

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 380 765**

51 Int. Cl.:  
**C12N 15/16** (2006.01)  
**C07K 14/605** (2006.01)  
**C12P 21/02** (2006.01)  
**A61K 38/26** (2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **04821092 .6**  
96 Fecha de presentación: **22.11.2004**  
97 Número de publicación de la solicitud: **1704234**  
97 Fecha de publicación de la solicitud: **27.09.2006**

54 Título: **Producción del péptido 2 de tipo glucagón y análogos**

30 Prioridad:  
**21.11.2003 US 523667 P**

45 Fecha de publicación de la mención BOPI:  
**18.05.2012**

45 Fecha de la publicación del folleto de la patente:  
**18.05.2012**

73 Titular/es:  
**NPS PHARMACEUTICALS, INC.**  
**550 HILLS DRIVE, 3RD FLOOR**  
**BEDMINSTER NJ 07921, US**

72 Inventor/es:  
**SASAKI, Ken;**  
**WILLIAMSON, Vanessa Jane;**  
**ARAUJO, Alberto de y**  
**WALCZYK, Ewa**

74 Agente/Representante:  
**de Elzaburu Márquez, Alberto**

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).

ES 2 380 765 T3

## DESCRIPCIÓN

Producción del péptido 2 de tipo glucagón y análogos

**Campo de la Invención**

- 5 Esta invención aplica la técnica de la biología molecular en el campo de la producción de proteínas. Más concretamente, la invención se refiere a la producción del péptido 2 de tipo glucagón recombinante, o GLP-2, y sus análogos.

**Antecedentes de la Invención**

- 10 El GLP-2 es un producto de 33 aminoácidos del gen del proglucagón. La evidencia reciente indica que el GLP-2 promueve la absorción de nutrientes a través de la expansión del epitelio de la mucosa mediante la estimulación de la proliferación de las células de las criptas y la inhibición de la apoptosis en el intestino delgado. El GLP-2 también reduce la permeabilidad epitelial, y disminuye la secreción de ácido gástrico y la motilidad gastrointestinal estimuladas por la comida. Muchos de estos efectos han sido atribuidos no solamente al péptido de tipo salvaje, sino también a sus análogos, incluyendo particularmente aquellos que se vuelven resistentes a la digestión por parte de las enzimas del suero, tales como DPP-IV, mediante la sustitución de la alanina residente en la posición 2 por glicina, por ejemplo. Se describen una variedad de análogos de GLP-2 bioactivos, por ejemplo, en la Patente de los Estados Unidos Núm. 5.789.379.

- 20 Con el reciente reconocimiento de sus propiedades farmacéuticas, existe una demanda de grandes cantidades de GLP-2 y análogos del mismo para permitir el desarrollo y el posterior uso médico de estos productos. Los métodos sintéticos en fase sólida o en fase de solución han sido empleados típicamente para producir las cantidades para investigación de GLP-2 y análogos utilizados hasta la fecha. La producción de GLP-2 como producto recombinante de anfitriones diseñados por ingeniería genética ha sido sugerida, por ejemplo en las Patentes de los Estados Unidos Núms. 5.789.379 y 6.287.806, y se describen en la Patente de los Estados Unidos Núm. 5.629.205. No obstante, los sistemas de producción de la técnica anterior tienen limitaciones en términos de rendimiento y calidad del producto, y sería deseable proporcionar un sistema que produzca péptido GLP-2 de calidad con un coste eficaz.

- 25 El documento WO 99/64611 A1 describe multímeros de péptidos antimicrobianos que son expresados con un compañero de fusión del gen purF. El gen que codifica el péptido se repite y se transcribe en un único ARNm, pero el producto de expresión del gen no está conectado. Por lo tanto, el ARNm es traducido en muchas copias separadas del producto proteico.

- 30 Los documentos WO 03/099848 A2 y WO 03/099854 A2 describen ambos un método de producción de polipéptidos de copia sencilla o de múltiples copias por medio de escisión con clostripaína a partir de una molécula precursora. La escisión específica de la proteína de fusión se logra seleccionando los aminoácidos que flanquean el sitio de escisión, evitando de ese modo la escisión dentro del péptido deseado.

- 35 Lidell et al. (2003) J. Biol. Chem. 278 (16): 13954-13951) describen la escisión autocatalítica de la mucina MUC2 en un enlace Asp-Pro a un pH inferior a 6,0.

El documento WO 03/10022 A2 describe un polipéptido en tándem que consiste en un compañero de fusión de un cuerpo de inclusión y un polipéptido preseleccionado cuyo polipéptido en tándem forma un cuerpo de inclusión que permite la producción y la purificación del polipéptido deseado.

- 40 Por consiguiente un objeto de la presente invención es proporcionar un procedimiento, y los intermedios y reactivos útiles en él, por medio del cual se puedan producir cantidades comerciales de GLP-2.

Otro objeto de la presente invención es proporcionar GLP-2 y análogos del mismo, concretamente el análogo [Gly<sup>2</sup>]hGLP-2, en una forma auténtica estructuralmente.

**Compendio de la Invención**

- 45 De acuerdo con la presente invención, se proporciona un procedimiento por medio del cual se producen GLP-2 y sus análogos no solamente con un rendimiento relativamente elevado, sino también como productos estructuralmente auténticos, que comprenden solamente la forma natural de los aminoácidos de origen natural en la secuencia que constituye el péptido GLP-2. Preferiblemente, los residuos N- y C-terminales son "auténticos terminalmente". En particular, el presente procedimiento produce el GLP-2 deseado en forma de péptido que tiene residuos N- y C-terminales que están sin aminoácidos residuales y otros radicales químicos que a menudo resultan de los métodos recombinantes de la producción de proteínas, concretamente aquellos que dependen de la producción de la proteína como un precursor fusionado a partir del cual debe ser liberada la proteína diana.
- 50

Más concretamente, y de acuerdo con un aspecto de la presente invención, se proporciona un precursor polipeptídico de cadena sencilla en el cual dos o más copias del péptido GLP-2 están acopladas en tándem por medio de un conector que es escindible para liberar cada unidad del péptido GLP-2 en forma de un producto que tiene extremos N y C auténticos. Los péptidos GLP-2 se acoplan utilizando un conector que presenta sitios de escisión en cada uno de sus flancos. El conector presenta un sitio de escisión con ácido en un flanco, y un sitio de escisión con enzima en su otro flanco.

En otro aspecto, la presente invención proporciona un método para producir un péptido GLP-2 que tiene extremos N y C auténticos, en el que el presente multímero del péptido GLP-2 es escindido para liberar cada unidad del péptido GLP-2 residente en él.

En otros aspectos de la presente invención, se proporcionan adicionalmente polinucleótidos útiles en la producción de tales precursores de péptido GLP-2 multimérico.

Tanto la descripción general anterior como la breve descripción de los dibujos y la descripción detallada siguientes son ilustrativas y explicativas y se pretende que proporcionen una explicación adicional de la invención reivindicada. Otros objetos, ventajas, y características novedosas serán fácilmente evidentes para los expertos en la técnica a partir de la siguiente descripción detallada de la invención.

#### **Breve Referencia a los Dibujos**

La Figura 1 ilustra la construcción basada en la PCR de un gen que codifica una unidad de [Gly<sup>2</sup>]hGLP-2 flanqueada por un sitio de escisión con trombina y un sitio de escisión con ácido;

La Figura 2 ilustra la secuencia de ADN esperada del producto de la amplificación de la Figura 1. La secuencia de la unidad de [Gly<sup>2</sup>]hGLP-2 está subrayada;

La Figura 3 es un mapa plasmídico de pKS58 que porta un gen que codifica un hexámero de [Gly<sup>2</sup>]hGLP-2;

La Figura 4 proporciona la secuencia de nucleótidos de pKS58, que porta un constructo que codifica un hexámero de [Gly<sup>2</sup>]hGLP-2, donde la secuencia de aminoácidos también es ilustrada, mostrando las unidades del péptido GLP-2 en negrita; y

La Figura 5 proporciona un análisis espectrométrico de masas de un péptido GLP-2 producido como se describe en la presente memoria.

