

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 661 704**

51 Int. Cl.:

C07K 14/705 (2006.01)

C07K 14/72 (2006.01)

C07K 19/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **09.08.2012** **PCT/GB2012/051940**

87 Fecha y número de publicación internacional: **14.02.2013** **WO13021206**

96 Fecha de presentación y número de la solicitud europea: **09.08.2012** **E 12762356 (9)**

97 Fecha y número de publicación de la concesión europea: **27.12.2017** **EP 2742066**

54 Título: **Proteínas estables**

30 Prioridad:

10.08.2011 US 201161522147 P

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

03.04.2018

73 Titular/es:

HEPTARES THERAPEUTICS LIMITED (100.0%)
BioPark Broadwater Road
Welwyn Garden City, Hertfordshire AL7 3AX, GB

72 Inventor/es:

JAZAYERI-DEZFULY, SEYED ALI y
MARSHALL, FIONA HAMILTON

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 661 704 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Proteínas estables

- 5 La invención se refiere a proteínas que no son fácilmente cristalizables, y particularmente a los GPCR, que no son fácilmente estabilizados, y por lo tanto no son fácilmente cristalizados. La invención también se refiere a métodos para la cristalización de dichas proteínas y a diversos usos de las mismas. Las proteínas son útiles para estudios de descubrimiento y desarrollo de fármacos.
- 10 Los GPCR constituyen una familia muy amplia de proteínas que controlan muchos procesos fisiológicos y que son los objetivos de muchos fármacos eficaces. Se hace referencia particularmente a Overington et al (2006) Nature Rev. Drug Discovery 5, 993-996 que indican que más de un cuarto de los presentes fármacos tienen un GPCR como objetivo. Tienen una considerable importancia farmacológica. Una lista de los GPCR se proporciona en Foord et al (2005) Pharmacol Rev. 57, 279-288.
- 15 Los GPCR son generalmente inestables cuando se aíslan, y a pesar de los considerables esfuerzos, únicamente ha sido posible cristalizar unos pocos GPCR que incluyen la rodopsina bovina, que de forma natural es excepcionalmente estable, y el receptor adrenérgico beta 2, que se cristalizó en forma de una proteína de fusión o en un complejo con un fragmento de anticuerpo.
- 20 Se cree que los GPCR existen en múltiples conformaciones distintas que están asociadas con diferentes clases farmacológicas de ligandos, tales como agonistas y antagonistas, y que se ciclan entre estas conformaciones con objeto de funcionar (Kenakin T. (1997) Ann N Y Acad Sci 812, 116-125). El cambio entre las conformaciones también contribuye a la dificultad en la obtención de las estructuras cristalinas de los receptores.
- 25 Tomando como base la homología en la secuencia y la arquitectura molecular, los GPCR pueden clasificarse en tres familias (A, B y C), aunque todos ellos comparten los característicos siete dominios transmembranarios (TM). La Familia A, el grupo más grande, consiste en receptores que son homólogos de la rodopsina. La Familia B, conocida también como la familia de receptores de la secretina, son receptores que están regulados por grandes hormonas peptídicas tales como la familia de hormonas del glucagón; los miembros de esta familia se caracterizan por un extremo N extracelular relativamente grande que contiene varias cisteínas que forman una red de puentes de disulfuro y es parte del bolsillo de unión al ligando. La Familia C consiste en receptores que son homólogos de los receptores metabotrópicos del glutamato; estos receptores se caracterizan por un extremo N extracelular muy largo, así como una larga cola de carboxi, y el extremo N forma el bolsillo de unión al ligando, que se ha demostrado que
- 30 forma un dímero unido por disulfuro, con una forma semejante a la de una *Dionaea*.
- 35 Durante los últimos pocos años se ha resuelto la estructura de varios de los GPCR de la Familia A, y estos hitos se han conseguido desarrollando varias técnicas clave. Una de dichas tecnologías es la inserción de la lisozima T4 (T4L) en el bucle citoplasmático intracelular (ICL) 3, que se cree que forma una gran área hidrófila que permite la
- 40 formación de contactos en el cristal [2] [3]. La aplicación de esta tecnología junto con la cristalografía en fase cúbica lipídica ha permitido la determinación de la estructura con una elevada resolución de los receptores Beta2, A2a, CXCR4 y D3 [2]. Por lo tanto, se ha recopilado una información significativa a partir de estas estructuras sobre la orientación y la organización del conjunto de receptores TM de la Familia A. Sin embargo, hay poca información disponible sobre los miembros de los receptores de las Familias B y C, y dada la elevada divergencia en la
- 45 secuencia, es probable que existan diferencias significativas en la arquitectura y la organización de los dominios TM entre las familias [1].
- 50 En los receptores de la Familia A, se insertó la T4L en el ICL3 porque se cree que la distancia entre las hélices 5 y 6 es similar a la distancia entre los extremos N y C de la T4L. por lo tanto, es posible acomodar la proteína de fusión en esta posición, mientras que se cree que las distancias entre otras hélices no son favorables para la inserción de un compañero de fusión. De hecho, la T4L se ha fusionado con varios receptores diferentes de la Familia A en el ICL3, y la proteína funcional ha sido expresada en cada caso con el beneficio añadido de reducir la flexibilidad del receptor, aumentando por lo tanto la estabilidad global. El documento WO 2009/051769 se refiere a métodos para la cristalización de un GPCR. El documento WO 2009/051769 se refiere a una proteína de fusión que comprende a)
- 55 una primera porción de un GPCR, en el que la primera porción comprende las regiones TM1, TM2, TM3, TM4 y TM5 del GPCR; b) una inserción de una proteína plegada estable; y c) una segunda porción del GPCR, en la que la segunda porción comprende las regiones TM6 y TM7 del GPCR. El documento WO 2009/051769 se refiere específicamente a la inserción de una proteína plegada estable entre las regiones TM5 y TM6 de un GPCR.
- 60 Ensayamos el efecto de la inserción de la T4L en los bucles internos de los receptores de la Familia B, y particularmente en el ICL3. Nuestros datos indican que los receptores de la Familia B no pueden tolerar la fusión de la T4L en el ICL3, sin embargo, sorprendente e inesperadamente, en vista de la arquitectura de los receptores de la Familia A, la adición de la T4L al ICL2 mejora las propiedades bioquímicas de los receptores de la Familia B. El ICL2 conecta la porción del GPCR que comprende la hélice transmembranaria (TM)-1, la TM2 y la TM3 con la porción del
- 65 GPCR que comprende la TM4, la TM5, la TM6 y la TM7. Los datos actuales sugieren que, al contrario que los receptores de la Familia A, la distancia entre las hélices 3 y 4 en los receptores de la Familia B es más similar a la

distancia entre los extremos N y C de la T4L, que la distancia entre las hélices 5 y 6. Por lo tanto, se cree que la inserción de un dominio de proteína estable entre estas dos porciones de un GPCR representa una nueva técnica para facilitar la cristalización de los GPCR que podría no haber sido predicha previamente.

- 5 Consecuentemente, un primer aspecto de la invención proporciona una proteína de fusión que comprende, desde el extremo N hasta el C- terminal:

- a) una primera porción de un receptor acoplado a proteínas G de la Familia B (GPCR) que comprende la hélice transmembranaria (TM)-1, la TM2 y la TM3 del GPCR;
- 10 b) un dominio de proteína estable seleccionado entre el grupo que consiste en un dominio de lisozima, del citocromo b562, de la flavodoxina, de la β -lactamasa y de la ATPasa de choque térmico de 70 kDa; y
- c) una segunda porción del GPCR que comprende la TM4, la TM5, la TM6 y la TM7 del GPCR.

15 Por "GPCR" queremos significar un receptor o un polipéptido acoplado a una proteína G que tiene la actividad de señalización de un GPCR y que conserva una región de 7TM intacta. La nomenclatura habitual en la materia denomina las hélices transmembranarias de un GPCR desde del extremo N hacia el extremo C como la TM1, la TM2, la TM3, la TM4, la TM5, la TM6 y la TM7. Las hélices transmembranarias están unidas por tramos de aminoácidos extracelularmente entre la TM2 y la TM3, entre la TM4 y la TM5, y entre la TM6 y la TM7, denominados bucle extracelular (ECL) 1, 2 y 3, respectivamente. Las hélices transmembranarias están unidas por tramos de aminoácidos intracelularmente entre la TM1 y la TM2, entre la TM3 y la TM4, y entre la TM5 y la TM6, denominados bucle intracelular (ICL) 1, 2 y 3, respectivamente. Por lo tanto, la primera y la segunda porción del GPCR como se ha definido anteriormente están unidas de forma natural por la región del ICL2, es decir, el ICL2 conecta la primera porción N-terminal con el ICL2 que comprende la TM1, la TM2 y la TM3 con la segunda porción C-terminal al ICL2 que comprende la TM4, la TM5, la TM6 y la TM7.

25 El GPCR deriva preferentemente de secuencias naturales completas que incluyen polimorfismos naturales o moléculas mutantes del GPCR que han sido alteradas, por ejemplo, de forma que se mejore una o más propiedades del GPCR, por ejemplo, la estabilidad.

30 El GPCR puede derivar de cualquier GPCR natural y mutante, en el que los GPCR mutantes pueden ser GPCR estabilizados desviados hacia una conformación particular tal como agonista o antagonista. Por ejemplo, después puede insertarse un dominio de proteína estable entre la TM3 y la TM4 de un GPCR estabilizado conformacionalmente.

35 Previamente hemos desarrollado una metodología para la estabilización de un GPCR en una conformación biológica pertinente (véase el documento WO 2008/114020) que describe la producción de GPCR estabilizados conocidos como StaRs™, que permite la purificación de los receptores acoplados a proteínas G recombinantes que mantienen su conformación, su estabilidad y su función cuando son purificados a partir de la membrana celular. Además, esta plataforma tecnológica también proporciona el medio para diseñar receptores desviados hacia la conformación agonista o la conformación antagonista (véase también Magnani et al, 2008; Serrano-Vega et al, 2008; Shibata et al, 2009), es decir, tiene un aumento en la estabilidad en una conformación en particular. Dichos receptores estabilizados pueden ser usados en la presente invención y tienen varias ventajas, por ejemplo, estabilidad, un elevado rendimiento de la proteína purificada, una reducción en la desnaturalización y una reducción en la unión no específica. Cuando se usa un GPCR mutante estable en la presente invención, se selecciona y se prepara preferentemente mediante el uso de cualquiera de los métodos descritos en las solicitudes PCT WO 2008/114020, WO 2009/114020 y WO 2009/081136. Preferiblemente, la primera y la segunda porción del GPCR son de un GPCR que tiene un aumento en la estabilidad en una conformación en particular con respecto al GPCR parental (es decir, una estabilidad de conformación aumentada). Por estabilidad de conformación aumentada incluimos el significado de que una conformación en particular de un GPCR mutante tiene, en comparación con la misma conformación del GPCR parental, una estabilidad aumentada (por ejemplo, un tiempo de vida prolongado) cuando es expuesto a condiciones desnaturalizantes. Algunos ejemplos de condiciones desnaturalizantes incluyen el calor, un detergente, un agente caotrópico y un pH extremo. Como es bien conocido en la materia, dichas condiciones desnaturalizantes pueden afectar a la estructura secundaria y terciaria de una proteína, pero no a la secuencia primaria.

55 Algunos GPCR adecuados para su uso en la práctica de la invención incluyen cualquier GPCR de la Familia B tal como cualquiera del receptor peptídico glucagonoide 1 (GLP1 R), el receptor peptídico glucagonoide 2 (GLP2R), el receptor de la calcitonina (CT), el receptor de la amilina/CGRP (AMY_{1a}), el receptor de la amilina (AMY_{2a}), el receptor de la amilina/CGRP (AMY_{3a}), el receptor de la CGRP/adrenomedulina (CGRP_{1a}), el receptor de la adrenomedulina/CGRP (AM_{1a}), el receptor de la adrenomedulina/CGRP (el receptor de la AM_{2a}), el receptor del factor liberador de la corticotropina (CRF₁), el receptor de las urocortinas (CRF₂), el receptor del factor liberador de la hormona de crecimiento (GHRH), el receptor del polipéptido inhibidor gástrico (GIP), el receptor del glucagón, el receptor de la secretina, el receptor del TIP-39 (PTH2), el receptor de la hormona paratiroidea (PTH1), el receptor del VIP/PACAP (VPAC₁), el receptor del PACAP (PAC₂) y el receptor del VIP/PACAP (VPAC₂). En una realización particularmente preferida, el GPCR es el GLP1R. Otros GPCR adecuados son bien conocidos en la materia e incluyen los enumerados en Overington et al, *supra*. Además, la International Union of Pharmacology produce una lista de GPCR que incluye los GPCR de la Familia B (Foord et al (2005) Pharmacol. Rev. 57, 279-288, y esta lista se

actualiza periódicamente en <http://www.iuphar-db.org/GPCR/ReceptorFamiliesForward>; los GPCR de la Familia B están recogidos en la Tabla 2 como los GPCR de la Clase 2).

Las secuencias de aminoácidos (y las secuencias de nucleótidos de los ADNc que los codifican) de muchos GPCR están fácilmente disponibles, por ejemplo, mediante referencia al GenBank. En particular, Foord et al *supra* proporcionan los símbolos de los genes humanos y las ID de genes humanos, de ratón y de rata en Entrez Gene (<http://www.ncbi.nlm.nih.gov/entrez>). También debería apreciarse que, debido a que la secuencia del genoma humano está sustancialmente completa, las secuencias de aminoácidos de los GPCR humanos pueden ser deducidas a partir de la misma.

