>ASNRAT</u> GIPARFS GSGSGT DFTLTIS SLEPEDF AVYYCQ <u>Q</u> <u>RTNWPL</u> <u>TFGGGT</u> KVEIK	8181	9202, 9225, 9226
	10B4	CD70	QIQLV ESGGG VVQP GRSLR LSCAA SGFTF <u>GYYA</u> <u>MHWV</u> RQAPG KGLE WVAV <u>ISYDG</u> <u>SIKYY</u>	8085	8972, 8973, 8974	AIQLTQS PSSLSAS VGDRV ITC <u>RASQ</u> <u>GISSALA</u> WYQQKP GKAPKF LIY <u>DASS</u> <u>LESGVP</u> SRFSGSG SGTDFTL TISSLQP EDFATY	8182	9227, 9228, 9229

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>ADSV</u> <u>KGRFT</u> ISRDN SKNTL YLQM NSLRA ED TAVY YCAR <u>EGPYS</u> <u>NYLD</u> <u>YWGQ</u> GTLVT VSS			<u>YCQQ</u> <u>ENSYPF</u> <u>TFGPGT</u> KVDIK		
	8B5	CD70	QVQL VESGG GVVQ PGRSL RLSCA TSGFT <u>FSDYG</u> <u>MHWV</u> RQAPG KGLE WVAV <u>IWYD</u> <u>GSNK</u> <u>YYAD</u> <u>SVKG</u> RFTISR DNSK KTLSL QMNS LRAED TAVY YCAR <u>DSIMV</u> <u>RGDY</u> WGQG TLVTV SS	8086	8975, 8976, 8977	DIQMTQ SPSSLSA SVGDRV TITC <u>RAS</u> <u>QGISSW</u> <u>LAWYQ</u> QKPEKA PKSLIYA <u>ASSLQS</u> GVPSRFS GSGSGT DFTLTIS SLQPEDF ATYYC <u>Q</u> <u>Q</u> <u>YNSYPL</u> <u>TFGGGT</u> KVEIK	8183	9230, 9231, 9232
	18E7	CD70	QVQL VESGG GVVQ PGRSL RLSCA ASGFT <u>FSDHG</u> <u>MHWV</u> RQAPG	8087	8978, 8976, 8977	DIQMTQ SPSSLSA SVGDRV TITC <u>RAS</u> <u>QGISSW</u> <u>LAWYQ</u> QKPEKA PKSLIYA <u>ASSLQS</u>	8183	9230, 9231, 9232

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			KGLE WVAV <u>V</u> <u>IWYD</u> <u>GSNK</u> <u>YYAD</u> <u>SVKG</u> RFTISR DNSK NTLYL QMNS LRAED TAVY YCAR <u>DSIMV</u> <u>RGDY</u> WGQG TLVTV SS			GVPSRFS GSGSGT DFTLTIS SLQPEDF ATYYC <u>Q</u> <u>Q</u> <u>YNSYPL</u> <u>TFGGGT</u> KVEIK		
	69A7	CD70	QVQL QESGP GLVKP SETLS LTCTV SGGSV <u>SSDYY</u> <u>YWSW</u> IRQPP GKGLE WLGY <u>I</u> <u>YYSGS</u> <u>TNYN</u> <u>PSLKS</u> RVTIS VDTSK NQFSL KLRSV TTA DTAV YYCA <u>RGDG</u> <u>DYGG</u> <u>NCFD</u> YWGQ GTLVT VSS	8088	8979, 8980, 8981	EIVLTQS PATLSLS PGERAT LSC <u>RAS</u> <u>QSVSSY</u> <u>LAWYQ</u> QKPGQA PRLLI <u>F</u> <u>ASNRAT</u> GIPARFS GSGSGT DFTLTIS SLEPEDF AVYYC <u>Q</u> <u>Q</u> <u>RSNWPL</u> <u>TFGGGT</u> KVEIK	8184	9202, 9203, 9233
CE- 355621		cMET	QVQL VQSG AEVK KPGAS VKVSC	8089	8982, 8983, 8984	DIQMTQ SPSSVSA SVGDRV TITC <u>RAS</u> <u>QGINTW</u>	8185	9234, 9235, 9236

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			KASG YTFT <u>S</u> <u>YGFS</u> WVRQ APGQ GLEW MGW <u>I</u> <u>SASNG</u> <u>NTYY</u> <u>AOKL</u> <u>QGRV</u> TMTT DTSTS TAYM ELRSL RSDDT AVYY CAR <u>V</u> <u>YADY</u> <u>ADYW</u> GQGT L VTVSS			<u>L</u> AWYQ QKPGKA PKLLIYA <u>ASSLKS</u> GVPSRFS GSGSGT DFTLTIS SLQPEDF ATYYC <u>Q</u> <u>QANSFP</u> <u>L</u> TFGGG TKVEIK		
LY287 5358	emibe tuzum ab	cMET	QVQL VQSG AEVK KPGAS VKVSC KAS <u>G</u> <u>YTFTD</u> <u>YYMH</u> WVRQ APGQ GLEW MGR <u>V</u> <u>NPNR</u> <u>RGTT</u> YNQK FEGRV TMTT DTSTS TAYM ELRSL RSDDT AVYY <u>CARA</u> <u>NWLD</u> <u>YWGQ</u> GTTVT VSS	8090	8985, 8986, 8987	DIQMTQ SPSSLSA SVGDRV TITCSVS <u>SSVSSIY</u> LHWYQ QKPGKA PKLLIY <u>S</u> <u>TSN</u> LAS GVPSRFS GSGSGT DFTLTIS SLQPEDF ATYYC <u>Q</u> <u>VYSGYP</u> <u>L</u> TFGGG TKVEIK	8186	9237, 9238, 9239
MetMA	onartu	cMET	EVQLV	8091	8988,	DIQMTQ	8187	9240,

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
b	zumab		ESGGG LVQPG GSLRL SCAAS GYTFT <u>SYWL</u> <u>HWVR</u> QAPG KGLE WVG <u>M</u> <u>IDPSN</u> <u>SDTRF</u> <u>NPNE</u> <u>KDRFT</u> ISADT SKNTA YLQM NSLRA EDTAV YY <u>C</u> <u>A</u> <u>TYRSY</u> <u>VTPLD</u> <u>YWGQ</u> GTLVT VSS		8989, 8990	SPSSLSA SVGDRV TITC <u>KSS</u> <u>QSLLYT</u> <u>SSQKNY</u> <u>LAWYQ</u> QKPGKA PKLLIY <u>WASTRE</u> <u>SGVPSRF</u> SGSGSG TDFTLTI SSLQPED FATYYC <u>QQYYA</u> <u>YPWTFG</u> QGTKVE IK		9241, 9242
	tremel imum ab (CP- 67520 6, or 11.2.1)	CTLA 4	QVQL VESGG GVVQ PGRSL RLSCA AS <u>GFT</u> <u>FSSYG</u> <u>MHWV</u> RQAPG KGLE WVAV <u>IWYD</u> <u>GSNK</u> <u>YYAD</u> <u>SVKGR</u> FTISR DNSK NTLYL QMNS LRAED TAVY YCAR <u>DPRG</u> <u>ATLY</u> <u>YYYY</u>	8092	8991, 8992, 8993	DIQMTQ SPSSLSA SVGDRV TITC <u>RAS</u> <u>QSINSYL</u> <u>DWYQQ</u> KPGKAP KLLIYA <u>ASSLQS</u> GVPSRFS GSGSGT DFTLTIS SLQPEDF ATYYC <u>Q</u> <u>QYYSTP</u> <u>FTFGPG</u> TKVEIK	8188	9243, 9231, 9244

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>GMDV</u> WGQG TTVTV SS					
Yervoy	Ipilim umab 10D1	CTLA 4	QVQL VESGG GVVQ PGRSL RLSCA ASGFT FSSYT <u>MHWV</u> RQAPG KGLE WVTEI <u>SYDG</u> <u>NNKY</u> <u>YADS</u> <u>VKGR</u> FTISR DNSK NTLYL QMNS LRAED TAIYY CART <u>GWLG</u> <u>PFDY</u> WGQG TLVTV SS	8093	8994, 8995, 8996	EIVLTQS PGTSLS PGERAT LSC <u>RAS</u> <u>QSVGSS</u> <u>YLA</u> WY QQKPGQ APRLLIY <u>GAFSRA</u> <u>TG</u> IPDRF SGSGSG TDFTLTI SRLEPED FAVYYC <u>QQYGSS</u> <u>PWT</u> FGQ GTKVEI K	8189	9245, 9246, 9247
AGS16 F	H16- 7.8	ENPP 3	QVQL QESGP GLVKP SQTLS LTCTV SGGSI <u>SSGGY</u> <u>YWSW</u> IRQHP GKGLE WIG <u>I</u> <u>YYSGS</u> <u>TYYN</u> <u>PSLKS</u> RVTIS VDTSK NQFSL KLNSV TAADT	8094	8997, 8998, 8999	EIVLTQS PDFQSV TPKEKV TITC <u>RAS</u> <u>QSIGISL</u> <u>H</u> WYQQ KPDQSP KLLIK <u>Y</u> <u>ASQSFS</u> GVPSRFS SGSGGT DFTLTIN SLEAED AATYYC <u>HQSRSE</u> <u>PWT</u> FGQ GTKVEI K	8190	9248, 9249, 9250

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			AVFYC AR <u>VAI</u> <u>VTTIP</u> <u>GGMD</u> <u>VWGQ</u> GTTVT VSS					
MT110	solito mab	EpCA M	EVQLL EQSGA ELVRP GTSVK ISCKA SGYAF <u>TNYW</u> <u>LGWV</u> KQRPG HGLE WIG <u>DI</u> <u>FPGSG</u> <u>NIHYN</u> <u>EKFK</u> <u>GKAT</u> LTADK SSSTA YMQ LSSLTF EDSAV YFCAR <u>LRNW</u> <u>DEPM</u> <u>DYWG</u> QGTTV TVSS	8095	9000, 9001, 9002	ELVMTQ SPSSLTV TAGEKV TMS <u>CKS</u> <u>SQSLLN</u> <u>SGNQKN</u> <u>YLTWY</u> QQKPGQ PPKLLIY <u>WASTRE</u> <u>SGVPDR</u> FTGSGS GTDFTL TISSVQA EDLAVY YC <u>QNDY</u> <u>SYPLTF</u> GAGTKL EIK	8191	9251, 9241, 9252
MT201	Adeca- tumu mab	EpCA M	EVQLL ESGGG VVQP GRSLR LSCAA SGFTF <u>SSYG</u> <u>MHWV</u> RQAPG KGLE WVAV <u>ISYDG</u> <u>SNKY</u> <u>YADS</u> <u>VKGR</u> FTISR DNSK	8096	9003, 9004, 9005	ELQMTQ SPSSLSA SVGDRV TITC <u>RTS</u> <u>QSISSYL</u> <u>NWYQQ</u> KPGQPP KLLIY <u>W</u> <u>ASTRES</u> GVPDRF SGSGSG TDFTLTI SSLQPED SATYYC <u>QQSYDI</u> <u>PYTFGQ</u> GTKLEIK	8192	9253, 9241, 9254

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			NTRYL QMNS LRAED TAVY YCAK <u>DMG</u> <u>WGSG</u> <u>WRPY</u> <u>YYYG</u> <u>MDVW</u> GQGT TVSS					
Panore x	Edrec oloma b Mab CO17 -1A	EpCA M	QVQL QQSG AELVR PGTSV KVSCK ASGY <u>AFTN</u> <u>YLIEW</u> VKQRP GQGLE WIGVI <u>NPGS</u> <u>GGTN</u> YNEKF KGKA TLTAD KSSST AYMQ LSSLT SDDSA VYFCA <u>RDGP</u> <u>WFAY</u> WGQG TLVTV SA	8097	9006, 9007, 9008	NIVMTQ SPKSMS MSVGER VTLTCK ASENVV <u>TYVSWY</u> QQKPEQ SPKLLIY <u>GASNRY</u> TGVPRD FTGSGS ATDFTL TISSVQA EDLADY HCGQG <u>YSYPYT</u> FGGGTK LEIK	8193	9255, 9256, 9257
	tucotu zumab	EpCA M	QIQLV QSGPE LKKPG ETVKI SCKAS <u>GYTF</u> <u>TNYG</u> <u>MNWV</u> RQAPG KGLK WMG <u>WINT</u>	8098	8963, 9009, 9010	QILLTQS PAIMSAS PGEKVT MTC <u>SAS</u> <u>SSVSYM</u> LWYQQ KPGSSPK PWIF <u>DTS</u> <u>NLASGF</u> PARFSGS GSGTSY SLIISSM	8194	9258, 9259, 9260

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>YTGE</u> <u>PTYA</u> <u>DDFK</u> GRFVF SLETS ASTAF LQLNN LRSED TATYF CVRFI <u>SKGD</u> <u>YWGQ</u> GTSVT VSS			EAEDAA TYY <u>CHQ</u> <u>RSGYPY</u> <u>TFGGGT</u> KLEIK		
UBS- 54		EpCA M	VQLQ QSDAE LVKPG ASVKI SCKAS <u>GYTF</u> <u>TDHAI</u> <u>HWVK</u> QNPEQ GLEWI <u>GYFSP</u> <u>GNDD</u> <u>FKYN</u> <u>ERFKG</u> KATLT ADKSS STAYV QLNSL TSEDS AVYFC <u>TRSLN</u> <u>MAYW</u> GQGTS VTVSS	8099	9011, 9012, 9013	DIVMTQ SPDSLA VSLGER ATINCK <u>S</u> <u>SQSVLY</u> <u>SSNNKN</u> <u>YLA</u> WY QQKPGQ PPKLLIY <u>WASTRE</u> <u>SGVPDR</u> FSGSGSG TDFTLTI SSLQAE DVAVYY <u>CQQYYS</u> <u>YPLTFG</u> GGTKVK ES	8195	9261, 9241, 9262
3622W 94	323/A 3	EpCA M	EVQLV QSGPE VKKP GASV KVSCK <u>ASGY</u> <u>TFTNY</u> <u>GMNW</u> VRQAP GQGLE WMG <u>WINT</u>	8100	8963, 9014, 9015	DIVMTQ SPLSLPV TPGEPAS ISCR <u>RSSI</u> <u>NKKGSN</u> <u>GITYLY</u> WYLQKP GQSPQL LIYQMS NLASGV PDRF <u>SGS</u> <u>GSGTDF</u>	8196	9263, 9264, 9265

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>YTGE</u> <u>PTYG</u> <u>EDFKG</u> RFAFS LDTSA STAY MELSS LRSED TAVYF CARFG <u>NYVD</u> <u>YWGQ</u> GSLVT VSS			TLKISRV EAEDVG VYYCA <u>Q</u> <u>NLEIPR</u> <u>TFGQGT</u> KVEIK		
4D5M OCBv2		EpCA M	EVQLV QSGPG LVQPG GSVRI SCAAS GYTFT <u>NYGM</u> <u>NWVK</u> QAPG KGLE WMG <u>WINT</u> <u>YTGES</u> <u>TYADS</u> <u>FKGRF</u> TFSLD TSASA AYLQI NSLRA EDTAV YYCA <u>RFAIK</u> <u>GDYW</u> GQGTL LTVSS	8101	9016, 9017, 9018	DIQMTQ SPSSLSA SVGDRV TITC <u>RST</u> <u>KSLLHS</u> <u>NGITYL</u> <u>YWYQQ</u> KPGKAP KLLIY <u>Q</u> <u>MSNLAS</u> GVPSRFS SSGSGT DFTLTIS SLQPEDF ATYYCA <u>A</u> <u>QNLEIP</u> <u>RTFGQG</u> TKVEIK	8197	9266, 9267, 9265
4D5M OCB		EpCA M	EVQLV QSGPG LVQPG GSVRI SCAAS GYTFT <u>NYGM</u> <u>NWVK</u> QAPG KGLE WMG	8101	9016, 9017, 9019	DIQMTQ SPSSLSA SVGDRV TITC <u>RST</u> <u>KSLLHS</u> <u>NGITYL</u> <u>YWYQQ</u> KPGKAP KLLIY <u>Q</u> <u>MSNLAS</u> GVPSRFS	8198	9266, 9267, 9265

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>WINT</u> <u>YTGES</u> <u>TYADS</u> <u>FKGRF</u> TFSLD TSASA AYLQI NSLRA EDTAV YYCA <u>RFAIK</u> <u>GDYW</u> GQCTL LTVSS			SSGSGT DFTLTIS SLQPEDF ATYYCA <u>QNLEIP</u> <u>RTFGQG</u> TKVELK		
MEDI- 547	1C1	EphA 2	EVQLL ESGGG LVQPG GSLRL SCAAS GFTFS <u>HYMM</u> <u>AWVR</u> QAPG KGLE <u>WVSRI</u> <u>GPSG</u> <u>GPTH</u> <u>YADS</u> <u>VKGR</u> FTISR DNSK NTLYL QMNS LRAED TAVY YCAG YDSG <u>YDYV</u> <u>AVAG</u> <u>PAEYF</u> <u>QHWG</u> QGTLV TVSS	8102	9020, 9021, 9022	DIQMTQ SPSSLSA SVGDRV TITCRAS <u>QSISTW</u> <u>LAWYQ</u> QKPGKA PKLLIYK <u>ASNLHT</u> GVPSRFS GSGSGT EFSLTIS GLQPDD FATYYC <u>QQYNSY</u> <u>SRTFGQ</u> GTKVEI K	8199	9268, 9269, 9270
MORA b-003	farletu zumab	FOLR 1	EVQLV ESGGG VVQP GRSLR LSCSA <u>SGFTE</u> <u>SGYG</u>	8103	9023, 9024, 9025	DIQLTQS PSSLSAS VGDRV ITCSVSS <u>SISSNNL</u> HWYQQ KPGKAP	8200	9271, 9272, 9273

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			LSWV RQAPG KGLE WVAM <u>ISSGG</u> <u>SYTTY</u> ADSV KGRFA ISRDN AKNTL FLQM DSLRP EDTGV YFC <u>AR</u> <u>HGDD</u> <u>PAWF</u> AYWG QGTPV TVSS			KPWIIY <u>G</u> <u>TS</u> NLAS GVPSRFS GSGSGT DYTFTIS SLQPEDI ATYYC <u>Q</u> <u>QWSSYP</u> <u>YMYT</u> FG QGTKVE IK		
M9346 A	huMO V19 (vLCv 1.00)	FOLR 1	QVQL VQSG AEVV KPGAS VKISC KASG YTFT <u>G</u> <u>YFMN</u> WVKQ SPGQS LEWIG <u>RIHPY</u> <u>DGDT</u> <u>FYNQ</u> <u>KFQG</u> KATLT VDKSS NTAH MELLS LTSED FAVY YCT <u>RY</u> <u>DGSR</u> <u>AMDY</u> WGQG TTVTV SS	8104	9026, 9027, 9028	DIVLTQS PLSLAVS LGQPAII SC <u>KASQ</u> <u>SVSFAG</u> <u>TSLMH</u> WYHQKP GQQPRL LIY <u>RASN</u> <u>LEAGVP</u> DRFSGS GSKTDF TLNISPV EAEDAA TYYC <u>QQ</u> <u>SREYPY</u> <u>TFGGGT</u> KLEIK	8201	9274, 9275, 9276
M9346 A	huMO V19 (vLCv 1.60)	FOLR 1	QVQL VQSG AEVV KPGAS	8105	9026, 9027, 9029	DIVLTQS PLSLAVS LGQPAII SC <u>KASQ</u>	8202	9274, 9275, 9276

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			VKISC KASG YTFT G YFMN WVKQ SPGQS LEWIG RIHPY DGDT FYNQ KFOG KATLT VDKSS NTAH MELLS LTSED FAVY YCTRY DGSR AMDY WGQG TTVTV SS			SVSFAG TSLMH WYHQKP GQQPRL LIY RASN LEAGVP DRFSGS GSKTDF TLTISPV EAEDAA TYYC QQ SREYPY TFGGGT KLEIK		
26B3.F 2		FOLR 1	GPELV KPGAS VKISC KASD YSFT G YFMN WVMQ SHGKS LEWIG RIFPY NGDT FYNQ KFKG RATLT VDKSS STAH MELRS LASED SAVYF CARG THYF DYWG QGTTL TVSS	8106	9026, 9030, 9031	PASLSAS VGETVTI TC RTSE NIFSILA WYQQK QGISPQL LVY NAK TLAEGV PSRFSGS GSGTQF SLKINSL QPEDFG SYYC QH HYAFP WTFGGG SKLEIK	8203	9277, 9278, 9279
RG768 6	GC33	GPC3	QVQL VQSG AEVK	8107	9032, 9033, 9034	DVVM MTQ SPLSLPV TPGEPAS	8204	9280, 9281, 9282

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			KPGAS VKVSC KASG YTFT <u>D</u> <u>YEMH</u> WVRQ APGQ GLEW MG <u>AL</u> <u>DPKT</u> <u>GDTA</u> <u>YSQK</u> <u>FKGR</u> VTLTA DKSTS TAYM ELSSL TSED TAVY YCTRF <u>YSYTY</u> WGQG TLVTV SS			ISCR <u>SSQ</u> <u>SLVHSN</u> <u>GNTYLH</u> WYLQKP GQSPQL LIY <u>KVS</u> <u>NRFS</u> GV PDRFSGS GSGTDF TLKISRV EAEDVG V YYC <u>SON</u> <u>THVPPT</u> FGQGTK LEIK		
	4A6	GPC3	EVQLV QSGAE VKKP GESLK ISCKG SGYSF T <u>SYWI</u> <u>AWVR</u> QMPG KGLE WMG <u>II</u> <u>FPGDS</u> <u>DTRY</u> S <u>PSFQG</u> QVTIS ADRSI RTAYL QWSSL KASD TALYY CART <u>REGY</u> <u>FDYW</u> GQGTL VTVSS	8108	9035, 9036, 9037	EIVLTQS PGTLSLS PGERAT LSC <u>RAV</u> <u>QSVSS</u> <u>YLAWY</u> QQKPGQ APRLLIY <u>GASSRA</u> <u>TGIPDRF</u> SGSGSG TDFTLTI SRLEPED FAVYYC <u>Q</u> <u>QYGSSP</u> <u>TFGGGT</u> KVEIK	8205	9283, 9182, 9284
	11E7	GPC3	EVQLV	8109	9038,	EIVLTQS	8206	9285,

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			QSGAE VKKP GESLK ISCKG SGYSF <u>TNYW</u> <u>IAWV</u> RQMP GKGLE WMG <u>I</u> <u>YPGDS</u> <u>DTRY</u> <u>S</u> <u>PSFQG</u> QVTIS ADKSI RTAYL QWSSL KASD TAMY YCAR <u>TREG</u> <u>YFDY</u> WGQG TLVTV SS		9039, 9037	PGTLSLS PGERAT LSC <u>RAS</u> <u>QSVSSS</u> <u>YLA</u> WY QQKPGQ APRLLIY <u>GASSRA</u> <u>T</u> GIPDRF SGSGSG TDFTLTI SRLEPED FAVYYC <u>Q</u> <u>QYGSSP</u> <u>T</u> FGGGT KVEIK		9182, 9284
	16D1 0	GPC3	EVQLV QSGA DVTKP GESLK ISCKV SGYRF <u>TNYW</u> <u>IGWM</u> RQMS GKGLE WMG <u>I</u> <u>YPGDS</u> <u>DTRY</u> <u>S</u> <u>PSFQG</u> HVTIS ADKSI NTAYL RWSSL KASD TAIYY CART <u>I</u> <u>REGFF</u> <u>DY</u> WG QGTPV TVSS	8110	9040, 9039, 9041	EILLTQS PGTLSLS PGERAT LSC <u>RAS</u> <u>QSVSSS</u> <u>YLA</u> WY QQKPGQ APRLLIY <u>GASSRA</u> <u>T</u> GIPDRF SGSGSG TDFTLTI SRLEPED FAVYYC <u>Q</u> <u>QYGSSP</u> <u>T</u> FGQGT KVEIK	8207	9285, 9182, 9284

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
AMG- 595		HER1 (EGF R)	QVQL VESGG GVVQ SGRSL RLSCA <u>ASGFT</u> <u>FRNY</u> <u>GMH</u> WVRQ APGK GLEW <u>VAVI</u> <u>WYDG</u> <u>SDKY</u> <u>YADS</u> <u>VRGR</u> FTISR DNSK NTLYL QMNS LRAED TAVY YCAR DGY <u>DI</u> <u>LTGN</u> <u>PRDFD</u> <u>Y</u> WGQ GTLVT VSS	8111	9042, 9043, 9044	DTVMTQ TPLSSHV TLGQPA SISCR <u>SS</u> <u>QSLVHS</u> <u>DGNTYL</u> <u>SWLQQR</u> PGQPPRL LIY <u>RISR</u> <u>R</u> FSGVP DRFSGS GAGTDF TLEISRV EAEDVG VYY <u>CM</u> <u>QSTHVP</u> <u>RT</u> FGQG TKVEIK	8208	9286, 9287, 9288
Erubitu x™	cetutx imab	HER1 (EGF R)	QVQL KQSGP GLVQP SQLSI TCTVS <u>GFSLT</u> <u>NYGV</u> HWVR QSPGK GLEW LGVI <u>WSGG</u> <u>NTDY</u> NTPFT SRLSI NKDN SKSQV FFKM NSLQS NDTAI YYC <u>A</u>	8112	9045, 9046, 9047	DILLTQS PVILSVS PGERVSF SCRAS <u>Q</u> <u>SIGTNIH</u> WYQQRT NGSPRL LIK <u>Y</u> ASE SISGIPSR FSGSGSG TDFTLSI NSVESE DIADYY C <u>QQNNN</u> <u>WPTTFG</u> AGTKLE LK	8209	9289, 9290, 9291

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>RALT</u> <u>YYDY</u> <u>EFAY</u> WGQG TLVTV SA					
GA201	Imgat uzuma b	HER1 (EGF R)	QVQL VQSG AEVK KPGSS VKVSC KASGF TFT <u>DY</u> <u>KIHW</u> VRQAP GQGLE WMGY <u>FNPNS</u> <u>GYST</u> <u>YAQK</u> <u>FQGR</u> VTITA DKSTS TAYM ELSSL RSED TAVYY CARL <u>S</u> <u>PGGY</u> <u>YVMD</u> <u>AWGQ</u> GTTVT VSS	8113	9048, 9049, 9050	DIQMTQ SPSSLSA SVGDRV TITCR <u>AS</u> <u>QGINNY</u> <u>LNWYQ</u> QKPGKA PKRLIYN <u>TNNLQT</u> GVPSRFS GSGSGT EFTLTIS SLQPEDF ATYYCL <u>QHNSFP</u> <u>TFGQGT</u> KLEIK	8210	9292, 9293, 9294
Humax	zalutu muma b	HER1 (EGF R)	QVQL VESGG GVVQ PGRSL RLSCA ASGFT FST <u>TYG</u> <u>MHWV</u> RQAPG KGLE WVAV <u>IWDD</u> <u>GSYK</u> <u>YYGD</u> <u>SVKG</u> RFTISR DNSK	8114	9051, 9052, 9053	AIQLTQS PSSLSAS VGDRV ITCR <u>ASQ</u> <u>DISSALV</u> WYQQKP GKAPKL LIYD <u>ASS</u> <u>LESGVP</u> SRFSGSE SGTDFTL TISSLQP EDFATY YC <u>QQFN</u> <u>SYPLTF</u> GGGTKV EIK	8211	9295, 9228, 9296

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			NTLYL QMNS LRAED TAVY YCAR <u>DGIT</u> <u>MVRG</u> <u>VMKD</u> <u>YFDY</u> WGQG TLVTV SS					
IMC- 11F8	necitu muma b	HER1 (EGF R)	QVQL QESGP GLVKP SQTLS LTCTV SGGSI <u>SSGDY</u> <u>YWSW</u> IRQPP GKGLE WIG <u>YI</u> <u>YYSGS</u> <u>TDYN</u> <u>PSLKS</u> RVTM SVDTS KNQFS LKVNS VTAA DTAV YYCA <u>RVSIF</u> <u>GVGT</u> <u>FDYW</u> GQGT LTVSS	8115	9054, 9055, 9056	EIVMTQ SPATLSL SPGERA TLSC <u>RA</u> <u>SQSVSS</u> <u>YLA</u> WY QQKPGQ APRLLIY <u>DASNRA</u> <u>TG</u> IPARF SGSGSG TDFTLTI SSLEPED FAVYYC <u>HQYGST</u> <u>PLT</u> FGG GTKAEI K	8212	9202, 9203, 9297
MM- 151	P1X	HER1 (EGF R)	QVQL VQSG AEVK KPGSS VKVSC KASG GTFSS <u>Y</u> AI <u>SW</u> VRQAP GQGLE WMGS <u>I</u> PI <u>FG</u>	8116	9057, 9058, 9059	DIQMTQ SPSTLSA SVGDRV TITC <u>RAS</u> <u>QSISSW</u> <u>W</u> AWYQ QKPGKA PKLLIY <u>D</u> <u>ASSLES</u> GVPSRFS GSGSGT EFTLTIS	8213	9298, 9228, 9299

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>TVNY</u> <u>AQKF</u> <u>QGRV</u> TITAD ESTST AYME LSSLR SEDTA VYYC <u>ARDPS</u> <u>VNLY</u> <u>WYFD</u> <u>LWGR</u> GTLVT VSS			SLQPDD FATYYC <u>QQYHA</u> <u>HPTTFG</u> GGTKVE IK		
MM- 151	P2X	HER1 (EGF R)	QVQL VQSG AEVK KPGSS VKVSC KASG GTFGS <u>YAI</u> SW VRQAP GQGLE WMGS <u>IPI</u> FG <u>AANP</u> <u>AQKS</u> <u>QGRV</u> TITAD ESTST AYME LSSLR SEDTA VYYC <u>AKMG</u> <u>RGKV</u> <u>AFDI</u> W GQGT MVTV SS	8117	9057, 9060, 9061	DIVMTQ SPDSLA VSLGER ATINCK <u>S</u> <u>SQSVLY</u> <u>SPNNKN</u> <u>YLA</u> WY QQKPGQ PPKLLIY <u>WASTRE</u> <u>SGV</u> PDR FSGSGSG TDFTLTI SSLQAE DVAVYY <u>CQQYY</u> <u>GSPIT</u> FG GGTKVE IK	8214	9300, 9241, 9301
MM- 151	P3X	HER1 (EGF R)	QVQL VQSG AEVK KPGAS VKVSC KASG YAFT <u>S</u> <u>YGIN</u>	8118	9062, 9063, 9064	EIVMTQ SPATLSV SPGERA TLSC <u>RA</u> <u>SQSVSS</u> <u>NLA</u> WY QQKPGQ APRLLIY	8215	9302, 9303, 9304

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Objeto	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			WVRQ APGQ GLEW MGW <u>I</u> <u>SAYN</u> <u>GNTY</u> <u>YAQK</u> <u>LRGR</u> VTMT TDTST STAY MELRS LRSDD TAVY YCAR <u>DLGG</u> <u>YGSGS</u> <u>VPEDP</u> WGQG TLVTV SS			<u>GASTRA</u> <u>TGIPARF</u> SGSGSG TEFTLTI SSLQSED FAVYYC <u>QDYRT</u> <u>WPRRVF</u> GGGTKV EIK		
TheraC IM	nimot uzuma b	HER1 (EGF R)	QVQL QQSG AEVK KPGSS VKVSC KASG YTFT <u>N</u> <u>YYIY</u> WVRQ APGQ GLEWI G <u>GINP</u> <u>TSGGS</u> <u>NFNE</u> <u>KFKT</u> RVTIT ADESS TTAY MELSS LRSED TAFYF CTR <u>Q</u> <u>GLWF</u> <u>DSDG</u> <u>RGFD</u> <u>FWGQ</u> GTTVT VSS	8119	9065, 9066, 9067	DIQMTQ SPSSLSA SVGDRV TITC <u>RSS</u> <u>QNIVHS</u> <u>NGNTYL</u> <u>DWYQQ</u> TPGKAP KLLIY <u>K</u> <u>VSNRES</u> GVPSRFS GSGSGT DFTFTIS SLQPEDI ATYYC <u>F</u> <u>QYSHVP</u> <u>WTFGQG</u> TKLQIT	8216	9305, 9281, 9306
Vectibi	panitu	HER1	QVQL	8120	9068,	DIQMTQ	8217	9307,