#### **Descripción Detallada de la Invención**

En un aspecto, la presente invención proporciona un constructo genético, en forma de un polinucleótido, adaptado para producir el péptido GLP-2 en forma de un precursor multimérico, de cadena sencilla que comprende al menos dos copias de un péptido GLP-2. Cada uno de tales péptidos está acoplado al siguiente por medio de un conector que tiene flancos que presentan sitios de escisión que permiten la liberación de los péptidos GLP-2 en forma de monómeros que tienen extremos N y C que son auténticos, y de este modo están esencialmente libres de residuos químicos que se originan a partir del conector o del procedimiento de escisión. En cuanto al producto recombinante, el péptido GLP-2 resultante también está libre de radicales químicos tales como grupos bloqueadores utilizados en la síntesis de péptidos en fase de solución y sólida.

En la presente invención, las unidades del péptido GLP-2 dentro del multímero están acopladas utilizando un conector que presenta sitios de escisión en los extremos N y C de las unidades del péptido GLP-2 residentes. Estos sitios, y los agentes utilizados para escindirlos, se seleccionan de manera que el péptido GLP-2 permanezca intacto durante el procedimiento de escisión, para que así el aislamiento y la purificación produzcan un péptido GLP-2 que tenga los residuos N- y C-terminales deseados sin ningún requerimiento de procesamiento adicional.

En una realización preferida de la presente invención, el conector es una secuencia peptídica relativamente corta, que consiste en no más de alrededor de 25 residuos, deseablemente menos de alrededor de 20 residuos, adecuadamente menos de alrededor de 15 residuos, y muy adecuadamente menos de alrededor de 10 residuos. La secuencia del conector se selecciona para evitar la formación de estructuras secundarias complejas que enmascaran el conector al agente de escisión seleccionado. El sitio de escisión presentado por el conector puede ser un sitio que sea vulnerable a la escisión por enzimas o por condiciones químicas tales como el pH.

El conector es aquél que presenta un sitio de escisión con enzima en un flanco, y un sitio de escisión con ácido en el otro flanco. El sitio sensible a la escisión por la enzima puede ser cualquier sitio que no se reproduzca en ninguna parte en el multímero del péptido GLP-2 y sea escindido por cualquier enzima no presente por lo demás durante el proceso de fabricación. Las enzimas adecuadas para dicha escisión, y las secuencias reconocidas y escindidas por esas enzimas, incluyen la enteroquinasa y la secuencia Asp-Asp-Asp-Asp-Lys, y el Factor Xa y la secuencia Ile-Glu-

Gly-Arg. En una realización preferida, el sitio de escisión con enzima es aquél escindido por la trombina, y la secuencia de escisión de la trombina es ValSerGlyProArg.

Un sitio de escisión con ácido presentado en el multímero del péptido GLP-2 es adecuadamente la secuencia Asp-Pro, que es cortada en condiciones de pH bajo entre los residuos Asp y Pro.

5 En las realizaciones de la presente invención, el conector proporciona, dentro del multímero, un sitio de escisión con ácido en su extremo N y un sitio de escisión con trombina en su extremo C. En una realización específica, el conector tiene la secuencia de aminoácidos ProValSerGlyProArg. Alternativamente, se apreciará que el residuo Pro N-terminal y el sitio de escisión con trombina C-terminal pueden estar separados por una secuencia de aminoácidos adicional que no menoscabe la vulnerabilidad de los flancos a las condiciones de escisión deseadas.

10 Cuando el conector concreto observado es incorporado al multímero, las unidades del péptido GLP-2 son aquellas que incorporan Asp como residuo C-terminal, y que carecen por otra parte tanto de un sitio de escisión con ácido como de un sitio de escisión con trombina. Cuando están conectados entre tales unidades del péptido GLP-2, el residuo Pro N-terminal del conector, junto con el residuo Asp C-terminal de la unidad del péptido GLP-2 aguas arriba, forman el sitio Asp-Pro que es escindible en ácido, esto es, a un pH bajo, para producir el auténtico extremo C del péptido GLP-2. Por otra parte, la secuencia del conector ValSerGlyProArg presenta una secuencia de reconocimiento por trombina que es escindida por trombina en el lado C-terminal de su residuo Arg, para producir un auténtico residuo N-terminal en la unidad del péptido GLP-2 aguas abajo del mismo. Si bien se muestra una secuencia de escisión con trombina específica, se entenderá que se puede incorporar al conector cualquier secuencia equivalente reconocida y escindida por trombina, incluyendo aquellas secuencias referidas por Chang, J. (1985) Eur. J. Biochem. 151, 217-224. Asimismo se apreciará que cualquier unidad del péptido GLP-2 dentro del multímero no debe incorporar ninguna secuencia de escisión con trombina dentro de la estructura primaria de la unidad de GLP-2.

25 De este modo, en un aspecto valioso de la presente invención, se proporciona un polipéptido de cadena sencilla que incorpora al menos dos unidades de péptido GLP-2 acopladas en tándem por medio de un conector que tiene una secuencia ProValSerGlyProArg, donde el péptido GLP-2 incorpora un residuo Asp C-terminal, y carece por otra parte tanto de una secuencia de escisión con trombina como de una secuencia de escisión con ácido.

30 En una realización preferida de este aspecto de la presente invención, la unidad del péptido GLP-2 incorporada dentro del multímero es el análogo de GLP-2 humano en el que la Ala de la posición 2 es sustituida por Gly, esto es, [Gly<sup>2</sup>]hGLP-2, que tiene la secuencia de aminoácidos ilustrada en la Figura 2. En la alternativa, el péptido GLP-2 puede ser el GLP-2 humano de tipo salvaje que tiene la secuencia de aminoácidos referida por Buhl et al. en J. Biol. Chem., 1988, 263(18):8621, un homólogo del mismo, o cualquier otro análogo del mismo que conserve un residuo Asp C-terminal y que carezca por otra parte de sitios de escisión tanto con trombina como con ácido. Se pueden seleccionar análogos adecuados por ejemplo entre aquellos descritos en las Patentes de los Estados Unidos del mismo cesionario Núms. 5.789.379 y 6.184.201.

35 En otras realizaciones, el precursor del péptido GLP-2 multimérico comprende al menos dos unidades del péptido GLP-2, y tantas como 10 o más de tales unidades, p. ej. hasta alrededor de 30 unidades y más adecuadamente hasta alrededor de 20 unidades, conectadas en tándem por medio del conector observado. En realizaciones específicas, el número de unidades de péptido GLP-2 en el precursor es de 2, 3, 4, 5, 6 o 7. En una realización preferida, el precursor multimérico incorpora seis unidades del péptido GLP-2. En otra realización preferida, el precursor incorpora siete unidades de péptido GLP-2.

40 Se apreciará que el multímero del péptido GLP-2, para su expresión como producto recombinante, incorporará una prolongación N-terminal que incorpora al menos un residuo de Metionina inicial. En las realizaciones, la prolongación N-terminal se incorpora en forma de péptido portador que porta el residuo de metionina N-terminal y es escindible del multímero per se. El péptido portador puede ser de este modo una señal de secreción que es escindida por el anfitrión en el proceso de secreción del multímero maduro. Alternativamente y deseablemente, el péptido portador no es una señal de secreción, y el producto multimérico se acumula en el citoplasma del anfitrión donde es recuperado opcionalmente en forma de cuerpos de inclusión. Cuando el péptido portador no está diseñado para ser eliminado de la célula anfitriona, el péptido portador incorpora deseablemente además aminoácidos que constituyen el mismo sitio de escisión con enzima presentado en el multímero en el flanco N-terminal de cada unidad del péptido GLP-2. En esta disposición, el tratamiento del multímero de GLP-2 expresado con la enzima seleccionada no solamente corta el portador del multímero, sino que también corta el multímero en el extremo N de cada unidad del péptido GLP-2 residente allí. En una realización, el péptido portador se inicia con un residuo Met y termina con un sitio de escisión con trombina, tal como ValSerGlyProArg. El péptido portador N-terminal del multímero de GLP-2 puede incorporar además otras secuencias funcionales intermedias, por ejemplo, en la purificación del multímero tal como la denominada Etiqueta His, en el aumento del nivel de expresión del multímero por el anfitrión seleccionado, o en la promoción de la formación del multímero en forma de cuerpos de inclusión tales como secuencias de aminoácidos hidrófobas.