Aunque el GPCR puede proceder de cualquier fuente, particularmente se prefiere que sea de una fuente eucariota. Particularmente se prefiere que proceda de una fuente de un vertebrado, tal como un mamífero. Particularmente se prefiere que el GPCR proceda de rata, de ratón, de conejo o de perro, o de un primate no humano, o del hombre. Para evitar dudas, incluimos en el significado "procedente de" que un ADNc o un gen se obtuvo originalmente usando el material genético de la fuente, pero que la proteína puede ser expresada posteriormente en cualquier célula hospedadora. Por lo tanto, será evidente que un GPCR eucariota (tal como un GPCR de ave o de mamífero) puede ser expresado en una célula hospedadora procariota, tal como *E. coli*, pero ser considerado como procedente de un ave o de un mamífero, según sea el caso.

En algunos casos, el GPCR puede estar formado por más de una subunidad diferente. Por ejemplo, el receptor del péptido relacionado con el gen de la calcitonina requiere la unión de una única proteína de la hélice transmembranaria (RAMP1) para adquirir sus características de unión al ligando fisiológico. Las proteínas efectoras, accesorias, auxiliares o de interacción con un GPCR que se combinan con el GPCR para formar o modular un complejo funcional, son bien conocidas en la materia e incluyen, por ejemplo, las cinasas de receptor, las proteínas G y las arrestinas (Bockaert et al (2004) *Curr Opin Drug Discov and Dev* 7, 649-657). En algunos casos, el GPCR puede estar unido a un ligando del GPCR. Por "ligando" incluimos cualquier molécula que se une al GPCR. Se conocen muchos ligandos, por ejemplo, a partir del documento WO 2008/114020 y de Neubig et al (2003) *Pharmacol. Rev.* 55, 597-606. Por lo tanto, la proteína de fusión puede comprender una porción del GPCR que comprende la TM1, la TM2 y la TM3 conectada a una porción del GPCR que comprende la TM4, la TM5, la TM6 y la TM7, en la que el GPCR está unido a un compañero de unión del GPCR. De esta forma es posible obtener una percepción estructural de las interacciones del GPCR al ser capaces de cristalizar complejos del GPCR con otras moléculas. Se prefiere que las moléculas nos sean aquellas que se unen al ICL2 del GPCR.

Para cualquier GPCR dado, las hélices TM pueden ser determinadas por la persona experta usando las técnicas convencionales en la materia. Por ejemplo, hay disponibles programas informáticos que modelan las regiones transmembranarias de los GPCR basándose en la hidrofobicidad (Kyle & Dolittle (1982) *J. Mol. Biol.* 157, 105-132). Asimismo, hay ampliamente disponibles servidores de algoritmos de predicción transmembranaria en Internet (por ejemplo, ExPASy), muchos de los cuales se basan en un análisis de la hidropatía. El TMHMM es un método de predicción de la topología de la proteína de membrana que puede usarse, basado en un modelo oculto de Markov (TMHMM Server v. 2.0; <http://www.cbs.dtu.dk/services/TMHMM/>). Cuando ya se conocen las regiones transmembranarias de un GPCR dado, por ejemplo, mediante un análisis estructural o mediante un análisis de la hidropatía, también pueden identificarse las regiones análogas en otro GPCR mediante una alineación de la secuencia múltiple o por parejas. Por ejemplo, la alineación puede llevarse a cabo usando el programa Clustal W (Thompson *et al.*, 1994). Los parámetros usados pueden ser como sigue: parámetros de la alineación por emparejamiento rápido: tamaño de K-tuple (palabra); 1, tamaño de ventana; 5, penalización por un hueco; 3, número de diagonales superiores; 5. Método de puntuación: x por ciento. Parámetros de la alineación múltiple: penalización por hueco abierto; 10, penalización por extensión del hueco; 0,05. Matriz de puntuación: BLOSUM.

La Figura 6 recoge las secuencias de aminoácidos de los GPCR de la Familia B y destaca la posición de la TM3, del ICL2 y de la TM4. Por ejemplo, para el GLP1R humano, la TM3 termina en la Phe 257, el ICL2 se corresponde con la Ser 258 hasta la Ser 261, y la TM4 comienza en la Glu 262. Por lo tanto, convenientemente, cuando el GPCR es un receptor de la Familia B, las posiciones de la TM3, del ICL2 y de la TM4 pueden ser identificadas localizando los residuos de aminoácidos que se corresponden con los aminoácidos que definen los límites de la TM3, del ICL2 y de la TM4 en la Figura 6, cuando las secuencias están alineadas, por ejemplo, usando el CLUSTAL W.

Sin embargo, se apreciará que los límites no son absolutos y que pueden depender bastante del modelo proporcionado del GLP1R que se haya usado para definirlos. En la Figura 1, por ejemplo, el TM3 del GLP1R humano finaliza en la Leu 254, el ICL2 se corresponde con la Leu 255 hasta el Trp 264, y la TM4 comienza en la Ile 265. También, las regiones de los bucles pueden estar definidas como estructuras de aminoácidos que se unen a hélices alfa o estructuras de aminoácidos que se predice que estarán fuera de la membrana, y dependiendo de qué definición se use, los límites cambiarán.

En una realización, el dominio de proteína estable está insertado en el ICL2. Por lo tanto, la invención proporciona un GPCR en el que se ha insertado un dominio de proteína estable en el ICL2. Por 'insertado en el ICL2' incluimos tanto la adición de la secuencia de aminoácidos que define el dominio de proteína estable en la secuencia de aminoácidos del ICL2 sin la delección de ninguno de los aminoácidos del ICL2, como también la sustitución de uno o

más, o de todos, los aminoácidos del ICL2 por la secuencia de aminoácidos que codifica el dominio de proteína estable. Se apreciará que, en esta realización, la primera y/o la segunda porción del GPCR puede comprender al menos parte del ICL2, además de las necesarias hélices transmembranarias. La primera porción del GPCR puede comprender la TM1, la TM2 y la TM3 y una parte N-terminal del ICL2. La segunda porción del GPCR puede comprender la TM4, la TM5, la TM6 y la TM7 y una parte C-terminal del ICL2.

Se aprecia que el dominio de proteína estable puede ser insertado en el ICL2 y estar flanqueado por una o dos fracciones separadoras en sus extremos N y/o C. De esta forma, el dominio de proteína estable no está unido directamente al ICL2, sino que está unido indirectamente. Las fracciones separadoras pueden usarse para ayudar a reducir la tensión en las hélices.

Preferiblemente, el dominio de proteína estable es insertado en el ICL2 mediante la sustitución de uno o más aminoácidos consecutivos (por ejemplo, 2, 3, 4 o 5, o más aminoácidos) en la secuencia de aminoácidos del ICL2 por la secuencia de aminoácidos del dominio de proteína estable. En una realización, el uno o más aminoácidos que son sustituidos es/son al menos uno o dos aminoácidos del extremo C de la TM3 y/o del extremo N de la TM4. En otras palabras, la proteína de fusión puede tener al menos uno o dos aminoácidos del ICL2 en al menos uno de los lados del dominio de proteína estable.

Según se describe en el Ejemplo 1, hemos insertado la lisozima T4 en varias posiciones del ICL2 del GLP1 R, y las inserciones entre la Phe 257 y la Ser 261 dieron como resultado unos receptores de fusión GLP1 R productivos. Por lo tanto, particularmente se prefiere que el dominio de proteína estable sea insertado en la región ICL2 del GPCR en una posición entre los residuos de aminoácidos que se corresponden con los aminoácidos Phe 257 y Ser 261 según la numeración del GLP1 R humano según se establece en la Figura 6.

Consecuentemente, el aminoácido del dominio de proteína estable puede ser insertado en la región del ICL2 del GPCR después de un aminoácido correspondiente al aminoácido Phe 257 y antes de un aminoácido correspondiente al aminoácido Ser 261 o Phe 260 o Val 259, según la numeración del GLP1 R humano según se establece en la Figura 6. Por ejemplo, la secuencia de aminoácidos del dominio de proteína estable puede sustituir al aminoácido correspondiente a la Ser 258, o puede sustituir a los aminoácidos correspondientes a la Ser 258 y a la Val 259, o puede sustituir a los aminoácidos correspondientes a la Ser 258, a la Val 259 y a la Phe 260, según la numeración del GLP1 R humano según se establece en la Figura 6. La inserción del dominio de proteína estable en dichas posiciones se corresponde con las construcciones de fusión de la lisozima GLP1R-T4 1 c, 2c y 3c mostradas en la Figura 1.

De forma análoga, el aminoácido del dominio de proteína estable puede ser insertado en la región ICL2 del GPCR después de un aminoácido correspondiente al aminoácido Ser 258 y antes de un aminoácido correspondiente al aminoácido Ser 261 o Phe 260 o Val 259, según la numeración del GLP1 R humano según se establece en la Figura 6. Por ejemplo, la secuencia de aminoácidos del dominio de proteína estable puede sustituir al aminoácido correspondiente a la Val 259, o puede sustituir a los aminoácidos correspondientes a la Val 259 y a la Phe 260, según la numeración del GLP1 R humano según se establece en la Figura 6. La inserción del dominio de proteína estable en dichas posiciones se corresponde con las construcciones de fusión de la lisozima GLP1 R-T4 1d, 2d y 3d mostradas en la Figura 1.

Por "residuo de aminoácido correspondiente" incluimos el significado del residuo de aminoácido en otro GPCR que se alineaba con un residuo de aminoácido dado del GLP1 R humano cuando se comparan el receptor GLP1 R humano y el otro GPCR usando MacVector y CLUSTAL W.

Aunque se prefiere que el dominio de proteína estable sea insertado en la región ICL2 del GPCR en una posición entre los residuos de aminoácidos que se corresponden con los aminoácidos Phe 257 y Ser 261 según la numeración del GLP1R humano según se establece en la Figura 6, se aprecia que puede ser insertado fuera de esta región.

Se aprecia que la función del dominio de proteína estable es aumentar la superficie hidrófila para el contacto del cristal y reducir la inherente flexibilidad de los GPCR con objeto de, por ejemplo, mejorar las propiedades de cristalización del GPCR. Consecuentemente, por 'dominio de proteína estable' incluimos el significado de cualquier polipéptido soluble plegado que proporciona una superficie hidrófila para el contacto de la red cristalina. Además, el dominio de la proteína es estable de forma que, en su forma plegada, es resistente a la desnaturalización (por ejemplo, es estable frente al calor, los detergentes y los agentes caotrópicos etc.). Los ensayos de estabilidad de la proteína son bien conocidos en la materia e incluyen los descritos en el documento WO2008/114020.

Normalmente, el dominio de proteína estable es aquel que se pliega de forma autónoma con respecto a las porciones del GPCR de la proteína de fusión en la célula. En el contexto de la presente invención, el dominio de proteína estable es según se define en las reivindicaciones.

Convenientemente, el dominio de proteína estable es aquel que es fácilmente cristizable. Por lo tanto, el dominio de proteína estable puede ser una proteína cuya estructura cristalina ha sido resuelta, por ejemplo, una cuyas

coordenadas han sido depositadas en el Protein Data Bank (<http://www.pdb.org/>).

Las características particularmente preferidas del dominio de proteína estable son:

- 5 1. el dominio es soluble, está bien plegado y puede ser expresado con facilidad en uno o más sistemas de expresión;
2. los extremos N y C del dominio están próximos entre sí en el espacio; normalmente en el intervalo de 5-17 Å, **por ejemplo**, de 6-16 Å, de 7-15 Å, de 7-10 Å, de 10-13 Å o de 12-15 Å;
- 10 3. el dominio es resistente a una desnaturalización térmica y química, así como a una degradación proteolítica;
4. el dominio es muy cristizable en diversos grupos espaciales y disposiciones de empaquetamiento cristalino.

- 15 Se prefiere que el dominio no contenga residuos de cisteína con el fin de evitar la formación de puentes de disulfuro tanto en el dominio o como en la porción del GPCR de la proteína de fusión. Se entenderá que, dado que el dominio es soluble, no debe ser hidrófobo ni tener propensión a la agregación de una forma desordenada.

- 20 En una realización, la longitud del dominio de proteína estable es de entre 50 y 1000 aminoácidos, preferentemente de entre 50 y 300 aminoácidos o de entre 100 y 300 aminoácidos, o de entre 150 y 250 aminoácidos. En el contexto de la presente invención, el dominio de proteína estable es según se define en las reivindicaciones. Siempre que en el presente documento se divulguen otros dominios de proteína estables, están incluidos simplemente con fines de referencia.

- 25 Una vez que se ha encontrado un polipéptido adecuado para el dominio de proteína estable, puede ser necesario modificar el polipéptido mediante la delección o la adición de residuos de aminoácidos de o en el extremo N, el extremo C o en ambos terminales del polipéptido, de forma que los átomos de carbono alfa más próximos en el esqueleto en los terminales del polipéptido estén separados por una distancia en el intervalo de 5-17 Å, **por ejemplo**, de 6-16 Å, de 7-15 Å, de 7-10 Å, de 10-13 Å o de 12-15 Å.