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
x TM	mima b	(EGF R)	QESGP GLVKP SETLS LTCTV <u>SGGSV</u> <u>SSGDY</u> <u>YWTW</u> IRQSP GKGLE WIGH <u>I</u> <u>YYSG</u> <u>NTNY</u> NPSLK SRLTIS IDTSK TQFSL KLSSV TAADT AIYYC <u>VRDR</u> <u>VTGA</u> <u>FDIWG</u> QGTM VTVSS		9069, 9070	SPSSLSA SVGDRV TITCQAS <u>QDISNY</u> LNWYQ QKPGKA PKLLI <u>YD</u> <u>ASNLET</u> GVPSRFS GSGSGT DFTFTIS SLQPEDI ATYFC <u>Q</u> <u>HFDHLP</u> <u>LA</u> FGGG TKVEIK		9218, 9308
07D06		HER1 (EGF R)	QIQLV QSGPE LKKPG ETVKI SCKAS <u>GYTF</u> <u>TEYPI</u> HWVK QAPG KGFK WMG <u>MIYTD</u> <u>IGKPT</u> YAEFF KGRFA FSLET SASTA YLQIN NLKNE DTATY FC <u>VRD</u> <u>RYDSL</u> <u>FDYW</u> GQGT LTVSS	8121	9071, 9072, 9073	DVVM <u>TQ</u> TPLSLPV SLGDQA SISCRSS <u>QSLVHS</u> <u>NGNTYL</u> HWYLQ KPGQSP KLLI <u>YK</u> <u>VSNRFS</u> GVPDRF SGSGSG TDFTLKI SRVEAE DLGVYF <u>CSQSTH</u> <u>VPWTFG</u> GGTKLEI K	8218	9309, 9310, 9311
12D03		HER1	EMQL	8122	9074,	DVVM <u>TQ</u>	8219	9280,

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
		(EGF R)	VESGG GFVKP GGSLK LSCAA SGFAF <u>SHYD</u> <u>MSWV</u> RQTPK QRLE WVAY <u>IASGG</u> <u>DITYY</u> <u>ADTV</u> <u>KGRFT</u> ISRDN AQNTL YLQM SSLKS EDTA MFYCS <u>RSSYG</u> <u>NNGD</u> <u>ALDE</u> WGQG TSVTV SS		9075, 9076	TPLSLPV SLGDQA SISCR <u>SS</u> <u>QSLVHS</u> <u>NGNTYL</u> <u>HWYLQ</u> KPGQSP KLLIY <u>K</u> <u>VSNRFS</u> GVPDRF SGSGSG TDFTLKI SRVEAE DLGVYF <u>CSQSTH</u> <u>VLTFGS</u> GTKLEIK		9281, 9312
	C1	HER2	QVQL VESGG GLVQP GGSLR LSCAA SGFTF <u>SSYA</u> <u>MGWV</u> RQAPG KGLE WVSSI <u>SGSSR</u> <u>YIYYA</u> <u>DSVK</u> <u>GRFTI</u> SRDNS KNTLY LQMN SLRAE DTAV YYCA <u>KMDA</u> <u>SGSYF</u> <u>NFWG</u>	8123	9077, 9078, 9079	QSPSFLS AFVGDR ITITC <u>RA</u> <u>SPGIRN</u> <u>YLAWY</u> QQKPGK APKLLIY <u>AASTLQ</u> <u>SGVPSRF</u> SGSGSG TDFTLTI SSLQPED FATYYC <u>QQYNSY</u> <u>PLSF</u> GG GTKVEI K	8220	9313, 9314, 9315

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			QGTLV TVSS					
Erbicin		HER2	QVQLL QSAAE VKKP GESLK ISCKG SGYSF T <u>SYWI</u> <u>GWVR</u> QMPG KGLE WMG <u>II</u> <u>YPGDS</u> <u>DTRY</u> <u>PSFOG</u> QVTIS ADKSI STAYL QWSSL KASDT AVYY CAR <u>W</u> <u>RDSPL</u> WGQG TLVTV SS	8124	9080, 9039, 9081	QAVVTQ EPSFSVS PGGTVT LTC <u>GLS</u> <u>SGSVSTS</u> <u>YYP</u> SWY QQTPGQ APRTLIIY <u>STNTRS</u> <u>SGVPDR</u> FSGSILG NKAALT ITGAQA DDESDY YCV <u>LY</u> <u>MGSGQ</u> <u>YVFGGG</u> TKLTVL	8221	9316, 9317, 9318
Hercept in	trastuz umab	HER2	EVQLV ESGGG LVQPG GSLRL SCAAS <u>GFNIK</u> <u>DTYIH</u> WVRQ APGK GLEW VARI <u>Y</u> <u>PTNG</u> <u>YTRY</u> ADSV KGRFT ISADT SKNTA YLQM NSLRA EDTAV YYC <u>SR</u> <u>WGGD</u> <u>GFYA</u>	8125	9082, 9083, 9084	DIQMTQ SPSSLSA SVGDRV TITCRAS <u>QDVNTA</u> VAWYQ QKPGKA PKLLI <u>YS</u> <u>ASFLYS</u> GVPSRFS GSRSGT DFTLTIS SLQPEDF ATYYC <u>Q</u> <u>QHYTTP</u> <u>PTFGQG</u> TKVEIK	8222	9319, 9320, 9321

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>MDYW</u> GQGT LVTSS					
MAGH 22	marge tuxim ab	HER2	QVQL QQSGP ELVKP GASLK LSCTA <u>SGFNI</u> <u>KDTYI</u> HWVK QRPEQ GLEWI <u>GRIYP</u> <u>TNGY</u> <u>TRYDP</u> KFQD KATIT ADTSS NTAYL QVSRL TSED T AVYY <u>CSRW</u> <u>GGDG</u> <u>FYAM</u> <u>DYWG</u> QGAS VTVSS	8126	9082, 9083, 9084	DIVMTQ SHKFMS TSVGDR VSITCKA <u>SQDVNT</u> <u>AVAWY</u> QQKPGH SPKLLIY <u>SAS</u> FRYT GVPDRF TGSRS G TDFTFTI SSVQAE DLAVYY <u>CQQHYT</u> <u>TPPT</u> FG GGTKVE IK	8223	9319, 9320, 9321
MM- 302	F5	HER2	QVQL VESGG GLVQP GGSLR LSCAA SGFTF <u>RSYA</u> <u>MSWV</u> RQAPG KGLE WVSAI <u>SGRG</u> <u>DNTY</u> <u>YADS</u> <u>VKGR</u> FTISR DNSK NTLYL QMNS LRAED TAVY	8127	9085, 9086, 9087	QSVLTQ PPSVSGA PGQRVTI <u>SCTGSSS</u> <u>NIGAGY</u> <u>GVHWY</u> QQLPGT APKLLIY <u>GNTNRP</u> <u>SGVPDR</u> FSGFKSG TSASLAI TGLQAE DEADYY <u>CQFYDS</u> <u>SLSGWV</u> FGGGTK LTVL	8224	9322, 9323, 9324

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>YCAK</u> <u>MTSN</u> <u>AFAFD</u> <u>YWGQ</u> GTLVT VSS					
Perjeta	pertuz umab	HER2	EVQLV ESGGG LVQPG GSLRL SCAAS <u>GFTFT</u> <u>DYTM</u> DWVR QAPG KGLE WVAD <u>VNPNS</u> <u>GGSIY</u> NQRFK GRFTL SVDRS KNTLY LQMN SLRAE DTAV YYCA <u>RNLG</u> <u>PSFYF</u> <u>DYWG</u> QGTLV TVSS	8128	9088, 9089, 9090	DIQMTQ SPSSLSA SVGDRV TITCKAS <u>QDVSIG</u> VAWYQ QKPGKA PKLLIYS <u>ASYRYT</u> GVPSRFS GSGSGT DFTLTIS SLQPEDF ATYYCQ <u>QYYIYP</u> <u>YTFGQG</u> TKVEIK	8225	9325, 9320, 9326
MM- 121/ SAR25 6212		HER3	EVQLL ESGGG LVQPG GSLRL SCAAS GFTFS <u>HYVM</u> <u>AWVR</u> QAPG KGLE WVSSI <u>SSSGG</u> <u>WTLY</u> <u>ADSV</u> <u>KGRFT</u> ISRDN SKNTL YLQM	8129	9091, 9092, 9093	QSALTQ PASVSGS PGQSITIS <u>CTGTSS</u> <u>DVGSYN</u> <u>VVSWYQ</u> QHPGKA PKLIIYE <u>VSQRPS</u> GVSNRF SGSKSG NTASLTI SGLQTE DEADYY <u>CCSYAG</u> <u>SSIEVIF</u> GGGTKV TVL	8226	9327, 9328, 9329

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			NSLRA EDTAV YYCTR <u>GLKM</u> <u>ATIED</u> <u>YWGQ</u> GTLVT VSS					
MEHD 7945A	Dulig otuma b	HER1 (EGF R)/HE R3	EVQLV ESGGG LVQPG GSLRL SCAAS GFTLS <u>GDWI</u> <u>HWVR</u> QAPG KGLE WVGE <u>ISAAG</u> <u>GYTD</u> <u>YADS</u> <u>VKGR</u> FTISA DTSKN TAYLQ MNSL RAEDT AVYY CARE <u>S</u> <u>RVSFE</u> <u>AAMD</u> <u>YWGQ</u> GTLVT VSS	8130	9094, 9095, 9096	DIQMTQ SPSSLSA SVGDRV TITC <u>RAS</u> <u>QNIATD</u> <u>VAWYQ</u> QKPGKA PKLLIY <u>S</u> <u>ASFLYS</u> GVPSRFS GSGSGT DFTLTIS SLQPEDF ATYYC <u>Q</u> <u>QSEPEP</u> <u>YTFGQG</u> TKVEIK	8227	9330, 9331, 9332
MM- 111		HER2 /3	QVQL QESGG GLVKP GGSLR LSCAA SGFTF <u>SSYW</u> <u>MSWV</u> RQAPG KGLE WVAN <u>INRDG</u> <u>SASY</u> <u>VDSV</u> <u>KGRFT</u>	8131	9097, 9098, 9099	QSALTQ PASVSGS PGQSITIS CT <u>GTSS</u> <u>DVGGY</u> <u>NEVSWY</u> QQHPGK APKLMI Y <u>DVSDR</u> <u>PSGVSD</u> RFSGSKS GNTASLI ISGLQAD DEADYY C <u>SSYGSS</u>	8228	9333, 9334, 9335

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			ISRDD AKNSL YLQM NSLRA EDTAV YYCA <u>RDRG</u> <u>VG YF</u> <u>DL</u> WG RGTLV TVSS			<u>STHV</u> IFG GGTKVT VL		
MM- 111		HER2 /3	QVQL VQSG AEVK KPGES LKISC KGSG YSFTS <u>YWIA</u> WVRQ MPGK GLE Y MGLIY <u>PGDSD</u> <u>TKYSP</u> <u>SFQG</u> QVTIS VDKS VSTAY LQWSS LKPSD SAVYF CARH <u>DVG Y</u> <u>CTDR</u> <u>TCAK</u> <u>WPEW</u> <u>LG</u> VW GQGT L VTVSS	8132	9035, 9100, 9101	QSVLTQ PPSVSAA PGQ KVTISCS <u>GSSSNIG</u> <u>NNYVSW</u> YQQLPG TAPKLLI YDHTNR <u>PAG</u> VPD RFSGSKS GTSASL AISGFRS EDEADY YC <u>ASWD</u> <u>YT</u> LSGW <u>V</u> FGGGT KLTVL	8229	9336, 9337, 9338
	Hu3S 193	Lewis -Y	EVQLV ESGGG VVQP GRSLR LSCST SGFTF <u>SDYY</u> <u>MY</u> WV RQAPG KGLE	8133	9102, 9103, 9104	DIQMTQ SPSSLSA SVGDRV TITCRSS <u>QRIVHS</u> <u>NGNTYL</u> <u>E</u> WYQQT PGKAPK LLIY <u>KVS</u> <u>NR</u> FSGV	8230	9339, 9281, 9340

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			WVAY <u>MSNV</u> <u>GAITD</u> <u>YPDT</u> <u>VKGR</u> FTISR DNSK NTLFL QMDS LRPED TGVYF CARG <u>TRDG</u> <u>SWFA</u> <u>YWGQ</u> GTPVT VSS			PSRFSGS GSGTDF TFTISSL QPEDIAT YYCEQG <u>SHVPFT</u> FGQGTK LQIT		
BAY 94- 9343	anetu mab ravtan sine	Mesot helin	QVELV QSGAE VKKP GESLK ISCKG <u>SGYSE</u> <u>TSYWI</u> GWVR QAPG KGLE WMGII <u>DPGDS</u> <u>RTRY</u> PSFQG QVTIS ADKSI STAYL QWSSL KASDT AMYY CARG <u>QLYG</u> <u>GTYM</u> <u>DGWG</u> QGTLV TVSS	8134	9105, 9106, 9107	DIALTQP ASVSGSP GQSITIS CTGTSS <u>DIGGYN</u> <u>SVSWYQ</u> QHPGKA PKLMIY <u>GVNNRP</u> SGVSNR FSGSKSG NTASLTI SGLQAE DEADYY <u>CSSYDIE</u> <u>SATPVF</u> GGGTKL TVL	8231	9341, 9342, 9343
	SS1	Mesot helin	QVQL QQSGP ELEKP GASV KISCK ASGYS FTGYT	8135		DIELTQS PAIMSAS PGEKVT MTCSAS SSVSYM HWYQQ KSGTSP	8232	

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Objeto	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			MNVV KQSH GKSLE WIGLI TPYNG ASSYN QKFRG KATLT VDKSS STAY MDLLS LTSED SAVYF CARG GYDG RGFDY WGQG TTVTV SS			KRWIYD TSKLAS GVPGRF SGSGSG NSYSLTI SSVEAE DDATYY CQQWSG YPLTFG AGTKLEI K		
		Mesot helin	QVYL VESGG GVVQ PGRSL RLSCA ASGIT FSIYG MHWV RQAPG KGLE WVAV IWYD GSHE YYAD SVKG RFTISR DNSK NTLYL LMNSL RAED TAVY YCAR DGDY YDSGS PLDY WGQG TLVTV SS	8136	9108, 9109, 9110	EIVLTQS PATLSLS PGERAT LSCRAS QSVSSY LAWYQ QKPGQA PRLLIYD ASNRAT GIPARFS GSGSGT DFTLTIS SLEPEDF AVYYCQ Q RSNWPL TFGGGT KVEIK	8233	9202, 9203, 9233
		Mesot helin	QVHL VESGG GVVQ	8137	9111, 9112, 9110	EIVLTQS PATLSLS PGERAT	8233	9202, 9203, 9233

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			PGRSL RLSCV ASGIT <u>FRIYG</u> <u>MHWV</u> RQAPG KGLE WVA <u>V</u> <u>LWYD</u> <u>GSHE</u> <u>YYAD</u> <u>SVKG</u> RFTISR DNSK NTLYL QMNS LRAED TAIYY CARD <u>GDYY</u> <u>DSGSP</u> <u>LDYW</u> GQCTL VTVSS			LSCR <u>AS</u> <u>QSVSSY</u> <u>LAWYQ</u> QKPGQA PRLLIY <u>D</u> <u>ASNRAT</u> GIPARFS GSGSGT DFTLTIS SLEPEDF AVYYC <u>Q</u> <u>Q</u> <u>RSNWPL</u> <u>TFGGGT</u> KVEIK		
		Mesot helin	EVHLV ESGGG LVQPG GSLRL SCAAS GFTFS <u>RYWM</u> <u>SWVR</u> QAQG KGLE WVA <u>SI</u> <u>KQAG</u> <u>SEKTY</u> <u>VDSV</u> <u>KGRFT</u> ISRDN AKNSL SLQM NSLRA ED TAVY YCAR <u>EGAY</u> <u>YYDS</u> <u>ASYYP</u> <u>YYYY</u>	8138	9113, 9114, 9115	EIVLTQS PGTSLS PGERAT LSCR <u>AS</u> <u>QSVSSS</u> <u>YLAWY</u> QQKPGQ APRLLIY <u>GASSRA</u> <u>TGIPDRF</u> SGSGSG TDFTLTI SRLEPED FAVYYC <u>Q</u> <u>QYGSSQ</u> <u>YTFGQG</u> TKLEIK	8234	9285, 9182, 9344

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>YSMD</u> <u>VWGQ</u> GTTVT VSS					
MORA b-009	amatu ximab	Mesot helin	QVQL QQSGP ELEKP GASV KISCK ASGYS FT <u>GYT</u> <u>MNWV</u> KQSH GKSLE WIG <u>LI</u> <u>TPYN</u> <u>GASSY</u> <u>NQKF</u> <u>RGKA</u> TLTVD KSSST AYMD LLSLT SEDSA VYFCA <u>RGGY</u> <u>DGRG</u> <u>FDYW</u> GSGTP VTVSS	8139	9116, 9117, 9118	DIELTQS PAIMSAS PGEKVT MTC <u>SAS</u> <u>SSVSYM</u> <u>HWYQQ</u> KSGTSP KRWI <u>YD</u> <u>TSKLAS</u> GVPGRF SGSGSG NSYSLTI SSVEAE DDATYY <u>CQQWS</u> <u>KHPLTF</u> GSGTKV EIK	8235	9345, 8909, 9346
hPAM4		MUC- 1	EVQLQ ESGPE LVKPG ASVK MSCK ASGYT FPS <u>YV</u> <u>LHWV</u> KQKP GQGLE WIG <u>YI</u> <u>NPYN</u> <u>DGTQ</u> <u>YNEK</u> <u>FKGK</u> ATLTS DKSSS TAYM ELSRL TSED	8140	9119, 9120, 9121	DIVMTQ SPAIMSA SPGEKV TMT <u>CSA</u> <u>SSSVSSS</u> <u>YLYWY</u> QQKPGS SPKLWI Y <u>STSNL</u> <u>ASGVPA</u> RFSGSGS GTSYSLT ISSMEAE DAASYF <u>CH</u> <u>QOWNRY</u> <u>PYT</u> FGG GTKLEIK	8236	9347, 9348, 9349

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			SAVY YCAR <u>GFGG</u> <u>SYGFA</u> <u>YWGQ</u> GTLIT VSA					
hPAM4 -Cide	clivat uzuma b	MUC 1	QVQL QQSG AEVK KFGAS VKVSC EASGY TFPS <u>Y</u> <u>VLHW</u> VKQA PGQGL EWIG <u>Y</u> <u>INPYN</u> <u>DGTQ</u> <u>TNKK</u> <u>FKGK</u> ATLTR DTSIN TAYM ELSRL RSDDT AVYY CARG <u>FGGS</u> <u>YGFA</u> <u>YNGQ</u> GTLVT VSS	8141	9119, 9122, 9121	DIQLTQS PSSLSAS VGDRV TMTCS <u>SAS</u> <u>SSVSSSY</u> <u>LYWYQ</u> QKPGKA PKLWIY <u>STSNLA</u> <u>SGV</u> PAR FSGSGSG TDFTLTI SSLQPED SASYFC <u>HQWNR</u> <u>YPYTFG</u> GGTRLEI K	8237	9347, 9348, 9349
SAR56 6658	huDS 6v1.0 1	MUC 1	QAQL QVSG AEVV KPGAS VKMS CKASG YTFT <u>S</u> <u>YNMH</u> WVKQ TPGQG LEWIG <u>YIYPG</u> <u>NGAT</u> <u>NYNQ</u> <u>KFOG</u> KATLT	8142	8947, 9123, 9124	EIVLTQS PATMSA SPGERV TITCS <u>SAH</u> <u>SSVSFM</u> <u>HW</u> FQQ KPGTSP KLWIY <u>S</u> <u>TSSLAS</u> GVPARF GGSGSG TSYSLTI SSMEAE DAATYY <u>CQQRSS</u> <u>FPLTFG</u>	8238	9350, 9351, 9352

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			ADTSS STAY MQISS LTSED SAVYF CARG <u>DSVPE</u> <u>AYWG</u> QGTLV TVSA			AGTKLE LK		
Therag yn	Pemtu moma b muH MFG1	MUC 1	QVQL QQSG AELM KPGAS VKISC KATG YTFSA <u>YWIE</u> WVKQ RPGHG LEWIG <u>EILPG</u> <u>SNNSR</u> <u>YNEK</u> <u>FKGK</u> ATFTA DTSSN TAYM QLSSL TSEDS AVYY CSRSY <u>DFAW</u> <u>FAYW</u> GQGTP VTVSA	8143	9125, 9126, 9127	DIVMSQ SPSSLAV SVGEKV TMSCK <u>S</u> <u>SQSLLY</u> <u>SSNQKI</u> <u>YLAWY</u> QQKPGQ SPKLLIY W <u>ASTRE</u> <u>SGVPDR</u> FTGGGS GTDFTL TISSVKA EDLAVY YC <u>QQYY</u> <u>RYPRTF</u> GGGTKL EIK	8239	9353, 9354, 9355
Therex	Sontu zumab huHM FG1 AS14 02 R1150	MUC 1	QVQL VQSG AEVK KPGAS VKVSC KASG YTFSA <u>YWIE</u> WVRQ APGK GLEW VGE <u>IL</u> <u>PGSNN</u> <u>SRYNE</u>	8144	9125, 9126, 9127	DIQMTQ SPSSLSA SVGDRV TITCK <u>S</u> <u>QSLLYS</u> <u>SNQKIY</u> <u>LAWYQ</u> QKPGKA PKLLIY W <u>ASTRE</u> <u>SGVPSRF</u> SGSGSG TDFTFTI SSLQPED	8240	9353, 9354, 9355

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			KFKG RVTVT RDTST NTAY MELSS LRSED TAVY YCARS YDFA WFAY WGQG TLVTV SS			IATYYC QQYYR YPRTFG QGTKVE IK		
<i>MDX- 1105 or BMS- 936559</i>		PD- L1	QVQL VQSG AEVK KPGSS VKVSC KTSGD TFST Y AISWV RQAPG QGLE WMGG IPIFG KAHY AQKE QGRV TITAD ESTST AYME LSSLR SEDTA VYFCA RKFH FVSGS PFGM DVWG QGTTV TVSS	8145	9128, 9129, 9130	EIVLTQS PATLSLS PGERAT LSCRAS QSVSSY LAWYQ QKPGQA PRLLIY D ASNRAT GIPARFS GSGSGT DFTLTIS SLEPEDF AVYYC Q QRSNWP TFGQGT KVEIK	8241	9202, 9203, 9356
MEDI- 4736	durval umab	PD- L1	EVQLV ESGGG LVQPG GSLRL SCAAS GFTFS RYWM SWVR QAPG KGLE	8146	9131, 9132, 9133	EIVLTQS PGTSLSL PGERAT LSCRAS QRVSSS YLA WY QQKPGQ APRLLIY DASSRA TGIPDRF	8242	9357, 9218, 9358

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			WVAN <u>IKQD</u> <u>GSEK</u> YYVD SVKGR FTISR DNAK NSLYL QMNS LRAED TAVY YCAR <u>EGGW</u> <u>FGEL</u> <u>AFDY</u> WGQG TLVTV SS			SGSGSG TDFTLTI SRLEPED FAVYYC <u>QQYGSL</u> <u>PWTFGQ</u> GTKVEI K		
MPDL 3280A	atezoli zumab	PD- L1	EVQLV ESGGG LVQPG GSLRL SCAAS <u>GFTFS</u> <u>DSWIH</u> WVRQ APGK GLEW VAWIS <u>PYGG</u> <u>STYYA</u> DSVK GRFTI SADTS KNTA YLQM NSLRA EDTAV YYCA <u>RRHW</u> <u>PGGF</u> <u>DYWG</u> QGTLV TVSS	8147	9134, 9135, 9136	DIQMTQ SPSSLSA SVGDRV TITCRAS <u>QDVSTA</u> VAWYQ QKPGKA PKLLIYS <u>ASFLYS</u> GVPSRFS GSGSGT DFTLTIS SLQPEDF ATYYCQ <u>QYLYHP</u> <u>ATFGQG</u> TKVEIK	8243	9359, 9320, 9360
MSB00 10718C	avelu mab	PD- L1	EVQLL ESGGG LVQPG GSLRL SCAAS <u>GFTFS</u>	8148	9137, 9138, 9139	QSALTQ PASVSGS PGQSITIS CTGT <u>SS</u> <u>DVGGY</u> <u>NYVSWY</u>	8244	9361, 9362, 9363

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>SYIM</u> MWVR QAPG KGLE WVSSI <u>YPSG</u> <u>GITFY</u> ADTV KGRFT ISRDN SKNTL YLQM NSLRA EDTAV YYCA <u>RIKLG</u> <u>TVTT</u> <u>VDYW</u> GQGT LTVSS			QQHPGK APKLMI Y <u>DV</u> SNR PSGVSN RFSGSKS GNTASL TISGLQA EDEADY YC <u>SSYT</u> <u>SSSTRV</u> F GTGTKV TVL		
MLN59 1		PSM A	EVQLV QSGPE VKKP GATV KISCK TSGYT <u>FTEYT</u> <u>IHWV</u> KQAP GKGLE WIGNI <u>NPNN</u> <u>GGTT</u> <u>YNQK</u> <u>FEDK</u> ATLTV DKSTD TAYM ELSSL RSED TAVYY CAAG <u>WNFD</u> <u>YWGQ</u> GTLT VSS	8149	9140, 9141, 9142	DIQMTQ SPSSLST SVGDRV TLTCKA <u>SQDVGT</u> <u>AVDWY</u> QQKPGP SPKLLIY <u>WASTR</u> <u>HTGIPSR</u> FSGSGSG TDFTLTI SSLQPED FADYYC <u>QQYNSY</u> <u>PLTFGP</u> GTKVDI K	8245	9364, 9365, 9322
MT112	pasotu xizum ab	PSM A	QVQL VESGG GLVKP GESLR	8150	9143, 9144, 9145	DIQMTQ SPSSLSA SVGDRV TITCKAS	8246	9366, 9320, 9367

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			LSCAA <u>SGFTE</u> <u>SDYY</u> MYWV RQAPG KGLE WVAII <u>SDGG</u> <u>YYTY</u> YSDIIK GRFTI SRDNA KNSLY LQMN SLKAE DTAV YYCA <u>RGFPL</u> <u>LRHG</u> <u>AMDY</u> WGQG TLVTV SS			<u>QNVDTN</u> VAWYQ QKPGQA PKSLIYS <u>ASYRYS</u> DVPSRFS GSASGT DFTLTIS SVQSED FATYYC <u>QQYDSY</u> <u>PYTFGG</u> GTKLEIK		
		ROR1	QEQLV ESGGR LVTPG GSLTL SCKAS GFDFS <u>AYYM</u> <u>SWVR</u> QAPG KGLE WIATI <u>YPSSG</u> <u>KTY Y</u> <u>ATWV</u> <u>NGRFT</u> ISSDN AQNT VDLQ MNSLT AAD RATYF CARDS <u>YADD</u> <u>GALE</u> <u>NIWGP</u> GTLVT ISS	8151	9146, 9147, 9148	ELVLTQ SPSVSAA LGSPAKI TCTLSS <u>AHKTDI</u> <u>IDWYQQ</u> LQGEAP RYLMQV QSDGSY <u>TKRPGV</u> PDRFSGS SSGADR YLIIPSV QADDEA DY YCGADY <u>IGGYVF</u> GGGTQL TVTG	8247	9368, 9369, 9370

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
		ROR1	EVKLV ESGGG LVKPG GSLKL SCAAS GFTFS <u>SYAM</u> <u>SWVR</u> QIPEK RLEW <u>VASIS</u> <u>RGGT</u> <u>TYYP</u> <u>DSVK</u> <u>GRFTI</u> SRDNV RNILY LQMSS LRSED T AMYY CGR <u>Y</u> <u>DYDG</u> <u>YYAM</u> <u>DYWG</u> QGTSV TVSS	8152	9085, 9149, 9150	DIKMTQ SPSSMY ASLGER VTIT <u>C</u> <u>ASPDINS</u> <u>YLSWFQ</u> QKPGKS PKTLI <u>R</u> <u>ANRLVD</u> GVPSRFS GGGSGQ DYSLTIN SLEYED MGIYYC <u>LQ</u> <u>YDEFPY</u> <u>TFGGGT</u> KLEMK	8248	9371, 9372, 9373
		ROR1	QSLEE SGGRL VTPGT PLTLT CTVSG IDL <u>N</u> <u>HWMS</u> WVRQ APGK GLEWI <u>GIIAA</u> <u>SGSTY</u> <u>YANW</u> <u>AKGR</u> FTISKT STTVD LRIAS PTTED TATY FCAR <u>D</u> <u>YGDY</u> <u>RLVTF</u> <u>NIWGP</u>	8153	9151, 9152, 9153	ELVMTQ TPSSVSA AVGGTV TIN <u>Q</u> <u>AS</u> <u>QSIGSY</u> <u>LAWYQ</u> QKPGQP PKLLI <u>Y</u> <u>ASNLAS</u> GVPSRFS GSGSGT EYTLTIS GVQRED AATYYC <u>LG</u> <u>SLSNSD</u> <u>NVFGGG</u> TELEIL	8249	9374, 9375, 9376

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			GTLVT VSS					
		ROR1	QSVKE SEGD VTPAG NLTLT CTASG SDIND <u>YPISW</u> VRQAP GKGLE WIGFI <u>NSGGS</u> <u>TWYA</u> <u>SWVK</u> <u>GRFTI</u> SRTST TVDLK MTSLT TDDTA TY FCARG <u>YSTYY</u> <u>CDFNI</u> WGPG TLVTI SS	8154	9154, 9155, 9156	ELVMTQ TPSSTSG AVGGTV TINC <u>QAS</u> <u>QSIDSNL</u> <u>AWFQQK</u> PGQPPTL LIY <u>RASN</u> <u>LASGVP</u> SRFSGSR SGTEYT LTISGVQ REDAAT YYC <u>LG</u> <u>GVGNVS</u> <u>YRTSFG</u> GGTEVV VK	8250	9377, 9378, 9379
CC49 (Huma- nized)		TAG- 72	QVQL VQSG AEVV KPGAS VKISC KASG YTFTD <u>HAIH</u> WVKQ NPGQR LEWIG <u>YFSPG</u> <u>NDDF</u> <u>KYNE</u> <u>RFKG</u> KATLT ADTSA STAYV ELSSL RSED AVYFC TR <u>SLN</u> <u>MAYW</u>	8155	9157, 9158, 9013	DIVMSQ SPDSLA VSLGER VTLNCK SS <u>QSL</u> <u>YSGNQK</u> <u>NYLAW</u> YQQKPG QSPKLLI <u>YWASA</u> <u>RESGVP</u> DRFSGS GSGTDF TLTISSV QAEDVA VYYC <u>QQ</u> <u>YYSYPL</u> <u>TFGAGT</u> KLELK	8251	9380, 9381, 9262