Asimismo se apreciará que el multímero del péptido GLP-2 puede terminar con una unidad del péptido GLP-2 o, si se desea, puede terminar con una prolongación peptídica del mismo útil, por ejemplo, en la purificación del multímero. Si se incorpora un péptido de prolongación C-terminal, se incorpora deseablemente un residuo de Pro en forma de su residuo inicial, de manera que el tratamiento del multímero resultante con ácido escinde no solamente la prolongación C-terminal sino también en el extremo C de cada unidad del péptido GLP-2 dentro del multímero.

En una realización muy preferida de la invención, se proporciona un multímero del péptido GLP-2 que tiene la secuencia ilustrada en la Figura 4, que comprende 6 unidades de  $[\text{Gly}^2]\text{hGLP-2}$  y que incorpora, como conector, la secuencia ProValSerGlyProArg.

La producción de dicho multímero se puede lograr en cualquier anfitrión celular para el cual se han desarrollado sistemas de expresión. El GLP-2 y sus análogos no requieren modificación post-traducciona para su actividad, y de ese modo pueden ser producidos en una variedad de anfitriones bacterianos, así como eucarióticos.

En una realización, el multímero se expresa en células bacterianas, tales como células de *E. coli*, utilizando sistemas de expresión adaptados y bien establecidos para este fin. Se puede incorporar, por ejemplo un polinucleótido que codifica el multímero para la expresión en casetes que dirigen la expresión a partir de promotores tales como lac, tac, trp, T7 y similares. La cepa de *E. coli* seleccionada como anfitrión también puede variar ampliamente, e incluye DH5, JM101 y BL21 entre otras. Los vectores útiles en la transformación del anfitrión seleccionado incluirán típicamente plásmidos que incorporan orígenes de replicación y marcadores seleccionables que permiten la detección y la supervivencia selectiva de los transformantes.

De un modo similar, se pueden explotar una variedad de anfitriones y sistemas de expresión eucarióticos. Estos incluyen *Saccharomyces cerevisiae* y sistemas de expresión basados en el sistema del factor de apareamiento alfa, anfitriones de *Aspergillus nidulans* que utilizan el sistema de la alcohol deshidrogenasa (alcA), o *Aspergillus nidulans* que utiliza el sistema de expresión basado en el gen de la glucoamilasa, así como sistemas de células de mamífero tales como los sistemas de células COS y los sistemas basados en CHO.

Los polinucleótidos que codifican el multímero de GLP-2 pueden ser producidos por supuesto sintéticamente *de novo*, o pueden ser preparados a partir de ADN que codifica la unidad del péptido GLP-2 siguiendo una serie de etapas de amplificación y ligación, todo de acuerdo con la práctica convencional, y como se ilustra en la presente memoria.

Las condiciones de cultivo seleccionadas para el anfitrión celular transformado también dependerán por supuesto de la especie del anfitrión, y del sistema de expresión utilizado. En una realización, cuando el anfitrión es una especie de *E. coli* y el sistema de expresión depende del promotor tac, el transformante será cultivado a escala comercial en presencia de antibiótico para mantener la presión selectiva sobre los transformantes. En o cerca de la fase de crecimiento logarítmico, el cultivo recibirá IPTG para des-reprimir el promotor y permitir que comience la expresión. El cultivo se puede llevar a cabo a escala comercial de 200 litros o de más de alrededor de 200 litros.

Después del cultivo, el multímero de GLP-2 expresado se puede aislar mediante cromatografía de selección por tamaños, mediante cromatografía de intercambio iónico, o mediante cromatografía de afinidad, concretamente en el caso en el que se incorpora una etiqueta de afinidad al multímero. Cuando el multímero se produce en forma de un producto intracelular, las células cultivadas pueden ser tratadas en una primera etapa para lisar las células y liberar el multímero y otros productos intracelulares, por ejemplo utilizando urea 8M o hidroclocluro de guanidina 6M o rotura celular mecánica tal como un homogeneizador o un sonicador. No es necesario separar los contenidos para un tratamiento adicional. En una realización de la invención, los productos de la lisis se tratan *in situ* para establecer las condiciones de la disociación, por ejemplo mediante la adición de cloruro de guanidinio, y después se ajusta el pH de la mezcla con HCl, o un ácido equivalente, para introducir condiciones ácidas, en un intervalo de pH de alrededor de 1 a 3. A este pH, el sitio Asp-Pro es roto en cada interfase entre el extremo C de la unidad de péptido GLP-2 y el extremo N del conector. Los productos de escisión resultantes, incluyendo las unidades de péptido GLP-2 que portan residuos de conector en el extremo N, pueden ser aislados después mediante cualquier método conveniente por ejemplo mediante HPLC, mediante cromatografía de exclusión por tamaños, mediante cromatografía de intercambio iónico, o mediante cromatografía de afinidad. Los productos recuperados pueden ser sometidos después a una etapa de escisión con enzima en la que la exposición a trombina da como resultado la eliminación del conector residual en el extremo N de cada unidad de GLP-2. El resultado es un rendimiento multi-molar de péptidos GLP-2 a partir de un único multímero de GLP-2, teniendo cada péptido GLP-2 extremos N y C que, según se desee, son auténticos y carentes de cualquier modificación química no deseada.

Como se observa en los siguientes ejemplos, la producción mediante este método ha producido  $[\text{Gly}^2]\text{hGLP-2}$  como producto auténtico terminalmente que tiene una masa (3752,59) que es esencialmente equivalente a la teórica (3751,99).

El péptido GLP-2 producido de este modo se formula, en un aspecto de la presente invención, para su uso farmacéutico mediante la formación de una composición farmacéutica en la cual se combina una cantidad

terapéuticamente útil del péptido con un portador farmacéuticamente aceptable. En una realización, la composición se formula para la administración parenteral, y comprende una dosis unitaria de péptido GLP-2 y un vehículo acuoso que es tamponado para que esté dentro de un intervalo de pH y una tonicidad fisiológicamente tolerables, p. ej., pH 4-8, utilizando por ejemplo solución salina tamponada con fosfato como vehículo. La formulación también puede comprender un agente estabilizante, tal como histidina, como se describe en el documento WO 01/49314, o un agente de depósito tal como gelatina como se describe en la Patente de los Estados Unidos Núm. 5.789.379. Las dosis unitarias del péptido GLP-2 se encuentran por lo general dentro del intervalo de 0,1 a 50 mg en un volumen de inyección de alrededor de 1 mL.

Los siguientes ejemplos se proporcionan para ilustrar la presente invención. Se debe entender, sin embargo, que la invención no está limitada a las condiciones específicas o los detalles descritos en estos ejemplos.

### **Ejemplos**

Se pueden elaborar varios constructos multiméricos del gen [Gly<sup>2</sup>]hGLP-2 en una reacción directa sacando provecho de la endonucleasa de restricción, Bsal. Esta endonucleasa reconoce la secuencia no palindrómica (GGTCTC), de manera que el conector y los genes [Gly<sup>2</sup>]hGLP-2 puedan ser ligados solamente en una dirección, ligación cabeza a cola.

Para obtener el nivel máximo de expresión, los constructos del gen multimérico se insertaron en un plásmido bajo el control del promotor del bacteriófago T7. Utilizando esta estrategia, se obtuvieron siete constructos multiméricos, que contenían de 2 a 7 unidades del gen [Gly<sup>2</sup>]hGLP-2 (de dímero a heptámero). Los genes multiméricos se expresaron después de la inducción por IPTG. El nivel de expresión más alto se encontró en los constructos de hexámeros y heptámeros.

Asimismo se desarrolló una lisis celular conveniente y un método de escisión con ácido. Después de la inducción, el sedimento celular se lisó con hidrocloreuro de guanidina 6M y se centrifugó. Se ajustó el pH de la solución de sobrenadante a 1,8 por medio de la adición de HCl. De este modo, se completaron la lisis celular y la escisión con ácido en etapas muy simples sin purificación alguna entre la lisis y la escisión con ácido. Puede ser posible lograr la lisis celular y la escisión con ácido en una única reacción, si el pH del hidrocloreuro de guanidina 6M se ajusta a 1,8 con HCl y después se añade al sedimento celular de *E. coli*.

Los productos escindidos con ácido se purificaron mediante HPLC utilizando una columna C 18 y después se trataron con trombina para obtener [Gly<sup>2</sup>]hGLP-2 maduro, que se purificó adicionalmente mediante HPLC utilizando la misma columna C<sub>18</sub>. Solamente se necesitaron dos etapas de HPLC (primera después de la escisión con ácido y segunda después de la escisión con trombina) para purificar [Gly<sup>2</sup>]hGLP-2, y se confirmó que el [Gly<sup>2</sup>]hGLP-2 purificado era [Gly<sup>2</sup>]hGLP-2 auténtico por medio de espectrometría de masas.