- 30 Se prefiere que la inserción del dominio de proteína estable no afecte a la actividad biológica del GPCR, tal como una actividad de unión o una actividad de modulación de una ruta de señalización. Idealmente, la proteína de fusión debería conservar al menos el 60 % o el 70 % o el 80 % o el 90 % de su actividad biológica, y lo más idealmente el 100 % de su actividad biológica con respecto al nivel de la misma actividad en ausencia del dominio de proteína estable. Los métodos para la evaluación de la unión al GPCR y de la señalización del GPCR son bien conocidos en la materia y se describen, **por ejemplo**, en el documento WO 2008/114020 y en el documento WO 2009/101383. Por lo tanto, cuando la actividad biológica es una actividad de unión, la unión a cualquier compañero de unión del GPCR puede ser evaluada usando los ensayos de unión rutinarios conocidos en la materia; cuando la actividad biológica es una actividad de modulación de una ruta de señalización, la actividad puede ser evaluada mediante cualquier ensayo adecuado para la ruta de señalización en particular (**por ejemplo**, un ensayo con genes indicadores).

- 40 Se aprecia que la conservación de la capacidad de unión al ligando es más importante para los fines de cristalización que la conservación de la actividad de señalización, y que puede ser deseable evaluar únicamente la capacidad de unión al ligando antes de la cristalización. Por lo tanto, en una realización particularmente preferida, el dominio de proteína estable no afecta a la actividad de unión de un GPCR.

- 45 Para evitar dudas, el dominio de proteína estable no es el ICL2 ni una parte del mismo del GPCR en particular.

- 50 En una realización preferida, el dominio de proteína estable es una lisozima. La lisozima es conocida por ser fácilmente cristizable, y las estructuras de varias lisozimas naturales y variantes han sido depositadas en el Protein Data Bank (www.rcsb.org). Algunos ejemplos adecuados incluyen 135L, 193L, 194L, 1AKI, 1GBS, 1 IEE, 1 LZ1, 1P7S, 1 REX, 1VDQ, 2ANV, 2ANX, 2D4K, 2FBB, 2IHL, 2NWD, 2XBR, 2XBS, 2Z2F, 2ZYP, 3A8Z, 3K2R, 3N9A, 3N9C, 3N9E y 3OD9

- 55 Aunque puede usarse una lisozima procedente de cualquier Fuente, se prefiere particularmente que la lisozima proceda del fago T4. En la Figura 7 se proporcionan dos secuencias de aminoácidos de la lisozima del fago T4, y puede usarse cualquiera de las secuencias en el contexto de la invención. La lisozima puede proceder de secuencias naturales completas que incluyen polimorfismos naturales, o puede ser una lisozima mutante que ha sido alterada, **por ejemplo**, para mejorar una o más propiedades. Por lo tanto, se entiende que pueden usarse variantes de las secuencias de aminoácidos proporcionadas en la Figura 7, tales como secuencias de aminoácidos con una identidad en la secuencia de al menos el 60 %, el 65 %, el 70 %, el 75 %, el 80 %, el 85 % o el 90 % con cualquiera de las secuencias establecidas en la Figura 7, y más preferentemente con una identidad en la secuencia de al menos el 95 % o el 99 % con cualquiera de las secuencias establecidas en la Figura 7.

- 65 La identidad en la secuencia puede medirse mediante el uso de algoritmos tales como BLAST o PSI-BLAST (Altschul et al, NAR (1997), 25, 3389-3402) o con métodos basados en modelos ocultos de Markov (Eddy S et al, J Comput Biol (1995) Spring 2 (1) 9-23). Normalmente, el porcentaje de identidad en la secuencia entre dos

polipéptidos puede determinarse usando cualquier programa informático adecuado, por ejemplo, el programa GAP del University of Wisconsin Genetic Computing Group, y se apreciará que el porcentaje de identidad se calcula en relación con los polipéptidos cuya secuencia ha sido alineada óptimamente. La alineación puede llevarse a cabo, como alternativa, usando el programa Clustal W (Thompson *et al.*, 1994), como se ha mencionado anteriormente.

Aunque la lisozima es un ejemplo preferido de un dominio de proteína estable, pueden usarse los principios generales para emplear cualquier cantidad de polipéptidos que tenga las características analizadas anteriormente. Por lo tanto, algunos candidatos adecuados incluyen aquellos que contienen la secuencia de aminoácidos de proteínas que son fácilmente cristalizables, por ejemplo, como se encuentra al interrogar el banco de datos de proteínas u otras bases de datos de cristalización conocidas en la materia. Otros ejemplos incluyen los mencionados en Engel *et al* (2002) BBA 1564: 38-46, tales como el citocromo_{b562}, la flavodoxina, la β -lactamasa y el dominio de la ATPasa de choque térmico de 70 kDa. En el contexto de la presente invención, el dominio de proteína estable se selecciona entre el grupo que consiste en la lisozima, el citocromo_{b562}, la flavodoxina, la β -lactamasa y el dominio de la ATPasa de choque térmico de 70 kDa.

La proteína de fusión puede modificarse para que sea detectada más fácilmente, por ejemplo, mediante una biotilación o incorporando cualquier marcador detectable conocido en la materia tal como radiomarcadores, marcadores fluorescentes o marcadores enzimáticos. En una realización particularmente preferida, el marcador es un marcador fluorescente tal como la EGFP. De forma análoga, la proteína de fusión puede modificarse para facilitar su purificación, por ejemplo, mediante la incorporación de cualquier fracción de afinidad conocida en la materia tal como una etiqueta de GST, una etiqueta de 6x His, una MBP u otra etiqueta de epítipo. Dichas modificaciones pueden ser en el extremo N o en el extremo C del GPCR, o en un bucle externo.

Como se muestra en el Ejemplo 1, se cree que las proteínas de fusión de la invención tienen unas propiedades bioquímicas mejoradas en comparación con las propiedades bioquímicas del GPCR sin la inserción del dominio de proteína estable. Dichas propiedades mejoradas hacen que la proteína de fusión sea más susceptible a la cristalización. Por lo tanto, se espera que la proteína de fusión tenga una mayor superficie hidrófila para los contactos del cristal. De forma análoga, se espera que la proteína de fusión sea más soluble, por ejemplo, mostrando una menor agregación en una solución detergente, que el GPCR sin la inserción del dominio de proteína estable. Los métodos para la evaluación de la solubilidad del GPCR son bien conocidos en la materia e incluyen la cromatografía de exclusión por tamaños, tal como una cromatografía de exclusión por tamaños fluorescente usada para evaluar la solubilidad en DDM según se describe en el Ejemplo 1. La proteína de fusión también puede ser más estable (por ejemplo, frente a cualquiera del calor, un detergente o agentes caotrópicos) que el GPCR sin la inserción del dominio de proteína estable. Los métodos para la evaluación de la estabilidad del GPCR son conocidos en la materia, incluyendo los descritos en el documento WO 2008/114020.

Convenientemente, la proteína de fusión es producida mediante las técnicas habituales de biología molecular y de ADN recombinante. Por ejemplo, pueden usarse fragmentos de ADN que codifican la primera y la segunda porción del GPCR, y el dominio de proteína estable puede elaborarse usando las técnicas de clonación habituales y una PCR, como es bien conocido en la materia. Después, los fragmentos pueden ligarse entre sí en marco según la práctica convencional, por ejemplo, empleando unos terminales en extremos romos o con extremos escalonados para la ligación, una digestión con enzimas de restricción para proporcionar los terminales apropiados, el relleno de los extremos cohesivos según sea apropiado, el tratamiento con fosfatasa alcalina para evitar uniones indeseadas, y una ligación enzimática.

Igualmente, las construcciones de fusión pueden elaborarse usando estrategias de ligación independientes de ligasa tales como InFusion o Gateway. La construcción también puede elaborarse sintéticamente a través de una síntesis génica *de novo*.

Se aprecia que uno o ambos del GPCR y del dominio de proteína estable pueden estar mutados de forma que se mejore cualquiera de la solubilidad, la estabilidad, la expresión y la cristalizabilidad.

Los métodos de biología molecular para la clonación y el diseño de genes y de ADNc, para la mutación del ADN y para la expresión de polipéptidos a partir de polinucleótidos en células hospedadoras son bien conocidos en la materia, como se ejemplifica en "Molecular cloning, a laboratory manual", tercera edición, Sambrook, J. & Russell, D. W. (eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Algunos sistemas de expresión adecuados incluyen sistemas de expresión constitutivos o inducibles en bacterias o levaduras, sistemas de expresión víricos tales como baculovirus, virus del bosque de semliki y lentivirus, o una transfección temporal en células de insecto o de mamífero. Algunas células hospedadoras adecuadas incluyen células de *E. coli*, de *Lactococcus lactis*, de *Saccharomyces cerevisiae*, de *Schizosaccharomyces pombe*, de *Pichia pastoris*, de *Spodoptera frugiperda* y de *Trichoplusia ni*. Algunas células hospedadoras animales adecuadas incluyen HEK 293, COS, S2, CHO, NSO, DT40 y similares. Se sabe que algunos GPCR requieren unos lípidos específicos (por ejemplo, colesterol) para funcionar. En ese caso, es deseable seleccionar una célula hospedadora que contenga el lípido. Adicionalmente o como alternativa, el lípido puede ser añadido durante el aislamiento y la purificación de la proteína de fusión. La purificación puede llevarse a cabo mediante las técnicas convencionales, tales como una cromatografía de afinidad.

Un segundo aspecto de la invención proporciona un polinucleótido que codifica la proteína de fusión según el primer aspecto de la invención. El polinucleótido puede ser ARN (por ejemplo, ARNm) o ADN, aunque normalmente es ADN.

- 5 Se apreciará que el polinucleótido puede ser incorporado en un vector, y por lo tanto la invención también proporciona un vector que comprende un polinucleótido según el segundo aspecto de la invención.

Los vectores adecuados son aquellos que se propagan en, y/o permiten la expresión de, la proteína de fusión en células procariotas (por ejemplo, bacterianas) o eucariotas (por ejemplo, de mamífero). Por ejemplo, el vector puede ser un plásmido, un cósmido, un fago o un cromosoma artificial bacteriano (BAC). La secuencia del polinucleótido del vector dependerá de la naturaleza de la célula hospedadora prevista, de la forma de introducción del polinucleótido del segundo aspecto de la invención en la célula hospedadora, y de si se desea un mantenimiento episomal o una integración. Convenientemente, el vector comprende al menos un marcador seleccionable tal como de resistencia a un antibiótico (por ejemplo, kanamicina o neomicina).

15 Los vectores son útiles para replicar el polinucleótido del segundo aspecto de la invención, y también son útiles para transfectar células con el polinucleótido, y también pueden promover la expresión de la proteína de fusión.

Algunos plásmidos de vectores procariotas son: pUC18, pUC19, pBR322 y pBR329 disponibles en Biorad Laboratories (Richmond, CA, EE.UU.); pTrc99A, pKK223-3, pKK233-3, pDR540 y pRIT5 disponibles en Pharmacia (Piscataway, NJ, EE.UU.); vectores pBS, vectores Phagescript, vectores Bluescript, pNH8A, pNH16A, pNH18A, pNH46A disponibles en Stratagene Cloning Systems (La Jolla, CA 92037, EE.UU.).

Un plásmido de un vector de célula de mamífero típico es pS-VL disponible en Pharmacia (Piscataway, NJ, EE.UU.). Este vector usa el promotor tardío del SV40 para dirigir la expresión de los genes clonados, encontrándose el mayor nivel de expresión en las células productoras del antígeno T, tales como las células COS-1. Otro ejemplo es pDNAC3.1 (neo) (Invitrogen) para su uso en células COS-1 o COS-7. Un ejemplo de un vector de expresión de mamífero inducible es pMSG, también disponible en Pharmacia (Piscataway, NJ, EE.UU.). Este vector usa el promotor inducible por glucocorticoides de la repetición terminal larga del virus del tumor de mama de ratón para dirigir la expresión del gen clonado.

Algunos vectores de plásmido de levadura útiles son pRS403-406 y pRS413-416, y están generalmente disponibles en Stratagene Cloning Systems (La Jolla, CA 92037, EE.UU.). Los plásmidos pRS403, pRS404, pRS405 y pRS406 son plásmidos de integración de levadura (YIp) e incorporan los marcadores seleccionables de levadura *HIS3*, *TRP1*, *LEU2* y *URA3*. Los plásmidos pRS413-416 son plásmidos de centrómero de levadura (YCp).

En una realización preferida, el vector que comprende el polinucleótido del segundo aspecto de la invención es pDNAC3.1 (<http://products.invitrogen.com/ivgn/product/V79020>).

40 Puede usarse cualquier método adecuado conocido en la materia para la construcción de los vectores que contienen el polinucleótido del segundo aspecto de la invención, incluyendo las técnicas de ligación descritas anteriormente.

Un tercer aspecto de la invención proporciona una célula que comprende un polinucleótido según el segundo aspecto de la invención, o un vector que comprende dicho polinucleótido. Dichas células pueden usarse para replicar el polinucleótido del segundo aspecto de la invención, o pueden usarse para expresar la proteína de fusión del primer aspecto de la invención.

La célula puede ser tanto procariota como eucariota.

50 Se aprecia que la construcción y la amplificación del polinucleótido del segundo aspecto de la invención se lleva a cabo convenientemente en células bacterianas. La expresión del polinucleótido puede llevarse a cabo en células tales como células de mamífero o células bacterianas.

Las células bacterianas son células hospedadoras procariotas preferidas, y normalmente son una cepa de *E. coli* tal como, por ejemplo, la cepa de *E. coli* DH5 disponibles en Bethesda Research Laboratories Inc., Bethesda, MD, EE.UU., y la RR1 disponible en la American Type Culture Collection (ATCC) de Rockville, MD, EE.UU. (ATCC nº 31343). Algunas células hospedadoras procariotas preferidas incluyen células o líneas celulares de levadura, de insecto y de mamífero, preferentemente células o líneas celulares de vertebrados tales como las de ratón, de rata, de mono o humanas. Las células particularmente preferidas son las células de riñón embrionario humanas tales como las células HEK293T.