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			GQGT LVT VSS					
	Murine A1	TPBG /5T4	QIQLV QSGPE LKKPG ETVKI SCKAS GYTFT NFGM NWVK QGPGE GLKW MGWI NTNTG EPRYA EEFKG RXAFS LETTA STAYL QINN LNEDT ATYFC ARDW DGAY FFDY WGQG TTLTV SS	8156	9159, 9160, 9161	SIVMTQ TPKFLLV SAGDRV TITCKAS <u>QSVSND</u> <u>VAWYQ</u> QKPGQS PKLLINE <u>ATNRYT</u> GVPNRF TGSGYG TDFTFTI STVQAE DLALYF <u>CQQ</u> <u>DYSSPW</u> <u>TFGGGT</u> KLEIK	8252	9382, 9383, 9384
	Murine A2	TPBG /5T4	QVQL QQSRP ELVKP GASV KMSC KASG YTFTD YVISW VKQR TGQGL EWIGE IYPGS NSIYY NEKFK GRATL TA DKSSS TAYM QLSSL TSEDS AVYFC AMGG	8157	9162, 9163, 9164	SVIMSR GQIVLT QSPAIMS ASLGER VTLTCT <u>ASSSVNS</u> <u>NYLHW</u> YQQKPG SSPKLWI Y <u>STSNL</u> <u>ASGVPA</u> RFSGSGS GTSYSLT ISSMEAE DAATYY <u>CHQYH</u> <u>RSPLTF</u> GAGTKL ELK	8253	9385, 9348, 9386

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			NYGF DYWG QGTTL TVSS					
	Murine A3	TPBG /5T4	EVQLV ESGGG LVQPK GSLKL SCAAS GFTFN TYAM NWVR QAPG KGLE WVARI RSKSN NYAT YYAD SVKDR FTISR DDSQS MLYL QMNN LKTED TAMY XCVR QWDY DVRA MNYW GQGS VTVSS	8158	742, 9165, 9166	DIVMTQ SHIFMST SVGDRV SITCK <u>KAS</u> <u>QDVDTA</u> <u>VAWYQ</u> QKPGQS PKLLIY <u>WASTRL</u> <u>TGVPDR</u> FTGSGS GTDFTL TISNVQS EDLADY FC <u>QQ</u> <u>YSSYPY</u> <u>TFGGGT</u> KLEIK	8254	9387, 9388, 9389
IMMU- 132	hRS-7	TROP -2	QVQL QQSGS ELKKP GASV KVSCK ASGYT FT <u>NYG</u> <u>MNWV</u> KQAP GQGL KWMG <u>WINT</u> <u>YTGE</u> <u>PTYTD</u> <u>DFKG</u> RFAFS LDTSV STAYL QISSL	8159	9016, 9167, 9168	DIQLTQS PSSLSAS VGDRVS ITCK <u>KAS</u> <u>QDVSIA</u> <u>VAWYQ</u> QKPGKA PKLLIY <u>S</u> <u>ASYRYT</u> GVPDRF SGSGSG TDFTLTI SSLQPED FAVYYC <u>QQHYIT</u> <u>PLTFGA</u> GTKVEI K	8255	9390, 9391, 9392

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			KADD TAVYF CARG <u>GFGSS</u> <u>YWYF</u> <u>DVWG</u> QGSLV TVSS					
IMC- 18F1	icrucu mab	VEGF R1	QAQV VESGG GVVQ SGRSL RLSCA <u>ASGFA</u> <u>FSSYG</u> <u>MHW</u> VRQAP GKGLE WVAY <u>IWYD</u> <u>GSNK</u> <u>YYAD</u> <u>SVRG</u> RFTISR DNSEN TLYLQ MNSL RAEDT AVYY CARD <u>HYGS</u> <u>GVHH</u> <u>YFYY</u> <u>GLDV</u> WGQG TTVTV SS	8160	9169, 9170, 9171	EIVLTQS PGTSLS PGERAT LSC <u>RAS</u> <u>QSVSSS</u> <u>YLAWY</u> QQKPGQ APRLLIY <u>GASSRA</u> <u>TGIPDRF</u> SGSGSG TDFTLTI SRLEPED FAVYYC <u>QQYGSS</u> <u>PLTFGG</u> GTKVEI K	8256	9285, 9182, 9393
Cyramz a	ramuc iruma b	VEGF R2	EVQLV QSGG GLVKP GGSLR LSCAA <u>SGFTE</u> <u>SSYSM</u> <u>NWVR</u> QAPG KGLE WVSSI <u>SSSSS</u> <u>YIYYA</u>	8161	9172, 9173, 9174	DIQMTQ SPSSVSA SIGDRV ITC <u>RASQ</u> <u>GIDNWL</u> GWYQQ KPGKAP KLLIYD <u>ASNLDI</u> GVPSRFS GSGSGT YFTLTIS SLQAED	8257	9394, 9395, 9396

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen-cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			<u>DSVK</u> <u>GRFTI</u> SRDNA KNSLY LQMN SLRAE DTAV YYCA <u>RVTD</u> <u>AFDIW</u> GQGT MVTV SSA			FAVYFC <u>QQAKAF</u> <u>PPTFGG</u> GTKVDI K		
g165D FM- PEG	alaciz umab pegol	VEGF R2	EVQLV ESGGG LVQPG GSLRL SCAAS <u>GFTFS</u> <u>SYGM</u> SWVR QAPG KGLE WVATI <u>TSGGS</u> <u>YTY</u> VDSV KGRFT ISRDN AKNTL YLQM NSLRA EDTAV YYCV <u>RIGED</u> <u>ALDY</u> WGQG TLVTV SS	8162	9175, 9176, 9177	DIQMTQ SPSSLSA SVGDRV TITCRAS <u>QDIAGS</u> LNWLQQ KPGKAI KRLIYA <u>TSS</u> LD SG VPKRFS GSRSGS DYTLTIS SLQPEDF ATYYCL <u>QYGSFP</u> <u>PTFGQG</u> TKVEIK	8258	9397, 9212, 9398
Imclon e6.64		VEGF R2	KVQL QQSGT ELVKP GASV KVSCK ASGYI FTEYII HWVK QRSGQ GLEWI GWLY	8163		DIVLTQS PASLAV SLGQRA TISCRAS ESVDSY GNSFMH WYQQKP GQPPKL LIYRASN LESGIPA RFSGSGS	8259	

Nom- bre comer- cial	Nom- bre del anti- cuer- po	Obje- tivo	Secuen- cia de VH	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)	Secuen- cia de VL	SEQ ID NO: (longi- tud com- pleta)	SEQ ID NO: (CDR)
			PESNII KYNE KFKD KATLT ADKSS STVY MELSR LTSED SAVYF CTRHD GTNFD YWGQ GTTLT VSSA			RTDFTLT INPVEAD DVATYY CQQSNE DPLTFG AGTKLE LK		

* las secuencias subrayadas & en negrita, si están presentes, son CDR dentro del VL y VH

La molécula de adhesión de células epiteliales (EpCAM, también conocida como antígeno 17-1A) es una glucoproteína integrada en la membrana de 40 kDa compuesta por 314 aminoácidos que se expresa en ciertos epitelios y en muchos carcinomas humanos (*ver*, Balzar, The biology of the 17-1A antigen (Ep-CAM), J. Mol. Med. 1999, 77:699-712). Debido a su origen de células epiteliales, las células tumorales de la mayoría de los carcinomas expresan EpCAM en su superficie (más que las células sanas normales), que incluyen la mayoría de las células de carcinoma de pulmón de células no pequeñas primarias, metastásicas y diseminadas (Passlick, B. y otros, The 17-1A antigen is expressed on primary, metastatic and disseminated non-small cell lung carcinoma cells. Int. J. Cancer 87(4):548–552, 2000), adenocarcinomas gástricos y de la unión gastroesofágica (Martin, I.G., Expression of the 17-1A antigen in gastric and gastro-oesophageal junction adenocarcinomas: a potential immunotherapeutic target? J Clin Pathol 1999;52:701–704), y cáncer de mama y colorrectal (Packeisen J, y otros Detection of surface antigen 17-1A in breast and colorectal cancer. Hybridoma. 1999 18(1):37-40) en cáncer de mama, la sobreexpresión de EpCAM en células tumorales es un predictor de supervivencia (Gastl, Lancet. 2000, 356, 1981-1982). Debido a su origen en células epiteliales, las células tumorales de la mayoría de los carcinomas expresan EpCAM en su superficie.

En una modalidad proporcionada en la presente descripción, se proporcionan composiciones de ensamblaje de polipéptidos quiméricos biespecíficos con una primera porción que tiene un dominio de unión específico para EpCAM y un dominio de unión específico para CD3. El problema técnico a resolver era proporcionar medios y métodos para la generación de composiciones mejoradas que exhibieran las propiedades de ser

medicamentos bien tolerados y más convenientes (dosificación menos frecuente) para el tratamiento efectivo y/o la mejora de enfermedades tumorales. La solución a dicho problema técnico se logra mediante las modalidades descritas en la presente descripción y caracterizadas en las reivindicaciones.

En consecuencia, en algunas modalidades, la presente invención se refiere a composiciones de ensamblaje de polipéptidos quiméricos de manera que dicha composición comprende una primera porción que comprende una composición de anticuerpo de cadena sencilla biespecífico que comprende al menos dos dominios de unión, de manera que uno de dichos dominios se une a un antígeno de célula efectora, tal como el antígeno CD3 y un segundo dominio se une al antígeno EpCAM, en donde dichos dominios de unión comprenden VL y VH específicos para EpCAM y VL y VH específicos para el antígeno CD3 humano. Preferentemente, en la modalidad, dicho dominio de unión específico para EpCAM tiene un valor de K_d mayor que 10^{-7} a 10^{-10} M, según se determina en un ensayo de unión in vitro. En una modalidad de lo anterior, los dominios de unión están en formato scFv. En otra modalidad de lo anterior, los dominios de unión están en un formato de diacuerpo de cadena sencilla.

En algunas modalidades, la invención proporciona composiciones de ensamblaje de polipéptidos quiméricos que comprenden un dominio de unión de la primera porción con afinidad de unión por un marcador específico de tumor y un segundo dominio de unión se une a un antígeno de célula efectora, tal como el antígeno CD3. Los marcadores específicos de tumor que comprenden estas modalidades de la invención incluyen, pero no se limitan a, CCR5, CD19, HER-2, HER-3, HER-4, EGFR, PSMA, CEA, MUC1, MUC2, MUC3, MUC4, MUC5AC, MUC5B, MUC7, β hCG, Lewis-Y, CD-20, CD33, CD30, gangliósido GD3, 9-O-Acetil-GD3, Globo H, fucosil GM1, GD-2, anhidrasa carbónica IX, CD44v6, Sonic Hedgehog, Wue-1, antígeno de células plasmáticas 1, proteoglicano de sulfato de condroitina de melanoma, CCR8, antígeno epitelial 6-transmembrana de próstata (STEAP), mesotelina, antígeno A33, antígeno de células madre de próstata (PSCA), LY-6, SAS, desmogleína 4, receptor de acetilcolina fetal, CD-25, antígeno de cáncer 19-9 (CA 19-9), antígeno de cáncer 125 (CA-125), receptor de sustancia inhibidora de Mueller tipo II (MISIIR), antígeno Tn sialilado, antígeno de activación de fibroblastos (FAP), endosialina (CD248), variante III del receptor del factor de crecimiento epidérmico (EGFRvIII), antígeno L6 asociado a tumor (TAL6), CD-63, TAG-72, antígeno de Thomsen-Friedenreich (antígeno TF), receptor del factor I de crecimiento similar a la insulina (IGF-IR), antígeno Cora, CD7, CD22, CD79a, CD79b, G250, F19, EphA2 y MT-MM. En ciertas modalidades, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos que comprende un dominio de unión de la primera porción con afinidad de unión por el marcador específico de tumor que comprende secuencias VL y VH antimarcador. Ejemplos no limitantes ilustrativos de secuencias VL y VH específicas para ciertos de estos marcadores tumorales se exponen en la Tabla 6f. En otras modalidades, la invención proporciona una composición de ensamblaje de polipéptidos quiméricos que

comprende un dominio de unión de la primera porción con afinidad de unión por el marcador específico de tumor que comprende la región CDR-L1, la región CDR-L2, la región CDR-L3, la región CDR-H1 la región CDR-H2 y la región CDR-H3, en donde cada una se deriva de las secuencias VL y VH respectivas. Preferentemente, en las modalidades, dicha unión tiene un valor de K_D superior a 10^{-7} a 10^{-10} M, según lo determinado en un ensayo de unión *in vitro*.

Se contempla específicamente que la composición de ensamblaje de polipéptidos quiméricos puede comprender cualquiera de los dominios de unión anteriores o variantes de secuencia de estos siempre que las variantes exhiban especificidad de unión por el antígeno descrito. En una modalidad, una variante de secuencia se crearía por sustitución de un aminoácido en la secuencia VL o VH con un aminoácido diferente. En las variantes de eliminación, se eliminan uno o más residuos de aminoácidos en una secuencia VL o VH como se describe en la presente descripción. Las variantes de eliminación, por lo tanto, incluyen todos los fragmentos de una secuencia polipeptídica del dominio de unión. En las variantes de sustitución, se eliminan uno o más residuos de aminoácidos de un polipéptido VL o VH (o CDR) y se reemplazan con residuos alternativos. En un aspecto, las sustituciones son de naturaleza conservativa y las sustituciones conservativas de este tipo se conocen bien en la técnica. Además, se contempla específicamente que las composiciones que comprenden el primer y segundo dominios de unión descritos en la presente descripción pueden utilizarse en cualquiera de los métodos descritos en la presente descripción.

Conformación no estructurada

Típicamente, el componente polipeptídico XTEN de las proteínas de fusión descritas en la presente descripción se diseña de modo que se comporte como secuencias peptídicas desnaturalizadas en condiciones fisiológicas, a pesar de la longitud extendida del polímero. "Desnaturalizado" describe el estado de un péptido en solución que se caracteriza por una gran libertad conformacional de la cadena principal peptídica. La mayoría de los péptidos y proteínas adoptan una conformación desnaturalizada en presencia de altas concentraciones de desnaturalizantes o a temperatura elevada. Los péptidos en conformación desnaturalizada tienen, por ejemplo, espectros característicos de dicroísmo circular (CD) y se caracterizan por una falta de interacciones de largo alcance según lo determinado por NMR. "Conformación desnaturalizada" y "conformación no estructurada" se utilizan como sinónimos en la presente descripción. En algunos casos, la invención proporciona polipéptidos XTEN que, en condiciones fisiológicas, pueden parecerse a secuencias desnaturalizadas en gran parte desprovistas de estructura secundaria. En otros casos, los polipéptidos XTEN pueden estar sustancialmente desprovistos de estructura secundaria en condiciones fisiológicas. "En gran parte desprovisto", como se usa en este contexto, significa que menos del 50 % de los residuos de aminoácidos de XTEN de cada polipéptido XTEN contribuyen a la estructura secundaria medida o determinada por los medios descritos en la presente descripción. "Sustancialmente desprovisto", como se usa en este

contexto, significa que al menos aproximadamente 60 %, aproximadamente 70 %, aproximadamente 80 %, aproximadamente 90 %, aproximadamente 95 %, o aproximadamente 99 % de los residuos de aminoácidos de XTEN de la secuencia XTEN no contribuyen a la estructura secundaria, medida o determinada por los medios descritos en la presente descripción.

Se han establecido una variedad de métodos en la técnica para discernir la presencia o ausencia de estructuras secundarias y terciarias en un polipéptido dado. En particular, la estructura secundaria de XTEN puede medirse espectrofotométricamente, por ejemplo, mediante espectroscopía de dicroísmo circular en la región espectral de "UV lejano" (190-250 nm). Los elementos de la estructura secundaria, tales como la hélice alfa y la lámina beta, dan lugar cada uno a una forma y magnitud características de los espectros de CD. La estructura secundaria también puede predecirse para una secuencia polipeptídica a través de ciertos programas o algoritmos de computadora, tales como el bien conocido algoritmo de Chou-Fasman (Chou, PY, y otros (1974) *Biochemistry*, 13: 222-45) y el algoritmo Garnier-Osguthorpe-Robson ("GOR") (Garnier J, Gibrat JF, Robson B. (1996), el método GOR para predecir la estructura secundaria de proteínas a partir de la secuencia de aminoácidos. *Methods Enzymol* 266: 540-553), como se describe en la publicación de solicitud de patente de Estados Unidos núm. 20030228309A1. Para una secuencia dada, los algoritmos pueden predecir si existe alguna o ninguna estructura secundaria, expresada como el total y/o el porcentaje de residuos de la secuencia que forman, por ejemplo, hélices alfa o láminas beta o el porcentaje de residuos de la secuencia que se predice que dará como resultado la formación de espirales aleatorias (que carecen de estructura secundaria).

En algunos casos, los polipéptidos XTEN usados en las composiciones de proteína de fusión de la invención pueden tener un porcentaje de hélice alfa que varía de 0 % a menos de aproximadamente 5 %, según lo determinado por un algoritmo de Chou-Fasman. En otros casos, los polipéptidos XTEN que comprenden las composiciones de proteína de fusión pueden tener un porcentaje de lámina beta que varía de 0 % a menos de aproximadamente 5 %, según lo determinado por un algoritmo de Chou-Fasman. En algunos casos, las secuencias de XTEN de las composiciones de proteínas de fusión pueden tener un porcentaje de hélice alfa que varía de 0 % a menos de 5 % y un porcentaje de lámina beta que varía de 0 % a menos de 5 % según lo determinado por un algoritmo de Chou-Fasman. En modalidades preferidas, los polipéptidos XTEN que comprenden las composiciones de proteína de fusión pueden tener un porcentaje de hélice alfa menor al 2 % y un porcentaje de lámina beta de menos del 2 %. En otros casos, las secuencias XTEN de las composiciones de proteína de fusión pueden tener un alto grado de porcentaje de enrollamiento aleatorio, según lo determinado por un algoritmo GOR. En algunas modalidades, un polipéptido XTEN puede tener al menos aproximadamente 80 %, con mayor preferencia al menos aproximadamente 90 %, con mayor preferencia al menos aproximadamente 91 %, con mayor preferencia al menos aproximadamente 92 %, con mayor preferencia al menos aproximadamente 93 %, con mayor preferencia al menos

aproximadamente 94 %, con mayor preferencia al menos aproximadamente 95 %, con mayor preferencia al menos aproximadamente 96 %, con mayor preferencia al menos aproximadamente 97 %, con mayor preferencia al menos aproximadamente 98 %, y con la máxima preferencia al menos aproximadamente 99 % de enrollamiento aleatorio, según se determina por un algoritmo GOR.

Carga neta

En otros casos, los polipéptidos XTEN pueden tener una característica no estructurada impartida por la incorporación de residuos de aminoácidos con una carga neta y/o la reducción de la proporción de aminoácidos hidrófobos en el polipéptido XTEN. La carga neta global y la densidad de carga neta pueden controlarse mediante la modificación del contenido de aminoácidos cargados en los polipéptidos XTEN. En algunos casos, la densidad de carga neta del XTEN de las composiciones puede ser superior a +0,1 o inferior a -0,1 cargas/residuo. En otros casos, la carga neta de un polipéptido XTEN puede ser de aproximadamente 0 %, aproximadamente 1 %, aproximadamente 2 %, aproximadamente 3 %, aproximadamente 4 %, aproximadamente 5 %, aproximadamente 6 %, aproximadamente 7 %, aproximadamente 8 %, aproximadamente 9 %, aproximadamente 10 %, aproximadamente 11 %, aproximadamente 12 %, aproximadamente 13 %, aproximadamente 14 %, aproximadamente 15 %, aproximadamente 16 %, aproximadamente 17 %, aproximadamente 18 %, aproximadamente 19 % o aproximadamente 20 % o más.

Debido a que la mayoría de los tejidos y superficies en un ser humano o animal tienen una carga neta negativa, los polipéptidos XTEN pueden diseñarse de modo que tengan una carga neta negativa para minimizar las interacciones inespecíficas entre las composiciones que contienen el polipéptido XTEN y varias superficies tales como vasos sanguíneos, vasos sanguíneos, tejidos sanos o varios receptores. Sin estar ligado a una teoría particular, el polipéptido XTEN puede adoptar conformaciones abiertas debido a la repulsión electrostática entre los aminoácidos individuales del polipéptido XTEN que portan individualmente una carga negativa neta alta y que se distribuyen a lo largo de la secuencia del polipéptido XTEN. Tal distribución de carga negativa neta en las longitudes de secuencia extendidas del polipéptido XTEN puede conducir a una conformación no estructurada que, a su vez, puede dar como resultado un aumento efectivo en el radio hidrodinámico. En consecuencia, en una modalidad, la invención proporciona polipéptidos XTEN en los que los polipéptidos XTEN contienen aproximadamente 8, 10, 15, 20, 25 o incluso aproximadamente 30 % de ácido glutámico. Los polipéptidos XTEN de las composiciones de la presente invención generalmente tienen un contenido bajo o nulo de aminoácidos con carga positiva. En algunos casos, los polipéptidos XTEN pueden tener menos de aproximadamente 10 % de residuos de aminoácidos con carga positiva, menos de aproximadamente 7 %, menos de aproximadamente 5 % o menos de aproximadamente 2 % de residuos de aminoácidos con una carga positiva. Sin embargo, la invención contempla construcciones en las que puede incorporarse un número limitado de aminoácidos con carga

positiva, tales como la lisina, en polipéptidos XTEN para permitir la conjugación entre la amina épsilon de la lisina y un grupo reactivo en un péptido, un puente enlazador, o un grupo reactivo en un fármaco o molécula pequeña a conjugar con la cadena principal del polipéptido XTEN. En lo anterior, pueden construirse proteínas de fusión que comprenden uno o más polipéptidos XTEN, una proteína biológicamente activa, más un agente quimioterapéutico útil en el tratamiento de enfermedades o trastornos metabólicos, en donde el número máximo de moléculas del agente incorporado en el componente polipeptídico XTEN está determinado por el número de lisinas u otros aminoácidos con cadenas laterales reactivas (por ejemplo, cisteína) incorporados en el XTEN.

En algunos casos, un polipéptido XTEN puede comprender residuos cargados separados por otros residuos tales como serina o glicina, lo que puede conducir a un mejor comportamiento de expresión o purificación. En base a la carga neta, los polipéptidos XTEN de las composiciones humanas o animales pueden tener un punto isoeléctrico (pI) de 1,0, 1,5, 2,0, 2,5, 3,0, 3,5, 4,0, 4,5, 5,0, 5,5, 6,0 o incluso 6,5. En modalidades preferidas, el polipéptido XTEN tiene un punto isoeléctrico entre 1,5 y 4,5. En estas modalidades, el XTEN incorporado en las composiciones de proteína de fusión BPXTEN de la presente invención tendría una carga neta negativa en condiciones fisiológicas que puede contribuir a la conformación no estructurada y unión reducida del componente polipeptídico XTEN a proteínas y tejidos de mamíferos.

Dado que los aminoácidos hidrófobos pueden impartir estructura a un polipéptido, la invención establece que el contenido de aminoácidos hidrófobos en los polipéptidos XTEN es típicamente menor que 5 %, menor que 2 % o menor que el 1 % del contenido de aminoácidos hidrófobos. En una modalidad, el contenido de aminoácidos de metionina y triptófano en el componente XTEN de una proteína de fusión BPXTEN es típicamente menor que 5 % o menor que 2 % y con la máxima preferencia menor que el 1 %. En otra modalidad, el polipéptido XTEN tiene una secuencia que tiene menos del 10 % de residuos de aminoácidos con una carga positiva, o menos de aproximadamente 7 %, o menos de aproximadamente 5 %, o menos de aproximadamente 2 % de residuos de aminoácidos con una carga positiva, la suma de residuos de metionina y triptófano es inferior al 2 %, y la suma de residuos de asparagina y glutamina es inferior al 10 % del polipéptido XTEN total.

Mayor radio hidrodinámico

En algunas modalidades, el polipéptido XTEN puede tener un radio hidrodinámico elevado, lo que confiere un mayor peso molecular aparente correspondiente a la proteína de fusión BPXTEN que incorpora el polipéptido XTEN. La unión de un polipéptido XTEN a secuencias de BP puede dar como resultado composiciones de BPXTEN que pueden tener mayores radios hidrodinámicos, mayor peso molecular aparente y mayor factor de peso molecular aparente en comparación con una BP no unida a un polipéptido XTEN. Por ejemplo, en aplicaciones terapéuticas en las que se desea una semivida prolongada, las composiciones en las que se incorpora un polipéptido XTEN con un radio hidrodinámico elevado a una proteína de fusión que comprende una o más BP pueden aumentar de forma

efectiva el radio hidrodinámico de la composición más allá del tamaño de poro glomerular de aproximadamente 3-5 nm (correspondiente a un peso molecular aparente de aproximadamente 70 kDa) (Caliceti. 2003. Pharmacokinetic and biodistribution properties of poly(ethylene glycol)-protein conjugates. *Adv. Drug Deliv. Rev.* 55:1261-1277), lo que da como resultado la reducción de la eliminación renal de las proteínas en circulación. Sin estar ligado a una teoría particular, un polipéptido XTEN puede adoptar conformaciones abiertas debido a la repulsión electrostática entre cargas individuales del péptido o la flexibilidad inherente impartida por los aminoácidos particulares en la secuencia que carecen de potencial para conferir una estructura secundaria. La conformación abierta, extendida y no estructurada del polipéptido XTEN puede tener un radio hidrodinámico proporcional mayor en comparación con los polipéptidos de longitud de secuencia y/o peso molecular comparables que tienen una estructura secundaria y/o terciaria, tales como las proteínas globulares típicas. Los métodos para determinar el radio hidrodinámico se conocen bien en la técnica, tales como mediante el uso de cromatografía de exclusión por tamaño (SEC), como se describe en las patentes de Estados Unidos números 6,406,632 y 7,294,513. La adición de longitudes crecientes de polipéptidos XTEN da como resultado aumentos proporcionales en los parámetros de radio hidrodinámico, peso molecular aparente y factor de peso molecular aparente, lo que permite la adaptación de BPXTEN a los pesos moleculares aparentes o radios hidrodinámicos límites característicos deseados. En consecuencia, en ciertas modalidades, la proteína de fusión BPXTEN puede configurarse con un polipéptido XTEN, en donde la proteína de fusión puede tener un radio hidrodinámico de al menos aproximadamente 5 nm, o al menos aproximadamente 8 nm, o al menos aproximadamente 10 nm, o 12 nm, o al menos aproximadamente 15 nm. En las modalidades anteriores, el gran radio hidrodinámico conferido por los polipéptidos XTEN en una proteína de fusión BPXTEN puede conducir a una reducción de la eliminación renal de la proteína de fusión resultante, lo que conduce a un aumento correspondiente en la semivida terminal, un aumento en el tiempo medio de residencia y/ o una disminución en la velocidad de aclaramiento renal.

En otra modalidad, puede incorporarse selectivamente un polipéptido XTEN de una longitud y secuencia seleccionadas en una BPXTEN para crear una proteína de fusión que tiene, en condiciones fisiológicas, un peso molecular aparente de al menos aproximadamente 150 kDa, o al menos aproximadamente 300 kDa, o al menos aproximadamente 400 kDa, o al menos aproximadamente 500 kDa, o al menos aproximadamente 600 kDa, o al menos aproximadamente 700 kDa, o al menos aproximadamente 800 kDa, o al menos aproximadamente 900 kDa, o al menos aproximadamente 1 000 kDa, o al menos aproximadamente 1,200 kDa, o al menos aproximadamente 1 500 kDa, o al menos aproximadamente 1,800 kDa, o al menos aproximadamente 2 000 kDa, o al menos aproximadamente 2,300 kDa o más. En otra modalidad, un polipéptido XTEN de una longitud y secuencia seleccionadas puede unirse selectivamente a una BP para dar como resultado una proteína de fusión BPXTEN que

tiene, en condiciones fisiológicas, un factor de peso molecular aparente de al menos tres, alternativamente de al menos cuatro, alternativamente de al menos cinco, alternativamente de al menos seis, alternativamente de al menos ocho, alternativamente de al menos 10, alternativamente de al menos 15, o un factor de peso molecular aparente de al menos 20 o mayor. En otra modalidad, la proteína de fusión BPXTEN tiene, en condiciones fisiológicas, un factor de peso molecular aparente que es de aproximadamente 4 a aproximadamente 20, o es de aproximadamente 6 a aproximadamente 15, o es de aproximadamente 8 a aproximadamente 12, o es de aproximadamente 9 a aproximadamente 10 con relación al peso molecular real de la proteína de fusión. En algunas modalidades, el polipéptido (de fusión) exhibe un factor de peso molecular aparente en condiciones fisiológicas mayor que aproximadamente 6.

Semivida terminal aumentada

En algunas modalidades, el polipéptido (de fusión) tiene una semivida terminal que es al menos dos veces más larga, o al menos tres veces más larga, o al menos cuatro veces más larga, o al menos cinco veces más larga, en comparación con el polipéptido biológicamente activo no unido a un polipéptido XTEN. En algunas modalidades, el polipéptido (de fusión) tiene una semivida terminal que es al menos dos veces más larga en comparación con el polipéptido biológicamente activo no unido a un polipéptido XTEN.

La administración de una dosis terapéuticamente efectiva de cualquiera de las modalidades de proteínas de fusión BPXTEN descritas en la presente descripción a un ser humano o animal que lo necesite puede dar como resultado una ganancia en el tiempo de al menos dos veces, o al menos tres veces, o al menos cuatro veces, o al menos cinco veces o más empleado dentro de una ventana terapéutica para la proteína de fusión en comparación con la BP correspondiente no unida al polipéptido XTEN y administrada a una dosis comparable a un ser humano o animal.

Baja inmunogenicidad

En otro aspecto, la invención proporciona composiciones en las que los polipéptidos XTEN tienen un bajo grado de inmunogenicidad o son sustancialmente no inmunogénicos. Varios factores pueden contribuir a la baja inmunogenicidad de los polipéptidos XTEN, por ejemplo, la secuencia sustancialmente no repetitiva, la conformación no estructurada de la misma, el alto grado de solubilidad, el bajo grado o ausencia de autoagregación, el bajo grado o ausencia de sitios proteolíticos dentro de la secuencia, y el bajo grado o ausencia de epítomos en polipéptidos XTEN.

Un experto en la técnica entenderá que, en general, un polipéptido que tiene secuencias de aminoácidos cortas altamente repetitivas (por ejemplo, en donde una secuencia de 200 aminoácidos de longitud contiene en promedio 20 repeticiones o más de un conjunto limitado de 3 o 4 mer) y/o que tienen residuos de aminoácidos repetitivos contiguos (por ejemplo, en donde las secuencias de 5 o 6 mer tienen residuos de aminoácidos idénticos) tienen tendencia a agregarse o formar estructuras de orden superior o formar contactos que dan como resultado estructuras cristalinas o pseudocristalinas.

En algunas modalidades, los polipéptidos XTEN son sustancialmente no repetitivos, en donde la secuencia de aminoácidos de XTEN no tiene tres aminoácidos contiguos que sean tipos de aminoácidos idénticos, a menos que el aminoácido sea serina, en cuyo caso no más de tres aminoácidos contiguos pueden ser residuos de serina; y en donde la secuencia de aminoácidos de XTEN no contiene secuencias de 3 aminoácidos (3 mer) que aparezcan más de 16, más de 14, más de 12 o más de 10 veces dentro de una secuencia de 200 aminoácidos de longitud del polipéptido XTEN. Un experto en la técnica comprenderá que tales secuencias sustancialmente no repetitivas tienen menos tendencia a agregarse y, por lo tanto, permiten el diseño de XTEN de secuencia larga con una frecuencia relativamente baja de aminoácidos cargados que probablemente se agregarían si de cualquier otra manera, las secuencias o los residuos de aminoácidos fueran más repetitivos.

Los epítomos conformacionales están formados por regiones de la superficie de la proteína que están compuestas por múltiples secuencias de aminoácidos discontinuas del antígeno proteico. El plegamiento preciso de la proteína lleva estas secuencias a configuraciones espaciales estables y bien definidas, o epítomos, que pueden ser reconocidos como "extraños" por el sistema inmunitario humoral del huésped, lo que da como resultado la producción de anticuerpos contra la proteína o la activación de una respuesta inmunitaria mediada por células. En el último caso, la respuesta inmunitaria a una proteína en un individuo está fuertemente influenciada por el reconocimiento del epítomo por linfocitos T que es una función de la especificidad de unión al péptido del alotipo HLA-DR de ese individuo. El acoplamiento de un complejo peptídico MHC de clase II con un receptor de linfocitos T afín en la superficie del linfocito T, junto con la unión cruzada de otros correceptores, tales como la molécula CD4, puede inducir un estado activado dentro del linfocito T. La activación conduce a la liberación de citocinas que activan aún más otros linfocitos, tales como los linfocitos B, para producir anticuerpos o activan los linfocitos T citolíticos como una respuesta inmunitaria celular completa.