### **Ejemplo 1- Construcción del Gen**

Para construir multímeros del gen [Gly<sup>2</sup>]hGLP-2, un gen [Gly<sup>2</sup>]hGLP-2, como se muestra más abajo, se amplificó primero mediante PCR utilizando un plásmido, pG3M, que portaba un gen [Gly<sup>2</sup>]hGLP-2 de codón optimizado y se cambió el nombre a pEW3G.

Como se muestra en la Figura 1, la secuencia del cebador directo de la PCR (Cebador KS1-5) contenía los sitios de reconocimiento de las endonucleasas NdeI y Bsal, y el sitio de escisión de la trombina, que está seguido de 18 nucleótidos que codifican los seis primeros aminoácidos de [Gly<sup>2</sup>]hGLP-2.

El cebador inverso de la PCR (Cebador KS2-3) contenía los sitios de reconocimiento de las endonucleasas BamHI y Bsal, el sitio de escisión con ácido, y una secuencia de 18 nucleótidos, que codifica los seis últimos residuos de aminoácidos de [Gly<sup>2</sup>]hGLP-2.

Para la reacción PCR, se preparó en primer lugar la mezcla de reacción de PCR de la parte inferior en un tubo de PCR. La mezcla de la parte inferior contenía 41 µL de agua, 5 µL de tampón 10X TsgPlus, 2 µL de una mezcla de desoxinucleótidos (2,5 mM cada uno), 1 µL de cebador KS1-5 (100 µM), y 1 µL de cebador KS2-3 (100 µM). A la mezcla de la parte inferior, se le añadió una porción de Ampliwax® y se calentó a 65°C durante 5 min y a continuación se enfrió a temperatura ambiente en un banco. Después de formar una capa fina de cera, la mezcla de la parte superior contenía 43,5 µL de agua, 5 µL de tampón 10X TsgPlus, 0,2 ng de plásmido, pG3M, en 1 µL, y 0,5 µL de enzima Tsg Plus. La enzima Tsg Plus era una mezcla de ADN polimerasa Tsg y ADN polimerasa Pfu. El tampón 10X Tsg Plus contenía Tris-HCl 200 mM (pH 8,8), KCl 100 mM, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 100 mM, MgSO<sub>4</sub> 20 mM, Triton X-100 al 1%, y seralbúmina bovina de 1 mg/mL.

Las condiciones del termociclador fueron las siguientes:

Etapas 1: 95°C durante 2 min

Etapla 2: 95°C durante 1 min

Etapla 3: 50°C durante 1 min

Etapla 4: 72°C durante 15 seg

Etapla 5: Ir a la Etapla 2 y repetir de la Etapla 2 a la Etapla 4 nueve veces más

5 Etapla 6: 95°C durante 1 min

Etapla 7: 65°C durante 30 seg

Etapla 8: 72°C durante 15 seg

Etapla 9: Ir a la Etapla 6 y repetir de la Etapla 6 a la Etapla 8 diecinueve veces más

Etapla 10: 72°C durante 5 min

10 Etapla 11: 4°C durante la noche

Después del ciclo total de la reacción PCR, como se ha descrito anteriormente, el producto esperado (una banda de ADN de aproximadamente 140 pb), y como se muestra en la Figura 2, se confirmó mediante electroforesis en gel de agarosa al 1,5%.

15 Cuarenta microlitros de la mezcla de reacción de PCR de 100 µL se purificaron utilizando un kit QIA ExII de acuerdo con las instrucciones del fabricante, y a continuación se digirieron con las enzimas de restricción BamHI y NdeI. El ADN digerido se separó por medio de electroforesis en gel de agarosa al 2%, la banda de ADN se separó del gel mediante corte y después se purificó utilizando QIA ExII. El producto de PCR purificado digerido con las dos enzimas y purificado se ligó en pET29a, que fue digerido previamente con las dos mismas enzimas de restricción, NdeI y BamHI. La ligación se realizó utilizando ADN ligasa Quick T4 a temperatura ambiente durante 6 minutos.

20 A continuación, las células competentes de *E. coli* DH5α se transformaron con el producto de ligación. A 50 µL de las células competentes descongeladas en un tubo de microcentrífuga (1,8 mL de capacidad), se les añadieron 3 µL de los 21 µL de la mezcla de ligación. La mezcla celular competente se mantuvo en hielo durante 30 min, se sometió a tratamiento térmico a 37°C durante 20 seg, y a continuación se mantuvo en hielo durante 2 minutos. A las células sometidas a tratamiento térmico, se les añadieron 900 µL de medio Super Optimal Catabolite ("SOC") precalentado (37°C). Después de sacudir la suspensión celular a 225 rpm a 37°C durante 1 hora, se extendieron 50 µL y 200 µL de la suspensión celular sobre placas de agar LB que contenían kanamicina (50 µg/mL) y se incubaron a 37°C durante la noche.

25 Las colonias individuales se aislaron de las placas de agar al día siguiente, y se cultivaron en 7 mL de caldo LB que contenía kanamicina (50 µg/mL) a 37°C a 250 rpm durante la noche. Seis mL de los 7 mL del cultivo se centrifugaron a 3.000 rpm durante 15 min y el plásmido se aisló del sedimento celular utilizando un kit QIAprep Spin Plasmid Miniprep.

30 Para identificar si el plásmido portaba el inserto, el plásmido aislado se digirió mediante una enzima de restricción, PmlI, a 37°C durante 2 horas y a continuación se separó por medio de electroforesis en gel de agarosa al 0,8%. El plásmido también se digirió con la enzima BsaI a 50°C durante 2,5 horas y se analizó sobre gel de agarosa al 1,5%. Como se observa en la Figura 2, el inserto amplificado mediante PCR portaba un solo sitio PmlI y dos sitios BsaI, pero el vector, pET29a, no portaba esos sitios para las enzimas de restricción. Por lo tanto, solo el plásmido, que portaba el inserto, se digirió con PmlI y BsaI.

35 La porción del inserto del plásmido se secuenció a continuación a partir de ambas direcciones utilizando los dos cebadores mostrados más abajo (cebadores Directo e Inverso) para confirmar la secuencia correcta del inserto en el plásmido. Uno de los plásmidos, que portaba el inserto sencillo con una secuencia de nucleótidos correcta se denominó pKS35.

### **Ejemplo 2 – Construcción del Vector**

45 Para construir multímeros del gen [Gly<sup>2</sup>]hGLP-2, se digirió pKS35 con BsaI a 50°C durante 2,5 horas y se separó sobre gel de agarosa al 1,5%. La banda de ADN más grande (la porción del vector) se separó del gel mediante corte y el ADN se extrajo de la porción de gel utilizando el kit de extracción en gel QIA quick. La banda de ADN más pequeña (el inserto, aproximadamente 110 pb) se separó del gel mediante corte y el ADN se extrajo utilizando QIA ExII. La porción grande del vector se trató adicionalmente con fosfatasa alcalina de intestino de ternera (CIP) para minimizar la autoligación del vector y se purificó mediante el kit de purificación QIA PCR.

El ADN vector tratado con CIP y el ADN más pequeño del inserto se mezclaron y se ligaron utilizando ligasa Quick T4. La mezcla de ligación se utilizó para transformar DH5 $\alpha$ , como se ha descrito antes, y a continuación las células bacterianas se cultivaron en placa en placas de agar 2X Extracto de Levadura (2x YE) que contenían kanamicina (30  $\mu$ g/mL).

Para examinar el número de unidades génicas [Gly<sup>2</sup>]hGLP-2 presentes en el plásmido en cada transformante, los insertos se amplificaron directamente a partir de células de *E. coli* lisadas por calor mediante PCR y se examinaron por medio de electroforesis en gel de agarosa. Como se muestra más abajo, el cebador directo utilizado para la PCR (KS003-5) era un oligonucleótido de 20 bases, que se recocía con la región promotora del fago T7 en pET29a. El cebador inverso (KS004-3) era un oligonucleótido de 19 bases, que se unía a la región terminadora de la transcripción de T7 en el plásmido.