Las células usadas para la expresión de la proteína de fusión pueden ser transfectadas de forma estable o no estable.

65 Un cuarto aspecto de la invención proporciona un método de cristalización de una proteína de fusión según el primer aspecto de la invención, comprendiendo el método proporcionar una proteína de fusión según el primer aspecto de

la invención, y cristalizarla para obtener los cristales.

En una realización, la proteína de fusión se proporciona cultivando una célula hospedadora según el tercer aspecto de la invención, para expresar la proteína de fusión y aislar la proteína.

Puede usarse cualquier método de cristalización adecuado para cristalizar la proteína de fusión, tal como cualquiera de los revisados en "Crystallisation of Biological Macromolecules" (Alexander McPherson; ISBN: 0-87969-617-6).

En una realización preferida, la cristalización se realiza usando una cristalografía en fase cúbica lipídica (véase el documento US 2011/0031438).

Un quinto aspecto de la invención proporciona un cristal que comprende una proteína de fusión del primer aspecto de la invención.

Las proteínas de fusión divulgadas en el presente documento son útiles para estudios de cristalización, y son útiles en programas de descubrimiento de fármacos. Pueden usarse en las mediciones biofísicas de la cinética del receptor/ligando y de los parámetros hemodinámicos, por ejemplo, mediante una resonancia de plasmón superficial o técnicas basadas en fluorescencia. Pueden usarse en cribados de unión al ligando, y pueden ser acopladas a superficies sólidas para su uso en cribados de alto rendimiento o como chips de biosensores. Los chips de biosensores que contienen las proteínas de fusión pueden usarse para la detección de moléculas, especialmente de biomoléculas.

La invención se describirá ahora con la ayuda de las siguientes figuras y ejemplos.

Figura 1. Diseño de construcciones de fusión de la T4L con el GLP1 R. Se insertó la T4L después de los residuos indicados del ICL2 (Parte 1 de 2) y del ICL3 (Parte 2 de 2). Construcción 1a significa que la T4L fue insertada entre la L255 y la S261. El modelo de los dominios TM y los bucles se basa en la referencia [4].

Figura 2. Señal total de la EGFP de las construcciones de fusión GLP1 R-T4L en comparación con la muestra natural (WT) y las transfectadas en falso (U). Cada medición se llevó a cabo con 50 ug del material celular total por duplicado. Las barras de error representan la desviación típica con respecto a la media.

Figura 3. Perfil de elución típico de una fSEC del GLP1 R natural.

Figura 4. Perfiles de elución de una fSEC de las construcciones de fusión GLP1 R-T4L en el ICL2 solubilizadas en DDM solapados con el perfil natural. En cada caso se muestra el perfil natural con una línea continua, y las construcciones de fusión con una línea de puntos.

Figura 5. Perfiles de elución de una fSEC de las construcciones de fusión GLP1 R-T4L en el ICL3 solubilizadas en DDM solapados con el perfil natural. En cada caso se muestra el perfil natural con una línea continua, y las construcciones de fusión con una línea de puntos.

Figura 6. Secuencias de aminoácidos del GPCR de la Familia B que muestran la posición de la TM3, del ICL2 y de la TM4 (SEQ ID Nos: 1-22). La porción del ICL2 que fue sustituida por la T4L en la construcción de fusión GLP1 R 1c está destacada en otros receptores de la Familia B de ratón, de rata y humano.

Figura 7. Secuencias de aminoácidos de la lisozima del fago T4: (A) Secuencia insertada en el ICL3 de los receptores de la Familia A [2], [3] (SEQ ID No: 23); (B) Secuencia insertada en el ICL2 de los receptores de la Familia B (SEQ ID No: 24) (véanse los Ejemplos). Las diferencias están indicadas en los cuadros.

Ejemplo 1: la inserción de la T4L en el ICL2 mejora las propiedades bioquímicas del GLP1R

Sumario

Hemos ensayado el efecto de la inserción de la T4L en los bucles internos de los receptores de la Familia B. Nuestros datos indican que los receptores de la Familia B no pueden tolerar la fusión de la T4L en el ICL3, sin embargo, la adición de la T4L al ICL2 mejora las propiedades bioquímicas del receptor.

Resultados

Las regiones de bucle del GLP1 R se determinaron según el modelo de GLP1 R y se insertó el ADN que codifica la T4L en diferentes ubicaciones del ICL2 y del ICL3 (Figura 1). La construcción GLP1 R se etiquetó en el extremo C con la EGFP con objeto de monitorizar la expresión total, así como la monodispersidad, usando una cromatografía de exclusión por tamaños con detección fluorescente.

Después de la confirmación de la secuencia, estas construcciones fueron expresadas temporalmente en HEK293T.

Como análisis inicial se midió la señal de la EGFP en las células completas para evaluar los niveles de expresión totales. De forma interesante, las construcciones del ICL3 no consiguieron producir ninguna señal de la EGFP, lo que indica que la fusión de la T4L en esta región del GLP1 R es incompatible con la arquitectura global de este receptor. Sin embargo, las fusiones del ICL2 dieron como resultado una robusta expresión del GLP1R (Figura 2).

Con objeto de analizar las propiedades bioquímicas de las fusiones GLP1 R-T4L, se solubilizaron las células que expresan estas construcciones en dodecil maltósido (DDM) y se aplicaron a una cromatografía de exclusión por tamaños con detección fluorescente (fSEC). La fSEC se ha usado ampliamente para proporcionar datos relativos a la monodispersidad y el estado de agregación de las proteínas, particularmente en cribados previos a la cristalización [5]. En general, las condiciones más favorables darán como resultado una mayor monodispersidad y una agregación reducida. El perfil de elución de la fSEC del GLP1 R natural solubilizado en DDM muestra la presencia del pico principal monodispersado con un hombro de agregación, así como especies libres de EGFP que son el resultado de una degradación proteolítica (Fig 3).

Coherentes con los datos de la señal de la EGFP mostrados en la Figura 2, los perfiles de elución de las construcciones de fusión ICL2 T4L 1a, 1b, 2a, 2b, 3a y 3b indican que las células no consiguieron expresar estas fusiones (Fig 4). Lo más probable es que esto sea debido a la proximidad del extremo N de la T4L a cada TM III, dando lugar a una disrupción de la estructura global. Por el contrario, los perfiles de elución de las construcciones 1c, 1d, 2c, 2d, 3c y 3d revelaron que se habían expresado unos productivos receptores de fusión, y más significativamente, parece que la fusión de la T4L da como resultado una reducción en el pico de agregación, y una concomitante mejora en el pico monodispersado, lo que indica en conjunto que la inserción de la T4L en esta región del receptor tiene un efecto beneficioso sobre las propiedades bioquímicas del receptor solubilizado (Fig 4). Este efecto es más pronunciado en las construcciones 1c y 2c.

Se llevó a cabo el mismo análisis para las construcciones de fusión en el ICL3 y, de acuerdo con los datos de la señal de la EGFP mostrados en la Figura 2, ninguna de las fusiones de la T4L en el ICL3 dio como resultado la expresión de algún receptor de fusión productivo (Fig 5).

Tomados en conjunto, estos datos indican que la fusión de la T4L en el tercer bucle citoplasmático del GLP1R no es tolerada, sin embargo, la inserción de la T4L en ciertas posiciones del segundo bucle citoplasmático no solo es bien tolerada, sino que mejora las propiedades bioquímicas del receptor solubilizado. Dada la elevada homología en la secuencia entre los miembros de los GPCR de la Familia B, sugerimos que estas observaciones pueden ser extendidas a otros miembros de esta Familia. La porción del ICL2 que fue sustituida en la mejor construcción (1 c) está destacada en otros miembros de la Familia B, como se muestra en la Figura 6.

Métodos y material

Se insertó la lisozima T4 en el segundo y el tercer bucle citoplasmático del GLP1 R humano usando las técnicas de biología molecular habituales. Estas construcciones fueron expresadas temporalmente en un pDNAC3.1 modificado en células HEK293T, generando receptores fusionados con la EEGFP en sus extremos C. Las transfecciones se llevaron a cabo usando GeneJuice (Merck Biosciences) según las directrices del fabricante. Normalmente, se usaron 6 ug de ADN para transfectar 3×10^6 células adherentes en placas de 10 cm. Las células se recogieron aproximadamente 40 horas después de la transfección y se resuspendieron en HEPES 50 mM a pH 7,5 / NaCl 150 mM / EDTA 0,5 mM complementado con un cóctel inhibidor de la proteasa completo exento de EDTA (Roche). Normalmente se solubilizaron 650 ug de cada muestra con DDM al 1 % en un volumen total de 200 ul durante 1 hora a 4 °C, seguido de una centrifugación a 50000 rpm durante 30 minutos. Se cargaron 50 ul del sobrenadante en una columna BioSEep-SEC-S3000 (Phenomenex), previamente equilibrada con tampón SEC (HEPES 50 mM a pH 7,5 / NaCl 150 mM / EDTA 0,5 mM / DDM al 0,03 %) y se analizaron con un caudal de 1 ml/minuto durante 15 minutos. El eluyente se pasó a través de un fluorímetro en línea con los siguientes ajustes: excitación a 490 nm, emisión a 513 nm y una ganancia de 13.

Referencias

[1] Kristiansen, K. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharma & Therap* 103, 21-80 (2004).

[2] Bill RM, Henderson PJ, Iwata S, Kunji ER, Michel H, Neutze R, Newstead S, Poolman B, Tate CG y Vogel H. Overcoming barriers to membrane protein structure determination. *Nat Biotechnol.* 29 (4), 335-340 (2011).

[3] Kobilka BK, Kobilka TS, Daniel K, Regan JW, Caron MG y Lefkowitz RJ. Chimeric alpha 2-,beta 2-adrenergic receptors: delineation of domains involved in effector coupling and ligand binding specificity. *Science* 240 (4857) 1310-6 (1988).

[4] Frimurer TM y Bywater RP. Structure of the integral membrane domain of the GLP1 receptor. *Proteins* 35 (4), 375-86 (1999).

[5] Kawate T y Gouaux E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure 14 (4), 673-81 (2004).

REIVINDICACIONES

1. Una proteína de fusión que comprende, desde el extremo N hasta el extremo C:

- a) una primera porción de un receptor acoplado a proteínas G de la Familia B (GPCR) que comprende la hélice transmembranaria (TM)-1, la TM2 y la TM3 del GPCR;
- b) un dominio de proteína estable seleccionado entre el grupo que consiste en un dominio de lisozima, del citocromo_{b562}, de la flavodoxina, de la β -lactamasa y de la ATPasa de choque térmico de 70 kDa; y
- c) una segunda porción del GPCR que comprende la TM4, la TM5, la TM6 y la TM7 del GPCR.

2. Una proteína de fusión según la Reivindicación 1, en la que el GPCR es un GPCR de la Familia B tal como cualquiera del receptor del péptido glucagonoide 1 (GLP1 R), el receptor del péptido glucagonoide 2 (GLP2R), el receptor de la calcitonina (CT), el receptor de la amilina/CGRP (AMY_{1a}), el receptor de amilina (AMY_{2a}), el receptor de la amilina/CGRP (AMY_{3a}), el receptor de la CGRP/adrenomedulina (CGRP_{1a}), el receptor de la adrenomedulina/CGRP (AM_{1a}), el receptor de la adrenomedulina/CGRP (el receptor de la AM_{2a}), el receptor del factor liberador de la corticotropina (CRF₁), el receptor de las urocortinas (CRF₂), el receptor de la hormona liberadora de la hormona del crecimiento (GHRH), el receptor del polipéptido inhibidor gástrico (GIP), el receptor del glucagón, el receptor de la secretina, el receptor del TIP-39 (PTH2), el receptor de la hormona paratiroidea (PTH1), el receptor del VIP/PACAP (VPAC₁), el receptor del PAC-AP (PAC₂) y el receptor del VIP/PACAP (VPAC₂).

3. Una proteína de fusión según las Reivindicaciones 1 o 2, en la que el GPCR es un GPCR mutante que tiene una estabilidad conformacional aumentada con respecto a su GPCR parental.

4. Una proteína de fusión según cualquiera de las Reivindicaciones 1-3, en la que el dominio de la proteína estable está insertado en la región del bucle intracelular 2 (ICL2) del GPCR, bucle que une la primera y la segunda porción del GPCR.

5. Una proteína de fusión según la Reivindicación 4, en la que el dominio de la proteína estable se ha insertado en la región ICL2 del GPCR en una posición entre los residuos de aminoácidos que se corresponden con los aminoácidos Phe 257 y Ser 261 según la numeración del GLP1 R humano según se establece en la Figura 6, opcionalmente, en donde el dominio de la proteína estable está insertado en la región ICL2 del GPCR después de un aminoácido correspondiente al aminoácido Phe 257 y antes de un aminoácido correspondiente a los aminoácidos Ser 261 o Phe 260 o Val 259, según la numeración del GLP1 R humano según se establece en la Figura 6, o en donde el dominio de la proteína estable está insertado en la región ICL2 del GPCR después de un aminoácido correspondiente al aminoácido Ser 258 y antes de un aminoácido correspondiente al aminoácido Ser 261 o Phe 260 o Val 259, según la numeración del GLP1 R humano según se establece en la Figura 6.

6. Una proteína de fusión según cualquiera de las Reivindicaciones 1-5, en la que el dominio de la proteína estable comprende la secuencia de aminoácidos de la lisozima, tal como la lisozima T4.

7. Una proteína de fusión según cualquiera de las Reivindicaciones 1-6, que comprende adicionalmente una fracción detectable tal como cualquiera de un marcador fluorescente, un radiomarcador o un marcador enzimático; opcionalmente, en la que la fracción detectable es la EGFP.