La capacidad de un péptido para unirse a una molécula de MHC Clase II dada para su presentación en la superficie de una APC (célula presentadora de antígeno) depende de una serie de factores; más particularmente su secuencia primaria. En una modalidad, puede lograrse un menor grado de inmunogenicidad mediante el diseño de polipéptidos XTEN que resisten el procesamiento de antígenos en células presentadoras de antígenos y/o la elección de secuencias que no se unen bien a los receptores MHC. La invención proporciona proteínas de fusión BPXTEN con polipéptidos XTEN sustancialmente no repetitivos diseñados para reducir la unión con los receptores MHC II, así como también evitar la formación de epítomos para la unión de anticuerpos o receptores de linfocitos T, lo que da como resultado un bajo grado de inmunogenicidad. Evitar la inmunogenicidad es, en parte, un resultado directo de la flexibilidad conformacional de los polipéptidos XTEN; es decir, la ausencia de estructura secundaria debido a la selección y orden de los residuos de aminoácidos. Por ejemplo, son de particular interés las secuencias que tienen una baja tendencia a adaptarse a conformaciones plegadas de forma compacta en solución acuosa o

en condiciones fisiológicas que podrían dar como resultado epítomos conformacionales. La administración de proteínas de fusión que comprenden polipéptido(s) de XTEN, mediante el uso de prácticas terapéuticas y dosificación convencionales, generalmente no daría como resultado la formación de anticuerpos neutralizantes contra el polipéptido XTEN, y también puede reducir la inmunogenicidad de la pareja de fusión de BP en las composiciones de BPXTEN.

En una modalidad, los polipéptidos XTEN utilizados en las proteínas de fusión humanas o animales pueden estar sustancialmente libres de epítomos reconocidos por linfocitos T humanos. La eliminación de dichos epítomos con el propósito de generar proteínas menos inmunogénicas se ha descrito anteriormente; ver, por ejemplo, los documentos WO 98/52976, WO 02/079232 y WO 00/3317 que se incorporan en la presente descripción como referencia. Se han descrito ensayos para epítomos de linfocitos T humanos (Stickler, M., y otros (2003) *J Immunol Methods*, 281: 95-108). De particular interés son las secuencias peptídicas que pueden oligomerizarse sin generar epítomos de linfocitos T o secuencias no humanas. Esto puede lograrse mediante la prueba de repeticiones directas de estas secuencias para detectar la presencia de epítomos de linfocitos T y la aparición de secuencias de 6 a 15 mer y, en particular, secuencias de 9 mer que no son humanas, y después la alteración del diseño del polipéptido XTEN para eliminar o alterar la secuencia del epítomo. En algunos casos, los polipéptidos XTEN son sustancialmente no inmunogénicos por la restricción del número de epítomos del polipéptido XTEN predichos para unirse a los receptores MHC. Con una reducción en el número de epítomos capaces de unirse a los receptores del MHC, hay una reducción concomitante en el potencial de activación de linfocitos T así como también en la función auxiliar de los linfocitos T, reducción de la activación o regulación positiva de los linfocitos B y reducción de la producción de anticuerpos. El bajo grado de epítomos de linfocitos T predichos puede determinarse mediante algoritmos de predicción de epítomos tales como, por ejemplo, TEPITOPE (Sturniolo, T., y otros (1999) *Nat Biotechnol*, 17: 555-61), como se muestra en **Ejemplo 74** de la publicación de solicitud de patente internacional núm. WO 2010/144502 A2, que se incorpora como referencia en su totalidad. La puntuación TEPITOPE de un marco peptídico dado dentro de una proteína es el logaritmo de la K_d (constante de disociación, afinidad, tasa de disociación) de la unión de ese marco peptídico a múltiples de los alelos de MHC humanos más comunes, como se describe en Sturniolo, T. y otros (1999) *Nature Biotechnology* 17:555). La puntuación varía en al menos 20 log, de aproximadamente 10 a aproximadamente -10 (correspondiente a restricciones de unión de $10e^{10} k_d$ a $10e^{-10} k_d$), y puede reducirse al evitar los aminoácidos hidrófobos que pueden servir como residuos de anclaje durante la presentación de péptidos en MHC, tales como M, I, L, V, F. En algunas modalidades, un polipéptido XTEN incorporado en una BPXTEN no tiene un epítomo de linfocitos T predicho en una puntuación TEPITOPE de aproximadamente -5 o mayor, o -6 o mayor, o -7 o mayor, o -8 o mayor, o en una puntuación TEPITOPE de -9 o mayor. Como se usa en la presente descripción, una

puntuación de "-9 o mayor" incluiría puntuaciones TEPITOPE de 10 a -9, incluidas, pero no incluiría una puntuación de -10, ya que -10 es menor que -9.

En otra modalidad, los polipéptidos XTEN de la invención, que incluyen los incorporados en las proteínas de fusión BPXTEN humanas o animales, pueden volverse sustancialmente no inmunogénicos mediante la restricción de los sitios proteolíticos conocidos de la secuencia del polipéptido XTEN, lo que reduce el procesamiento de los polipéptidos XTEN en pequeños péptidos que pueden unirse a los receptores MHC II. En otra modalidad, el polipéptido XTEN puede volverse sustancialmente no inmunogénico mediante el uso de una secuencia que está sustancialmente desprovista de estructura secundaria, lo que confiere resistencia a muchas proteasas debido a la alta entropía de la estructura. En consecuencia, la puntuación TEPITOPE reducida y la eliminación de sitios proteolíticos conocidos de los polipéptidos XTEN pueden hacer que las composiciones polipeptídicas de XTEN, que incluyen los polipéptidos XTEN de las composiciones de proteína de fusión BPXTEN, sean sustancialmente incapaces de unirse a los receptores de mamíferos, que incluyen los del sistema inmunitario. En una modalidad, un polipéptido XTEN de una proteína de fusión BPXTEN puede tener una unión a un receptor de mamífero con una $K_d > 100$ nM, o K_d de más de 500 nM, o K_d de más de 1 μ M hacia la superficie de una célula de mamífero o un receptor polipeptídico en circulación.

Adicionalmente, la secuencia sustancialmente no repetitiva y la correspondiente ausencia de epítomos de dichas modalidades de polipéptidos XTEN pueden limitar la capacidad de los linfocitos B para unirse a o ser activados por polipéptidos XTEN. Si bien un polipéptido XTEN puede hacer contactos con muchos linfocitos B diferentes a lo largo de su secuencia extendida, cada linfocito B individual solo puede hacer uno o un pequeño número de contactos con un polipéptido XTEN individual. Como resultado, los polipéptidos XTEN típicamente pueden tener una tendencia mucho menor a estimular la proliferación de linfocitos B y, por lo tanto, una respuesta inmunitaria. En una modalidad, la BPXTEN puede tener una inmunogenicidad reducida en comparación con la BP correspondiente que no está fusionada. En una modalidad, la administración de hasta tres dosis parenterales de una BPXTEN a un mamífero puede dar como resultado IgG anti-BPXTEN detectable a una dilución sérica de 1:100 pero no a una dilución de 1:1 000. En otra modalidad, la administración de hasta tres dosis parenterales de una BPXTEN a un mamífero puede dar como resultado IgG anti-BP detectable a una dilución sérica de 1:100 pero no a una dilución de 1:1 000. En otra modalidad, la administración de hasta tres dosis parenterales de una BPXTEN a un mamífero puede dar como resultado IgG anti-XTEN detectable a una dilución sérica de 1:100 pero no a una dilución de 1:1 000. En las modalidades anteriores, el mamífero puede ser un ratón, una rata, un conejo o un macaco cangrejero.

Una característica adicional de ciertas modalidades de polipéptidos XTEN con secuencias sustancialmente no repetitivas con relación a aquellas secuencias menos no repetitivas (tal como una que tiene tres aminoácidos contiguos que son idénticos) puede ser que los polipéptidos XTEN no repetitivos formen contactos más débiles con los anticuerpos (por

ejemplo interacciones monovalentes), lo que da como resultado de esta manera una menor posibilidad de aclaramiento inmunitario en donde las composiciones de BPXTEN pueden permanecer en circulación durante un mayor período de tiempo.

En algunas modalidades, el polipéptido (de fusión) es menos inmunogénico en comparación con el polipéptido biológicamente activo no unido a un polipéptido XTEN, en donde la inmunogenicidad se determina por medición de la producción de anticuerpos IgG que se unen selectivamente al polipéptido biológicamente activo después de la administración de dosis comparables a un ser humano o animal.

Espaciadores y segmento de liberación de BP

En algunas modalidades, la BPXTEN intacta retiene al menos una porción de la actividad biológica de la BP respectiva. En algunas modalidades, el componente BP se vuelve biológicamente activo o tiene un aumento en la actividad biológica tras su liberación del(de los) polipéptido(s) XTEN mediante la escisión de una secuencia de escisión opcional incorporada dentro de las secuencias espaciadoras en la BPXTEN, como se describe con más detalle a continuación.

Cualquier grupo de secuencias espaciadoras es opcional en las proteínas de fusión incluidas en la invención. El espaciador puede proporcionarse para mejorar la expresión de la proteína de fusión de una célula huésped o para disminuir el impedimento estérico en donde el componente BP puede asumir su estructura terciaria deseada y/o interactuar adecuadamente con su molécula objetivo. Para espaciadores y métodos para identificar espaciadores convenientes, ver, por ejemplo, George, y otros, (2003) *Protein Engineering* 15:871-879, que se incorpora específicamente como referencia en la presente descripción. En una modalidad, el espaciador comprende una o más secuencias peptídicas que tienen entre 1 y 50 residuos de aminoácidos de longitud, o aproximadamente 1 a 25 residuos, o aproximadamente 1 a 10 residuos de longitud. Las secuencias espaciadoras, exclusivas de los sitios de escisión, pueden comprender cualquiera de los 20 L aminoácidos naturales, y preferentemente comprenden aminoácidos hidrófilos que no tienen impedimentos estéricos que pueden incluir, pero sin limitarse a, glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) y prolina (P). En algunas modalidades, el espaciador puede ser poliglicinas o polialaninas, o predominantemente una mezcla de combinaciones de residuos de glicina y alanina. El polipéptido espaciador exclusivo de una secuencia de escisión está desprovisto en gran parte o sustancialmente de estructura secundaria. En una modalidad, una o ambas secuencias espaciadoras en una composición de proteína de fusión BPXTEN pueden contener cada una además una secuencia de escisión, que puede ser idéntica o diferente, en donde una proteasa puede actuar sobre la secuencia de escisión para liberar la BP de la proteína de fusión.

En algunos casos, la incorporación de la secuencia de escisión en la BPXTEN se diseña de modo que permita la liberación de una BP que se vuelve activa o más activa tras su liberación del polipéptido XTEN. Las secuencias de escisión se ubican lo suficientemente cerca de las secuencias de BP, generalmente dentro de 18, o dentro de 12, o dentro de 6, o

dentro de 2 aminoácidos del extremo de la secuencia de BP, en donde cualquier residuo restante unido a la BP después de la escisión no interfiere apreciablemente con la actividad (por ejemplo, tal como la unión a un receptor) de la BP, y sin embargo proporciona suficiente acceso a la proteasa para que pueda efectuar la escisión de la secuencia de escisión. En algunas modalidades, el sitio de escisión es una secuencia que puede escindirse por una proteasa endógena del mamífero humano o animal en donde la BPXTEN puede escindirse después de la administración a un ser humano o animal. En tales casos, la BPXTEN puede servir como un profármaco o depósito en circulación de la BP. Los ejemplos de sitios de escisión contemplados por la invención incluyen, pero no se limitan a, una secuencia polipeptídica escindible por una proteasa endógena de mamíferos seleccionada de FXIa, FXIIa, calicreína, FVIIa, FIXa, FXa, FIIa (trombina), elastasa 2, granzima B, MMP-12, MMP-13, MMP-17 o MMP-20, o por proteasas que no son de mamífero tales como TEV, enteroquinasa, proteasa PreScission™ (proteasa 3C de rinovirus) y sortasa A. Las secuencias conocidas por escindirse por las proteasas anteriores se conocen en la técnica. Las secuencias de escisión ilustrativas y los sitios de escisión dentro de las secuencias se exponen en la Tabla 7a, así como también variantes de secuencia. Por ejemplo, la trombina (factor de coagulación II activado) actúa sobre la secuencia LTPRSLLV (SEQ ID NO: 222) {Rawlings N.D. y otros, (2008) *Nucleic Acids Res.*, 36: D320}, que se escindiría después de la arginina en la posición 4 de la secuencia. El FIIa activo se produce por escisión de FII por FXa en presencia de fosfolípidos y calcio y se encuentra aguas abajo del factor IX en la vía de la coagulación. Una vez activado, su función natural en la coagulación es escindir el fibrinógeno, que después, a su vez, comienza la formación del coágulo. La actividad de FIIa está estrictamente controlada y solo ocurre cuando la coagulación es necesaria para una hemostasia adecuada. Sin embargo, dado que la coagulación es un proceso continuo en los mamíferos, mediante la incorporación de la secuencia LTPRSLLV (SEQ ID NO: 223) en la BPXTEN entre la BP y el polipéptido XTEN, los polipéptidos XTEN se eliminarían de la BP adyacente al mismo tiempo que la activación de las vías de coagulación extrínseca o intrínseca cuando la coagulación se requiere fisiológicamente, para liberar de esta manera la BP con el tiempo. De manera similar, la incorporación de otras secuencias en BPXTEN sobre las que actúan las proteasas endógenas proporcionaría una liberación sostenida de BP que puede, en ciertos casos, proporcionar un mayor grado de actividad de la BP de la forma de "profármaco" de la BPXTEN.

En algunos casos, solo los dos o tres aminoácidos que flanquean ambos lados del sitio de escisión (cuatro a seis aminoácidos en total) se incorporarían a la secuencia de escisión. En otros casos, la secuencia de escisión conocida puede tener una o más deleciones o inserciones o una o dos o tres sustituciones de aminoácidos para cualquiera de uno o dos o tres aminoácidos en la secuencia conocida, en donde las deleciones, inserciones o sustituciones dan como resultado una reducción o mejora de la susceptibilidad pero no una ausencia de susceptibilidad a la proteasa, lo que da como resultado una capacidad para

adaptar la velocidad de liberación de la BP desde el XTEN. Las sustituciones ilustrativas se muestran en la **Tabla 7a**.

Tabla 7a: Secuencias de escisión por proteasa para la liberación de BP

Proteasa que actúa sobre la secuencia	SEQ ID NO	Secuencia de escisión ilustrativa	Sitio de escisión mínimo*
FXIa	224	KLTR↓VVGG	KD/FL/T/R↓VA/VE/GT/GV
FXIIa	225	TMTR↓IVGG	NA
Kallikrein	226	SPFR↓STGG	-/-/FL/RY↓SR/RT/-/-
FVIIa	227	LQVR↓IVGG	NA
FIXa	228	PLGR↓IVGG	-/-/G/R↓-/-/-/-
FXa	229	IEGR↓TVGG	IA/E/GFP/R↓STI/VFS/-/G
FIIa (thrombin)	230	LTPR↓SLLV	-/-/PLA/R↓SAG/-/-/-
Elastase-2	231	LGPV↓SGVP	-/-/-/VIAT↓-/-/-/-
Granzyme-B	232	VAGD↓SLEE	V/-/-/D↓-/-/-/-
MMP-12	233	GPAG↓LGGA	G/PA/-/G↓L/-/G/- (SEQ ID NO: 241)
MMP-13	234	GPAG↓LRGA	G/P/-/G↓L/-/GA/- (SEQ ID NO: 242)
MMP-17	235	APLG↓LRLR	-/PS/-/-↓LQ/-/LT/-
MMP-20	236	PALP↓LVAQ	NA
TEV	237	ENLYFQ↓G	ENLYFQ↓G/S (SEQ ID NO: 243)
Enterokinase	238	DDDK↓IVGG	DDDK↓IVGG (SEQ ID NO: 244)
Protease 3C (PreScission™)	239	LEVLFQ↓GP	LEVLFQ↓GP (SEQ ID NO: 245)
Sortase A	240	LPKT↓GSES	L/P/KEAD/T↓G/-/EKS/S (SEQ ID NO: 246)

↓ indica el sitio de escisión;

NA: no aplicable;

* la lista de múltiples aminoácidos antes, entre o después de una barra inclinada indica aminoácidos alternativos que pueden sustituirse en la posición;

“-” indica que cualquier aminoácido puede sustituirse por el aminoácido correspondiente indicado en la columna del medio

En algunas modalidades, la proteína de fusión BPXTEN puede comprender secuencias espaciadoras que pueden comprender además una o más secuencias de escisión configuradas para liberar la BP de la proteína de fusión cuando una proteasa actúa sobre ella. En algunas modalidades, la una o más secuencias de escisión pueden ser una secuencia que tiene al menos aproximadamente 80 % (por ejemplo, al menos aproximadamente 85 %, al menos aproximadamente 90 %, al menos aproximadamente 95 %, al menos aproximadamente 96 %, al menos aproximadamente 97 %, al menos aproximadamente 98 %, al menos aproximadamente 99 % o 100 %) de identidad de secuencia con una secuencia expuesta en la Tabla 7a.

En algunas modalidades, la descripción proporciona péptidos de los segmentos de liberación (o segmento de liberación (RS)) que son sustratos de una o más proteasas de mamífero asociadas con o producidas por tejidos o células enfermos que se encuentran

cerca de tejidos enfermos. Tales proteasas pueden incluir, pero no limitarse a, las clases de proteasas tales como metaloproteinasas, cisteína proteasas, aspartato proteasas y serina proteasas que incluyen, pero sin limitarse a, las proteasas expuestas en la Tabla 7b. El RS es útil, entre otras cosas, para la incorporación en polipéptidos recombinantes humanos o animales, que confiere un formato de profármaco que puede activarse mediante la escisión del RS por proteasas de mamíferos. Como se describe en la presente descripción, los RS se incorporan en las composiciones polipeptídicas recombinantes humanas o animales, donde enlazan los restos de unión incorporados al XTEN (cuyas configuraciones se describen con más detalle, a continuación) en donde tras la escisión del RS por acción de la una o más proteasas cuyo sustrato es el RS, los restos de unión y XTEN se liberan de la composición y los restos de unión, que ya no están bloqueados por el XTEN, recuperan todo su potencial para unirse a sus ligandos. En aquellas composiciones de polipéptidos recombinantes que comprenden un primer y un segundo fragmento de anticuerpo, las composiciones también se denominan en la presente descripción composiciones de anticuerpos activables (AAC).

Tabla 7b: Proteasas de tejidos objetivo

Clase de proteasas	Proteasa
Metaloproteinasas	Meprina
	Neprilisina (CD10)
	PSMA
	BMP-1
	Una desintegrina y metaloproteinasas (ADAM)
	ADAM8
	ADAM9
	ADAM10
	ADAM12
	ADAM15
	ADAM17 (TACE)
	ADAM19
	ADAM28 (MDC-L)
	ADAM con motivos de trombospondina (ADAMTS)
	ADAMTS1
	ADAMTS4
	ADAMTS5
	Metaloproteinasas de matriz (MMP)
	MMP-1 (Colagenasa 1)
	MMP-2 (Gelatinasa A)
	MMP-3 (m1)
	MMP-7 (Matrilisina 1)
	MMP-8 (Colagenasa 2)
	MMP-9 (Gelatinasa B)
	MMP-10 (Estromalisina 2)
	MMP-11 (Estromalisina 3)
	MMP-12 (Elastasa de macrófagos)
	MMP-13 (Colagenasa 3)
	MMP-14 (MT1-MMP)
	MMP-15 (MT2-MMP)
	MMP-19
	MMP-23 (CA-MMP)

Clase de proteasas	Proteasa
	MMP-24 (MT5-MMP)
	MMP-26 (Matrilisina 2)
	MMP-27 (CMMP)
Proteasas de cisteína	Legumaína
	Catepsinas de cisteína
	Catepsina B
	Catepsina C
	Catepsina K
	Catepsina L
	Catepsina S
Proteasas de aspartato	Catepsina X
	Catepsina D
	Catepsina E
Serina proteasas	Secretasa
	Uroquinasa (uPA)
	Activador del plasminógeno de tipo tisular (tPA)
	Plasmina
	Trombina
	Antígeno prostático específico (PSA, KLK3)
	Elastasa de neutrófilos humanos (HNE)
	Elastasa
	Triptasa
	Serina proteasas transmembrana tipo II (TTSP)
	DESC1
	Hepsina (HPN)
	Matriptasa
	Matriptasa-2
	TMPRSS2
	TMPRSS3
	TMPRSS4 (CAP2)
	Proteína de activación de fibroblastos (FAP)
	peptidasa relacionada con la calicreína (familia KLK)
	KLK4
	KLK5
	KLK6
	KLK7
	KLK8
	KLK10
	KLK11
	KLK13
	KLK14

En una modalidad, la descripción proporciona polipéptidos recombinantes activables que comprenden una secuencia del primer segmento de liberación (RS1) que tiene al menos 88 %, o al menos 94 %, o 100 % de identidad de secuencia, cuando se alinea de manera óptima, con una secuencia seleccionada de las secuencias expuestas en la Tabla 8a, en donde el RS1 es un sustrato de una o más proteasas de mamíferos. En otras modalidades, la descripción proporciona polipéptidos recombinantes activables que comprenden un RS1 y una secuencia de un segundo segmento de liberación (RS2), cada uno con al menos 88 %, o al menos 94 % o 100 % de identidad de secuencia, cuando se alinea de forma óptima, con

una secuencia seleccionada de las secuencias expuestas en la Tabla 8a, en donde RS1 y RS2 son cada uno un sustrato de una o más proteasas de mamíferos. En otra modalidad, la descripción proporciona polipéptidos recombinantes activables que comprenden una primera secuencia RS (RS1) que tiene al menos 90 %, al menos 93 %, al menos 97 % o 100 % de identidad, cuando se alinea de forma óptima, con una secuencia seleccionada de las secuencias expuestas en la Tabla 8b, en donde RS es un sustrato de una o más proteasas de mamíferos. En otras modalidades, la descripción proporciona polipéptidos recombinantes activables que comprenden un RS1 y una secuencia de un segundo segmento de liberación (RS2), cada uno con al menos 88 %, o al menos 94 % o 100 % de identidad de secuencia, cuando se alinea de forma óptima, con una secuencia seleccionada de las secuencias expuestas en la Tabla 8b, en donde RS1 y RS2 son cada uno un sustrato de una o más proteasas de mamíferos. En las modalidades de polipéptidos recombinantes activables que comprenden RS1 y RS2, los dos segmentos de liberación pueden ser idénticos o las secuencias pueden ser diferentes.

La presente descripción contempla segmentos de liberación que son sustratos de una, dos o tres clases diferentes de proteasas seleccionadas de metaloproteinasas, cisteína proteasas, aspartato proteasas y serina proteasas, que incluyen las proteasas expuestas en la Tabla 7b. En una característica particular, el RS sirve como sustratos de proteasas que se encuentran en asociación cercana o se colocan con tejidos o células enfermas, tales como pero sin limitarse a, tumores, células cancerosas, y tejidos inflamatorios, y tras la escisión del RS, los fragmentos de unión que de cualquier otra manera están bloqueados por el XTEN de las composiciones de polipéptidos recombinantes humanos o animales (y por lo tanto tienen una afinidad de unión más baja por sus respectivos ligandos) se liberan de la composición y recuperan todo su potencial para unirse a ligandos en células objetivo y/o efectoras. En otra modalidad, el RS de las composiciones polipeptídicas recombinantes humanas o animales comprende una secuencia de aminoácidos que es un sustrato para una proteasa celular ubicada dentro de una célula objetivo, que incluye pero sin limitarse a las proteasas expuestas en la Tabla 7b. En otra característica particular de las composiciones polipeptídicas recombinantes humanas o animales, los RS que son sustratos de dos o tres clases de proteasas se diseñaron con secuencias que pueden escindirse en diferentes ubicaciones de la secuencia RS por las diferentes proteasas. Por lo tanto, los RS que son sustratos de dos, tres, o más clases de proteasas tienen dos, tres, o una pluralidad de sitios de escisión distintos en la secuencia RS, pero la escisión por una única proteasa de todos modos da como resultado la liberación de los restos de unión y el XTEN de la composición polipeptídica recombinante que comprende el RS.

En una modalidad, el RS de la descripción para la incorporación en las composiciones polipeptídicas recombinantes humanas o animales es un sustrato de una o más proteasas que incluyen meprina, neprilisina (CD10), PSMA, BMP-1, disintegrina A y metaloproteinasas (ADAM), ADAM8, ADAM9, ADAM10, ADAM12, ADAM15, ADAM17 (TACE), ADAM19, ADAM28 (MDC-L), ADAM con motivos de

trombospondina (ADAMT), ADAMTS1, ADAMTS4, ADAMTS5, MMP-1 (colagenasa 1), metaloproteinasa de matriz 1 (MMP-1), metaloproteinasa de matriz 2 (MMP-2, gelatinasa A), metaloproteinasa de matriz 3 (MMP-3, estromalisina 1), metaloproteinasa de matriz 7 (MMP-7, matrilisina 1), metaloproteinasa de matriz 8 (MMP-8, colagenasa 2), metaloproteinasa de matriz 9 (MMP-9, gelatinasa B), metaloproteinasa de matriz 10 (MMP-10, estromalisina 2), metaloproteinasa de matriz 11 (MMP-11, estromalisina 3), metaloproteinasa de matriz 12 (MMP-12, elastasa de macrófagos), metaloproteinasa de matriz 13 (MMP-13, colagenasa 3), metaloproteinasa de matriz 14 (MMP-14, MT1-MMP), metaloproteinasa de matriz 15 (MMP-15, MT2-MMP), metaloproteinasa de matriz 19 (MMP-19), metaloproteinasa de matriz 23 (MMP-23, CA-MMP), metaloproteinasa de matriz 24 (MMP-24, MT5-MMP), metaloproteinasa de matriz 26 (MMP-26, matrilisina 2), metaloproteinasa de matriz 27 (MMP-27, CMMP), legumaína, catepsina B, catepsina C, catepsina K, catepsina L, catepsina S, catepsina X, catepsina D, catepsina E, secretasa, uroquinasa (uPA), activador de plasminógeno de tipo tisular (tPA), plasmina, trombina, antígeno prostático específico (PSA, KLK3), elastasa de neutrófilos humanos (HNE), elastasa, triptasa, serina proteasas transmembrana de tipo II (TTSP), DESC1, hepsina (HPN), matriptasa, matriptasa-2, TMPRSS2, TMPRSS3, TMPRSS4 (CAP2), proteína de activación de fibroblastos (FAP), peptidasa relacionada con calicreína (familia KLK), KLK4, KLK5, KLK6, KLK7, KLK8, KLK10, KLK11, KLK13 y KLK14. En una modalidad, el RS es un sustrato de ADAM17. En una modalidad, el RS es un sustrato de BMP-1. En una modalidad, el RS es un sustrato de la catepsina. En una modalidad, el RS es un sustrato de HtrA1. En una modalidad, el RS es un sustrato de legumaína. En una modalidad, el RS es un sustrato de MMP-1. En una modalidad, el RS es un sustrato de MMP-2. En una modalidad, el RS es un sustrato de MMP-7. En una modalidad, el RS es un sustrato de MMP-9. En una modalidad, el RS es un sustrato de MMP-11. En una modalidad, el RS es un sustrato de MMP-14. En una modalidad, el RS es un sustrato de uPA. En una modalidad, el RS es un sustrato de matriptasa. En una modalidad, el RS es un sustrato de MT-SP1. En una modalidad, el RS es un sustrato de la elastasa de neutrófilos. En una modalidad, el RS es un sustrato de la trombina. En una modalidad, el RS es un sustrato de TMPRSS3. En una modalidad, el RS es un sustrato de TMPRSS4. En una modalidad, el RS de las composiciones polipeptídicas recombinantes humanas o animales es un sustrato de al menos dos proteasas que son leguminosas, MMP-1, MMP-2, MMP-7, MMP-9, MMP-11, MMP-14, uPA y matriptasa. En otra modalidad, el RS de las composiciones polipeptídicas recombinantes humanas o animales es un sustrato de legumaína, MMP-1, MMP-2, MMP-7, MMP-9, MMP-11, MMP-14, uPA y matriptasa.

Tabla 8a: Secuencias de segmento de liberación de BP.