Cebador Directo: TAATACGACTCACTATAGGG

Cebador Inverso: GCTAGTTATTGCTCAGCGG

La mezcla de PCR de la parte inferior contenía 42  $\mu$ L de agua, 5  $\mu$ L de tampón 10X Tsg Plus, 2  $\mu$ L de mezcla de desoxinucleótidos (2,5 mM cada uno), 0,5  $\mu$ L de cebador directo KS003-5 100  $\mu$ M, y 0,5  $\mu$ L de cebador inverso KS004-3 100  $\mu$ M en un total de 50  $\mu$ L. Se añadió una porción de Ampliwax a la mezcla de la parte inferior en un tubo de PCR, se calentó a 63°C durante 5 minutos, y a continuación se solidificó a temperatura ambiente. A la parte de arriba de la cera solidificada, se le añadió la mezcla de la parte superior (50  $\mu$ L). La mezcla de la parte superior contenía 44,5  $\mu$ L de agua, 5  $\mu$ L de tampón 10X Tsg Plus, y 0,5  $\mu$ L de enzima Tsg Plus. A continuación, se recogió una sola colonia entre muchos transformantes con un mondadientes estéril de una placa de agar y se suspendió en la mezcla de la parte superior. El tubo de PCR se sometió a continuación a los ciclos de calentamiento de PCR utilizando un termociclador, como se describe más abajo.

Las condiciones del termociclador fueron las siguientes:

Etapas 1: 95°C durante 5 min

Etapas 2: 95°C durante 1 min

Etapas 3: 55°C durante 30 seg

Etapas 4: 72°C durante 1 min

Etapas 5: Ir a la Etapa 2 y repetir de la Etapa 2 a la Etapa 4 veintinueve veces más

Etapas 6: 72°C durante 10 min

Etapas 7: 4°C durante la noche

Los productos de la PCR se separaron mediante electroforesis en gel de agarosa al 1,5% y se detectaron siete tamaños diferentes de productos de la PCR sobre el gel. Mediante comparación con los marcadores de tamaño de ADN (escala de 100 pb), se identificaron como monómero, dímero, trímero, tetrámero, pentámero, hexámero y heptámero. Estos multímeros se sometieron también a análisis de la secuencia de nucleótidos, que demostró que todos tenían las secuencias correctas de los multímeros.

### **Ejemplo 3 - Transformación y Cultivo**

Los constructos del multímero [Gly<sup>2</sup>]hGLP-2 se clonaron en un plásmido pET29a de tal manera que se expresaron bajo el control del promotor del fago T7. La ARN polimerasa de *E. coli* no puede reconocer el promotor de T7. La ARN polimerasa de T7 es requerida para la transcripción a partir del promotor de T7. La cepa de *E. coli*, BLR(DE3), porta un gen de la ARN polimerasa del fago T7 en su cromosoma. Por otra parte, el gen recA en BLR(DE3) es inactivado de manera que la posibilidad de pérdida de las unidades génicas [Gly<sup>2</sup>]hGLP-2 en los constructos de multímero por la recombinación homóloga es mínima en esta cepa. Las dos cepas DH5 $\alpha$  y BLR(DE3) están disponibles en el mercado, como lo está el sistema de T7 utilizado en la presente memoria.

El pET29a que porta un constructo hexamérico de [Gly<sup>2</sup>]hGLP-2 se designó pKS58 y se aisló de las células transformantes utilizando el kit Qiagen Plasmid Midi Prep. Las células congeladas competentes de BLR(DE3) (20  $\mu$ L) se descongelaron, se mezclaron con 1  $\mu$ L de pKS58, se mantuvieron en hielo durante 5 minutos, se sometieron a tratamiento térmico a 42°C durante 30 seg, y a continuación se mantuvieron en hielo durante 2 minutos. A la mezcla celular, se le añadieron 80  $\mu$ L de medio SOC y se incubó a 37°C a 250 rpm durante 1 hora. Se cultivaron en placa porciones de la suspensión celular (20 y 50  $\mu$ L) sobre placas de agar 2x YE que contenían kanamicina (30  $\mu$ g/mL) y se incubaron a 37°C durante la noche.

Para la expresión, una sola colonia de cada una de las placas de transformación de BLR(DE3), que portaba una unidad multimérica del gen [Gly<sup>2</sup>]hGLP-2, se suspendió en 50 mL de caldo 2x YE que contenía kanamicina (30 µg/mL) en un matraz Erlenmeyer de 250 mL y se sacudió a 37°C a 300 rpm durante la noche. Una alícuota (200 µL) del cultivo se añadió a 50 mL de caldo 2x YE precalentado que contenía kanamicina (30 µg/mL) y se sacudió a 37°C a 300 rpm. Al cabo de 2 horas y 10 minutos cuando la D.O. a 600 nm fue aproximadamente 0,35, se añadió IPTG para producir una concentración final 2 mM para inducir el gen multimérico.

Dos y tres horas después de la adición de IPTG, se cosecharon 2 mL de suspensión celular y se introdujeron en una microcentrífuga a 15.000 rpm durante 15 minutos. Los sedimentos celulares se lisaron con 50 µL de tampón de lisis celular a 100°C durante 5 minutos. Una porción del producto lisado celular (12 µL) se mezcló con 3 µL de tampón de carga de SDS-PAGE y las proteínas se separaron mediante SDS-PAGE. Las proteínas sobre los geles se tiñeron con Azul de Coomassie. La expresión de los constructos multiméricos se examinó mediante comparación con los marcadores de peso molecular de la proteína y el perfil proteico de las células no inducidas.

#### **Ejemplo 4 – Procesamiento del Multímero y Aislamiento del Péptido**

Después de la inducción, las células se cosecharon mediante centrifugación y un gramo de los sedimentos celulares frescos se lisaron en 20 mL de hidrocloreto de guanidina 6M. La suspensión celular se incubó sobre hielo durante 1 hora mezclando ocasionalmente y se centrifugó a 12.000 xg durante 30 min. Después de la adición de 30 mL de hidrocloreto de guanidina 6M a la solución de sobrenadante, el pH de la solución de sobrenadante se ajustó a 1,8 añadiendo gotas de HCl 1 N en primer lugar y HCl 0,1 N y a continuación se incubó a 65°C durante 12-14 horas removiendo suavemente. La mezcla de reacción se separó después mediante HPLC utilizando una columna C<sub>18</sub> y la elución mediante un gradiente de acetonitrilo de 30 a 60% en ácido trifluoroacético al 0,1%.

El pico del producto escindido con ácido ([Gly<sup>2</sup>]hGLP-2 con un conector peptídico corto) se recogió y se secó. A continuación, el material seco se disolvió en tampón de trombina (Tris 20 mM, NaCl 150 mM, CaCl<sub>2</sub> 2,5 mM, pH 8,4) y se trató con trombina a 37°C durante la noche y la reacción se detuvo a continuación por medio de la adición de ACN al 20% hasta el volumen final. El producto de la digestión se purificó después mediante HPLC utilizando las mismas condiciones descritas anteriormente.

El producto de la digestión se sometió a continuación a análisis mediante espectrometría de masas, utilizando un espectrómetro de masas Micromass Quattro Micro™ equipado con una fuente Z-spray funcionando en el modo ión positivo con los siguientes parámetros: Rango de Datos: m/z 400-1600; Voltaje del cono: 30-35 V; Temperatura de la Fuente: 80°C; Temperatura de Desolvatación: 200°C; La Inyección de flujo fue a través de HP1100; Disolvente: Acetonitrilo 50:50: Agua + ácido fórmico al 0,1%; Soporte Lógico: Los datos se adquirieron utilizando MassLynx 4.0. La calibración se realizó utilizando un espectro MS de mioglobina e histatina 5.

Como se observa en la Figura 5, la masa del pico predominante, que representa [Gly<sup>2</sup>]hGLP-2 auténtico, recombinante (producido genéticamente) tiene una masa que es 3752,59, que es esencialmente la misma que la masa teórica de 3751,99.

La recuperación del monómero GLP-2 a partir del multímero también se puede lograr convenientemente en una reacción "directa" utilizando el multímero como reactivo y proporcionando el monómero auténtico, maduro como producto final, sin requerir la numerosa separación de productos intermedios y etapas de transferencia.