8. Un polinucleótido que codifica la proteína de fusión según cualquiera de las Reivindicaciones 1-7.

9. Una célula hospedadora que comprende un polinucleótido según la Reivindicación 8.

10. Un cristal que comprende una proteína de fusión de cualquiera de las Reivindicaciones 1-7.

11. Una proteína de fusión según cualquiera de las Reivindicaciones 1-7, que está en una forma solubilizada o que está sustancialmente exenta de otras proteínas, o que está inmovilizada sobre un soporte sólido.

12. Uso de una proteína de fusión según cualquiera de las Reivindicaciones 1-7 para la cristalización o para el descubrimiento de fármacos, tal como en el cribado de unión a ligandos o en el desarrollo de un ensayo, o como un biosensor.

13. Un método de cristalización de una proteína de fusión según cualquiera de las Reivindicaciones 1-7, método que comprende proporcionar una proteína de fusión según cualquiera de las Reivindicaciones 1-7 y cristalizarla para obtener cristales.

14. Un método según la Reivindicación 13, en el que la proteína de fusión se obtiene mediante el cultivo de una célula hospedadora según la Reivindicación 9 para expresar la proteína de fusión y aislar la proteína.

15. Un método según las Reivindicaciones 13 o 14, en el que la cristalización se lleva a cabo usando una cristalografía en fase cúbica lipídica.

Figura 1 (Parte 1 de 2)

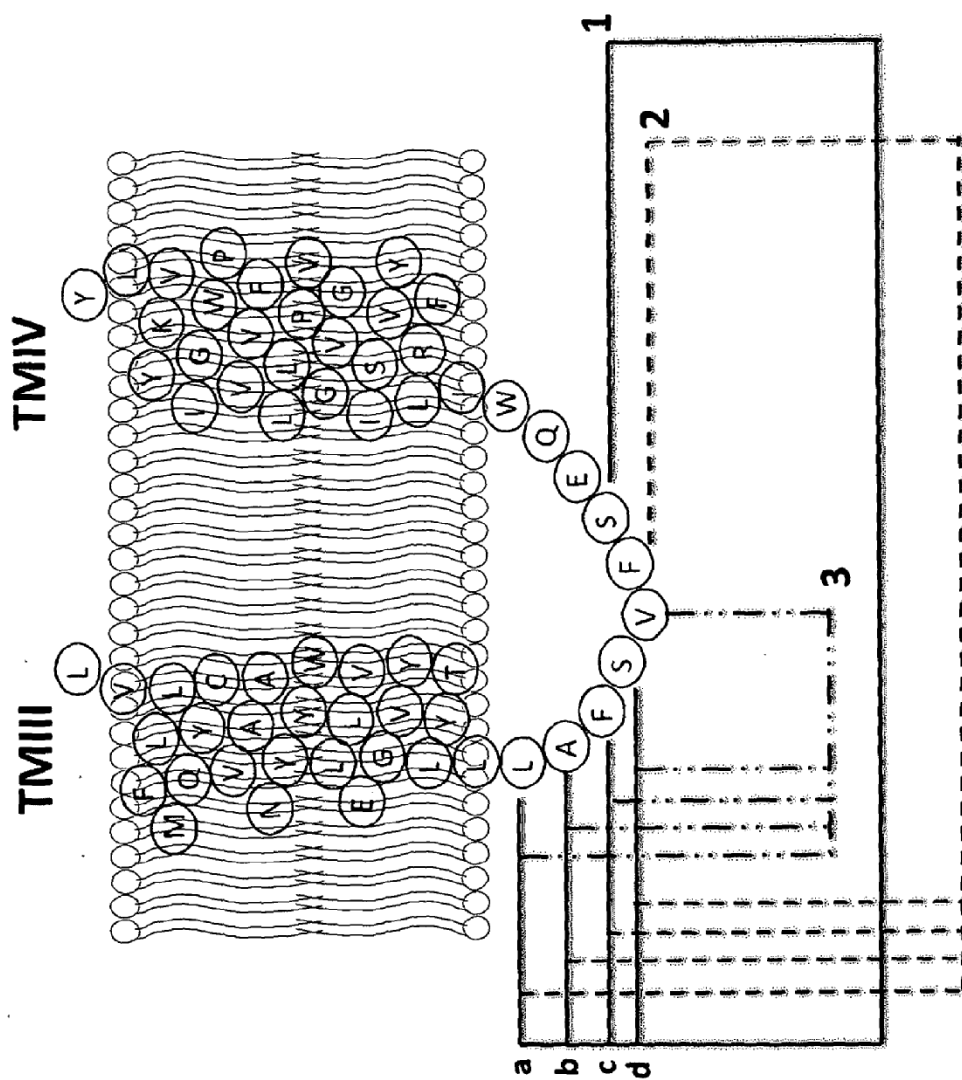


Figura 1 (Parte 2 de 2)