Nombre	ID de la construcción	Secuencia de aminoácidos	SEQ ID NO:
RSR-1517	AC1611	EAGRSANHEPLGLVAT	8261
BSRS-A1	AC1605	ASGRSTNAGPSGLAGP	8262
BSRS-A2	AC1606	ASGRSTNAGPQGLAGQ	8263

Nombre	ID de la construcción	Secuencia de aminoácidos	SEQ ID NO:
BSRS-A3	AC1607	ASGRSTNAGPPGLTGP	8264
VP-1	AC1608	ASSRGTNAGPAGLTGP	8265
RSR-1752	AC1609	ASSRTTNTGPSTLTGP	8266
RSR-1512	AC1610	AAGRSDNGTPLELVAP	8267
RSR-1517	AC1611	EAGRSANHEPLGLVAT	8261
VP-2	AC1612	ASGRGTNAGPAGLTGP	8268
RSR-1018	AC1613	LFGRNDNHEPLELGGG	8269
RSR-1053	AC1614	TAGRSDNLEPLGLVFG	8270
RSR-1059	AC1615	LDGRSDNFHPPELVAG	8271
RSR-1065	AC1616	LEGRSDNEEPENLVAG	8272
RSR-1167	AC1617	LKGRSDNNAPLALVAG	8273
RSR-1201	AC1618	VYSRGTNAGPHGLTGR	8274
RSR-1218	AC1619	ANSRGTNKGFAGLIGP	8275
RSR-1226	AC1620	ASSRLTNEAPAGLTIP	8276
RSR-1254	AC1621	DQSRGTNAGPEGLTDP	8277
RSR-1256	AC1622	ESSRGTNIGQGGLTGP	8278
RSR-1261	AC1623	SSSRGTNQDPAGLTIP	8279
RSR-1293	AC1624	ASSRGQNHSPMGLTGP	8280
RSR-1309	AC1625	AYSRGPNAGPAGLEGR	8281
RSR-1326	AC1626	ASERGNNAGPANLTGF	8282
RSR-1345	AC1627	ASHRGTNPKPAILTGP	8283
RSR-1354	AC1628	MSSRRTNANPAQLTGP	8284
RSR-1426	AC1629	GAGRTDNHEPLELGAA	8285
RSR-1478	AC1630	LAGRSENTAPLELTAG	8286
RSR-1479	AC1631	LEGRPDNHEPLALVAS	8287
RSR-1496	AC1632	LSGRSDNEEPLALPAG	8288
RSR-1508	AC1633	EAGRTDNHEPLELSAP	8289
RSR-1513	AC1634	EGGRSDNHGPLELVSG	8290
RSR-1516	AC1635	LSGRSDNEAPLELEAG	8291
RSR-1524	AC1636	LGGRADNHEPPELGAG	8292
RSR-1622	AC1637	PPSRGTNAEPAGLTGE	8293
RSR-1629	AC1638	ASTRGENAGPAGLEAP	8294
RSR-1664	AC1639	ESSRGTNGAPEGLTGP	8295
RSR-1667	AC1640	ASSRATNESPAGLTGE	8296
RSR-1709	AC1641	ASSRGENPPPGLTGP	8297
RSR-1712	AC1642	AASRGTTNTGPAELTGS	8298
RSR-1727	AC1643	AGSRTTNAGPGGLEGP	8299
RSR-1754	AC1644	APSRGENAGPATLTGA	8300
RSR-1819	AC1645	ESGRAANTGPPTLTAP	8301
RSR-1832	AC1646	NPGRAANEGPPGLPGS	8302
RSR-1855	AC1647	ESSRAANLTPPELTGP	8303
RSR-1911	AC1648	ASGRAANETPPGLTGA	8304
RSR-1929	AC1649	NSGRGENLGAPGLTGT	8305
RSR-1951	AC1650	TTGRAANLTPAGLTGP	8306
RSR-2295	AC1761	EAGRSANHTPAGLTGP	8307
RSR-2298	AC1762	ESGRAANTTPAGLTGP	8308
RSR-2038	AC1679	TTGRATEAANLTPAGLTGP	8309
RSR-2072	AC1680	TTGRAEEAANLTPAGLTGP	8310
RSR-2089	AC1681	TTGRAGEAANLTPAGLTGP	8311
RSR-2302	AC1682	TTGRATEAANATPAGLTGP	8312
RSR-3047	AC1697	TTGRAGEAEGATSAGATGP	8313
RSR-3052	AC1698	TTGEAGEAANATSAGATGP	8314
RSR-3043	AC1699	TTGEAGEAAGLTPAGLTGP	8315
RSR-3041	AC1700	TTGAAGEAANATPAGLTGP	8316
RSR-3044	AC1701	TTGRAGEAAGLTPAGLTGP	8317
RSR-3057	AC1702	TTGRAGEAANATSAGATGP	8318

Nombre	ID de la construcción	Secuencia de aminoácidos	SEQ ID NO:
RSR-3058	AC1703	TTGEAGEAAGATSAGATGP	8319
RSR-2485	AC1763	ESGRAANTEPPELGAG	8320
RSR-2486	AC1764	ESGRAANTAPEGLTGP	8321
RSR-2488	AC1688	EPGRAANHEPSGLTEG	8322
RSR-2599	AC1706	ESGRAANHTGAPPGGLTGP	8323
RSR-2706	AC1716	TTGRTGEGANATPGGLTGP	8324
RSR-2707	AC1717	RTGRSGEAAANETPEGLEGP	8325
RSR-2708	AC1718	RTGRTGESANETPAGLGGP	8326
RSR-2709	AC1719	STGRTGEPANETPAGLSGP	8327
RSR-2710	AC1720	TTGRAGEPANATPTGLSGP	8328
RSR-2711	AC1721	RTGRPGEGANATPTGLPGP	8329
RSR-2712	AC1722	RTGRGGEEAANATPSGLGGP	8330
RSR-2713	AC1723	STGRSGESANATPGGLGGP	8331
RSR-2714	AC1724	RTGRTGEEANATPAGLPGP	8332
RSR-2715	AC1725	ATGRPGEPANTTPEGLEGP	8333
RSR-2716	AC1726	STGRSGEPANATPGGLTGP	8334
RSR-2717	AC1727	PTGRGGEGANTTPTGLPGP	8335
RSR-2718	AC1728	PTGRSGEGANATPSGLTGP	8336
RSR-2719	AC1729	TTGRASEGANSTPAPLTEP	8337
RSR-2720	AC1730	TYGRAAEAANTTPAGLTAP	8338
RSR-2721	AC1731	TTGRATEGANATPAELTEP	8339
RSR-2722	AC1732	TVGRASEEANTTPASLTGP	8340
RSR-2723	AC1733	TTGRAPEAANATPAPLTGP	8341
RSR-2724	AC1734	TWGRATEPANATPAPLTSP	8342
RSR-2725	AC1735	TVGRASESANATPAELTSP	8343
RSR-2726	AC1736	TVGRAPEGANSTPAGLTGP	8344
RSR-2727	AC1737	TWGRATEAPNLEPATLTTP	8345
RSR-2728	AC1738	TTGRATEAPNLTPAPLTEP	8346
RSR-2729	AC1739	TQGRATEAPNLSPAALTSP	8347
RSR-2730	AC1740	TQGRAAEAPNLTPATLTAP	8348
RSR-2731	AC1741	TSGRAPEATNLAPAPLTGP	8349
RSR-2732	AC1742	TQGRAAEAANLTPAGLTEP	8350
RSR-2733	AC1743	TTGRAGSAPNLPPTGLTTP	8351
RSR-2734	AC1744	TTGRAGGAENLPPEGLTAP	8352
RSR-2735	AC1745	TTSRAGTATNLTPEGLTAP	8353
RSR-2736	AC1746	TTGRAGTATNLPSPGLTTP	8354
RSR-2737	AC1747	TTARAGEAENLSPSGLTAP	8355
RSR-2738	AC1748	TTGRAGGAGNLAPGGLTEP	8356
RSR-2739	AC1749	TTGRAGTATNLPPEGLTGP	8357
RSR-2740	AC1750	TTGRAGGAANLAPTGLTEP	8358
RSR-2741	AC1751	TTGRAGTAENLAPSGLTTP	8359
RSR-2742	AC1752	TTGRAGSATNLGPGGLTGP	8360
RSR-2743	AC1753	TTARAGGAENLTPAGLTEP	8361
RSR-2744	AC1754	TTARAGSAENLSPSGLTGP	8362
RSR-2745	AC1755	TTARAGGAGNLAPEGLTTP	8363
RSR-2746	AC1756	TTSRAGAAENLTPTGLTGP	8364
RSR-2747	AC1757	TYGRTTTPGNEPPASLEAE	8365
RSR-2748	AC1758	TYSRGESGPNEPPPGLTGP	8366
RSR-2749	AC1759	AWGRTGASENETPAPLGGE	8367
RSR-2750	AC1760	RWGRAETTPNTPPEGLETE	8368
RSR-2751	AC1765	ESGRAANHTGAEPPELGAG	8369
RSR-2754	AC1801	TTGRAGEAANLTPAGLTES	8370
RSR-2755	AC1802	TTGRAGEAANLTPAALTES	8371
RSR-2756	AC1803	TTGRAGEAANLTPAPLTES	8372
RSR-2757	AC1804	TTGRAGEAANLTPEPLTES	8373
RSR-2758	AC1805	TTGRAGEAANLTPAGLTGA	8374

Nombre	ID de la construcción	Secuencia de aminoácidos	SEQ ID NO:
RSR-2759	AC1806	TTGRAGEAANLTPEGLTGA	8375
RSR-2760	AC1807	TTGRAGEAANLTPEPLTGA	8376
RSR-2761	AC1808	TTGRAGEAANLTPAGLTEA	8377
RSR-2762	AC1809	TTGRAGEAANLTPEGLTEA	8378
RSR-2763	AC1810	TTGRAGEAANLTPAPLTEA	8379
RSR-2764	AC1811	TTGRAGEAANLTPEPLTEA	8380
RSR-2765	AC1812	TTGRAGEAANLTPEPLTGP	8381
RSR-2766	AC1813	TTGRAGEAANLTPAGLTGG	8382
RSR-2767	AC1814	TTGRAGEAANLTPEGLTGG	8383
RSR-2768	AC1815	TTGRAGEAANLTPEALTGG	8384
RSR-2769	AC1816	TTGRAGEAANLTPEPLTGG	8385
RSR-2770	AC1817	TTGRAGEAANLTPAGLTEG	8386
RSR-2771	AC1818	TTGRAGEAANLTPEGLTEG	8387
RSR-2772	AC1819	TTGRAGEAANLTPAPLTEG	8388
RSR-2773	AC1820	TTGRAGEAANLTPEPLTEG	8389

Tabla 8b: Secuencias de segmento de liberación

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
RSN-0001	GSAPGSAGGYAE LRMGGAIATSGS ETPGT	8390	RSC-0001	GTAEAASASG GSAGGYAELR MGGAIPGSP	8635
RSN-0002	GSAPGTGGGYAP LRMGGGAATSG SETPGT	8391	RSC-0002	GTAEAASASG GTGGGYAPLR MGGGAPGSP	8636
RSN-0003	GSAPGAEGGYA ALRMGGEIATSG SETPGT	8392	RSC-0003	GTAEAASASG GAEGGYAALR MGGEIPGSP	8637
RSN-0004	GSAPGGPGGYAL LRMGGPAATSGS ETPGT	8393	RSC-0004	GTAEAASASG GGPGGYALLR MGGPAPGSP	8638
RSN-0005	GSAPGEAGGYAF LRMGGSIATSGS ETPGT	8394	RSC-0005	GTAEAASASG GEAGGYAFLR MGSIPGSP	8639
RSN-0006	GSAPGPGGGYAS LRMGGTAATSGS ETPGT	8395	RSC-0006	GTAEAASASG GPGGGYASLR MGGTAPGSP	8640
RSN-0007	GSAPGSEGGYAT LRMGGAIATSGS ETPGT	8396	RSC-0007	GTAEAASASG GSEGGYATLR MGGAIPGSP	8641
RSN-0008	GSAPGTPGGYAN LRMGGGAATSG SETPGT	8397	RSC-0008	GTAEAASASG GTPGGYANLR MGGGAPGSP	8642
RSN-0009	GSAPGASGGYA HLRMGGEIATSG SETPGT	8398	RSC-0009	GTAEAASASG GASGGYAHLR MGGEIPGSP	8643
RSN-0010	GSAPGGTGGYGE LRMGGPAATSGS ETPGT	8399	RSC-0010	GTAEAASASG GGTGGYGELR MGGPAPGSP	8644
RSN-0011	GSAPGEAGGYPE LRMGGSIATSGS ETPGT	8400	RSC-0011	GTAEAASASG GEAGGYPELR MGSIPGSP	8645
RSN-0012	GSAPGPGGGYVE LRMGGTAATSGS	8401	RSC-0012	GTAEAASASG GPGGGYVELR	8646

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
	ETPGT			MGGTAPGSP	
RSN-0013	GSAPGSEGGYLE LRMGGAIATSGS ETPGT	8402	RSC-0013	GTAEAASASG GSEGGYLELR MGAIPGSP	8647
RSN-0014	GSAPGTPGGYSE LRMGGGAATSG SETPGT	8403	RSC-0014	GTAEAASASG GTPGGYSEL MGGGAPGSP	8648
RSN-0015	GSAPGASGGYTE LRMGGEIATSGS ETPGT	8404	RSC-0015	GTAEAASASG GASGGYTELR MGGEIPGSP	8649
RSN-0016	GSAPGGTGGYQE LRMGGPAATSGS ETPGT	8405	RSC-0016	GTAEAASASG GGTGGYQELR MGGPAPGSP	8650
RSN-0017	GSAPGEAGGYEE LRMGGSIATSGS ETPGT	8406	RSC-0017	GTAEAASASG GEAGGYEELR MGSIPGSP	8651
RSN-0018	GSAPGPGIGPAEL RMGGTAATSGSE TPGT	8407	RSC-0018	GTAEAASASG GPGIGPAELRM GGTAPGSP	8652
RSN-0019	GSAPGSEIGAAE LRMGGAIATSGS ETPGT	8408	RSC-0019	GTAEAASASG GSEIGAAELRM GGAIPGSP	8653
RSN-0020	GSAPGTPIGSAEL RMGGGAATSGS ETPGT	8409	RSC-0020	GTAEAASASG GTPIGSAELRM GGGAPGSP	8654
RSN-0021	GSAPGASIGTAE LRMGGEIATSGS ETPGT	8410	RSC-0021	GTAEAASASG GASIGTAEELRM GGEIPGSP	8655
RSN-0022	GSAPGGTIGNAE LRMGGPAATSGS ETPGT	8411	RSC-0022	GTAEAASASG GGTIGNAEELRM GGPAPGSP	8656
RSN-0023	GSAPGEAIGQAE LRMGGSIATSGS ETPGT	8412	RSC-0023	GTAEAASASG GEAIGQAEELRM GGSIPGSP	8657
RSN-0024	GSAPGPGGPYAE LRMGGTAATSGS ETPGT	8413	RSC-0024	GTAEAASASG GPGGPYAEELR MGGTAPGSP	8658
RSN-0025	GSAPGSEGAYAE LRMGGAIATSGS ETPGT	8414	RSC-0025	GTAEAASASG GSEGAYAEELR MGAIPGSP	8659
RSN-0026	GSAPGTPGVYAE LRMGGGAATSG SETPGT	8415	RSC-0026	GTAEAASASG GTPGVYAEELR MGGGAPGSP	8660
RSN-0027	GSAPGASGLYAE LRMGGEIATSGS ETPGT	8416	RSC-0027	GTAEAASASG GASGLYAEELR MGGEIPGSP	8661
RSN-0028	GSAPGGTGIYAE LRMGGPAATSGS ETPGT	8417	RSC-0028	GTAEAASASG GGTGIYAEELRM GGPAPGSP	8662
RSN-0029	GSAPGEAGFYAE LRMGGSIATSGS ETPGT	8418	RSC-0029	GTAEAASASG GEAGFYAEELR MGSIPGSP	8663
RSN-0030	GSAPGPGGYYAE LRMGGTAATSGS ETPGT	8419	RSC-0030	GTAEAASASG GPGGYYAEELR MGGTAPGSP	8664

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
RSN-0031	GSAPGSEGSYAE LRMGGAIATSGS ETPGT	8420	RSC-0031	GTAEAASASG GSEGSYAELR MGGAIPGSP	8665
RSN-0032	GSAPGTPGNYAE LRMGGGAATSG SETPGT	8421	RSC-0032	GTAEAASASG GTPGNYAELR MGGGAPGSP	8666
RSN-0033	GSAPGASGEYAE LRMGGEIATSGS ETPGT	8422	RSC-0033	GTAEAASASG GASGEYAELR MGGEIPGSP	8667
RSN-0034	GSAPGGTGHYAE LRMGGPAATSGS ETPGT	8423	RSC-0034	GTAEAASASG GGTGHYAELR MGGPAPGSP	8668
RSN-0035	GSAPGEAGGYAE ARMGGSIAATSGS ETPGT	8424	RSC-0035	GTAEAASASG GEAGGYAEAR MGSIPGSP	8669
RSN-0036	GSAPGPGGGYAE VRMGGTAATSG SETPGT	8425	RSC-0036	GTAEAASASG GPGGGYAEVR MGGTAPGSP	8670
RSN-0037	GSAPGSEGGYAE IRMGGAIATSGS ETPGT	8426	RSC-0037	GTAEAASASG GSEGGYAEIRM GGAIPGSP	8671
RSN-0038	GSAPGTPGGYAE FRMGGGAATSG SETPGT	8427	RSC-0038	GTAEAASASG GTPGGYAEFR MGGGAPGSP	8672
RSN-0039	GSAPGASGGYAE YRMGGEIATSGS ETPGT	8428	RSC-0039	GTAEAASASG GASGGYAEYR MGGEIPGSP	8673
RSN-0040	GSAPGGTGGYAE SRMGGPAATSGS ETPGT	8429	RSC-0040	GTAEAASASG GGTGGYAESR MGGPAPGSP	8674
RSN-0041	GSAPGEAGGYAE TRMGGSIAATSGS ETPGT	8430	RSC-0041	GTAEAASASG GEAGGYAETR MGSIPGSP	8675
RSN-0042	GSAPGPGGGYAE LAMGGTRATSGS ETPGT	8431	RSC-0042	GTAEAASASG GPGGGYAEAL MGGTRPGSP	8676
RSN-0043	GSAPGSEGGYAE LVMGGARATSG SETPGT	8432	RSC-0043	GTAEAASASG GSEGGYAEVL MGGARPGSP	8677
RSN-0044	GSAPGTPGGYAE LLMGGGRATSGS ETPGT	8433	RSC-0044	GTAEAASASG GTPGGYAEEL MGGGRPGSP	8678
RSN-0045	GSAPGASGGYAE LIMGGERATSGS ETPGT	8434	RSC-0045	GTAEAASASG GASGGYAEEL MGGGRPGSP	8679
RSN-0046	GSAPGGTGGYAE LWMGGPRATSG SETPGT	8435	RSC-0046	GTAEAASASG GGTGGYAEEL MGGPRPGSP	8680
RSN-0047	GSAPGEAGGYAE LSMGGSRAATSGS ETPGT	8436	RSC-0047	GTAEAASASG GEAGGYAEEL MGGSRPGSP	8681
RSN-0048	GSAPGPGGGYAE LTMGGTRATSGS ETPGT	8437	RSC-0048	GTAEAASASG GPGGGYAEEL MGGTRPGSP	8682
RSN-0049	GSAPGSEGGYAE	8438	RSC-0049	GTAEAASASG	8683

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
	LQMGGARATSG SETPGT			GSEGGYAELQ MGGARPGSP	
RSN-0050	GSAPGTPGGYAE LNMGGGRATSG SETPGT	8439	RSC-0050	GTAEAASASG GTPGGYAELN MGGGRPGSP	8684
RSN-0051	GSAPGASGGYAE LEMGGERATSGS ETPGT	8440	RSC-0051	GTAEAASASG GASGGYAELE MGGERPGSP	8685
RSN-0052	GSAPGGTGGYAE LRPGGPIATSGSE TPGT	8441	RSC-0052	GTAEAASASG GGTGGYAELR PGGPIPGSP	8686
RSN-0053	GSAPGEAGGYAE LRAGGSAATSGS ETPGT	8442	RSC-0053	GTAEAASASG GEAGGYAELR AGGSAPGSP	8687
RSN-0054	GSAPGPGGGYAE LRLGGTIATSGSE TPGT	8443	RSC-0054	GTAEAASASG GPGGGYAELR LGGTIPGSP	8688
RSN-0055	GSAPGSEGGYAE LRIGGAAATSGS ETPGT	8444	RSC-0055	GTAEAASASG GSEGGYAELRI GGAAPGSP	8689
RSN-0056	GSAPGTPGGYAE LRSGGGIATSGSE TPGT	8445	RSC-0056	GTAEAASASG GTPGGYAELRS GGGIPGSP	8690
RSN-0057	GSAPGASGGYAE LRNGGEAATSGS ETPGT	8446	RSC-0057	GTAEAASASG GASGGYAELR NGGEAPGSP	8691
RSN-0058	GSAPGGTGGYAE LRQGGPIATSGSE TPGT	8447	RSC-0058	GTAEAASASG GGTGGYAELR QGGPIPGSP	8692
RSN-0059	GSAPGEAGGYAE LRDGGSAATSGS ETPGT	8448	RSC-0059	GTAEAASASG GEAGGYAELR DGGSAPGSP	8693
RSN-0060	GSAPGPGGGYAE LREGGTIATSGSE TPGT	8449	RSC-0060	GTAEAASASG GPGGGYAELR EGGTIPGSP	8694
RSN-0061	GSAPGSEGGYAE LRHGGAAATSGS ETPGT	8450	RSC-0061	GTAEAASASG GSEGGYAELR HGGAAPGSP	8695
RSN-0062	GSAPGTPGGYAE LRMPGGIATSGS ETPGT	8451	RSC-0062	GTAEAASASG GTPGGYAELR MPGGIPGSP	8696
RSN-0063	GSAPGASGGYAE LRMAGEAATSGS ETPGT	8452	RSC-0063	GTAEAASASG GASGGYAELR MAGEAPGSP	8697
RSN-0064	GSAPGGTGGYAE LRMVGPIATSGS ETPGT	8453	RSC-0064	GTAEAASASG GGTGGYAELR MVGPIPGSP	8698
RSN-0065	GSAPGEAGGYAE LRMLGSAATSGS ETPGT	8454	RSC-0065	GTAEAASASG GEAGGYAELR MLGSAPGSP	8699
RSN-0066	GSAPGPGGGYAE LRMIGTIATSGSE TPGT	8455	RSC-0066	GTAEAASASG GPGGGYAELR MIGTIPGSP	8700
RSN-0067	GSAPGSEGGYAE LRMYGAIATSGS	8456	RSC-0067	GTAEAASASG GSEGGYAELR	8701

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
	ETPGT			MYGAIPGSP	
RSN-0068	GSAPGTPGGYAE LRMSGGAATSGS ETPGT	8457	RSC-0068	GTAEAASASG GTPGGYAELR MSGGAPGSP	8702
RSN-0069	GSAPGASGGYAE LRMNGEIATSGS ETPGT	8458	RSC-0069	GTAEAASASG GASGGYAELR MNGEIPGSP	8703
RSN-0070	GSAPGGTGGYAE LRMQGPAATSGS ETPGT	8459	RSC-0070	GTAEAASASG GGTGGYAELR MQGPAPGSP	8704
RSN-0071	GSAPGANHTPAG LTGPGARATSGS ETPGT	8460	RSC-0071	GTAEAASASG GANHTPAGLT GPGARPGSP	8705
RSN-0072	GSAPGANTAPEG LTGPSTRATSGSE TPGT	8461	RSC-0072	GTAEAASASG GANTAPEGLTG PSTRPGSP	8706
RSN-0073	GSAPGTGAPPGG LTGPGTRATSGS ETPGT	8462	RSC-0073	GTAEAASASG GTGAPPGGLTG PGTRPGSP	8707
RSN-0074	GSAPGANHEPSG LTEGSPRATSGSE TPGT	8463	RSC-0074	GTAEAASASG GANHEPSGLTE GSPRPGSP	8708
RSN-0075	GSAPGANTEPPE LGAGTERATSGS ETPGT	8464	RSC-0075	GTAEAASASG GANTEPPELGA GTERPGSP	8709
RSN-0076	GSAPGASGPPPG LTGPPGRATSGS ETPGT	8465	RSC-0076	GTAEAASASG GASGPPPGLTG PPGRPGSP	8710
RSN-0077	GSAPGASGTPAP LGGEPRATSGS ETPGT	8466	RSC-0077	GTAEAASASG GASGTPAPLGG EPGRPGSP	8711
RSN-0078	GSAPGPAGPPEG LETEAGRATSGS ETPGT	8467	RSC-0078	GTAEAASASG GPAGPPEGLET EAGRPGSP	8712
RSN-0079	GSAPGPTSGQGG LTGPESRATSGSE TPGT	8468	RSC-0079	GTAEAASASG GPTSGQGGLTG PESRPGSP	8713
RSN-0080	GSAPGSAGGAA NLVRGGAIATSG SETPGT	8469	RSC-0080	GTAEAASASG GSAGGAANLV RGGAIPEGSP	8714
RSN-0081	GSAPGTGGGAAP LVRGGGAATSGS ETPGT	8470	RSC-0081	GTAEAASASG GTGGGAAPLV RGGGAPGSP	8715
RSN-0082	GSAPGAEGGAA ALVRGGEIATSG SETPGT	8471	RSC-0082	GTAEAASASG GAEGGAAALV RGGEIPGSP	8716
RSN-0083	GSAPGGPGGAAL LVRGGPAATSGS ETPGT	8472	RSC-0083	GTAEAASASG GGPGGAALLV RGGPAPGSP	8717
RSN-0084	GSAPGEAGGAFL LVRGGSIATSGSE TPGT	8473	RSC-0084	GTAEAASASG GEAGGAFLV RGGSIPGSP	8718
RSN-0085	GSAPGPGGGAAS LVRGGTAATSGS ETPGT	8474	RSC-0085	GTAEAASASG GPGGGAASLV RGGTAPGSP	8719

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
RSN-0086	GSAPGSEGGAAT LVRGGAIATSGS ETPGT	8475	RSC-0086	GTAEAASASG GSEGGAATLV RGGAI PGSP	8720
RSN-0087	GSAPGTPGGAAG LVRGGGAATSGS ETPGT	8476	RSC-0087	GTAEAASASG GTPGGAAGLV RGGGAPGSP	8721
RSN-0088	GSAPGASGGAA DLVRGGEIATSG SETPGT	8477	RSC-0088	GTAEAASASG GASGGAADLV RGGEIPGSP	8722
RSN-0089	GSAPGGTGGAG NLVRGGPAATSG SETPGT	8478	RSC-0089	GTAEAASASG GGTGGAGNLV RGGPAPGSP	8723
RSN-0090	GSAPGEAGGAPN LVRGGS IATSGSE TPGT	8479	RSC-0090	GTAEAASASG GEAGGAPNLV RGS SIPGSP	8724
RSN-0091	GSAPGPGGGAV NLVRGGTAATSG SETPGT	8480	RSC-0091	GTAEAASASG GPGGGAVNLV RGGTAPGSP	8725
RSN-0092	GSAPGSEGGALN LVRGGAIATSGS ETPGT	8481	RSC-0092	GTAEAASASG GSEGGALNLV RGGAI PGSP	8726
RSN-0093	GSAPGTPGGASN LVRGGGAATSGS ETPGT	8482	RSC-0093	GTAEAASASG GTPGGASNLVR GGGAPGSP	8727
RSN-0094	GSAPGASGGATN LVRGGEIATSGS ETPGT	8483	RSC-0094	GTAEAASASG GASGGATNLV RGGEIPGSP	8728
RSN-0095	GSAPGGTGGAQ NLVRGGPAATSG SETPGT	8484	RSC-0095	GTAEAASASG GGTGGAQNLV RGGPAPGSP	8729
RSN-0096	GSAPGEAGGAEN LVRGGS IATSGSE TPGT	8485	RSC-0096	GTAEAASASG GEAGGAENLV RGS SIPGSP	8730
RSN-1517	GSAPEAGRSANH EPLGLVATATSG SETPGT	8486	RSC-1517	GTAEAASASGE AGRSANHEPLG LVATPGSP	8731
BSRS-A1	GSAPASGRSTNA GPSGLAGPATSG SETPGT	8487	BSRS-A1	GTAEAASASG ASGRSTNAGPS GLAGPPGSP	8732
BSRS-A2	GSAPASGRSTNA GPQGLAGQATSG SETPGT	8488	BSRS-A2	GTAEAASASG ASGRSTNAGPQ GLAGQPGSP	8733
BSRS-A3	GSAPASGRSTNA GPPGLTG PATSG SETPGT	8489	BSRS-A3	GTAEAASASG ASGRSTNAGPP GLTGPPGSP	8734
VP-1	GSAPASSRG TNA GPAGLTGPATSG SETPGT	8490	VP-1	GTAEAASASG ASSRG TNAGPA GLTGPPGSP	8735
RSN-1752	GSAPASSRTTNT GPSTLTGPATSGS ETPGT	8491	RSC-1752	GTAEAASASG ASSRTTNTGPS TLTGPPGSP	8736
RSN-1512	GSAPAAGRSDNG TPLELVAPATSG SETPGT	8492	RSC-1512	GTAEAASASG AAGRSDNGTPL ELVAPPGSP	8737
RSN-1517	GSAPEAGRSANH	8486	RSC-1517	GTAEAASASGE	8731

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
	EPLGLVATATSG SETPGT			AGRSANHEPLG LVATPGSP	
VP-2	GSAPASGRGTNA GPAGLTGPATSG SETPGT	8493	VP-2	GTAEAASASG ASGRGTNAGP AGLTGPPGSP	8738
RSN-1018	GSAPLFGRNDNH EPLELGGGATSG SETPGT	8494	RSC-1018	GTAEAASASGL FGRNDNHEPLE LGGGPGSP	8739
RSN-1053	GSAPTAGRSDNL EPLGLVFGATSG SETPGT	8495	RSC-1053	GTAEAASASGT AGRSDNLEPLG LVFGPGSP	8740
RSN-1059	GSAPLDGRSDNF HPPELVAGATSG SETPGT	8496	RSC-1059	GTAEAASASGL DGRSDNFHPPE LVAGPGSP	8741
RSN-1065	GSAPLEGRSDNE EPENLVAGATSG SETPGT	8497	RSC-1065	GTAEAASASGL EGRSDNEEPEN LVAGPGSP	8742
RSN-1167	GSAPLKGRSDNN APLALVAGATSG SETPGT	8498	RSC-1167	GTAEAASASGL KGRSDNNAPL ALVAGPGSP	8743
RSN-1201	GSAPVYSRGTNA GPHGLTGRATSG SETPGT	8499	RSC-1201	GTAEAASASG VYSRGTNAGP HGLTGRPGSP	8744
RSN-1218	GSAPANSRGTNK GFAGLIGPATSGS ETPGT	8500	RSC-1218	GTAEAASASG ANSRGTNKGF AGLIGPPGSP	8745
RSN-1226	GSAPASSRLTNE APAGLTIPATSGS ETPGT	8501	RSC-1226	GTAEAASASG ASSRLTNEAPA GLTIPPGSP	8746
RSN-1254	GSAPDQSRGTNA GPEGLTDPATSG SETPGT	8502	RSC-1254	GTAEAASASG DQSRGTNAGPE GLTDPPGSP	8747
RSN-1256	GSAPESSRGTNIG QGGLTGPATSGS ETPGT	8503	RSC-1256	GTAEAASASGE SSRGTNIGQGG LTGPPGSP	8748
RSN-1261	GSAPSSSRGTNQ DPAGLTIPATSGS ETPGT	8504	RSC-1261	GTAEAASASGS SSRGTNQDPAG LTIPPGSP	8749
RSN-1293	GSAPASSRGQNH SPMGLTGPATSG SETPGT	8505	RSC-1293	GTAEAASASG ASSRGQNHSP MGLTGPPGSP	8750
RSN-1309	GSAPAYSRGPN GPAGLEGRATSG SETPGT	8506	RSC-1309	GTAEAASASG AYSRGPNAGP AGLEGRPGSP	8751
RSN-1326	GSAPASERGNNA GPANLTGFATSG SETPGT	8507	RSC-1326	GTAEAASASG ASERGNNAGP ANLTGFPGSP	8752
RSN-1345	GSAPASHRGTP KPAILTGPATSGS ETPGT	8508	RSC-1345	GTAEAASASG ASHRGTPNPKP ILTGPPGSP	8753
RSN-1354	GSAPMSSRRTNA NPAQLTGPATSG SETPGT	8509	RSC-1354	GTAEAASASG MSSRRTNANP AQLTGPPGSP	8754
RSN-1426	GSAPGAGRTDN HEPLELGAAATS	8510	RSC-1426	GTAEAASASG GAGRTDNHEP	8755

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
	GSETPGT			LELGAAPGSP	
RSN-1478	GSAPLAGRSENT APLELTAGATSG SETPGT	8511	RSC-1478	GTAEAASASGL AGRSENTAPLE LTAGPGSP	8756
RSN-1479	GSAPLEGRPDNH EPLALVASATSG SETPGT	8512	RSC-1479	GTAEAASASGL EGRPDNHEPLA LVASPGSP	8757
RSN-1496	GSAPLSGRSDNE EPLALPAGATSG SETPGT	8513	RSC-1496	GTAEAASASGL SGRSDNEEPLA LPAGPGSP	8758
RSN-1508	GSAPEAGRSTDNH EPELSAPATSGS ETPGT	8514	RSC-1508	GTAEAASASGE AGRSTDNHEPLE LSAPPGSP	8759
RSN-1513	GSAPEGGRSDNH GPLELVSGATSG SETPGT	8515	RSC-1513	GTAEAASASGE GGRSDNHGPLE LVSGPGSP	8760
RSN-1516	GSAPLSGRSDNE APLELEAGATSG SETPGT	8516	RSC-1516	GTAEAASASGL SGRSDNEAPLE LEAGPGSP	8761
RSN-1524	GSAPLGGRADN HEPPELGAGATS GSETPGT	8517	RSC-1524	GTAEAASASGL GGRADNHEPPE LGAGPGSP	8762
RSN-1622	GSAPPPSRGTNA EPAGLTGEATSG SETPGT	8518	RSC-1622	GTAEAASASGP PSRGTNAEPAG LTGEPGSP	8763
RSN-1629	GSAPASTRGENA GPAGLEAPATSG SETPGT	8519	RSC-1629	GTAEAASASG ASTRGENAGPA GLEAPPGSP	8764
RSN-1664	GSAPESSRGTNG APEGLTGPATSG SETPGT	8520	RSC-1664	GTAEAASASGE SSRGTNGAPEG LTGPPGSP	8765
RSN-1667	GSAPASSRATNE SPAGLTGEATSG SETPGT	8521	RSC-1667	GTAEAASASG ASSRATNESPA GLTGEPGSP	8766
RSN-1709	GSAPASSRGENP PPGGLTGPATSG SETPGT	8522	RSC-1709	GTAEAASASG ASSRGENPPPG GLTGPPGSP	8767
RSN-1712	GSAPAASRGNT GPAELTGSATSG SETPGT	8523	RSC-1712	GTAEAASASG AASRGNTGPA ELTGSPGSP	8768
RSN-1727	GSAPAGSRRTNA GPGGLEGPATSG SETPGT	8524	RSC-1727	GTAEAASASG AGSRRTNAGPG GLEGPPGSP	8769
RSN-1754	GSAPAPSRGENA GPATLTGAATSG SETPGT	8525	RSC-1754	GTAEAASASG APSRGENAGPA TLTGAPGSP	8770
RSN-1819	GSAPESGRAANT GPPTLTAPATSGS ETPGT	8526	RSC-1819	GTAEAASASGE SGRAANTGPPT LTAPPGSP	8771
RSN-1832	GSAPNPGRAANE GPPGLPGSATSG SETPGT	8527	RSC-1832	GTAEAASASG NPGRAANEGPP GLPGSPGSP	8772
RSN-1855	GSAPESSRAANL TPPELTGPATSGS ETPGT	8528	RSC-1855	GTAEAASASGE SSRAANLTPPE LTGPPGSP	8773