Con referencia al ejemplo proporcionado anteriormente, el proceso directo elimina la etapa de lisis celular mediante HCl-guanidina 6M, y en primer lugar interrumpe mecánicamente la expresión de las células anfitrionas utilizando por ejemplo un homogenizador, o un sonicador. Después de la rotura celular, el pH de la suspensión se redujo, por ejemplo a pH 1-3, mediante adición por ejemplo de HCl. Como se ha descrito anteriormente, la suspensión se incubó a continuación a una temperatura apropiada, tal como 40-80°C p. ej., 65°C, para completar la escisión con ácido del multímero para producir los intermedios monoméricos que portan los conectores peptídicos N-terminales. El pH de la mezcla de reacción se eleva a continuación, por ejemplo utilizando Tris-HCl, al intervalo de pH adecuado para la actividad de la trombina p. ej., 7,5-9,0 y preferiblemente alrededor de 8,4. Después se añade la trombina para ocasionar la escisión de los conectores peptídicos N-terminales, generándose de ese modo el producto GLP-2 maduro que porta los extremos auténticos.

## REIVINDICACIONES

- 5 1. Un multímero proteico de cadena sencilla que comprende al menos dos unidades del péptido GLP-2 acopladas en tándem por un conector que proporciona en el multímero un sitio de escisión con ácido en su extremo N, y un sitio de escisión con enzima en su extremo C, donde la escisión de dicho multímero con un ácido y una enzima libera dichas unidades del péptido GLP-2, teniendo cada una residuos N y C terminales auténticos y donde el sitio de escisión con ácido tiene la secuencia Asp-Pro y el sitio de escisión con enzima se selecciona entre el sitio de escisión con enteroquinasa que tiene la secuencia Asp-Asp-Asp-Asp-Lys, un sitio de escisión con Factor Xa que tiene la secuencia Ile-Glu-Gly-Arg y un sitio de escisión con trombina que tiene la secuencia Val-Ser-Gly-Pro-Arg.
- 10 2. Un multímero proteico de cadena sencilla de acuerdo con la reivindicación 1, donde dicho conector consiste en no más de 25 residuos, menos de 20 residuos, menos de 15 residuos, o menos de 10 residuos.
3. Un multímero proteico de cadena sencilla de acuerdo con la reivindicación 1 o 2, donde dicho conector tiene la secuencia ProValSerGlyProArg.
4. Un multímero proteico de cadena sencilla de acuerdo con una cualquiera de las reivindicaciones 1 a 3, donde dicho péptido GLP-2 es [Gly<sup>2</sup>]hGLP-2.
- 15 5. Un multímero proteico de cadena sencilla de acuerdo con una cualquiera de las reivindicaciones 1 a 4, que comprende adicionalmente un péptido portador acoplado liberablemente al extremo N del mismo, proporcionando el péptido portador un residuo de Met en su extremo N y, en su extremo C, un sitio escindible por dicha enzima.
6. Un multímero proteico de cadena sencilla de acuerdo con una cualquiera de las reivindicaciones 1 a 5, que comprende adicionalmente un péptido de prolongación C-terminal.
- 20 7. Un procedimiento para preparar un multímero de acuerdo con una cualquiera de las reivindicaciones 1 a 6, que comprende la etapa de cultivar un anfitrión celular que incorpora un constructo de expresión en el que una molécula de ADN que codifica dicho multímero se conecta operablemente con ADN proporcionando la expresión del mismo.
8. El procedimiento de acuerdo con la reivindicación 7, donde dicho anfitrión es un anfitrión *E. coli*.
- 25 9. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 7 u 8, donde el constructo de expresión comprende adicionalmente el promotor del bacteriófago T7.
10. Un polinucleótido que codifica el multímero de acuerdo con una cualquiera de las reivindicaciones 1-6.
11. Un procedimiento para preparar un péptido GLP-2, que comprende las etapas de obtener un multímero de acuerdo con una cualquiera de las reivindicaciones 1 a 6, tratar el multímero con ácido y con enzima para escindir el conector residente allí, y a continuación aislar las unidades de péptido GLP-2 resultantes.
- 30 12. El procedimiento de acuerdo con la reivindicación 11, donde el multímero se escinde en primer lugar con ácido, el multímero escindido resultante es aislado y después escindido con enzima, y a continuación se aíslan las unidades de péptido GLP-2 resultantes que tienen extremos auténticos.
13. El procedimiento de acuerdo con la reivindicación 12, donde la etapa de escisión con ácido se realiza en el momento de extraer el multímero del péptido GLP-2 del anfitrión celular que produce dicho multímero.
- 35 14. El procedimiento de acuerdo con la reivindicación 11, donde la etapa de tratamiento del multímero con ácido y con enzima se realiza sin separación de los productos de reacción antes del tratamiento con enzima.

**Cebador KS1-5**

NdeI  
 ~~~~~  
 BsaI  
 ~~~~~  
 MetValSerGly ProArg**His** GlyAspGly  
 1 GGAATTCCAT ATGGTCTCAG GTCCGCGTCA TGGTGACGGT  
  
 SerPhe  
 41 TCTTTC

Sitio Trombina



**Cebador KS2-3**

BamHI  
 ~~~~~  
 BsaI  
 ~~~~~  
 1 CGCGGATCCT CATTAGCGCG GACCAGAGAC CGGGTCGGTG  
 ArgPro GlySerVal Pro**AspThrIle**  
 41 ATTTTGGTCT G  
 ..LysThrGln

Sitio Escisión Ácido



**Figura 1**

```

      NdeI
      ~~~~~~
      BsaI
      ~~~~~~
      MetValSerGly ProArgHis GlyAspGly
1  GGAATTCCAT ATGGTCTCAG GTCCGCGTCA TGGTGACGGT
   CCTTAAGGTA TACCAGAGTC CAGGCGCAGT ACCACTGCCA

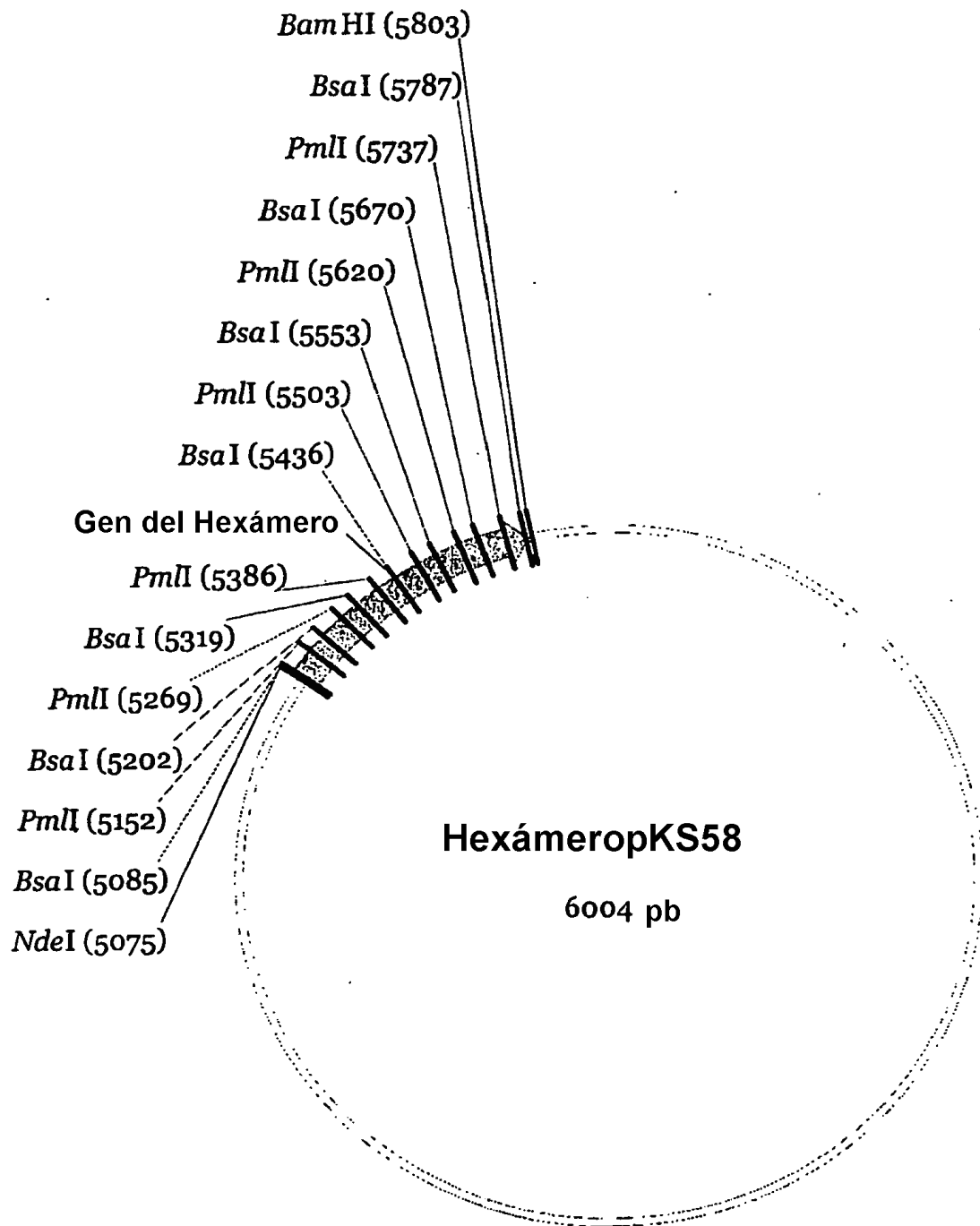
      SerPheSerAsp GluMetAsn ThrIleLeu AspAsnLeuAla
41 TCTTTCTCTG ACGAAATGAA CACCATCCTG GACAACCTGG
   AGAAAGAGAC TGCTTTACTT GTGGTAGGAC CTGTTGGACC

      PmlI
      ~~~~~~
      AAAlaArgAsp PheIleAsn TrpLeuIleGln ThrLysIle
81 CTGCACGTGA CTTCATCAAC TGGCTGATCC AGACCAAAT
   GACGTGCACT GAAGTAGTTG ACCGACTAGG TCTGGTTTTA

      BsaI                               BamHI
      ~~~~~~                               ~~~~~~
      ThrAspPro ValSerGlyPro Arg
121 CACCGACCCG GTCTCAGGTC CGCGCTAATG AGGATCCGCG
    GTGGCTGGGC CAGAGTCCAG GCGCGATTAC TCCTAGGCGC