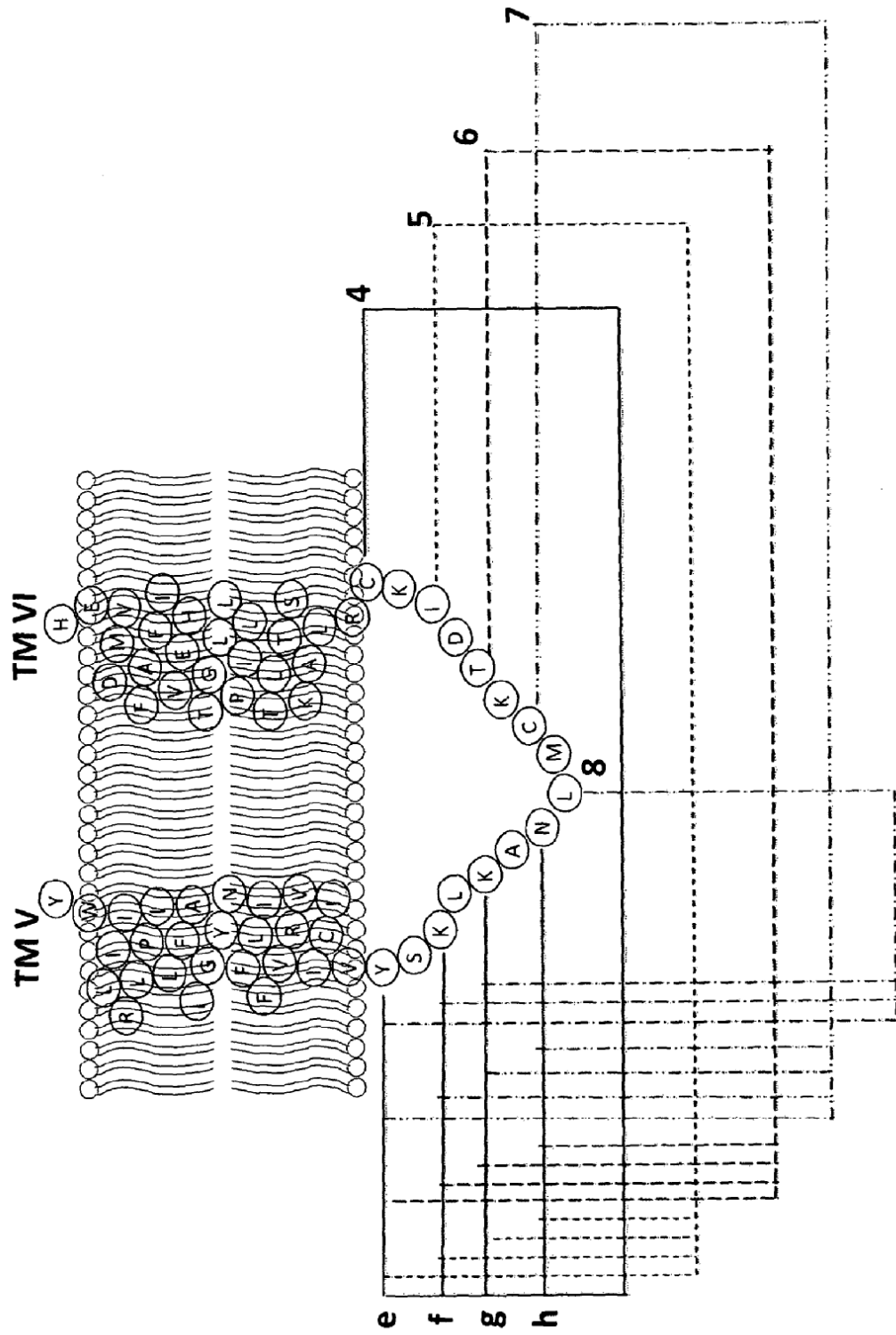


Figura 2

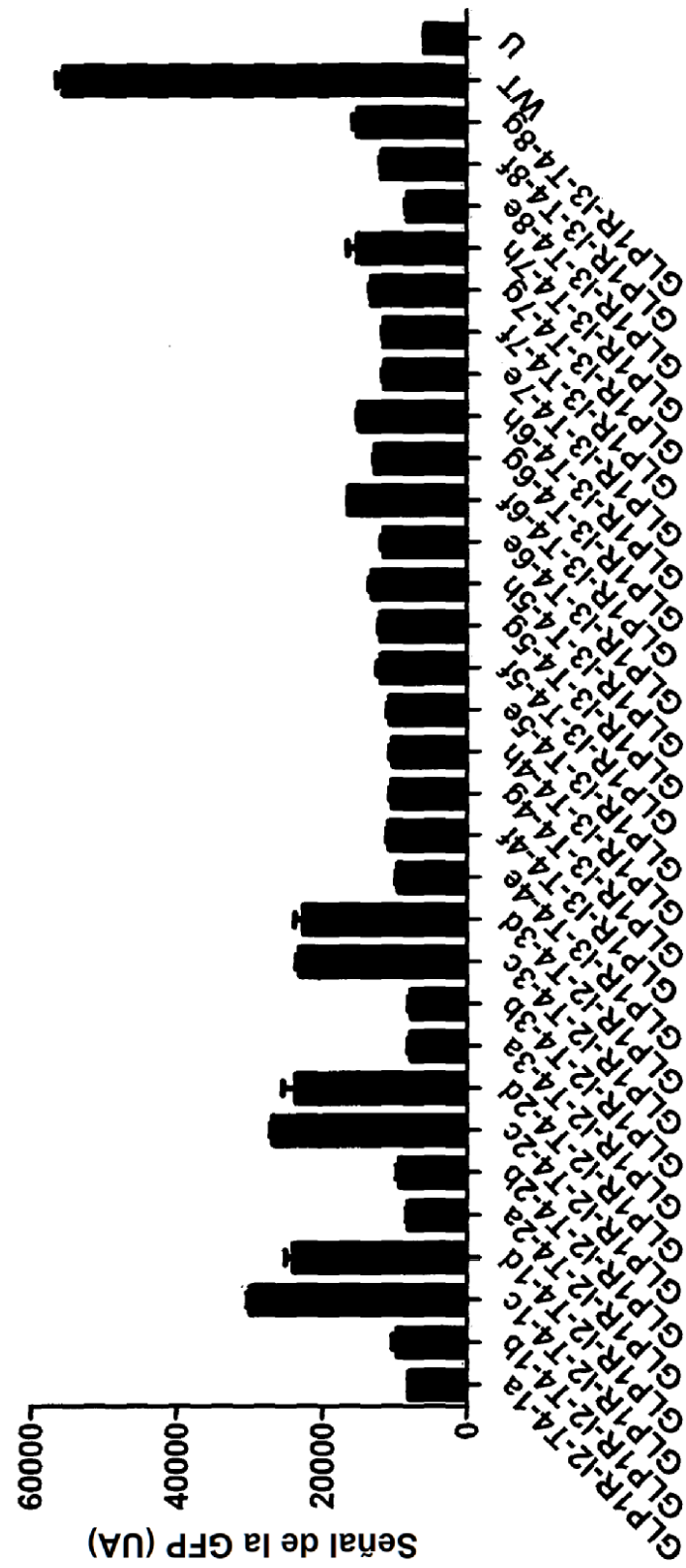


Figura 3

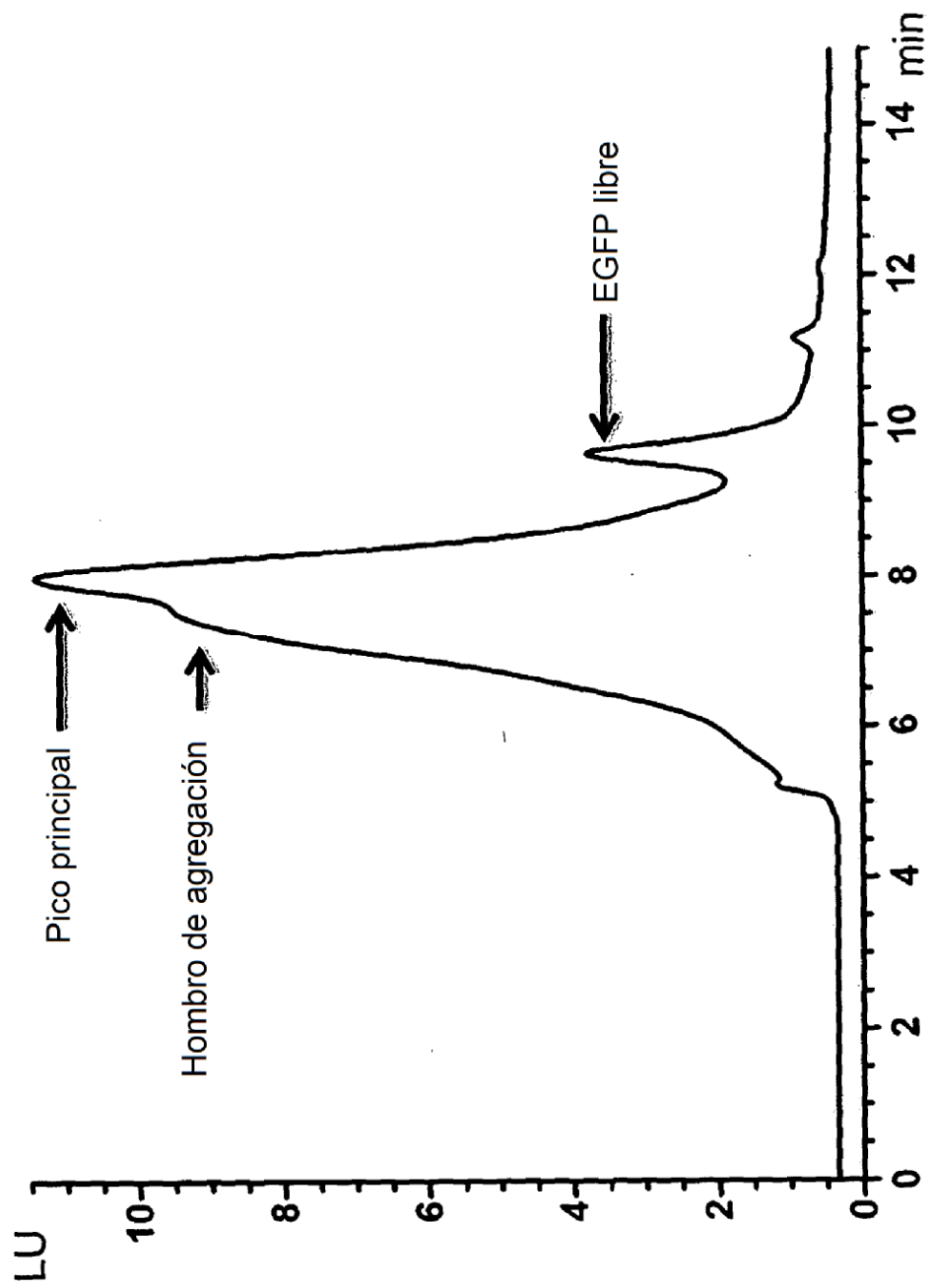


Figura 4 (Parte 1 de 3)

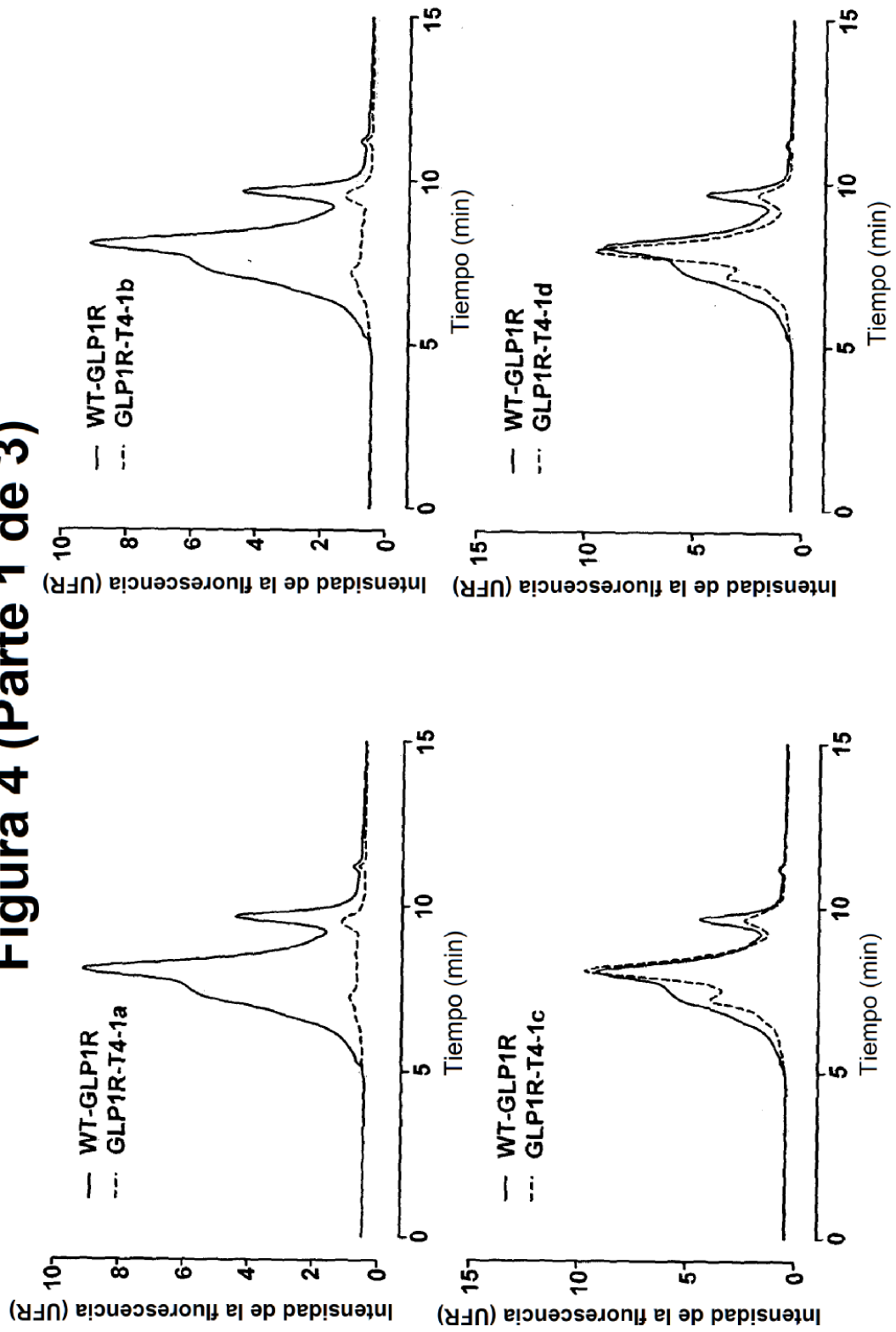


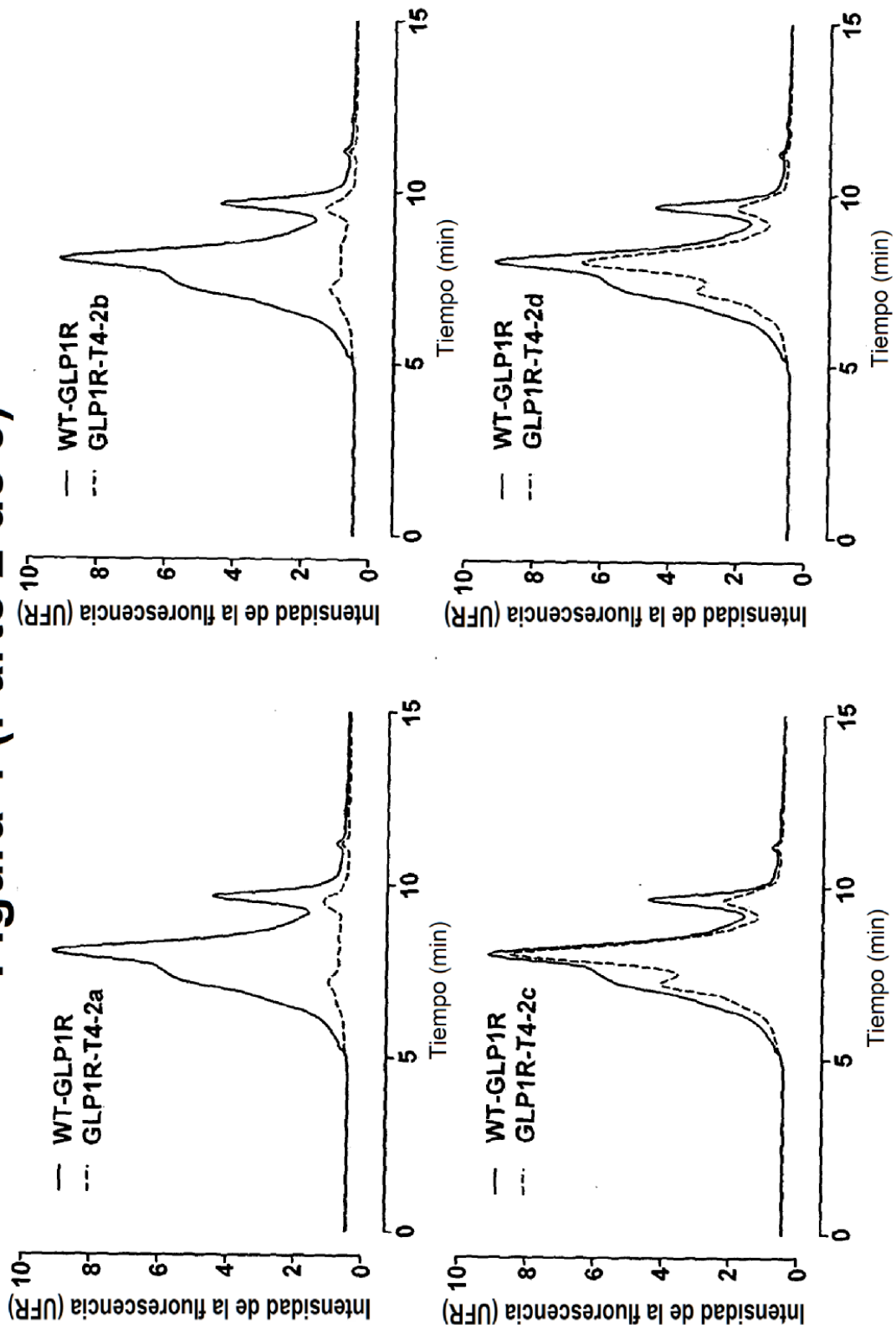
Figura 4 (Parte 2 de 3)

Figura 4 (Parte 3 de 3)

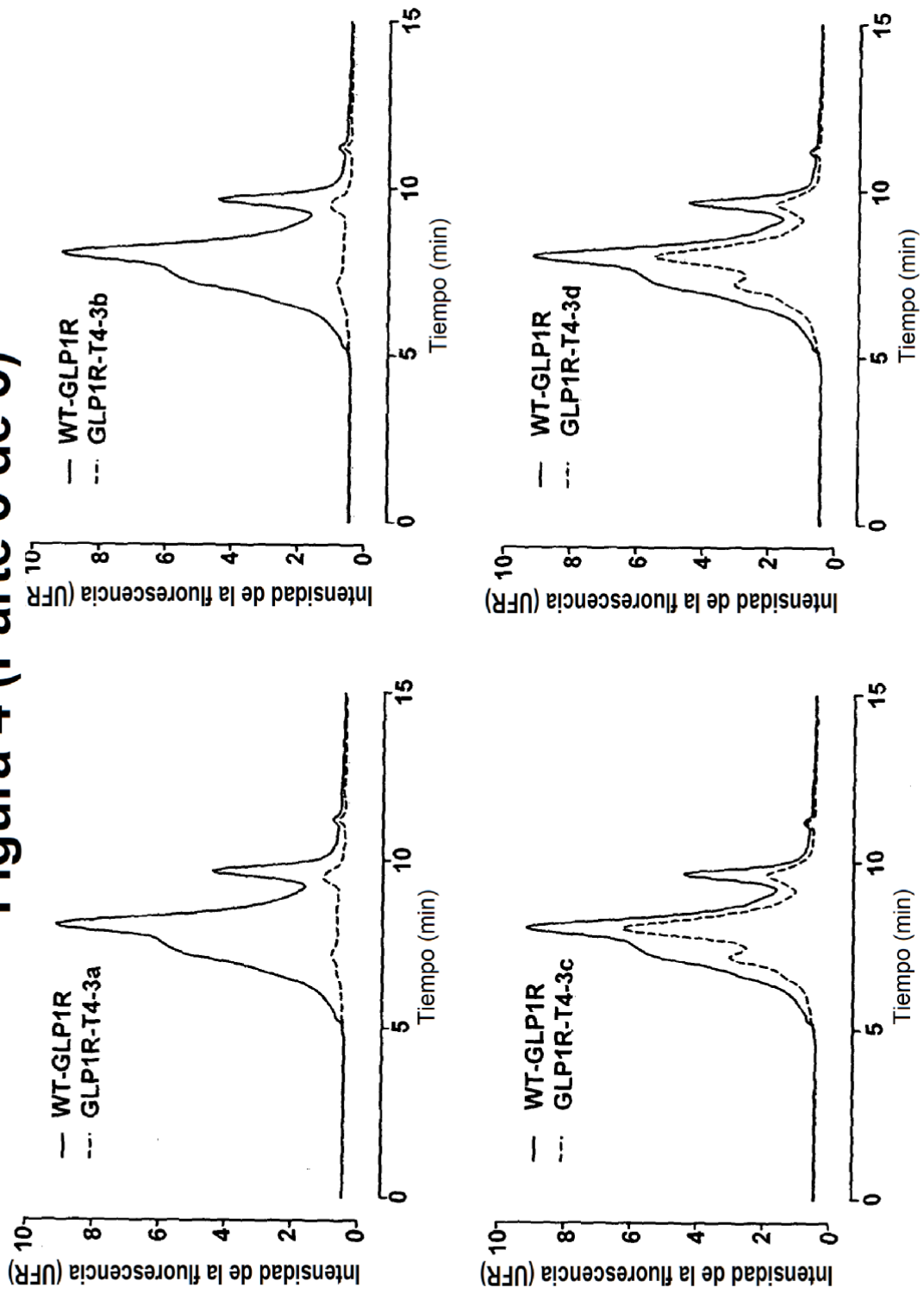


Figura 5 (Parte 1 de 5)