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
RSN-1911	GSAPASGRAANE TPPGLTGAATSG SETPGT	8529	RSC-1911	GTAEAASASG ASGRAANETPP GLTGAPGSP	8774
RSN-1929	GSAPNSGRGENL GAPGLTGTATSG SETPGT	8530	RSC-1929	GTAEAASASG NSGRGENLGAP GLTGTPGSP	8775
RSN-1951	GSAPTTGRAANL TPAGLTGPATSG SETPGT	8531	RSC-1951	GTAEAASASGT TGRAANLTPA GLTGPPGSP	8776
RSN-2295	GSAPEAGRSANH TPAGLTGPATSG SETPGT	8532	RSC-2295	GTAEAASASGE AGRSANHTPA GLTGPPGSP	8777
RSN-2298	GSAPESGRAANT TPAGLTGPATSG SETPGT	8533	RSC-2298	GTAEAASASGE SGRAANTTPAG LTGPPGSP	8778
RSN-2038	GSAPTTGRATEA ANLTPAGLTGPA TSGSETPGT	8534	RSC-2038	GTAEAASASGT TGRATEAANLT PAGLTGPPGSP	8779
RSN-2072	GSAPTTGRAEEA ANLTPAGLTGPA TSGSETPGT	8535	RSC-2072	GTAEAASASGT TGRAEEAANLT PAGLTGPPGSP	8780
RSN-2089	GSAPTTGRAGEA ANLTPAGLTGPA TSGSETPGT	8536	RSC-2089	GTAEAASASGT TGRAGEAANL TPAGLTGPPGS P	8781
RSN-2302	GSAPTTGRATEA ANATPAGLTGPA TSGSETPGT	8537	RSC-2302	GTAEAASASGT TGRATEAANA TPAGLTGPPGS P	8782
RSN-3047	GSAPTTGRAGEA EGATSAGATGPA TSGSETPGT	8538	RSC-3047	GTAEAASASGT TGRAGEAEGA TSAGATGPPGS P	8783
RSN-3052	GSAPTTGEAGEA ANATSAGATGPA TSGSETPGT	8539	RSC-3052	GTAEAASASGT TGEAGEAANA TSAGATGPPGS P	8784
RSN-3043	GSAPTTGEAGEA AGLTPAGLTGPA TSGSETPGT	8540	RSC-3043	GTAEAASASGT TGEAGEAAGL TPAGLTGPPGS P	8785
RSN-3041	GSAPTTGAAGEA ANATPAGLTGPA TSGSETPGT	8541	RSC-3041	GTAEAASASGT TGAAGEAANA TPAGLTGPPGS P	8786
RSN-3044	GSAPTTGRAGEA AGLTPAGLTGPA TSGSETPGT	8542	RSC-3044	GTAEAASASGT TGRAGEAAGL TPAGLTGPPGS P	8787
RSN-3057	GSAPTTGRAGEA ANATSAGATGPA TSGSETPGT	8543	RSC-3057	GTAEAASASGT TGRAGEAANA TSAGATGPPGS P	8788
RSN-3058	GSAPTTGEAGEA AGATSAGATGPA	8544	RSC-3058	GTAEAASASGT TGEAGEAAGA	8789

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
	TSGSETPGT			TSAGATGPPGSP	
RSN-2485	GSAPESGRAANT EPPELGAGATSG SETPGT	8545	RSC-2485	GTAEAASASGE SGRAANTEPPE LGAGPGSP	8790
RSN-2486	GSAPESGRAANT APEGLTG PATSG SETPGT	8546	RSC-2486	GTAEAASASGE SGRAANTAPEG LTGPPGSP	8791
RSN-2488	GSAPEPGRAANH EPSGLTEGATSG SETPGT	8547	RSC-2488	GTAEAASASGE PGRAANHEPSG LTEGPGSP	8792
RSN-2599	GSAPESGRAANH TGAPPGGLTGPA TSGSETPGT	8548	RSC-2599	GTAEAASASGE SGRAANHTGA PPGGLTGPPGSP	8793
RSN-2706	GSAPTTGRTGEG ANATPGGLTGPA TSGSETPGT	8549	RSC-2706	GTAEAASASGT TGR TGEGANA TPGGLTGPPGSP	8794
RSN-2707	GSAPRTGRSGEA ANETPEGLEGPA TSGSETPGT	8550	RSC-2707	GTAEAASASGR TGRSGEAANET PEGLEGPPGSP	8795
RSN-2708	GSAPRTGRTGES ANETPAGLG GPA TSGSETPGT	8551	RSC-2708	GTAEAASASGR TGR TGESANET PAGLGPPGSP	8796
RSN-2709	GSAPSTGRTGEP ANETPAGLSGPA TSGSETPGT	8552	RSC-2709	GTAEAASASGS TGR TGEPANET PAGLSGPPGSP	8797
RSN-2710	GSAPTTGRAGEP ANATPTGLSGPA TSGSETPGT	8553	RSC-2710	GTAEAASASGT TGRAGEPANA TPTGLSGPPGSP	8798
RSN-2711	GSAPRTGRPGEG ANATPTGLPGPA TSGSETPGT	8554	RSC-2711	GTAEAASASGR TGRPGEGANA TPTGLPGPPGSP	8799
RSN-2712	GSAPRTGRGGEA ANATPSGLGGPA TSGSETPGT	8555	RSC-2712	GTAEAASASGR TGRGGEAANA TPSGLGGPPGSP	8800
RSN-2713	GSAPSTGRSGES ANATPGGLGGPA TSGSETPGT	8556	RSC-2713	GTAEAASASGS TGRSGESANAT PGGLGGPPGSP	8801
RSN-2714	GSAPRTGRTGEE ANATPAGLP GPA TSGSETPGT	8557	RSC-2714	GTAEAASASGR TGR TGEEANAT PAGLP GPPGSP	8802
RSN-2715	GSAPATGRPGEP ANTTPEGLEGPA TSGSETPGT	8558	RSC-2715	GTAEAASASG ATGRPGEPANT TPEGLEGPPGSP	8803
RSN-2716	GSAPSTGRSGEP ANATPGGLTGPA TSGSETPGT	8559	RSC-2716	GTAEAASASGS TGRSGEPANAT PGGLTGPPGSP	8804
RSN-2717	GSAPPTGRGGEG ANTTPTGLPGPA TSGSETPGT	8560	RSC-2717	GTAEAASASGP TGRGGEGANT TPTGLPGPPGSP	8805
RSN-2718	GSAPPTGRSGEG	8561	RSC-2718	GTAEAASASGP	8806

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
	ANATPSGLTGPA TSGSETPGT			TGRSGEGANA TPSGLTGPPGSP	
RSN-2719	GSAPTTGRASEG ANSTPAPLTEPA TSGSETPGT	8562	RSC-2719	GTAEAASASGT TGRASEGANST PAPLTEPPGSP	8807
RSN-2720	GSAPTYGRAAEA ANTTPAGLTAPA TSGSETPGT	8563	RSC-2720	GTAEAASASGT YGRAAEAANT TPAGLTAPPGS P	8808
RSN-2721	GSAPTTGRATEG ANATPAELTEPA TSGSETPGT	8564	RSC-2721	GTAEAASASGT TGRATEGANA TPAELTEPPGSP	8809
RSN-2722	GSAPTVGRASEE ANTTPASLTGPA TSGSETPGT	8565	RSC-2722	GTAEAASASGT VGRASEEANTT PASLTGPPGSP	8810
RSN-2723	GSAPTTGRAPEA ANATPAPLTGPA TSGSETPGT	8566	RSC-2723	GTAEAASASGT TGRAPEAANA TPAPLTGPPGSP	8811
RSN-2724	GSAPTWGRATEP ANATPAPLTSPA TSGSETPGT	8567	RSC-2724	GTAEAASASGT WGRATEPANA TPAPLTSPPGSP	8812
RSN-2725	GSAPTVGRASES ANATPAELTSPA TSGSETPGT	8568	RSC-2725	GTAEAASASGT VGRASESANAT PAELTSPPGSP	8813
RSN-2726	GSAPTVGRAPEG ANSTPAGLTGPA TSGSETPGT	8569	RSC-2726	GTAEAASASGT VGRAPEGANST PAGLTGPPGSP	8814
RSN-2727	GSAPTWGRATE APNLEPATLTTP ATSGSETPGT	8570	RSC-2727	GTAEAASASGT WGRATEAPNL EPATLTTPPGSP	8815
RSN-2728	GSAPTTGRATEA PNLTPAPLTEPAT SGSETPGT	8571	RSC-2728	GTAEAASASGT TGRATEAPNLT PAPLTEPPGSP	8816
RSN-2729	GSAPTQGRATEA PNLSPAALTSPAT SGSETPGT	8572	RSC-2729	GTAEAASASGT QGRATEAPNLS PAALTSPPGSP	8817
RSN-2730	GSAPTQGRAAEA PNLTPATLTAPA TSGSETPGT	8573	RSC-2730	GTAEAASASGT QGRAAEAPNL TPATLTAPPGS P	8818
RSN-2731	GSAPTSGRAPEA TNLAPAPLTGPA TSGSETPGT	8574	RSC-2731	GTAEAASASGT SGRAPEATNLA PAPLTGPPGSP	8819
RSN-2732	GSAPTQGRAAEA ANLTPAGLTEPA TSGSETPGT	8575	RSC-2732	GTAEAASASGT QGRAAEAANL TPAGLTEPPGS P	8820
RSN-2733	GSAPTTGRAGSA PNLPPTGLTTPAT SGSETPGT	8576	RSC-2733	GTAEAASASGT TGRAGSAPNLP PTGLTTPPGSP	8821
RSN-2734	GSAPTTGRAGGA ENLPPEGLTAPA TSGSETPGT	8577	RSC-2734	GTAEAASASGT TGRAGGAENL PPEGLTAPPGSP	8822
RSN-2735	GSAPTTSRAGTA TNLTPEGLTAPA	8578	RSC-2735	GTAEAASASGT TSRAGTATNLT	8823

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
	TSGSETPGT			PEGLTAPPGSP	
RSN-2736	GSAPTTGRAGTA TNLPPSGLTTPAT SGSETPGT	8579	RSC-2736	GTAEAASASGT TGRAGTATNLP PSGLTTPPGSP	8824
RSN-2737	GSAPTTARAGEA ENLSPSGLTAPA TSGSETPGT	8580	RSC-2737	GTAEAASASGT TARAGEAENLS PSGLTAPPGSP	8825
RSN-2738	GSAPTTGRAGGA GNLAPGGLTEPA TSGSETPGT	8581	RSC-2738	GTAEAASASGT TGRAGGAGNL APGGLTEPPGS P	8826
RSN-2739	GSAPTTGRAGTA TNLPPEGLTGPA TSGSETPGT	8582	RSC-2739	GTAEAASASGT TGRAGTATNLP PEGLTGPPGSP	8827
RSN-2740	GSAPTTGRAGGA ANLAPTGLTEPA TSGSETPGT	8583	RSC-2740	GTAEAASASGT TGRAGGAANL APTGLTEPPGS P	8828
RSN-2741	GSAPTTGRAGTA ENLAPSGLTTPA TSGSETPGT	8584	RSC-2741	GTAEAASASGT TGRAGTAENL APSGLTTPPGSP	8829
RSN-2742	GSAPTTGRAGSA TNLGPGGLTGPA TSGSETPGT	8585	RSC-2742	GTAEAASASGT TGRAGSATNL GPGGLTGPPGS P	8830
RSN-2743	GSAPTTARAGGA ENLTPAGLTEPA TSGSETPGT	8586	RSC-2743	GTAEAASASGT TARAGGAENL TPAGLTEPPGS P	8831
RSN-2744	GSAPTTARAGSA ENLSPSGLTGPA TSGSETPGT	8587	RSC-2744	GTAEAASASGT TARAGSAENLS PSGLTGPPGSP	8832
RSN-2745	GSAPTTARAGGA GNLAPEGLTTPA TSGSETPGT	8588	RSC-2745	GTAEAASASGT TARAGGAGNL APEGLTTPPGS P	8833
RSN-2746	GSAPTTSRAGAA ENLTPTGLTGPA TSGSETPGT	8589	RSC-2746	GTAEAASASGT TSRAGAAENLT PTGLTGPPGSP	8834
RSN-2747	GSAPTYGRITTP GNEPPASLEAEA TSGSETPGT	8590	RSC-2747	GTAEAASASGT YGRITTPGNEP PASLEAEPGSP	8835
RSN-2748	GSAPTYSRGESG PNEPPPGLTGPA TSGSETPGT	8591	RSC-2748	GTAEAASASGT YSRGESGPNEP PPGLTGPPGSP	8836
RSN-2749	GSAPAWGRTGA SENETPAPLGGE ATSGSETPGT	8592	RSC-2749	GTAEAASASG AWGRTGASEN ETPAPLGGEPS P	8837
RSN-2750	GSAPRWGRAETT PNTPEGLETEAT SGSETPGT	8593	RSC-2750	GTAEAASASGR WGRAETTPNTP PEGLETEPGSP	8838
RSN-2751	GSAPESGRAANH TGAEPPELGAGA TSGSETPGT	8594	RSC-2751	GTAEAASASGE SGRAANHTGA EPPELGAGPGS	8839

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
				P	
RSN-2754	GSAPTTGRAGEA ANLTPAGLTESA TSGSETPGT	8595	RSC-2754	GTAEAASASGT TGRAGEAANL TPAGLTESPGS P	8840
RSN-2755	GSAPTTGRAGEA ANLTPAALTESA TSGSETPGT	8596	RSC-2755	GTAEAASASGT TGRAGEAANL TPAALTESPGS P	8841
RSN-2756	GSAPTTGRAGEA ANLTPAPLTESA TSGSETPGT	8597	RSC-2756	GTAEAASASGT TGRAGEAANL TPAPLTESPGSP	8842
RSN-2757	GSAPTTGRAGEA ANLTPEPLTESAT SGSETPGT	8598	RSC-2757	GTAEAASASGT TGRAGEAANL TPEPLTESPGSP	8843
RSN-2758	GSAPTTGRAGEA ANLTPAGLTGAA TSGSETPGT	8599	RSC-2758	GTAEAASASGT TGRAGEAANL TPAGLTGAPGS P	8844
RSN-2759	GSAPTTGRAGEA ANLTPEGLTGAA TSGSETPGT	8600	RSC-2759	GTAEAASASGT TGRAGEAANL TPEGLTGAPGS P	8845
RSN-2760	GSAPTTGRAGEA ANLTPEPLTGAA TSGSETPGT	8601	RSC-2760	GTAEAASASGT TGRAGEAANL TPEPLTGAPGS P	8846
RSN-2761	GSAPTTGRAGEA ANLTPAGLTEAA TSGSETPGT	8602	RSC-2761	GTAEAASASGT TGRAGEAANL TPAGLTEAPGS P	8847
RSN-2762	GSAPTTGRAGEA ANLTPEGLTEAA TSGSETPGT	8603	RSC-2762	GTAEAASASGT TGRAGEAANL TPEGLTEAPGS P	8848
RSN-2763	GSAPTTGRAGEA ANLTPAPLTEAA TSGSETPGT	8604	RSC-2763	GTAEAASASGT TGRAGEAANL TPAPLTEAPGS P	8849
RSN-2764	GSAPTTGRAGEA ANLTPEPLTEAA TSGSETPGT	8605	RSC-2764	GTAEAASASGT TGRAGEAANL TPEPLTEAPGSP	8850
RSN-2765	GSAPTTGRAGEA ANLTPEPLTGPA TSGSETPGT	8606	RSC-2765	GTAEAASASGT TGRAGEAANL TPEPLTGPPGSP	8851
RSN-2766	GSAPTTGRAGEA ANLTPAGLTGGA TSGSETPGT	8607	RSC-2766	GTAEAASASGT TGRAGEAANL TPAGLTGGPGS P	8852
RSN-2767	GSAPTTGRAGEA ANLTPEGLTGGA TSGSETPGT	8608	RSC-2767	GTAEAASASGT TGRAGEAANL TPEGLTGPGS P	8853
RSN-2768	GSAPTTGRAGEA ANLTPEALTGGA	8609	RSC-2768	GTAEAASASGT TGRAGEAANL	8854

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
	TSGSETPGT			TPEALTGGPGSP	
RSN-2769	GSAPTTGRAGEA ANLTPEPLTGGA TSGSETPGT	8610	RSC-2769	GTAEAASASGT TGRAGEAANL TPEPLTGGPGSP	8855
RSN-2770	GSAPTTGRAGEA ANLTPAGLTEGA TSGSETPGT	8611	RSC-2770	GTAEAASASGT TGRAGEAANL TPAGLTEGPGSP	8856
RSN-2771	GSAPTTGRAGEA ANLTPEGLTEGA TSGSETPGT	8612	RSC-2771	GTAEAASASGT TGRAGEAANL TPEGLTEGPGSP	8857
RSN-2772	GSAPTTGRAGEA ANLTPAPLTEGA TSGSETPGT	8613	RSC-2772	GTAEAASASGT TGRAGEAANL TPAPLTEGPGSP	8858
RSN-2773	GSAPTTGRAGEA ANLTPEPLTEGA TSGSETPGT	8614	RSC-2773	GTAEAASASGT TGRAGEAANL TPEPLTEGPGSP	8859
RSN-3047	GSAPTTGRAGEA EGATSAGATGPA TSGSETPGT	8538	RSC-3047	GTAEAASASGT TGRAGEAEGA TSAGATGPPGSP	8783
RSN-2783	GSAPEAGRSAEA TSAGATGPATSG SETPGT	8615	RSC-2783	GTAEAASASGE AGRSAEATSAG ATGPPGSP	8860
RSN-3107	GSAPSASGTYSR GESGPGSPATSG SETPGT	8616	RSC-3107	GTAEAASASGS ASGTYSRGESG PGSPPGSP	8861
RSN-3103	GSAPSASGEAGR TDTHPGSPATSG SETPGT	8617	RSC-3103	GTAEAASASGS ASGEAGRDTDT HPGSPPGSP	8862
RSN-3102	GSAPSASGEPGR AAEHGPGSPATSG SETPGT	8618	RSC-3102	GTAEAASASGS ASGEPGRAAEH PGSPPGSP	8863
RSN-3119	GSAPSPAGESSR GTTIAGSPATSGS ETPGT	8619	RSC-3119	GTAEAASASGS PAGESSRGTTI AGSPPGSP	8864
RSN-3043	GSAPTTGEAGEA AGLTPAGLTGPA TSGSETPGT	8540	RSC-3043	GTAEAASASGT TGEAGEAAGL TPAGLTGPPGSP	8785
RSN-2789	GSAPEAGESAGA TPAGLTGPATSG SETPGT	8620	RSC-2789	GTAEAASASGE AGESAGATPA GLTGPPGSP	8865
RSN-3109	GSAPSASGAPLE LEAGPGSPATSG SETPGT	8621	RSC-3109	GTAEAASASGS ASGAPLELEAG PGSPPGSP	8866
RSN-3110	GSAPSASGEPPEL GAGPGSPATSGS ETPGT	8622	RSC-3110	GTAEAASASGS ASGEPPELGAG PGSPPGSP	8867
RSN-3111	GSAPSASGEPSP LTEGPGSPATSGS	8623	RSC-3111	GTAEAASASGS ASGEPSPGLTEG	8868

Nombre	Secuencia de aminoácidos	SEQ ID NO:	Nombre	Secuencia de aminoácidos	SEQ ID NO:
	ETPGT			PGSPPGSP	
RSN-3112	GSAPSASGTPAP LTEPPGSPATSGS ETPGT	8624	RSC-3112	GTAEAASASGS ASGTPAPLTEP PGSPPGSP	8869
RSN-3113	GSAPSASGTPAE LTEPPGSPATSGS ETPGT	8625	RSC-3113	GTAEAASASGS ASGTPAELTEP PGSPPGSP	8870
RSN-3114	GSAPSASGPPPGL TGPPGSPATSGSE TPGT	8626	RSC-3114	GTAEAASASGS ASGPPPGLTGP PGSPPGSP	8871
RSN-3115	GSAPSASGTPAP LGGEPPGSPATSG SETPGT	8627	RSC-3115	GTAEAASASGS ASGTPAPLGGE PGSPPGSP	8872
RSN-3125	GSAPSPAGAPEG LTGPAGSPATSG SETPGT	8628	RSC-3125	GTAEAASASGS PAGAPEGLTGP AGSPPGSP	8873
RSN-3126	GSAPSPAGPPEG LETEAGSPATSG SETPGT	8629	RSC-3126	GTAEAASASGS PAGPPEGLETE AGSPPGSP	8874
RSN-3127	GSAPSPTSGQGG LTGPGSEPATSGS ETPGT	8630	RSC-3127	GTAEAASASGS PTSGQGGLTGP GSEPPGSP	8875
RSN-3131	GSAPSESAPPEGL ETESTEPATSGSE TPGT	8631	RSC-3131	GTAEAASASGS ESAPPEGLETES TEPPGSP	8876
RSN-3132	GSAPSEGSEPLEL GAASETPATSGS ETPGT	8632	RSC-3132	GTAEAASASGS EGSEPLELGAA SETPPGSP	8877
RSN-3133	GSAPSEGSGPAG LEAPSETPATSGS ETPGT	8633	RSC-3133	GTAEAASASGS EGSGPAGLEAP SETPPGSP	8878
RSN-3138	GSAPSEPTPPASL EAEPGSPATSGSE TPGT	8634	RSC-3138	GTAEAASASGS EPTPPASLEAEP GSPPGSP	8879

En otro aspecto, los RS para su incorporación en los polipéptidos recombinantes humanos o animales pueden diseñarse de modo que sean selectivamente sensibles para tener diferentes tasas de escisión y diferentes eficiencias de escisión para las diversas proteasas de las cuales son sustratos. Puesto que una proteasa dada puede encontrarse en diferentes concentraciones en tejidos enfermos, que incluyen, pero sin limitarse a, un tumor, un cáncer de la sangre, o un tejido inflamatorio o sitio de inflamación en comparación con tejidos sanos o en la circulación, la descripción proporciona RS a los que se han modificado genéticamente secuencias de aminoácidos individuales de modo que tengan una eficiencia de escisión mayor o menor para una proteasa dada para garantizar que el polipéptido recombinante se convierta preferentemente de la forma de profármaco a la forma activa (es decir, mediante la separación y liberación de los restos de unión y XTEN del polipéptido recombinante después de la escisión del RS) cuando está cerca de la célula o tejido objetivo y sus proteasas colocalizadas en comparación con la velocidad de escisión del RS en tejido

sano o en la circulación en donde los restos de unión de fragmento de anticuerpo liberados tienen una mayor capacidad de unirse a los ligandos en los tejidos enfermos en comparación con la forma de profármaco que permanece en circulación. Mediante tales diseños selectivos, el índice terapéutico de las composiciones resultantes puede mejorarse, lo que da como resultado efectos secundarios reducidos con relación a los tratamientos convencionales que no incorporan tal activación específica del sitio.

Como se usa en la presente descripción, la eficiencia de escisión se define como el valor de $\log_2$ de la relación del porcentaje del sustrato de prueba que comprende el RS escindido y el porcentaje del sustrato de control AC1611 escindido cuando cada uno es humano o animal para la enzima proteasa en ensayos bioquímicos (con más detalle en los Ejemplos) en los que se realiza la reacción en donde la concentración inicial del sustrato es 6 μM , las reacciones se incuban a 37 °C durante 2 horas antes de detenerlas por adición de EDTA, donde la cantidad de productos de digestión y sustrato no escindido se analiza por SDS-PAGE no reductora para establecer la relación del porcentaje escindido. La eficiencia de escisión se calcula de la siguiente manera:

$$\text{Log}_2 \left(\frac{\% \text{ escindido para el sustrato de interés}}{\% \text{ escindido para AC1611 en el mismo experimento}} \right).$$

Por lo tanto, una eficiencia de

escisión de -1 significa que la cantidad de sustrato de prueba escindido fue del 50 % en comparación con la del sustrato de control, mientras que una eficiencia de escisión de +1 significa que la cantidad de sustrato de prueba escindido fue del 200 % en comparación con la del sustrato de control. Una velocidad de escisión más alta por la proteasa de prueba con relación al control daría como resultado una mayor eficiencia de escisión, y una velocidad más lenta de escisión por la proteasa de prueba con relación al control daría como resultado una menor eficiencia de escisión. Como se detalla en los Ejemplos, se estableció que una secuencia de RS de control AC1611 (RSR-1517), que tiene la secuencia de aminoácidos EAGRSANHEPLGLVAT (SEQ ID NO: 8261), tiene una eficiencia de escisión inicial adecuada por las proteasas legumina, MMP-2, MMP-7, MMP-9, MMP-14, uPA y matriptasa, cuando se evalúa en ensayos bioquímicos *in vitro* para las velocidades de escisión por las proteasas individuales. Mediante la sustitución selectiva de aminoácidos en ubicaciones individuales en los péptidos RS, se crearon bibliotecas de RS y se evaluaron contra el panel de las 7 proteasas (detalladas con más detalle en los Ejemplos), lo que da como resultado perfiles que se usaron para establecer pautas para las sustituciones de aminoácidos adecuadas para lograr RS con eficiencias de escisión deseadas. Al producir RS con eficiencias de escisión deseadas, se prefieren las sustituciones mediante el uso de los aminoácidos hidrófilos A, E, G, P, S y T, sin embargo, otros L-aminoácidos pueden sustituirse en posiciones dadas para ajustar la eficiencia de escisión siempre que el RS retenga al menos cierta susceptibilidad a la escisión por una proteasa. Las sustituciones conservativas de aminoácidos en un péptido para retener o efectuar la actividad están dentro del conocimiento y las capacidades de un experto en la técnica. En una modalidad, la descripción proporciona RS en el que el RS se escinde mediante una proteasa

seleccionada de legumaína, MMP-1, MMP-2, MMP-7, MMP-9, MMP-11, MMP-14, uPA o matriptasa con una eficiencia de escisión al menos $0,2 \log_2$, o $0,4 \log_2$, o $0,8 \log_2$, o $1,0 \log_2$ mayor en un ensayo competitivo bioquímico *in vitro* en comparación con la escisión por la misma proteasa de una secuencia de control RSR-1517 que tiene la secuencia EAGRSANHEPLGLVAT (SEQ ID NO: 8261). En otra modalidad, la descripción proporciona RS en el que el RS se escinde mediante una proteasa seleccionada de legumaína, MMP-1, MMP-2, MMP-7, MMP-9, MMP-11, MMP-14, uPA o matriptasa con al menos una eficiencia de escisión al menos $0,2 \log_2$, o $0,4 \log_2$, o $0,8 \log_2$, o $1,0 \log_2$ menor en un ensayo competitivo bioquímico *in vitro* en comparación con la escisión por la misma proteasa de una secuencia de control RSR-1517 que tiene la secuencia EAGRSANHEPLGLVAT (SEQ ID NO: 8261). En una modalidad, la descripción proporciona RS en el que la velocidad de escisión del RS por una proteasa seleccionada de legumaína, MMP-1, MMP-2, MMP-7, MMP-9, MMP-11, MMP-14, uPA o matriptasa es al menos 2 veces, o al menos 4 veces, o al menos 8 veces, o al menos 16 veces más rápida en comparación con la secuencia de control RSR-1517 que tiene la secuencia EAGRSANHEPLGLVAT (SEQ ID NO: 8261). En otra modalidad, la descripción proporciona RS en el que la velocidad de escisión del RS por una proteasa seleccionada de legumaína, MMP-1, MMP-2, MMP-7, MMP-9, MMP-11, MMP-14, uPA o matriptasa es al menos 2 veces, o al menos 4 veces, o al menos 8 veces, o al menos 16 veces más lenta en comparación con la secuencia de control RSR-1517 que tiene la secuencia EAGRSANHEPLGLVAT (SEQ ID NO: 8261).

En otro aspecto, la descripción proporciona AAC que comprende múltiples RS en donde cada secuencia RS se selecciona del grupo de secuencias expuestas en la Tabla 8a y los RS se unen entre sí por 1 a 6 aminoácidos seleccionados de glicina, serina, alanina y treonina. En una modalidad, la AAC comprende una primera RS y una segunda RS diferente de la primera RS en donde cada secuencia de RS se selecciona del grupo de secuencias expuestas en la Tabla 8a y los RS se unen entre sí por 1 a 6 aminoácidos seleccionados de glicina, serina, alanina y treonina. En otra modalidad, la AAC comprende un primer RS, un segundo RS diferente del primer RS y un tercer RS diferente del primer y el segundo RS en donde cada secuencia se selecciona del grupo de secuencias expuesto en la Tabla 8a y el primer y el segundo y el tercer RS se unen entre sí por 1 a 6 aminoácidos seleccionados de glicina, serina, alanina y treonina. Se pretende específicamente que los múltiples RS de la AAC puedan concatenarse para formar una secuencia que pueda escindirse por múltiples proteasas a diferentes velocidades o eficiencia de escisión. En otra modalidad, la descripción proporciona AAC que comprende un RS1 y un RS2 seleccionados del grupo de secuencias expuestas en las Tablas 8a-8b y un XTEN 1 y XTEN 2, tales como los descritos anteriormente o descritos en otra parte en la presente descripción, en donde el RS1 está fusionado entre el XTEN1 y los restos de unión y el RS2 está fusionado entre el XTEN2 y los restos de unión. Se contempla que dichas composiciones se escindirían más fácilmente en tejidos objetivo enfermos que expresan múltiples proteasas, en comparación con tejidos

sanos o cuando están en circulación normal, con el resultado de que los fragmentos resultantes que portan los restos de unión penetrarían más fácilmente en el tejido objetivo; por ejemplo, un tumor, y tendrían una capacidad mejorada para unirse y vincular la célula objetivo y la célula efectora (o solo la célula objetivo en el caso de la AAC diseñada con un único resto de unión).

Los RS de la descripción son útiles para su inclusión en polipéptidos recombinantes como agentes terapéuticos para el tratamiento de cánceres, enfermedades autoinmunitarias, enfermedades inflamatorias y otras afecciones en las que es conveniente la activación localizada del polipéptido recombinante. Las composiciones humanas o animales abordan una necesidad insatisfecha y son superiores en uno o más aspectos que incluyen una semivida terminal mejorada, suministro dirigido y una relación terapéutica mejorada con toxicidad reducida para los tejidos sanos en comparación con los agentes terapéuticos tipo anticuerpos convencionales o los agentes terapéuticos tipo anticuerpos biespecíficos que son activos tras la inyección.