```

Figura 2



**Figura 3**

**Figura 4**

```

1  TGGCGAATGG GACGCGCCCT GTAGCGGCGC ATTAAGCGCG
   ACCGCTTACC CTGCGCGGGA CATCGCCGCG TAATTCGCGC
41  GCGGGTGTGG TGGTTACGCG CAGCGTGACC GCTACACTTG
   CGCCACACACC ACCAATGCGC GTCGCACTGG CGATGTGAAC
81  CCAGCGCCCT AGCGCCCGCT CCTTTCGCTT TCTTCCCTTC
   GGTGCGGGGA TCGCGGGCGA GGAAAGCGAA AGAAGGGAAG
121 CTTTCTCGCC ACGTTCGCCG GCTTTCCCCG TCAAGCTCTA
   GAAAGAGCGG TGCAAGCGGC CGAAAGGGGC AGTTCGAGAT
161 AATCGGGGGC TCCCTTTAGG GTTCCGATTT AGTGCTTTAC
   TTAGCCCCCG AGGGAAATCC CAAGGCTAAA TCACGAAATG
201 GGCACCTCGA CCCCAAAAAA CTTGATTAGG GTGATGGTTC
   CCGTGGAGCT GGGGTTTTTT GAACTAATCC CACTACCAAG
241 ACGTAGTGGG CCATCGCCCT GATAGACGGT TTTTCGCCCT
   TGCATCACCC GGTAGCGGGA CTATCTGCCA AAAAGCGGGA
281 TTGACGTTGG AGTCCACGTT CTTTAATAGT GGACTCTTGT
   AACTGCAACC TCAGGTGCAA GAAATTATCA CCTGAGAACA
321 TCCAAACTGG AACAACACTC AACCTATCT CGGTCTATTC
   AGGTTTGACC TTGTTGTGAG TTGGGATAGA GCCAGATAAG
361 TTTTGATTTA TAAGGGATTT TGCCGATTTT GGCCTATTGG
   AAAACTAAAT ATTCCTAAA ACGGCTAAAG CCGGATAACC
401 TTAAAAAATG AGCTGATTTA ACAAAAATTT AACCGGAATT
   AATTTTTTAC TCGACTAAAT TGTTTTTAAA TTGCGCTTAA
441 TTAACAAAAT ATTAACGTTT ACAATTTTCA GTGGCACTTT
   AATTGTTTTA TAATTGCAA TGTTAAAGTC CACCGTGAAA
481 TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTT
   AGCCCCTTTA CACGCGCCTT GGGGATAAAC AAATAAAAAG
521 TAAATACATT CAAATATGTA TCCGCTCATG AATTAATTCT
   ATTTATGTAA GTTTATACAT AGGCGAGTAC TTAATTAAGA
561 TAGAAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT
   ATCTTTTTGA GTAGCTCGTA GTTTACTTTG ACGTTAAATA
601 TCATATCAGG ATTATCAATA CCATATTTTT GAAAAAGCCG
   AGTATAGTCC TAATAGTTAT GGTATAAAAA CTTTTTCGGC
641 TTTCTGTAAT GAAGGAGAAA ACTCACCGAG GCAGTTCCAT
   AAAGACATTA CTCCTCTTT TGAGTGGCTC CGTCAAGGTA
681 AGGATGGCAA GATCCTGGTA TCGGTCTGCG ATTCCGACTC
   TCCTACCGTT CTAGGACCAT AGCCAGACGC TAAGGCTGAG
721 GTCCAACATC AATACAACCT ATTAATTTCC CCTCGTCAAA
   CAGGTTGTAG TTATGTTGGA TAATTAAAGG GGAGCAGTTT
761 AATAAGGTTA TCAAGTGAGA AATCACCATG AGTGACGACT
   TTATTCCAAT AGTTCACTCT TTAGTGGTAC TCACTGCTGA
801 GAATCCGGTG AGAATGGCAA AAGTTTATGC ATTTCTTTCC
   CTTAGGCCAC TCTTACCGTT TTCAAATACG TAAAGAAAGG
841 AGACTTGTTT AACAGGCCAG CCATTACGCT CGTCATCAAA

```

```

      TCTGAACAAG TTGTCCGGTC GGTAAATGCGA GCAGTAGTTT
881 ATCACTCGCA TCAACCAAAC CGTTATTCAT TCGTGATTGC
      TAGTGAGCGT AGTTGGTTTG GCAATAAGTA AGCACTAACG
921 GCCTGAGCGA GACGAAATAC GCGATCGCTG TAAAAGGAC
      CGGACTCGCT CTGCTTTATG CGCTAGCGAC AATTTTCCTG
961 AATTACAAAC AGGAATCGAA TGCAACCGGC GCAGGAACAC
      TTAATGTTTG TCCTTAGCTT ACGTTGGCCG CGTCCTTGTC
1001 TGCCAGCGCA TCAACAATAT TTTCACCTGA ATCAGGATAT
      ACGGTCGCGT AGTTGTTATA AAAGTGGACT TAGTCCTATA
1041 TCTTCTAATA CCTGGAATGC TGTTTTCCCG GGGATCGCAG
      AGAAGATTAT GGACCTTACG ACAAAGGGC CCCTAGCGTC
1081 TGGTGAGTAA CCATGCATCA TCAGGAGTAC GGATAAAATG
      ACCACTCATT GGTACGTAGT AGTCCTCATG CCTATTTTAC
1121 CTTGATGGTC GGAAGAGGCA TAAATTCCGT CAGCCAGTTT
      GAACTACCAG CCTTCTCCGT ATTTAAGGCA GTCGGTCAAA
1161 AGTCTGACCA TCTCATCTGT AACATCATTG GCAACGCTAC
      TCAGACTGGT AGAGTAGACA TTGTAGTAAC CGTTGCGATG
1201 CTTTGCCATG TTTCAGAAAC AACTCTGGCG CATCGGGCTT
      GAAACGGTAC AAAGTCTTTG TTGAGACCGC GTAGCCCGAA
1241 CCCATACAAT CGATAGATTG TCGCACCTGA TTGCCCCGACA
      GGGTATGTTA GCTATCTAAC AGCGTGGACT AACGGGCTGT
1281 TTATCGCGAG CCCATTTATA CCCATATAAA TCAGCATCCA
      AATAGCGCTC GGGTAAATAT GGGTATATTT AGTCGTAGGT
1321 TGTTGGAATT TAATCGCGGC CTAGAGCAAG ACGTTTCCCG
      ACAACCTTAA ATTAGCGCCG GATCTCGTTC TGCAAAGGGC
1361 TTGAATATGG CTCATAACAC CCCTTGATT ACTGTTTATG
      AACTTATACC GAGTATTGTG GGGAACATAA TGACAAATAC
1401 TAAGCAGACA GTTTTATTGT TCATGACCAA AATCCCTTAA
      ATTCGTCTGT CAAAATAACA AGTACTGGTT TTAGGGAATT
1441 CGTGAGTTTT CGTTCCACTG AGCGTCAGAC CCCGTAGAAA
      GCACTCAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT
1481 AGATCAAAGG ATCTTCTTGA GATCCTTTTT TTCTGCGCGT
      TCTAGTTTCC TAGAAGAACT CTAGGAAAAA AAGACGCGCA
1521 AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG
      TTAGACGACG AACGTTTGTT TTTTGGTGG CGATGGTCGC
1561 GTGGTTTGTT TGCCGGATCA AGAGCTACCA ACTCTTTTTC
      CACCAAACAA ACGGCCTAGT TCTCGATGGT TGAGAAAAAG
1601 CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA TACCAAATAC
      GCTTCCATTG ACCGAAGTCG TCTCGCGTCT ATGGTTTATG
1641 TGTCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG
      ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC
1681 AACTCTGTAG CACCGCCTAC ATACCTCGCT CTGCTAATCC
      TTGAGACATC GTGGCGGATG TATGGAGCGA GACGATTAGG
1721 TGTTACCAGT GGCTGCTGCC AGTGGCGATA AGTCGTGTCT
      ACAATGGTCA CCGACGACGG TCACCGCTAT TCAGACAGAG
1761 TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG
      ATGGCCCAAC CTGAGTTCTG CTATCAATGG CCTATTCCGC