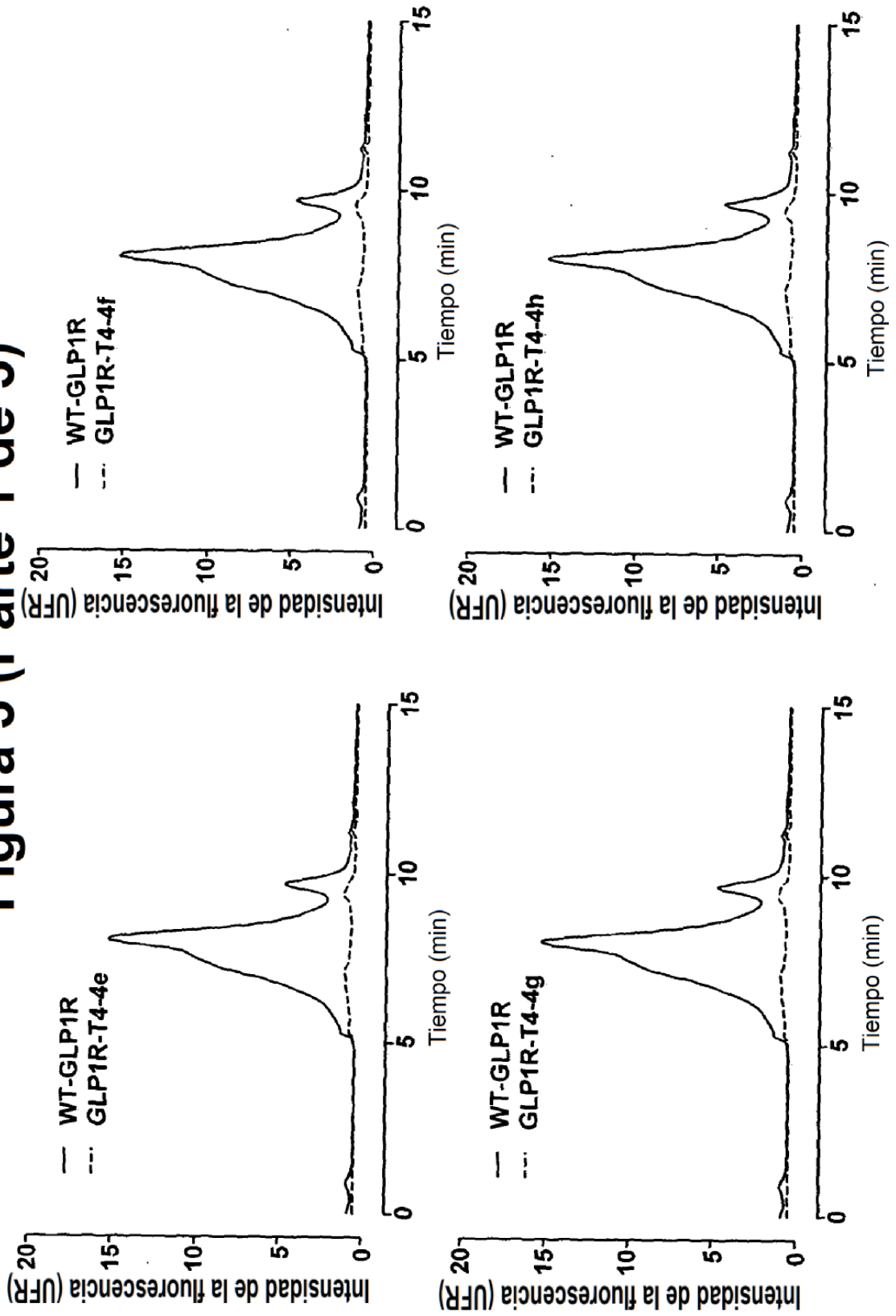


Figura 5 (Parte 2 de 5)

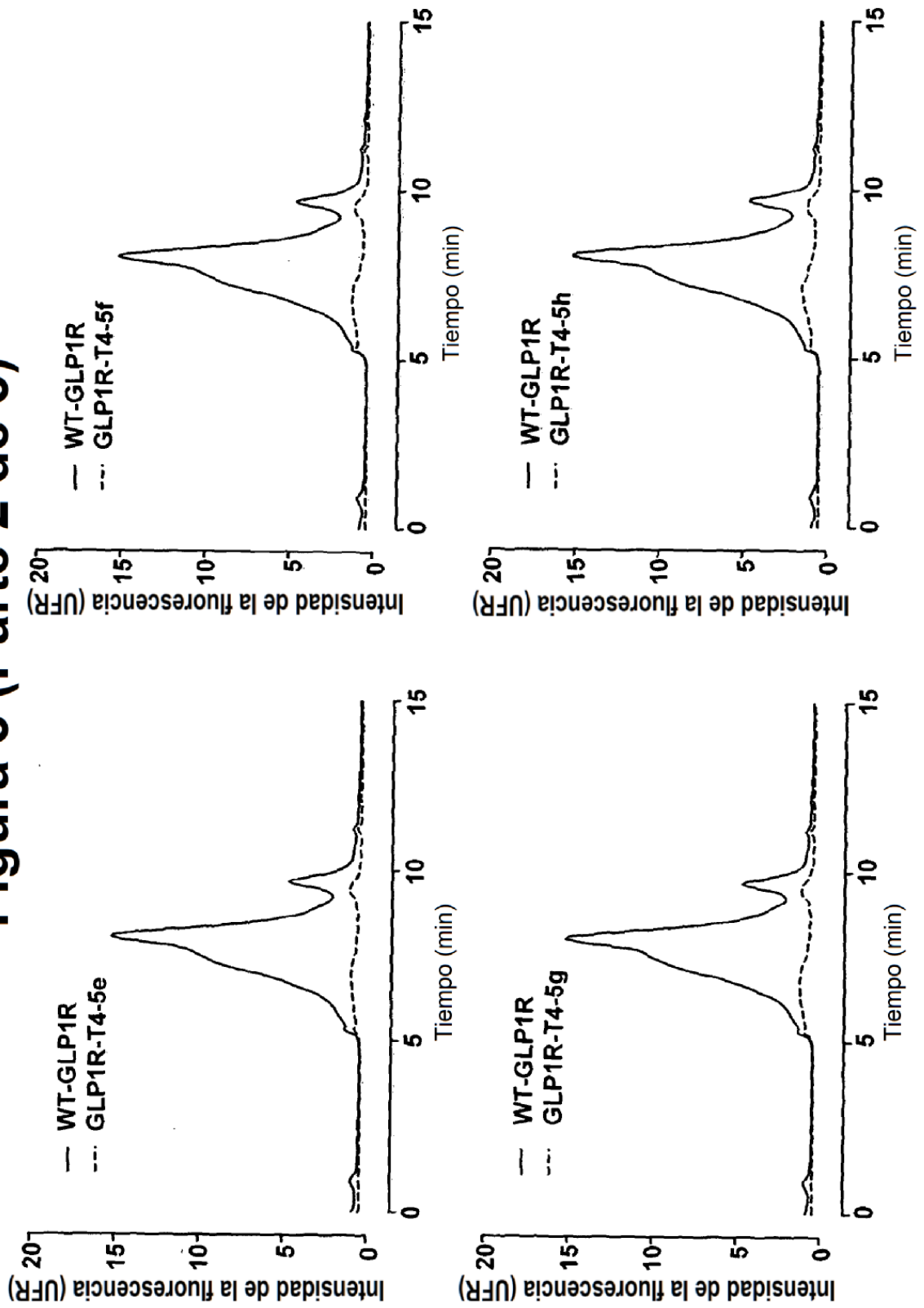


Figura 5 (Parte 3 de 5)

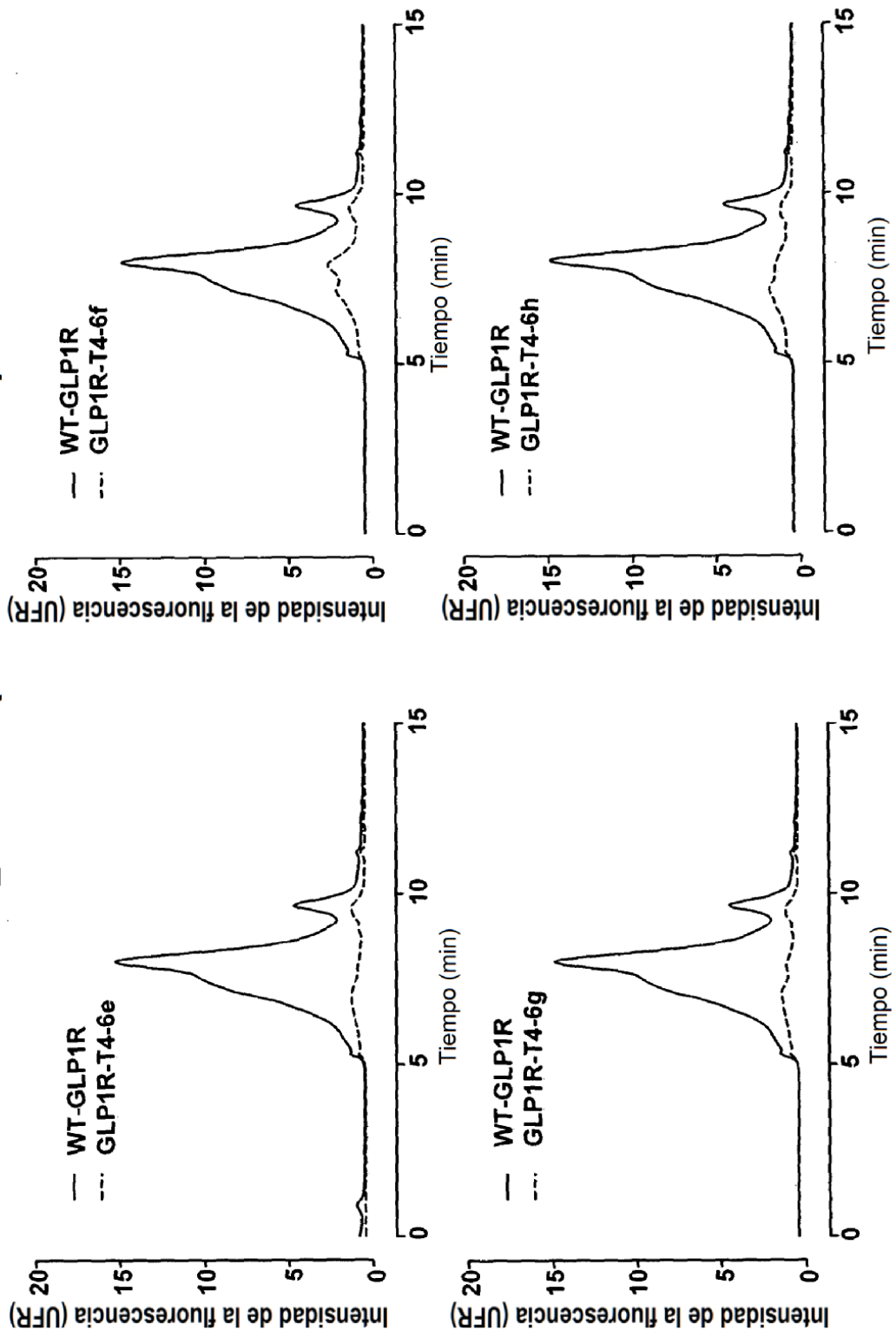


Figura 5 (Parte 4 de 5)