En algunas modalidades, el polipéptido (de fusión) comprende un primer segmento de liberación (RS1) ubicado entre el (primer) XTEN y el polipéptido biológicamente activo. En algunas modalidades, el polipéptido comprende además un segundo segmento de liberación (RS2) ubicado entre el polipéptido biológicamente activo y el segundo XTEN. En algunas modalidades, RS1 y RS2 son idénticos en secuencia. En algunas modalidades, RS1 y RS2 no son idénticos en secuencia. En algunas modalidades, el RS1 comprende una secuencia de aminoácidos que tiene al menos 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una secuencia expuesta en las Tablas 8a-8b. En algunas modalidades, el RS2 comprende una secuencia de aminoácidos que tiene al menos 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con una secuencia expuesta en las Tablas 8a-8b. En algunas modalidades, el RS1 y el RS2 son cada uno un sustrato para la escisión por múltiples proteasas en uno, dos o tres sitios de escisión dentro de cada secuencia de segmento de liberación.

Fragmento de referencia

En algunas modalidades, el polipéptido (de fusión) comprende además uno o más fragmentos de referencia liberables del polipéptido tras la digestión por la proteasa. En algunas modalidades, el uno o más fragmentos de referencia comprenden cada uno una porción del polipéptido biológicamente activo. En algunas modalidades, el uno o más fragmentos de referencia es un único fragmento de referencia que difiere en secuencia y peso molecular de todos los demás fragmentos peptídicos que pueden liberarse del polipéptido tras la digestión del polipéptido por la proteasa.

Mezcla de polipéptidos

Lo descrito en la presente descripción incluye una mezcla que comprende una pluralidad de polipéptidos de longitud variable; la mezcla comprende un primer conjunto de polipéptidos y un segundo conjunto de polipéptidos. En algunas modalidades, cada polipéptido del

primer conjunto de polipéptidos comprende un fragmento de código de barras que (a) puede liberarse del polipéptido por digestión con una proteasa y (b) tiene una secuencia y peso molecular que difieren de la secuencia y el peso molecular de todos otros fragmentos que pueden liberarse del primer conjunto de polipéptidos. En algunas modalidades, el segundo conjunto de polipéptidos carece del fragmento de código de barras del primer conjunto de polipéptidos. En algunas modalidades, tanto el primer conjunto de polipéptidos como el segundo conjunto de polipéptidos comprenden cada uno un fragmento de referencia que (a) es común al primer conjunto de polipéptidos y al segundo conjunto de polipéptidos y (b) liberable por digestión con la proteasa. En algunas modalidades, la relación del primer conjunto de polipéptidos respecto a los polipéptidos que comprenden el fragmento de referencia es mayor que 0,70. En algunas modalidades, la relación del primer conjunto de polipéptidos respecto a los polipéptidos que comprenden el fragmento de referencia es mayor que 0,8, 0,9, 0,95 o 0,98. En algunas modalidades, el fragmento de referencia aparece no más de una vez en cada polipéptido del primer conjunto de polipéptidos y el segundo conjunto de polipéptidos. En algunas modalidades, la proteasa es una proteasa que escinde en el lado C-terminal de los residuos de ácido glutámico. En algunas modalidades, la proteasa es una proteasa Glu-C. En algunas modalidades, la proteasa no es tripsina. En algunas modalidades, los polipéptidos de longitudes variables comprenden polipéptidos que comprenden al menos un polipéptido recombinante extendido (XTEN), tal como cualquiera descrito anteriormente o descrito en cualquier otra parte en la presente descripción. En algunas modalidades, el primer conjunto de polipéptidos comprende un polipéptido de longitud completa, en donde el fragmento de código de barras es una porción del polipéptido de longitud completa. En algunas modalidades, el polipéptido de longitud completa es un polipéptido (de fusión), tal como cualquiera descrito anteriormente o descrito en cualquier otra parte en la presente descripción. En algunas modalidades, el fragmento de código de barras carece (no comprende) tanto el aminoácido N-terminal como el aminoácido C-terminal del polipéptido de longitud completa. En algunas modalidades, la mezcla de polipéptidos de longitudes variables difieren entre sí debido al truncamiento N-terminal, el truncamiento C-terminal o el truncamiento tanto N-como C-terminal de un polipéptido de longitud completa. En algunas modalidades, el primer conjunto de polipéptidos y el segundo conjunto de polipéptidos pueden diferir en una o más propiedades farmacológicas. Las propiedades ilustrativas no limitantes incluyen.

Método de caracterización de polipéptidos

Lo descrito en la presente descripción incluye un método para evaluar, en una mezcla que comprende polipéptidos de longitud variable, una cantidad relativa de un primer conjunto de polipéptidos en la mezcla con respecto a un segundo conjunto de polipéptidos en la mezcla, en donde (1) cada polipéptido del primer conjunto de polipéptidos comparte un fragmento de código de barras que aparece una y solo una vez en el polipéptido y (2) cada polipéptido del segundo conjunto de polipéptidos carece del fragmento de código de barras que comparten los polipéptidos del primer conjunto, en donde los polipéptidos individuales

tanto del primero de los polipéptidos como del primero el segundo conjunto de polipéptidos comprenden cada uno un fragmento de referencia. El método puede comprender poner en contacto la mezcla con una proteasa para producir una pluralidad de fragmentos proteolíticos que son el resultado de la escisión del primer conjunto de polipéptidos y el segundo conjunto de polipéptidos, en donde la pluralidad de fragmentos proteolíticos comprende una pluralidad de fragmentos de referencia y una pluralidad de fragmentos de código de barras. El método puede comprender además la determinación de una relación de la cantidad de fragmentos de código de barras respecto a la cantidad de fragmentos de referencia, para evaluar de esta manera las cantidades relativas del primer conjunto de polipéptidos respecto al segundo conjunto de polipéptidos. En algunas modalidades, el fragmento de código de barras aparece no más de una vez en cada polipéptido del primer conjunto de polipéptidos. En algunas modalidades, el fragmento de referencia aparece no más de una vez en cada polipéptido del primer conjunto de polipéptidos y el segundo conjunto de polipéptidos. En algunas modalidades, la pluralidad de fragmentos proteolíticos comprende una pluralidad de fragmentos de referencia y una pluralidad de fragmentos de código de barras. En algunas modalidades, la proteasa escinde el primer y el segundo conjuntos de polipéptidos (o los polipéptidos de longitud variable) en el lado C-terminal de los residuos de ácido glutámico que no van seguidos por un residuo de prolina. En algunas modalidades, la proteasa es una proteasa Glu-C. En algunas modalidades, la proteasa no es tripsina. En algunas modalidades, la etapa de determinación de una relación de la cantidad de fragmentos de código de barras respecto a la cantidad de fragmentos de referencia comprende cuantificar los fragmentos de código de barras y los fragmentos de referencia de la mezcla después de ponerlos en contacto con la proteasa. En algunas modalidades, los fragmentos de código de barras y los fragmentos de referencia se identifican en base a sus respectivas masas. En algunas modalidades, los fragmentos de código de barras y los fragmentos de referencia se identifican mediante espectrometría de masas. En algunas modalidades, los fragmentos de código de barras y los fragmentos de referencia se identifican mediante cromatografía líquida-espectrometría de masas (LC-MS). En algunas modalidades, la etapa de determinación de una relación de los fragmentos de código de barras respecto a los fragmentos de referencia comprende el marcaje isobárico. En algunas modalidades, la etapa de determinación de una relación de los fragmentos de código de barras respecto a los fragmentos de referencia comprende enriquecer la mezcla con uno o ambos de un fragmento de referencia marcado con isótopo y un fragmento de código de barras marcado con isótopo. En algunas modalidades, los polipéptidos de longitudes variables comprenden polipéptidos que comprenden al menos un polipéptido recombinante extendido (XTEN), como se describió anteriormente o se describe en cualquier otra parte en la presente descripción. En algunas modalidades, el XTEN se caracteriza porque (i) comprende al menos 150 aminoácidos; (ii) al menos 90 % de los residuos de aminoácidos del XTEN se seleccionan de glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) y prolina (P); y (iii) comprende al menos 4 tipos diferentes de aminoácidos seleccionados

de G, A, S, T, E y P. En algunas modalidades, el fragmento de código de barras, cuando está presente, es una porción del XTEN. En algunas modalidades, la mezcla de polipéptidos de longitudes variables comprende un polipéptido como cualquiera descrito anteriormente en la presente descripción o descrito en cualquier otra parte en la presente descripción. En algunas modalidades, los polipéptidos de longitud variable comprenden un polipéptido de longitud completa y fragmentos truncados de este. En algunas modalidades, los polipéptidos de longitud variable consisten esencialmente en el polipéptido de longitud completa y fragmentos truncados de este. En algunas modalidades, la mezcla de polipéptidos de longitudes variables difieren entre sí debido al truncamiento N-terminal, el truncamiento C-terminal o el truncamiento tanto N- como C-terminal de un polipéptido de longitud completa. En algunas modalidades, el polipéptido de longitud completa es un polipéptido como se describió anteriormente o se describe en cualquier otra parte en la presente descripción. En algunas modalidades, la relación entre la cantidad de fragmentos de código de barras y fragmentos de referencia es mayor que 0,5, 0,6, 0,7, 0,8, 0,9, 0,95, 0,98 o 0,99.

Cuantificación de péptidos basada en marcaje isobárico

En algunas modalidades, puede usarse el marcaje isobárico para determinar una relación de los fragmentos de código de barras respecto a los fragmentos de referencia. Un experto entenderá que el marcaje isobárico es una estrategia de espectrometría de masas usada en la proteómica cuantitativa, en donde los péptidos o proteínas (o porciones de estos) se marcan con varios grupos químicos que son isobáricos (idénticos en masa) pero que varían en términos de distribución de isótopos pesados alrededor de su estructura. Estas etiquetas, comúnmente denominadas etiquetas de masa en tándem, están diseñadas para que la etiqueta de masa se escinda en una región enlazadora específica tras la disociación inducida por colisión (CID) de alta energía durante la espectrometría de masas en tándem, para producir de esta manera iones indicadores de diferentes masas. Un experto entenderá que una de las etiquetas isobáricas más comunes son las etiquetas reactivas con amina.

La capacidad mejorada para detectar y cuantificar productos truncados (por ejemplo, mediante marcaje isobárico) puede generar conocimientos que pueden ayudar a diseñar procesos de fabricación para incluir etapas de purificación para minimizar la presencia de variantes no deseadas en el producto/sustancia farmacológica purificada.

Producción recombinante

La descripción en la presente descripción incluye un ácido nucleico. El ácido nucleico puede comprender un polinucleótido (o una secuencia polinucleotídica) que codifica un polipéptido (de fusión), tal como cualquiera descrito anteriormente o descrito en cualquier otra parte en la presente descripción; o el ácido nucleico puede comprender el complemento inverso de un polinucleótido (o secuencia polinucleotídica) de este tipo.

La descripción en la presente descripción proporciona un vector de expresión que comprende una secuencia polinucleotídica tal como cualquiera descrita en el párrafo anterior y una secuencia reguladora unida operativamente a la secuencia polinucleotídica.

La descripción en la presente descripción incluye una célula huésped que comprende un vector de expresión, tal como se describió en el párrafo anterior. En algunas modalidades, la célula huésped es una procariota. En algunas modalidades, la célula huésped es *E. coli*. En algunas modalidades, la célula huésped es una célula de mamífero.

En otro aspecto, la descripción proporciona métodos para fabricar las composiciones humanas o animales. En una modalidad, el método comprende cultivar una célula huésped que comprende una construcción de ácido nucleico que codifica un polipéptido o una composición que contiene XTEN de cualquiera de las modalidades descritas en la presente descripción en condiciones que promueven la expresión del polipéptido o el polipéptido de fusión BPXTEN, seguido de la recuperación del polipéptido o el polipéptido de fusión BPXTEN mediante el uso de métodos de purificación estándar (por ejemplo, cromatografía en columna, HPLC y similares) en donde la composición se recupera en donde al menos 70 %, o al menos 80 %, o al menos 90 %, o al menos 95 %, o al menos 97 %, o al menos 99 % de los fragmentos de unión del polipéptido expresado o el polipéptido de fusión BPXTEN se pliegan correctamente. En otra modalidad del método de preparación, el polipéptido o el polipéptido de fusión BPXTEN expresado se recupera en el que al menos o al menos 90 %, o al menos 95 %, o al menos 97 %, o al menos 99 % del polipéptido o el polipéptido de fusión BPXTEN se recupera en forma monomérica, soluble.

En otro aspecto, la descripción se refiere a métodos de preparación del polipéptido y el polipéptido de fusión BPXTEN a altos niveles de expresión en fermentación de la proteína funcional mediante el uso de una célula huésped de *E. coli* o mamífero, así como también proporcionar vectores de expresión que codifican las construcciones útiles en los métodos para producir las composiciones de construcciones de polipéptidos con actividad citotóxica a altos niveles de expresión. En una modalidad, el método comprende las etapas de 1) preparar el polinucleótido que codifica los polipéptidos de cualquiera de las modalidades descritas en la presente descripción, 2) clonar el polinucleótido en un vector de expresión, que puede ser un plásmido u otro vector bajo el control de secuencias de transcripción y traducción adecuadas para la expresión de proteínas a alto nivel en un sistema biológico, 3) transformar una célula huésped apropiada con el vector de expresión, y 4) cultivar la célula huésped en medios de nutrientes convencionales en condiciones adecuadas para la expresión de la composición polipeptídica. Cuando sea conveniente, la célula huésped es *E. coli*. Mediante el método, la expresión del polipéptido da como resultado títulos de fermentación de al menos 0,05 g/l, o al menos 0,1 g/l, o al menos 0,2 g/l, o al menos 0,3 g/l, o al menos 0,5 g/l, o al menos 0,6 g/l, o al menos 0,7 g/l, o al menos 0,8 g/l, o al menos 0,9 g/l, o al menos 1 g/l o al menos 2 g/l, o al menos 3 g/l, o al menos 4 g/l, o al menos 5 g/l del producto de expresión de la células huésped y en donde al menos 70 %, o al menos 80 %, o al menos 90 %, o al menos 95 %, o al menos 97 %, o al menos 99 % de la proteína expresada se pliega correctamente. Como se usa en la presente descripción, la expresión “plegado correctamente” significa que el componente de fragmentos de unión a antígeno de la composición tiene la capacidad de unirse específicamente a su ligando objetivo. En otra

modalidad, la descripción proporciona un método para producir un polipéptido o polipéptido de fusión BPXTEN, el método comprende cultivar en una reacción de fermentación una célula huésped que comprende un vector que codifica un polipéptido que comprende el polipéptido o polipéptido de fusión BPXTEN en condiciones efectivas para expresar el producto polipeptídico a una concentración de más de aproximadamente 10 miligramos/gramo de célula huésped en peso en seco (mg/g), o al menos aproximadamente 250 mg/g, o aproximadamente 300 mg/g, o aproximadamente 350 mg/g, o aproximadamente 400 mg/g, o aproximadamente 450 mg/g, o aproximadamente 500 mg/g de dicho polipéptido cuando la reacción de fermentación alcanza una densidad óptica de al menos 130 a una longitud de onda de 600 nm, y en donde los fragmentos de unión a antígeno de la proteína expresada se pliegan correctamente. En otra modalidad, la descripción proporciona un método para producir un polipéptido o polipéptido de fusión BPXTEN, el método comprende cultivar en una reacción de fermentación una célula huésped que comprende un vector que codifica la composición en condiciones efectivas para expresar el producto polipeptídico a una concentración de más de aproximadamente 10 miligramos/gramo de célula huésped en peso en seco (mg/g), o al menos aproximadamente 250 mg/g, o aproximadamente 300 mg/g, o aproximadamente 350 mg/g, o aproximadamente 400 mg/g, o aproximadamente 450 mg/g, o aproximadamente 500 mg/g de dicho polipéptido cuando la reacción de fermentación alcanza una densidad óptica de al menos 130 a una longitud de onda de 600 nm y en donde el producto polipeptídico expresado es soluble.

Composición farmacéutica

Lo descrito en la presente descripción incluye una composición farmacéutica que comprende un polipéptido BPXTEN, tal como cualquiera descrito anteriormente o descritos en cualquier otra parte en la presente descripción, y uno o más excipientes farmacéuticamente aceptables. En algunas modalidades, la composición farmacéutica se formula para la administración intradérmica, subcutánea, intravenosa, intraarterial, intraabdominal, intraperitoneal, intravítrea, intratecal o intramuscular. En algunas modalidades, la composición farmacéutica está en forma líquida. En algunas modalidades, la composición farmacéutica se encuentra en un dispositivo que se implanta en el ojo u otra parte del cuerpo. En algunas modalidades, la composición farmacéutica se encuentra en una jeringa precargada para una inyección única. En algunas modalidades, la composición farmacéutica se formula como un polvo liofilizado que se reconstituye antes de la administración.

En algunas modalidades, la dosis se administra por vía intradérmica, subcutánea, intravenosa, intravítrea (o se inyecta de cualquier otra manera en el ojo), intraarterial, intraabdominal, intraperitoneal, intratecal o intramuscular. En algunas modalidades, la composición farmacéutica se administra mediante el uso de un dispositivo implantado en el ojo u otra parte del cuerpo. En algunas modalidades, el ser humano o animal es un ratón, una rata, un mono o un ser humano.

Las composiciones farmacéuticas pueden administrarse para terapia por cualquier vía adecuada. Además, las composiciones farmacéuticas también pueden contener otros compuestos farmacéuticamente activos o una pluralidad de compuestos de la invención.

En algunas modalidades, la composición farmacéutica puede administrarse a una dosis terapéuticamente efectiva. En algunos casos de lo anterior, la dosis terapéuticamente efectiva da como resultado una ganancia en el tiempo empleado dentro de una ventana terapéutica para la proteína de fusión en comparación con la BP correspondiente de la proteína de fusión no unida a la proteína de fusión y administrada a una dosis comparable a un ser humano o animal.

En otra modalidad, la invención proporciona un método para tratar una enfermedad, trastorno o afección, que comprende administrar la composición farmacéutica descrita anteriormente a un ser humano o animal mediante el uso de múltiples dosis consecutivas de la composición farmacéutica administrada mediante el uso de un régimen de dosificación terapéuticamente efectivo

Los polipéptidos BPXTEN de la presente invención pueden formularse de acuerdo con métodos conocidos para preparar composiciones farmacéuticamente útiles, de manera que el polipéptido se combina mezclado con un vehículo portador farmacéuticamente aceptable, tal como soluciones acuosas o tampones, suspensiones y emulsiones farmacéuticamente aceptables. Las formulaciones terapéuticas se preparan para el almacenamiento mediante la mezcla del ingrediente activo que tiene el grado deseado de pureza con portadores, excipientes o estabilizadores fisiológicamente aceptables opcionales, como se describe en Remington's Pharmaceutical Sciences 16ta edición, Osol, A. Ed. (1980).

Kits farmacéuticos

En otro aspecto, la invención proporciona un kit para facilitar el uso de los polipéptidos BPXTEN. En una modalidad, el kit comprende, en al menos un primer contenedor: (a) una cantidad de una composición de proteína de fusión BPXTEN suficiente para tratar una enfermedad, afección o trastorno tras la administración a un ser humano o animal que lo necesite; y (b) una cantidad de un portador farmacéuticamente aceptable; juntos en una formulación lista para inyección o para reconstitución con agua estéril, tampón o dextrosa; junto con una etiqueta que identifique el fármaco BPXTEN y las condiciones de almacenamiento y manipulación, y una hoja de las indicaciones aprobadas para el fármaco, instrucciones para la reconstitución y/o administración del fármaco BPXTEN para el uso para la prevención y/o tratamiento de una indicación aprobada, dosis adecuada e información de seguridad, e información que identifica el lote y la caducidad del fármaco. En otra modalidad de lo anterior, el kit puede comprender un segundo contenedor que puede portar un diluyente adecuado para la composición de BPXTEN, que proporciona al usuario la concentración adecuada de BPXTEN para el suministro al ser humano o animal.

Método de tratamiento

Lo descrito en la presente descripción incluye el uso de un polipéptido, tal como cualquiera descrito anteriormente o descrito en cualquier otra parte en la presente descripción, en la

preparación de un medicamento para el tratamiento de una enfermedad en un ser humano o animal. En algunas modalidades, la enfermedad particular a tratar depende de la elección de las proteínas biológicamente activas. En algunas modalidades, la enfermedad es cáncer.

Lo descrito en la presente descripción incluye un método para tratar una enfermedad en un ser humano o animal, el método comprende administrar al ser humano o animal que lo necesite una o más dosis terapéuticamente efectivas de la composición farmacéutica, tal como cualquiera descrita anteriormente o descrita en cualquier otra parte en la presente descripción. En algunas modalidades, la enfermedad es cáncer. En algunas modalidades, la composición farmacéutica se administra al ser humano o animal como una o más dosis terapéuticamente efectivas administradas de acuerdo con un régimen de dosificación. En algunas modalidades, el ser humano o animal es un ratón, una rata, un mono o un ser humano.

Los siguientes son ejemplos de composiciones y evaluaciones de composiciones de la descripción. Se entiende que pueden llevarse a la práctica diversas otras modalidades, dada la descripción general proporcionada anteriormente.

EJEMPLOS

Ejemplo 1. Diseño de XTEN con código de barras mediante mutaciones mínimas de XTEN de propósito general

Este ejemplo ilustra un enfoque de diseño ilustrativo del polipéptido XTEN con código de barras al hacer mutación(ones) mínima(s) de la secuencia de aminoácidos de un polipéptido XTEN de propósito general (tal como uno de la Tabla 3b anterior). Los criterios relevantes para hacer mutación(ones) mínima(s) incluyen uno o más de los siguientes: (a) minimizar el cambio de secuencia del polipéptido XTEN correspondiente; (b) minimizar el cambio de composición de aminoácidos en el polipéptido XTEN correspondiente; (c) mantener sustancialmente la carga neta en el polipéptido XTEN correspondiente; (d) mantener sustancialmente la inmunogenicidad baja del polipéptido XTEN correspondiente; (e) mantener sustancialmente las propiedades farmacocinéticas proporcionadas por el polipéptido XTEN.

Por ejemplo, los XTEN con código de barras se construyeron al realizar una o más mutaciones que comprenden la eliminación de un residuo de ácido glutámico, la inserción de un residuo de ácido glutámico, la sustitución de un residuo de ácido glutámico o la sustitución de un residuo de ácido glutámico, o cualquiera de sus combinaciones con los XTEN de propósito general en la Tabla 9.

Tabla 9. Cuatro XTEN de propósito general usados para la modificación genética de polipéptidos XTEN con código de barras

SEQ ID NO.	Nombre de XTEN	Secuencia de aminoácidos
676	AE144	GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPA GSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGS

SEQ ID NO.	Nombre de XTEN	Secuencia de aminoácidos
		ETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGS EPATSGSETPGTSTEPSEGSAP
686	AE288_1	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTS TEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETP GSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTE PSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS AP
688	AE576	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTST EPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPE SGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGS EPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEE GSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTE PSEGSAP
690	AE864	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTST EPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPE SGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGS EPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEE GSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSP AGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAP GTSESATPESGPGTSESATPESGPGTSESATPESGPGSEP ATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEG SAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAP

Ejemplo 2. Análisis de secuencia de polipéptidos XTEN con código de barras y selección de estos para la fusión con un polipéptido biológicamente activo ("BP")

Este ejemplo ilustra el diseño y la selección del polipéptido XTEN con código de barras (y el ensamblaje de más de un XTEN con código de barras en un conjunto) para la fusión a un polipéptido biológicamente activo. En dependencia de la ubicación del(de los) fragmento(s)

de código de barras dentro del(de los) XTEN y la forma en que el(los) polipéptido(s) XTEN se fusiona(n) con una proteína biológicamente activa para formar una construcción que contiene el polipéptido XTEN (por ejemplo, un acoplador de linfocitos T activadas por proteasa XTENilada (XPAT)), el(los) fragmento(s) de código de barras puede(n) indicar truncamiento(s) del polipéptido XTEN.

El análisis de digestión con GluC *in silico* se realizó en dos polipéptidos XTEN ilustrativos (XTEN864 y XTEN288_1) para cuantificar los fragmentos peptídicos liberables tras la digestión completa con GluC del polipéptido XTEN. El análisis *in silico* tiene en cuenta que, con respecto a un polipéptido XTEN que tiene residuos de ácido glutámico consecutivos (por ejemplo, "EE"), GluC puede escindirse después de cualquiera de los residuos de ácido glutámico. Como se muestra en los resultados que se resumen a continuación en la Tabla 10, cada una de una secuencia peptídica de 10 mer "TPGTSTEPSE (SEQ ID NO: 8880)" y una secuencia peptídica de 14 mer "GSAPGSEPATSGSE (SEQ ID NO: 8881)" aparece una y solo una vez en el XTEN864 más largo, mientras que todas las demás secuencias peptídicas aparecen dos o más veces en XTEN864. Y la secuencia peptídica de 14 mer "GSAPGSEPATSGSE (SEQ ID NO: 8881)" también aparece una y solo una vez en el XTEN288_1 más corto.

La exclusividad de un código de barras candidato se evalúa en relación con todos los demás fragmentos peptídicos liberables de la construcción que contiene el polipéptido XTEN. En consecuencia, una secuencia de código de barras en un polipéptido XTEN no puede aparecer en ninguna otra parte de la construcción que contiene el polipéptido XTEN, que incluye cualquier otro polipéptido XTEN contenido en el mismo, cualquier proteína biológicamente activa contenida en el mismo o cualquier conexión entre los componentes vecinos de este. Por ejemplo, la Tabla 11 muestra una tabla de "exclusividad" de péptidos para el conjunto de dos polipéptidos XTEN. Debido a su presencia tanto en XTEN864 como en XTEN288, la secuencia peptídica de 14 mer "GSAPGSEPATSGSE (SEQ ID NO: 8881)" no es exclusiva del conjunto de polipéptidos XTEN que comprende tanto XTEN864 como XTEN288 y, por lo tanto, no puede usarse como código de barras para detectar truncamientos en productos polipeptídicos que contienen los dos polipéptidos XTEN.

La selección de un código de barras (o un conjunto de códigos de barras) puede involucrar además la identificación y determinación de la(s) ubicación(ones) o posición(ones) adecuada(s) del(de los) código(s) de barras candidato(s) dentro del polipéptido XTEN. La ubicación o posición de un código de barras candidato puede asociarse con información farmacológicamente relevante del polipéptido XTEN (y la construcción que contiene el polipéptido XTEN como un todo), tal como el truncamiento del polipéptido XTEN más allá de una longitud crítica y/o eliminación(ones) en el polipéptido XTEN. El péptido de 10 mer "TPGTSTEPSE (SEQ ID NO: 8880)" podría servir como un fragmento de código de barras adecuado si XTEN864 se coloca en el extremo N-terminal del producto que contiene el polipéptido XTEN y si el truncamiento de 238 aminoácidos del extremo N-terminal del producto no afecta significativamente las propiedades farmacológicas del producto.

Tabla 10. Secuencias representativas de XTEN para el análisis de digestión con GluC

XTEN ilustrativo	Secuencia de aminoácidos	SEQ ID NO:
XTEN864	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGS PAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGS PTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS APGTSESATPESGPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSET TPGTSTEPSEGS APGTSTEPSEGSAPG TSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTS TEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEG SAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTS ESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAG SPTSTEEGTSTEPSEGSAPGTSTEPSE GSAPGSEPATSG GSETPGTSESATPESGPGTSTEPSEGSAP	8882
AE288_1	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGS EPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPESGPGTSESATPES GPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSE GSAPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAP	8883

Tabla 11. Análisis de "exclusividad" de péptidos

Fragmentos peptídicos	SEQ ID NO:	XTEN AE288_1	XTEN864	Ambos
SATPE	8884	9	23	32
SGPGSEPATSGSE	8885	4	9	13
GSAPGTSTEPSE	8886	1	10	11
TPGTSE	8887	4	7	11
SGPGTSTEPSE	8888	2	8	10
GSAPGTSE	8889	1	8	9
GTSTEPSE	8890	2	5	7
GSAPGSPAGSPTSTEE	8891	1	4	5
GSPAGSPTSTEE	186	1	4	5
SGPGTSE	8892	2	3	5
SGPGSPAGSPTSTEE	8893	1	3	4
TPGSEPATSGSE	8894	1	2	3
TPGSPAGSPTSTEE	8895	1	2	3
GSAPGSEPATSGSE	8881	1	1	2
TPGTSTEPSE	8880		1	1

Todas las secuencias subrayadas producen péptidos GluC únicos

Núcleo no XTEN subrayado y en cursiva

Los péptidos de código de barras están en negrita

Las secuencias de péptidos de código de barras ilustrativas se ilustran a continuación en la Tabla 12. Estas secuencias de códigos de barras deben estar flanqueadas de acuerdo con la fórmula estructural (I):

AAA–Glu–Péptido de código de barras–BBB,

en donde "AAA" representa Gly, Ala, Ser, Thr o Pro y "BBB" representa Gly, Ala, Ser o Thr configurados para facilitar la liberación eficiente del péptido de código de barras por digestión con GluC. En particular, la inserción de cada péptido de código de barras en el XTEN puede dar como resultado secuencias únicas adicionales que preceden o siguen directamente a los péptidos de código de barras insertados.

Tabla 12. Lista de péptidos de código de barras adecuados

Péptido(s) de código de barras candidato(s)	SEQ ID NO:	Aparición en XTEN864	Aparición en XTEN288_1	Aparición en XTEN1152
SPATSGSTPE	8020	0	0	
GSAPATSE	8021	0	0	0
GSAPGTATE	8022	0	0	0
GSAPGTE	8023	0	0	0
PATSGPTE	8024	0	0	0
SASPE	8025	0	0	0
PATSGSTE	8026	0	0	0
GSAPGTSAE	8027			
SATSGSE	8028	0	0	0
SGPGSTPAE	8029	0	0	
SGSE	8030	0		

Ejemplo 3: Diseño y selección de XTEN(s) en construcciones de polipéptidos xtenilados de secuencia completa

Este ejemplo ilustra el diseño de una construcción de polipéptido de secuencia completa que contiene dos polipéptidos XTEN, uno en el extremo N-terminal y el otro en el extremo C-terminal.

La Tabla 13 a continuación ilustra los polipéptidos XTEN usados en un BPXTEN con código de barras representativo (que contiene polipéptidos XTEN con código de barras en los extremos N- y C-terminal) y un BPXTEN de referencia (que contiene XTEN de propósito general en los extremos N- y C-terminal). En el BPXTEN con código de barras representativo, un polipéptido XTEN con código de barras (SEQ ID NO. 8014) se fusiona en el extremo N-terminal de la BP, y otro polipéptido XTEN con código de barras (SEQ ID NO. 8015) se fusiona en el extremo C-terminal de la BP. En el BPXTEN de referencia, un polipéptido XTEN "Ref-N" (SEQ ID NO. 8896) se fusiona en el extremo N-terminal de la BP y un polipéptido XTEN "Ref-C" (SEQ ID NO. 8897) se fusiona en el extremo C-terminal de la BP. El polipéptido "Ref-N" XTEN (SEQ ID NO. 8896) es comparable en longitud con el polipéptido XTEN con código de barras SEQ ID NO. 8014; y el polipéptido

XTEN "Ref-C" (SEQ ID NO. 8897) es comparable en longitud con el polipéptido XTEN con código de barras SEQ ID NO. 8015. Los BPXTEN con código de barras y de referencia contienen cada uno una secuencia de referencia en el componente BP. La secuencia de referencia es única y difiere en peso molecular de todos los demás fragmentos peptídicos que pueden liberarse del BPXTEN correspondiente tras la digestión completa por la proteasa GluC (por ejemplo, de acuerdo con el Ejemplo 5). La exclusividad de la secuencia de referencia se evalúa en relación con todos los demás fragmentos peptídicos que pueden liberarse de la construcción de BPXTEN.

Tabla 13. Conjuntos representativos de XTEN N- y C-terminales usados en construcciones de BPXTEN de longitud completa

SEQ ID NO.	Tipo de XTEN	Secuencia de aminoácidos	# Total de AA
8014 (de la Tabla 3a)	XTEN N-terminal	<u>SPAGSPTSTES</u> <u>GTSE</u> SATPESGPGTSTEPSEG SAPGTSESATPESGPGSEPATSGSETPGTSES ATPES <u>SGPGSTPAESGSE</u> TPGTSESATPESGP GTSTEPSEGSAPGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAP GTSESATPESGPGTSESATPESGPGTSESATP ESGPGSEPATSGSETPGSEPATSGSETPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSE <u>GSAP</u> <u>GGSAP</u>	292
8015 (de la Tabla 3a)	XTEN C-terminal	<u>PGSPAGSPTSTEE</u> GTSESATPESGPGSEPATSG GSETPGTSESATPESGPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP PGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSPAGSPTSTEEGSPAGSPTSTEEGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAP PGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGSPAGSPTSTEEGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTS ESATPESGPGTSESATPESGPGTSESATPESG PGSEPATSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSTEPSE <u>GSAPGTESTP</u> SEGSAPGS <u>EPATSGSETPGTSESATPESGPGTSTEPSEGS</u> <u>APGEPEA</u>	582
8896 Ref-N	XTEN N-terminal	<u>SPAGSPTSTEE</u> GTSESATPESGPGTSTEPSEG SAPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAPGSPAGSPTSTEEGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAP GTSESATPESGPGTSESATPESGPGTSESATP ESGPGSEPATSGSETPGSEPATSGSETPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSE <u>GSAP</u> <u>GGSAP</u>	292
8897 Ref-C	XTEN C-terminal	<u>PGSPAGSPTSTEE</u> GTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTS	584

SEQ ID NO.	Tipo de XTEN	Secuencia de aminoácidos	# Total de AA
		TEPSEGSAPGTSESATPESGPGSEPATSGSET PGSEPATSGSETPGSPAGSPTSTEEGTSESAT PESGPGTSTEPSEGSAPGTSTEPSEGSAPGSP AGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS PGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSET <u>TPGTSTEPSEGS</u> APGTS TEPSEGSAPGTSESATPESGPGTSESATPESG PGSPAGSPTSTEEGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS PGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSPAGSPTSTEEGSPAGSPTSTEEGSP AGSPTSTEEGTSESATPESGPGTSTEPSE <u>GSA</u> <u>PGAAEPEA</u>	

Ejemplo 4: Construcción recombinante y producción de polipéptidos de fusión XTENilados con código de barras

Los Ejemplos 4a-4b ilustran la construcción, producción y purificación recombinantes de polipéptidos de longitud completa que contienen polipéptido(s) XTEN con código de barras mediante el uso de los métodos descritos en la presente descripción.

Ejemplo 4a. Un polipéptido de fusión XTENilado que contiene un XTEN con código de barras en el extremo C-terminal

EXPRESIÓN: Las construcciones que codifican un polipéptido de fusión XTENilado que contiene un fragmento variable de cadena sencilla (scFv) anti-EpCAM y una secuencia XTEN con código de barras de 864 aminoácidos de longitud (SEQ ID NO: 8008) en el extremo C-terminal se expresan en una cepa de *E. coli* AmE098 patentada y se distribuyen en el periplasma mediante una secuencia líder secretora N-terminal (MKKNIAFLLASMFVFSIATNAYA-)(SEQ ID NO: 8898), que se escinde durante la translocación. Los cultivos de fermentación se cultivan con un medio complejo libre de componentes animal a 37 °C; y la temperatura se cambia a 26 °C antes del agotamiento del fosfato. Durante la cosecha, el caldo de fermentación completo se centrifuga para sedimentar las células. En la cosecha, se registra el volumen total y el peso celular en húmedo (WCW; relación de sedimento respecto al sobrenadante), y las células sedimentadas se recolectaron y se congelaron a -80 °C.

RECOBRADO: El sedimento celular congelado se resuspende en tampón de lisis (ácido cítrico 17,7 mM, HPO_4 22,3 mM, NaCl 75 mM, EDTA 2 mM, pH 4,0) dirigido al 30 % del peso celular en húmedo. Se permite que la resuspensión se equilibre a pH 4 y después se homogeneiza en dos pases a 800±50 bar mientras se monitorea y mantiene la temperatura de salida a 15±5 °C. Se confirma que el pH del homogeneizado está dentro del intervalo especificado (pH 4,0±0,2).

CLARIFICACIÓN: Para reducir las endotoxinas y las impurezas de las células huésped, se permite que el homogeneizado se someta a floculación ácida (pH $4,0\pm 0,2$) a baja temperatura (10 ± 5 °C) durante la noche (15-20 horas). Para eliminar la fracción insoluble, el homogeneizado floculado se centrifuga durante 40 minutos a 16,900 RCF a 2-8°C y se retiene el sobrenadante. El sobrenadante se diluye aproximadamente 3 veces con agua Milli-Q (MQ), después se ajusta a 7 ± 1 mS/cm con NaCl 5 M. Para eliminar el ácido nucleico, los lípidos y la endotoxina y para actuar como auxiliar de filtración, el sobrenadante se ajusta a 0,1 % (m/m) de tierra de diatomeas. Para mantener suspendido el auxiliar de filtración, el sobrenadante se mezcla mediante un impulsor y se deja equilibrar durante 30 minutos. Se ensambla un tren de filtros, que consiste en un filtro de profundidad seguido de un filtro de $0,22\ \mu\text{m}$, y después se enjuaga con MQ. El sobrenadante se bombea a través del tren de filtros mientras se modula el flujo para mantener una disminución de presión de 25 ± 5 psig. Para ajustar el sistema de tampón compuesto (en base a la relación de ácido cítrico y Na_2HPO_4) al intervalo deseado para la cromatografía de captura, el filtrado se ajusta con Na_2HPO_4 500 mM en donde la relación final de Na_2HPO_4 respecto al ácido cítrico es 9,33:1, y se confirma que el pH del filtrado tamponado está dentro del intervalo especificado (pH $7,0\pm 0,2$).

Purificación

Captura con AEX: Para separar dímeros, agregados y truncamientos grandes del producto monomérico, y para eliminar endotoxinas y ácidos nucleicos, se utiliza la cromatografía de intercambio aniónico (AEX) para capturar el dominio de XTEN electronegativo C-terminal. La fase estacionaria de AEX1 (GE Q Sepharose FF), la fase móvil de AEX1 A (Na_2HPO_4 12,2 mM, NaH_2PO_4 7,8 mM, NaCl 40 mM) y fase móvil B de AEX1 (Na_2HPO_4 12,2 mM, NaH_2PO_4 7,8 mM, NaCl 500 mM) se usan en la presente descripción. La columna se equilibra con la fase móvil A de AEX1. En base a la concentración de proteína total medida mediante el ensayo de ácido bicinconínico (BCA), el filtrado se carga en la columna con un objetivo de 28 ± 4 g/l de resina, seguido de la fase móvil A de AEX1 y después se lava con una etapa a 30 % de B. El material unido se eluye con un gradiente de 30 % de B a 60 % de B en 20 CV. Las fracciones se recolectan en alícuotas de 1 CV mientras $A_{220} \geq 100$ mAU por encima de la línea base (local). Las fracciones de elución se analizan y agrupan en base a SDS-PAGE y SE-HPLC.

Purificación intermedia por IMAC: Para garantizar la integridad del extremo C-terminal, se usa cromatografía de afinidad con metal inmovilizado (IMAC) para capturar la etiqueta de polihistidina del extremo C-terminal (His(6) (SEQ ID NO: 8031)). La fase estacionaria de IMAC (GE IMAC Sepharose FF), la fase móvil A de IMAC (Na_2HPO_4 18,3 mM, NaH_2PO_4 1,7 mM, NaCl 500 mM, imidazol 1 mM) y la fase móvil B de IMAC (Na_2HPO_4 18,3 mM, NaH_2PO_4 1,7 mM, NaCl 500 mM, imidazol 500 mM) se usan en la presente descripción. La columna se carga con una solución de zinc y se equilibra con la fase móvil A de IMAC. El grupo de AEX1 se ajusta a un pH de $7,8\pm 0,1$, 50 ± 5 mS/cm (con NaCl 5 M) e imidazol 1 mM, se carga en la columna de IMAC con un objetivo de 2 g/l de resina, y se sigue con la

fase móvil A de IMAC hasta que la absorbancia a 280 nm (A280) vuelve a la línea base (local). El material unido se eluye con una etapa 25 % de fase móvil B de IMAC. La recolección de la elución de IMAC se inicia cuando $A_{280} \geq 10$ mAU está por encima de la línea base (local), se dirige a un contenedor previamente enriquecido con EDTA suficiente para llevar 2 CV a EDTA 2 mM y se termina una vez que se recogen 2 CV. La elución se analiza por SDS-PAGE.

Purificación intermedia con proteína-L: Para garantizar la integridad del extremo N-terminal, se usa proteína-L para capturar los dominios kappa presentes cerca del extremo N-terminal de la molécula BPXTEN (específicamente el scFv contra aEpCAM). La fase estacionaria de proteína L (GE Capto L), la fase móvil A de proteína-L (ácido cítrico 16,0 mM, Na_2HPO_4 20,0 mM, pH $4,0 \pm 0,1$), la fase móvil B de proteína-L (ácido cítrico 29,0 mM, Na_2HPO_4 7,0 mM, pH $2,60 \pm 0,02$) y la fase móvil C de proteína-L (ácido cítrico 3,5 mM, Na_2HPO_4 32,5 mM, NaCl 250 mM, pH $7,0 \pm 0,1$) se usan en la presente descripción. La columna se equilibra con la fase móvil C de proteína-L. La elución de IMAC se ajusta a un pH de $7,0 \pm 0,1$ y 30 ± 3 mS/cm (con NaCl 5 M y MQ) y se carga en la columna de proteína-L con un objetivo de 2 g/l. Después, la resina se siguió con la fase móvil C de proteína-L hasta que la absorbancia a 280 nm (A280) volvió a la línea base (local). La columna se lava con la fase móvil A de proteína-L y las fases móviles A y B de proteína-L se usan para efectuar la elución a pH bajo. El material unido se eluye a un pH de aproximadamente 3,0 y se recolecta en un contenedor previamente enriquecido con una parte de Na_2HPO_4 0,5 M por cada 10 partes de volumen recolectado. Las fracciones se analizan por SDS-PAGE.

Pulido con HIC: Para separar las variantes N-terminales (4 residuos en el extremo N-terminal absoluto no son esenciales para la unión a proteína-L) y las variantes de conformación generales, se usa la cromatografía de interacciones hidrófobas (HIC). Fase estacionaria de HIC (GE Capto Phenyl ImpRes), la fase móvil A de HIC (histidina 20 mM, polisorbato 80 al 0,02 % (p/v), pH $6,5 \pm 0,1$) y fase móvil B de HIC (sulfato de amonio 1 M, histidina 20 mM, polisorbato 80 al 0,02 % (p/v), pH $6,5 \pm 0,1$) se usan en la presente descripción. La columna se equilibra con la fase móvil B de HIC. La elución de proteína-L ajustada se carga en la columna de HIC con un objetivo de 2 g/l de resina y se sigue con la fase móvil B de HIC hasta que la absorbancia a 280 nm (A280) vuelve a la línea base (local). La columna se lava con 50 % de B. El material unido se eluye con un gradiente de 50 % de B a 0 % de B en 75 CV. Las fracciones se recolectan en alícuotas de 1 CV mientras $A_{280} \geq 3$ mAU por encima del valor inicial (local). Las fracciones de elución se analizan y agrupan en base a SE-HPLC y HI-HPLC.

Formulación: Para intercambiar el producto en el tampón de formulación y llevar el producto a la concentración objetivo (0,5 g/L), se usa nuevamente el intercambio aniónico para capturar el XTEN C-terminal. La fase estacionaria de AEX2 (GE Q Sepharose FF), la fase móvil A de AEX2 (histidina 20 mM, NaCl 40 mM, polisorbato 80 al 0,02 % (p/v), pH $6,5 \pm 0,2$), la fase móvil B de AEX2 (histidina 20 mM, NaCl 1 M, polisorbato 80 al 0,02 %

(p/v), pH $6,5 \pm 0,2$) y la fase móvil C de AEX2 (Na_2HPO_4 12,2 mM, NaH_2PO_4 7,8 mM, NaCl 40 mM, polisorbato 80 al 0,02 % (p/v), pH $7,0 \pm 0,2$) se usan en la presente descripción. La columna se equilibra con la fase móvil C de AEX2. El grupo de HIC se ajusta a un pH de $7,0 \pm 0,1$ y 7 ± 1 mS/cm (con MQ) y se carga en la columna de AEX2 con un objetivo de 2 g/l de resina y después se sigue con la fase móvil C de AEX2 hasta que A280 vuelve a la línea base (local). La columna se lava con la fase móvil A de AEX2 (histidina 20 mM, NaCl 40 mM, polisorbato 80 al 0,02 % (p/v), pH $6,5 \pm 0,2$). Las fases móviles A y B de AEX2 se usan para generar una etapa de {NaCl} y efectuar la elución. El material unido se eluye con una etapa a 38 % de fase móvil B de AEX2. La recolección de la elución de AEX2 se inicia cuando $A_{280} \geq 5$ mAU por encima de la línea base (local) y finaliza una vez que se recolectan 2 volúmenes de columna. La elución de AEX2 se filtra a $0,22 \mu\text{m}$ dentro de una BSC, se divide en alícuotas, se etiqueta y se almacena a -80°C como sustancia farmacológica a granel (BDS). La sustancia farmacológica a granel (BDS) se confirma mediante diversos métodos analíticos para cumplir con todos los criterios de liberación de lotes. La calidad general se analiza por SDS-PAGE, la relación de monómero a dímero y agregado se analiza mediante SE-HPLC, y la calidad del N-terminal y la homogeneidad del producto se analizan mediante HI-HPLC.

Ejemplo 4b. Un polipéptido de fusión XTENilado que contiene un XTEN con código de barras en el extremo C-terminal y otro XTEN con código de barras en el extremo N-terminal

EXPRESIÓN: Una construcción que codifica un polipéptido de fusión XTENilado que contiene un fragmento variable de cadena sencilla (scFv) anti-EGFR, un XTEN con código de barras de 864 aminoácidos de longitud (SEQ ID NO: 8008) en el extremo C-terminal, y un XTEN con código de barras de 288 aminoácidos de longitud (SEQ ID NO: 8007) en el extremo N-terminal se expresa en una cepa de *E. coli* AmE098 patentada y se distribuye en el periplasma mediante una secuencia líder secretora N-terminal (MKKNIAFLLASMFVFSIATNAYA-)(SEQ ID NO: 8898), que se escinde durante la translocación. Los cultivos de fermentación se cultivaron con medio de complejo libre de componentes animales a 37°C y la temperatura se cambió a 26°C antes del agotamiento del fosfato. Durante la cosecha, el caldo de fermentación completo se centrifuga para sedimentar las células. En la cosecha, se registraron el volumen total y el peso celular en húmedo (WCW; relación de sedimento respecto al sobrenadante), y las células sedimentadas se recolectaron y se congelaron a -80°C .

RECOBRADO: El sedimento de células congeladas se resuspende en tampón de lisis (ácido cítrico 100 mM) con un peso celular en húmedo objetivo del 30 %. Se permite que la resuspensión se equilibre a pH 4,4 y después se homogeneiza a $17\,000 \pm 200$ bar mientras se monitorea y mantiene la temperatura de salida a $15 \pm 5^\circ\text{C}$. Se confirma que el pH del homogeneizado está dentro del intervalo especificado (pH $4,4 \pm 0,1$).

CLARIFICACIÓN: Para reducir las endotoxinas y las impurezas de la célula huésped, se permite que el homogeneizado se someta a floculación ácida (pH $4,4 \pm 0,1$) a baja

temperatura (10 ± 5 °C) durante la noche (15-20 horas). Para eliminar la fracción insoluble, el homogeneizado floculado se centrifuga durante 40 minutos a 8 000 RCF y $2-8$ °C, y se retiene el sobrenadante. Para eliminar el ácido nucleico, los lípidos y la endotoxina y para actuar como auxiliar de filtración, el sobrenadante se ajusta a 0,1 % (m/m) de tierra de diatomeas. Para mantener suspendido el auxiliar de filtración, el sobrenadante se mezcla mediante un impulsor y se deja equilibrar durante 30 minutos. Se ensambla un tren de filtros, que consiste en un filtro de profundidad seguido de un filtro de $0,22 \mu\text{m}$, y después se enjuaga con MQ. El sobrenadante se bombea a través del tren de filtros mientras se modula el flujo para mantener una disminución de presión de 25 ± 5 psig.

Purificación

Captura con proteína-L: Para eliminar las proteínas, la endotoxina y el ácido nucleico de la célula huésped, se usa proteína-L para capturar el dominio kappa presente dentro del scFv de aEGFR de la molécula BPXTEN. La fase estacionaria de proteína-L (Tosoh TP AF-rProtein L-650F), la fase móvil A de proteína-L (ácido cítrico 11,5 mM, Na_2HPO_4 24,5 mM, NaCl 125 mM, polisorbato 80 al 0,005 %, pH 5,0) y la fase móvil B de proteína-L (ácido fosfórico 11 mM, polisorbato 80 al 0,005 %, pH 2,0) se usan en la presente descripción. La columna se equilibra con la fase móvil A de proteína-L. El filtrado se ajusta a $\text{pH } 5,5 \pm 0,2$ y se carga en la columna de proteína-L con un objetivo de 2-4 g/l de resina, después se sigue con la fase móvil A de proteína-L hasta que la absorbancia a 280 nm (A_{280}) vuelve a la línea base (local). El material unido se eluye con la fase móvil B y se recolecta como una fracción de 2 CV previamente enriquecida con 0,4 CV de Na_2HPO_4 0,5 M y se analiza por SDS-PAGE.

Purificación intermedia por IMAC: Para garantizar la integridad del N-terminal, se usa la cromatografía de afinidad por metal inmovilizado (IMAC) para capturar la etiqueta de polihistidina N-terminal (His(6); (SEQ ID NO: 8031)) de la molécula polipeptídica de fusión. La fase estacionaria de IMAC (GE IMAC Sepharose FF), la fase móvil A de IMAC (Na_2HPO_4 12,2 mM, NaH_2PO_4 7,8 mM, NaCl 500 mM, polisorbato 80 al 0,005 %, pH 7,0) y la fase móvil B de IMAC (histidina 50 mM, NaCl 200 mM, polisorbato 80 al 0,005 %, pH 6,5) se usan en la presente descripción. La columna se equilibra con la fase móvil A de IMAC. La elución de proteína-L se ajusta a un pH de $7,8 \pm 0,1$ y 50 ± 5 mS/cm (con NaCl 5 M). El grupo de proteína-L ajustado se carga en la columna IMAC con un objetivo de 2 g/l de resina y se sigue con la fase móvil A de IMAC hasta que la absorbancia a 280 nm (A_{280}) vuelve a la línea base (local). El material unido se eluye con la fase móvil B de IMAC. La elución de IMAC se recolecta como una fracción de 2 CV previamente enriquecida con 0,02 CV de EDTA 200 mM y se analiza por SDS-PAGE.

Purificación intermedia con etiqueta C: Para garantizar la integridad del extremo C-terminal, se usa cromatografía de afinidad a etiqueta C para capturar la etiqueta EPEA C-terminal (SEQ ID NO: 8033). La fase estacionaria de etiqueta C (Thermo C-tagXL), la fase móvil A de etiqueta C (histidina 50 mM, NaCl 200 mM, polisorbato 80 al 0,005 %, pH 6,5) y la fase móvil B de etiqueta C (Tris 20 mM, MgCl_2 0,6 M, polisorbato 80 al 0,005 %, pH

7,0) se usan en la presente descripción. La columna se equilibra con la fase móvil A de etiqueta C. La elución de IMAC se carga en la columna de etiqueta C con un objetivo de 2 g/l de resina y se sigue con la fase móvil A de etiqueta C hasta que la absorbancia a 280 nm (A280) vuelve a la línea base (local). El material unido se eluye con una fase móvil B de etiqueta C. La elución de etiqueta C se recolecta como una fracción de 2 CV y se analiza por SDS-PAGE.

Pulido por AEX: Para separar el dímero y el agregado del producto monomérico, se utiliza la cromatografía de intercambio aniónico (AEX) para capturar los dominios de XTEN N- y C-terminales electronegativos. La fase estacionaria de AEX1 (BIA QA-80), la fase móvil A de AEX1 (histidina 50 mM, NaCl 200 mM, polisorbato 80 al 0,005 %, pH 6,5) y la fase móvil B de AEX1 (histidina 50 mM, NaCl 500 mM, polisorbato al 0,005 % 80, pH 6,5) se usan en la presente descripción. La columna se equilibra con la fase móvil A de AEX. La elución de la etiqueta C se diluye a 10 mS/cm con MQ, se carga con un objetivo de 2 g/l de resina y después se sigue con la fase móvil A de AEX hasta que la absorbancia a 280 nm vuelve a la línea base (local). El material unido se eluye con un gradiente de 0 % de B a 100 % de B en 60 CV. Las fracciones se recolectan en alícuotas de 1 CV mientras $A_{280} \geq 2$ mAU por encima de la línea base (local). Las fracciones de elución se analizan por SDS-PAGE y SE-HPLC, y las fracciones que son ≥ 98 % de monómero se agrupan (grupo de AEX) para su posterior procesamiento.

FORMULACIÓN: Para intercambiar el producto en el tampón de formulación y llevar el producto a la concentración objetivo (0,5 g/l), se usa ultrafiltración/diafiltración (UF/DF). Mediante el uso de una membrana de 10 kDa con un área de 0,1 m² y un objetivo de TMP de 15 psi, el grupo de AEX se concentra a 0,5 g/l, después se diluye 10 veces con tampón de formulación (histidina 50 mM, NaCl 200 mM, polisorbato 80 al 0,005 %, pH 6,5). El grupo de AEX se concentra 10 veces y se diluye 10 veces dos veces más. El producto formulado recuperado se filtra a 0,22 µm dentro de un BSC, se divide en alícuotas, se etiqueta y se almacena a -80 °C como sustancia farmacológica a granel (BDS). El BDS se confirma mediante diversos métodos analíticos para cumplir con todos los criterios de liberación de lotes. La calidad general se analiza por SDS-PAGE, la relación de monómero a dímero y agregado se analiza por SE-HPLC, y la calidad del N-terminal y la homogeneidad del producto se analizan por HI-HPLC. La identidad se confirma por ESI-MS.

Ejemplo 5. Liberación de péptidos de código de barras mediante digesto de proteasas

Este ejemplo ilustra la liberación de fragmento(s) de código de barras y fragmento(s) de referencia de una mezcla de polipéptidos que contienen longitudes variables o formas truncadas de la construcción que contiene XTEN mediante el uso de los métodos descritos en la presente descripción.

Una muestra de construcción que contiene XTEN se reduce y alquila mediante incubación en DTT y después en yodoacetamida, secuencialmente. Después, las muestras se intercambian con tampón y se desalinizan mediante el uso de un cartucho de centrífuga de

exclusión por tamaño. Se añade proteasa Glu-C a las muestras en una relación de enzima a sustrato de 1:5 y las muestras se incuban a 37 °C para la digestión. Después, las muestras se mueven a 4 °C para detener la reacción proteolítica y se colocan en viales de automuestreador para su análisis.

Ejemplo 6. Detección y cuantificación de péptido(s) de código de barras y péptido(s) de referencia

Este ejemplo ilustra métodos de espectrometría de masas usados para generar mediciones cuantitativas de péptidos de código de barras individuales. Se programa un método de LC-Monitoreo de reacción en paralelo (PRM) en un espectrómetro de masas de masa exacta de alta resolución (HRAM). A diferencia de los métodos tradicionales de espectrometría de masas de adquisición dependiente de datos (DDA), los métodos PRM se centran en un conjunto específico de 15-30 péptidos en una sola ejecución, con secuenciación de cada uno mediante MS-MS una vez por ciclo de trabajo. Como tal, este método genera cromatogramas de iones extraídos (XIC) para los iones precursores no fragmentados del péptido intacto, así como también para cada ion de fragmento del péptido para confirmar su secuencia. Los XIC de iones de fragmentos a menudo son más sensibles y selectivamente cuantitativos que los fragmentos de iones precursores. El método LC-PRM usado incluye las versiones ligera y pesada de siete péptidos de código de barras. Las áreas de los picos cromatográficos de todos los iones de fragmentos de estos 14 péptidos se miden después de la adquisición y el ion de fragmento más fuerte se usa para la medición cuantitativa. Las relaciones de áreas de los picos de los péptidos de código de barras XTEN respecto a los péptidos de código de barras PAT se calculan después para determinar la abundancia relativa de XTEN:PAT en diversos puntos a lo largo de las moléculas de XTEN.

Ejemplo 7. Marcaje de isótopos estables para cuantificar los péptidos por espectrometría de masas (MS)

Este ejemplo ilustra el esquema de marcaje de isótopos estables para permitir la cuantificación absoluta (en lugar de relativa) de péptidos de código de barras de polipéptidos que contienen XTEN. Se emplea un esquema cuantitativo estándar de aminoácidos marcados pesados en donde los análogos sintéticos de los péptidos de código de barras en los que el ácido glutámico C-terminal se reemplaza con el análogos marcado pesado ($^{13}\text{C}_5\text{H}_7(^{15}\text{N})\text{O}_3$) se adquieren de un proveedor especializado. Se prepara una curva de calibración en la que una cantidad conocida de polipéptido que contiene código de barras XTEN se diluye en serie en una matriz en la que el péptido sintético marcado pesado se mantiene a una concentración constante. Se puede realizar una cuantificación exacta mediante la calibración de las relaciones pesadas:ligeras del área de los picos cromatográficos de la curva frente a muestras de investigación que contienen el mismo nivel de picos de péptido marcado pesado.

Ejemplo 8. Cuantificación del truncamiento del polipéptido que contiene XTEN

Este ejemplo ilustra la cuantificación de variantes de longitud o variantes de truncamiento en una mezcla de polipéptidos que contienen XTEN.

Por ejemplo, un péptido de código de barras "SGPGSTPAESGSE" (SEQ ID NO: 8899) se ubica 76 aminoácidos en la secuencia representativa de BPXTEN con código de barras descrita y obtenida del Ejemplo 3 para indicar un truncamiento intenso de XTEN en el extremo N-terminal de BPXTEN. Se considera también un posible fragmento de código de barras "SPAGSPTSTESGTSE" (SEQ ID NO: 8260) que se ubica en el extremo N-terminal. La relación de medición de abundancia de cada péptido de código de barras con relación a una secuencia peptídica de referencia única de la proteína biológicamente activa (por ejemplo, un fragmento scFv) según el procedimiento del Ejemplo 6 indica la cantidad total de polipéptidos de longitud completa y las variantes que tienen truncamientos que podría afectar la eficacia farmacológica en la mezcla de muestra. La medición de la abundancia de al menos un fragmento de referencia se usa para indicar la cantidad total de todas las variantes del polipéptido en la mezcla de muestra. En consecuencia, la abundancia diferencial entre el fragmento de referencia y los fragmentos de código de barras informa la cantidad de variantes de polipéptido truncado. Los datos de LC-MS se analizan para determinar la relación entre la cantidad del fragmento de código de barras y el fragmento de referencia, que indica la cantidad relativa de variantes farmacológicamente eficaces en la mezcla de polipéptidos.

Se usa un conjunto de dos (o tres) códigos de barras para indicar diferentes niveles de truncamiento del polipéptido. Los datos de LC-MS se usan para determinar la relación entre la cantidad de cada fragmento de código de barras y la cantidad del fragmento de referencia, de esta manera se cuantifica la distribución de variantes de truncamiento en la mezcla de polipéptidos.

Ejemplo 9. Cuantificación del truncamiento del polipéptido que contiene XTEN

Una etapa en la purificación de AC2329 es la cromatografía de intercambio aniónico QIR. La segunda mitad del pico de elución principal de esta columna contiene la proteína completa, mientras que la primera mitad del pico contiene una mezcla de proteína completa así como también muchas formas truncadas. Para los propósitos de evaluación del péptido de código de barras, se tomaron dos fracciones. Una fracción incluía solo la segunda mitad del pico, hasta ahora denominada fracción de "longitud completa". La segunda fracción incluía la primera mitad del pico, hasta ahora denominada fracción de proteína sintética y truncamientos, que se expone en la presente descripción:

Cromatografía de exclusión por tamaño (SEC) analítica que superpone la fracción de longitud completa (azul) y la fracción de longitud completa + truncamientos (negro). Es evidente que la fracción de proteína sintética + truncamientos incluye fragmentos tan grandes como la proteína sintética intacta, pero también muchos componentes más pequeños como lo indica el gran hombro inclinado en el lado derecho del pico.

La proteína sintética de longitud completa y la proteína sintética de longitud completa y los truncamientos se mezclaron en las relaciones descritas en la matriz de dilución.

Matriz de dilución

Dilución	1	2	3	4	5	6	7	8	9
Compuesto 1 (nM)	400	350	300	250	200	150	100	50	0
Compuesto 1 trunc (nM)	0	50	100	150	200	250	300	350	400
nM totales	400	400	400	400	400	400	400	400	400
Vol total	60	60	60	60	60	60	60	60	60

Después, cada una de las muestras de 400 nM se digirió con 1 mg/ml de GluC (Roche 10791156001) en tampón de reacción que contenía Tris-HCl 50 mM, pH 7,5 + 0,1 % de Rapigest (Waters 186001861). La digestión con GluC se realizó durante la noche a 37 °C en una incubadora con agitación. Después de la digestión, se añadió ácido fórmico hasta una concentración final de 10 % y se incubó a 37 °C durante 45 min. Los precipitados de Rapigest se eliminaron mediante centrifugación a 16 000 x g durante 10 min y se añadió una mezcla estándar de péptidos pesados hasta una concentración final de 400 nM. El análisis LC-MS de cada digestión se realizó como se describió. Cada muestra se midió por duplicado.

Análisis por LC-MS

Las digestiones se analizaron mediante el uso de un método de gradiente de 30 minutos en una columna Waters HSS-T3 (176003994) mediante el uso de un sistema UHPLC Vanquish (Thermo) conectado a un espectrómetro de masas Q-Exactive Plus (Thermo). El método MS consistió en un método DDA de los diez principales en el que los diez péptidos principales se secuenciaron mediante análisis MSMS después de cada exploración MS. Los archivos de datos resultantes se procesaron mediante el uso del software Skyline (MacCoss Lab, UW) en el que se podían calcular y medir las concentraciones normalizadas de péptidos pesados de cada péptido de código de barras.

Cada medición, como se muestra en la Figura 5B, es el área de XIC del código de barras en N SGPGSTPAE (SEQ ID NO. 8029) y el código de barras en C GSAPGTE (SEQ ID NO. 8023) normalizados a un pico de 400 nM de su péptido sintético marcado con isótopo pesado correspondiente. La dilución 1 tiene la cantidad más baja de truncamientos y la dilución 9 tiene la cantidad más alta. Estos datos indican que el péptido de código de barras en N se mide a una abundancia relativamente similar en muestras que contienen truncamientos y no truncamientos. Sin embargo, se observa una disminución de la abundancia del péptido de código de barras en C en la fracción que contiene truncamientos. Esto sugiere que la terminación de la traducción es el factor que más contribuye a las especies truncadas, ya que la traducción en procariotas se inicia en el extremo N-terminal. Habiendo descrito la invención en detalle y con referencia a aspectos específicos y/o modalidades de la misma, será evidente que son posibles modificaciones y variaciones sin apartarse del alcance de la invención definido en las reivindicaciones adjuntas. Más específicamente, aunque algunos aspectos de la presente invención pueden identificarse en

la presente descripción como particularmente ventajosos, se contempla que la presente invención no se limita a estos aspectos particulares de la invención.