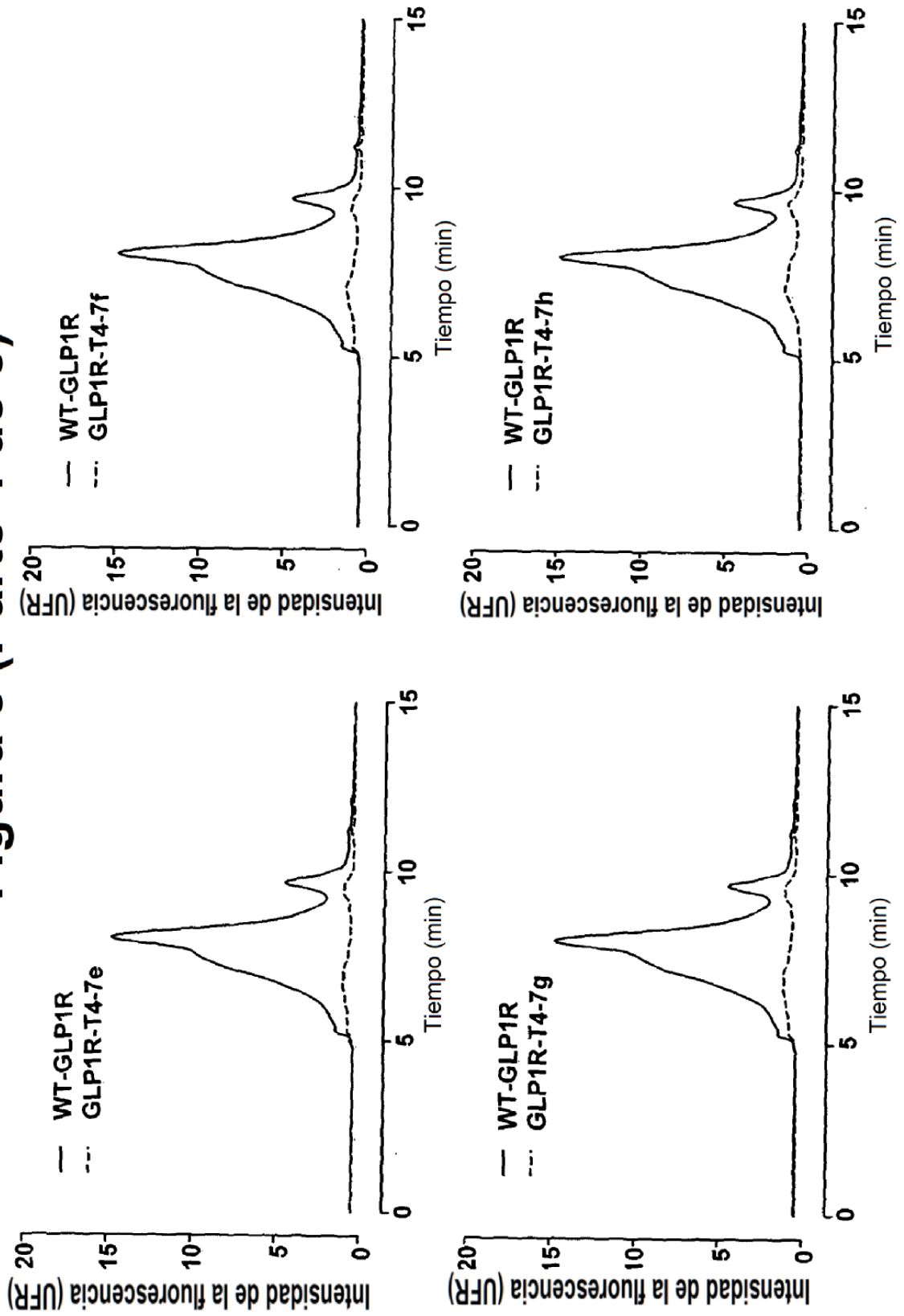


Figura 5 (Parte 5 de 5)

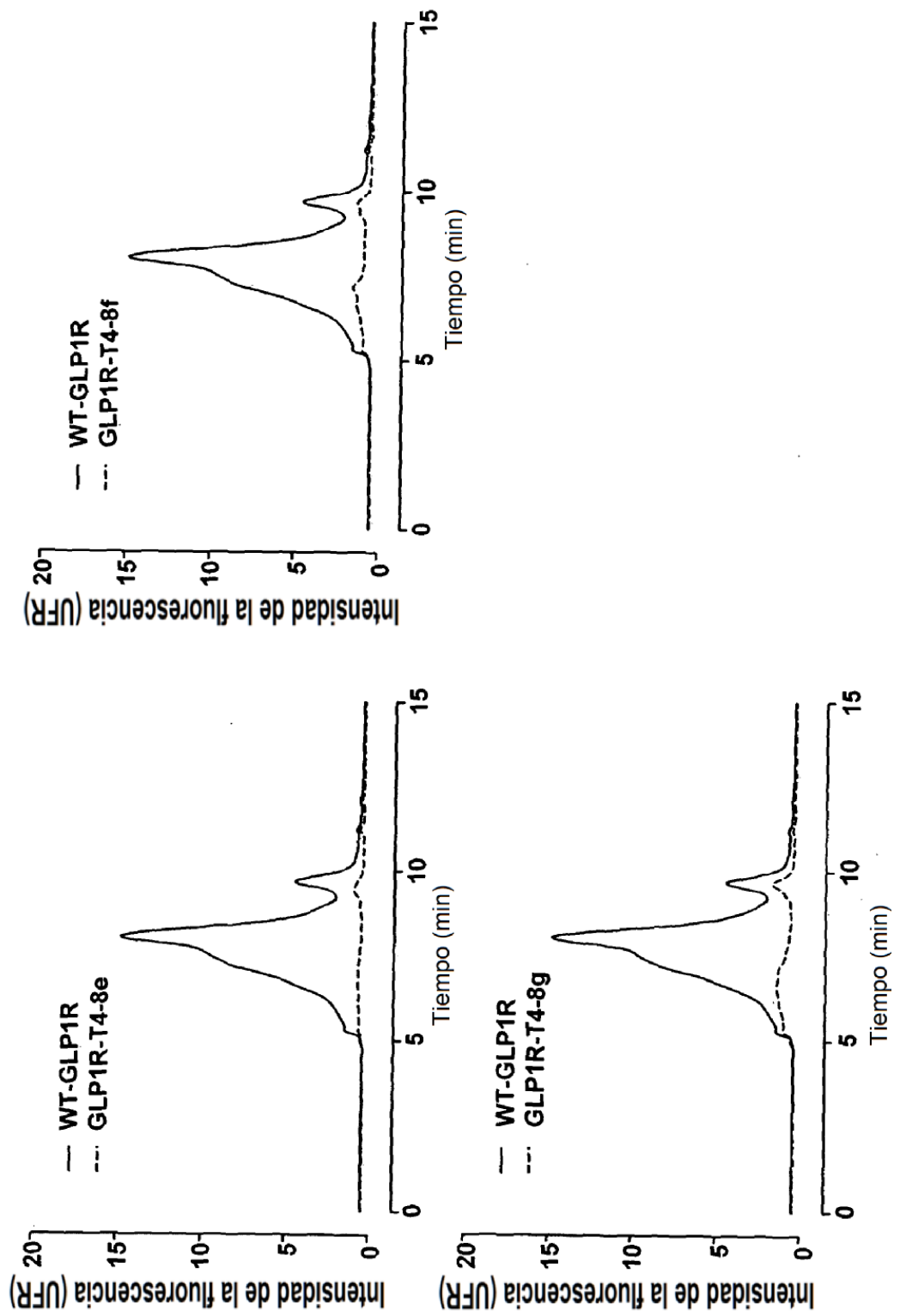


Figura 6

```

-----> <----- HLX4 -----<-----
PTHR1_RATA SLIFMAFFSEKKYLWGFTIFGWGLPAVFVAVVWGVVRATLANTGCWDLSSG--HKKWIIQV
PTHR1_RATÓN SLIFMAFFSEKKYLWGFTIFGWGLPAVFVAVVWGVVRATLANTGCWDLSSG--HKKWIIQV
PTHR1_HUMANO SLIFMAFFSEKKYLWGFTVFGWGLPAVFVAVVWVSVRATLANTGCWDLSSG--NKKWIIQV
PTHR2_RATÓN NLIFVSFFSDTKYLWGFTISIGWGFPVAVFVAVAVARATLADTRCWELSA---GDRWIYQA
PTHR2_RATA NLIFVSFFSDTKYLWGFTILIGWGFPVAVFVAVAVARATLADTRCWE---LSAGDRWIYXX
PTHR2_HUMANO NLIFVAFSSDTKYLWGFTILIGWGFPAAAFVAAVAVARATLADARCWEL--SAGDIKWIIYQA
GHRHR_RATÓN CLLASTSERSKPAFWWLVLGAGWGLPVLCGTGWVGCKHSFEDTECWDLDNS-SPCWWIIKG
GHRHR_RATA CLLASTSERSKPAFWWLVLGAGWGLPVLCGTGWVGCKLAFEDTACWDLDDS-SPYWWIIKG
GHRHR_HUMANO CLLASTSESSRAFWWLVLGAGWGLPVLTGTWVSCKLAFEDIACWDLDDT-SPYWWIIKG
PACR_RATÓN TLLVETFFPERRYFYWYTIIGWGTPTVCVTVWAVLRRLYFDDAGCWDMDNS-TALWWVIKG
PACR_RATA TLLVETFFPERRYFYWYTIIGWGTPTVCVTVWAVLRRLYFDDAGCWDMDNS-TALWWVIKG
PACR_HUMANO TLLVETFFPERRYFYWYTIIGWGTPTVCVTVWATLRRLYFDDTGCWDMNDNS-TALWWVIKG
VIPR2_RATÓN TLLVAIL-PPSRCFLAYLLIGWGIPSVICIGAWTATRLSLEDTCWDTNDH-SIPWWVIRM
VIPR2_RATA TLLVAIL-PPSRCFLAYLLIGWGIPSVICIGAWIATRLSLEDTCWDTNDH-SIPWWVIRM
VIPR2_HUMANO TLLVAML-PPRCFLAYLLIGWGLPTVCIGAWTAARLYLEDTCWDTNDH-SVPWWVIRI
VIPR1_RATÓN TLLAVSFFSERKYFWGYILIGWGVPSVFIMIWTIVRIHFEDFGCWDTIIN-SSLWWIIKG
VIPR1_RATA TLLAVSFFSERKYFWGYILIGWGVPSVFITWTVVRIYFEDFGCWDTIIN-SSLWWIIKA
VIPR1_HUMANO TLLAVSFFSERKYFWGYILIGWGVPSFTMTVWTIARIHFEDYGCWDT-IN-SSLWWIIKG
SCTR_HUMANO TLLAISFFSERKYLQGFVAFGWGSPAIFVALWAIARHFLEDVGCWDINAN-ASIWWIIRG
SCTR_RATA TLLAISFFSERKYLQAFVLLGWGSPAIFVALWAIARHFLENTGCWDINAN-ASVWWVIRG
GLR_RATÓN SLLSLATFSERSEFFSLYLIGWGAPLLEFVIPWVVVKCLFENVQCWTSNDN-MGFWWILRI
GLR_RATA SLLSITTFSEKSFFSLYLCIGWGSPLLEFVIPWVVVKCLFENVQCWTSNDN-MGFWWILRI
GLR_HUMANO NLLGLATIPERSFFSLYLIGWGAPMLFVVPWAVVKCLFENVQCWTSNDN-MGFWWILRF
GIPR_HUMANO SLLVLVGGSEEGHFRYLLLLGWGAPALFVIPWVIVRYLYENTQCWERNEV-KAIWWIIRT
GIPR_RATA HLLVVRSEKGFHRCYLLLLGWGAPALFVIPWVIVRYLYENTQCWERNEV-KAIWWIIRT
GLP1R_RATÓN TLLAFSVESEQRIFKLYLSIGWGVPLLEFVIPWGIVKLYEDEGCWTRNSN-MNYWLIIRL
GLP1R_RATA TLLAFSVESEQRIFKLYLSIGWGVPLLEFVIPWGIVKLYEDEGCWTRNSN-MNYWLIIRL
GLP1R_HUMANO TLLAFSVESEQWIFRLYVSIGWGVPLLEFVVPWGIVKLYEDEGCWTRNSN-MNYWLIIRL
GLP2R_HUMANO TLLEPTVLPERRLWPKYLVGWAFVLPVFPWGFARAHLENTRCWATNGN-LKIWWIIRG
GLP2R_RATA TAIVLTYSTDRLRKWMFICIGWGVFPPIIVAWAIGKLYYDNEKCWFGKRPGVYTDYIYQG
CRFR1_HUMANO TAIVLTYSTDRLRKWMFVCIGWGVFPPIIVAWAIGKLYYDNEKCWFGKRPGVYTDYIYQG
CRFR1_RATÓN TAIVLTYSTDRLRKWMFVCIGWGVFPPIIVAWAIGKLYYDNEKCWFGKRPGVYTDYIYQG
CRFR1_RATA TAIVLTYSTDRLRKWMFVCIGWGVFPPIIVAWAIGKLYYDNEKCWFGKRPGVYTDYIYQG
CRFR2_HUMANO TAIVMTYSTERLRKCLFLFIGWCIPFPIIVAWAIGKLYYENEQCWFGKEPGDLVDYIYQG
CRFR2_RATA TAIVMTYSTEHLRKWLFLFIGWCIPCPPIIWAIVAGKLYYENEQCWFGKEPGDLVDYIYQG
CRFR2_RATÓN TAIVMTYSTEHLRKWLFLFIGWCIPCPPIIWAIVAGKLYYENEQCWFGKEAGDLVDYIYQG
CALCR_RATÓN TLIVMAVFTEDQRLRWYYLLGWGFPIVPTIIHAITRALYNDNCWLSAET--HLLYIIHG
CALCR_RATA TLIVMAVFTEDQRLRWYYLLGWGFPIVPTIIHAITRAVYNDNCWLSTET--HLLYIIHG
CALCR_HUMANO TLIVVAVFTEKQRLRWYYLLGWGFPLVPTTIIHAITRAVYFNDNCWLSVET--HLLYIIHG
CALRL_RATÓN TLIVVAVFAEKQHLMWYYFLGWGFPLLPACIHAIARSLYYNDNCWISSDT--HLLYIIHG
CALRL_RATA TLIVVAVFAEKQHLMWYYFLGWGFPLLPACIHAIARSLYYNDNCWISSDT--HLLYIIHG
CALRL_HUMANO TLIVVAVFAEKQHLMWYYFLGWGFPLIPACIHAIARSLYYNDNCWISSDT--HLLYIIHG
: ** * : : ** :

```

Figura 7

A.

MNIFEMLRIDEGRLRLKIYKDTGEYTYTIGIGHLLTKSPSLNAAKSELDKAIGRN^TNGVITKDEAEKLFNQD
 VDAAVRGILRNAKLKPVYDSLDAVRR^AALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYN
 QTPNRAKRV^I^TTFRTGTWDAYKNL

B.

MNIFEMLRIDEGRLRLKIYKDTGEYTYTIGIGHLLTKSPSL^{SV}AKSELDKAIGRN^SNGVITKDEAEKLFNQD
 VDAAVRGILRNAKLKPVYDSLDAVRR^SALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYN
 QTPNRAKRV^I^ATFRTGTWDAYKNL