```

```

1801 CAGCGGTCGG GCTGAACGGG GGGTTCGTGC ACACAGCCCA
      GTCGCCAGCC CGACTTGCCC CCCAAGCACG TGTGTCGGGT
1841 GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA
      CGAACCTCGC TTGCTGGATG TGGCTTGACT CTATGGATGT
1881 GCGTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA
      CGCACTCGAT ACTCTTTCGC GGTGCGAAGG GCTTCCCTCT
1921 AAGGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG
      TTCCGCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGTC
1961 GAGAGCGCAC GAGGGAGCTT CCAGGGGGAA ACGCCTGGTA
      CTCTCGCGTG CTCCTCGAA GGTCCCCCTT TGCGGACCAT
2001 TCTTTATAGT CCTGTCGGGT TTCGCCACCT CTGACTTGAG
      AGAAATATCA GGACAGCCCA AAGCGGTGGA GACTGAACTC
2041 CGTCGATTTT TGTGATGCTC GTCAGGGGGG CGGAGCCTAT
      GCAGCTAAAA ACACTACGAG CAGTCCCCCC GCCTCGGATA
2081 GGAAAAACGC CAGCAACGCG GCCTTTTAC GGTTCCTGGC
      CCTTTTTCGC GTCGTTGCGC CGGAAAAATG CCAAGGACCG
2121 CTTTTGCTGG CTTTTGCTC ACATGTTCTT TCCTGCGTTA
      GAAAACGACC GGAAAACGAG TGTACAAGAA AGGACGCAAT
2161 TCCCCTGATT CTGTGGATAA CCGTATTACC GCCTTTGAGT
      AGGGGACTAA GACACCTATT GGCATAATGG CGGAAACTCA
2201 GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG
      CTCGACTATG GCGAGCGGCG TCGGCTTGCT GGCTCGCGTC
2241 CGAGTCAGTG AGCGAGGAAG CGGAAGAGCG CCTGATGCGG
      GCTCAGTCAC TCGCTCCTTC GCCTTCTCGC GGACTACGCC
2281 TATTTTCTCC TTACGCATCT GTGCGGTATT TCACACCGCA
      ATAAAAGAGG AATGCGTAGA CACGCCATAA AGTGTGGCGT
2321 TATATGGTGC ACTCTCAGTA CAATCTGCTC TGATGCCGCA
      ATATAACACG TGAGAGTCAT GTTAGACGAG ACTACGGCGT
2361 TAGTTAAGCC AGTATACACT CCGCTATCGC TACGTGACTG
      ATCAATTCGG TCATATGTGA GGCATAGCG ATGCACTGAC
2401 GGTGATGGCT GCGCCCCGAC ACCCGCCAAC ACCCGCTGAC
      CCAGTACCGA CGCGGGGCTG TGGGCGGTTG TGGGCGACTG
2441 GCGCCCTGAC GGGCTTGTCT GCTCCCGGCA TCCGCTTACA
      CGCGGGGACTG CCCGAACAGA CGAGGGCCGT AGGCGAATGT
2481 GACAAGCTGT GACCGTCTCC GGGAGCTGCA TGTGTCAGAG
      CTGTTTCGACA CTGGCAGAGG CCCTCGACGT ACACAGTCTC
2521 GTTTTACCG TCATCACCGA AACGCGCGAG GCAGCTGCGG
      CAAAAGTGGC AGTAGTGGCT TTGCGCGCTC CGTCGACGCC
2561 TAAAGCTCAT CAGCGTGGTC GTGAAGCGAT TCACAGATGT
      ATTTTCGAGTA GTCGCACCAG CACTTCGCTA AGTGTCTACA
2601 CTGCCTGTTT ATCCGCGTCC AGCTCGTTGA GTTTCTCCAG
      GACGGACAAG TAGGCGCAGG TCGAGCAACT CAAAGAGGTC
2641 AAGCGTTAAT GTCTGGCTTC TGATAAAGCG GGCCATGTTA
      TTCGCAATTA CAGACCGAAG ACTATTTCGC CCGGTACAAT
2681 AGGGCGGTTT TTTCTGTTT GGTCAGTGAT GCCTCCGTGT
      TCCCGCCAAA AAAGGACAAA CCAGTGACTA CGGAGGCACA
2721 AAGGGGGATT TCTGTTTCATG GGGGTAATGA TACCGATGAA

```

```

TTCCCCCTAA AGACAAGTAC CCCCATTACT ATGGCTACTT
2761 ACGAGAGAGG ATGCTCACGA TACGGGTAC TGATGATGAA
TGCTCTCTCC TACGAGTGCT ATGCCCAATG ACTACTACTT
2801 CATGCCCCGGT TACTGGAACG TTGTGAGGGT AAACAACCTGG
GTACGGGCCA ATGACCTTGC AACACTCCCA TTTGTTGACC
2841 CGGTATGGAT GCGGCGGGAC CAGAGAAAAA TCACTCAGGG
GCCATACCTA CGCCGCCCTG GTCTCTTTT AGTGAGTCCC
2881 TCAATGCCAG CGCTTCGTTA ATACAGATGT AGGTGTTCCA
AGTTACGGTC GCGAAGCAAT TATGTCTACA TCCACAAGGT
2921 CAGGGTAGCC AGCAGCATCC TGCATGCAG ATCCGGAACA
GTCCCATCGG TCGTCGTAGG ACGCTACGTC TAGGCCTTGT
2961 TAATGGTGCA GGGCGCTGAC TTCCGCGTTT CCAGACTTTA
ATTACCACGT CCCGCGACTG AAGGCGCAAA GGTCTGAAAT
3001 CGAAACACGG AAACCGAAGA CCATTCATGT TGTGCTCAG
GCTTTGTGCC TTTGGCTTCT GGTAAGTACA ACAACGAGTC
3041 GTCGCAGACG TTTTGCAGCA GCAGTCGCTT CACGTTGCT
CAGCGTCTGC AAAACGTCGT CGTCAGCGAA GTGCAAGCGA
3081 CGCGTATCGG TGATTCATTC TGCTAACCAG TAAGGCAACC
GCGCATAGCC ACTAAGTAAG ACGATTGGTC ATTCCGTTGG
3121 CCGCCAGCCT AGCCGGGTCC TCAACGACAG GAGCACGATC
GGCGGTCGGA TCGGCCCAGG AGTTGCTGTC CTCGTGCTAG
3161 ATGCGCACCC GTGGGGCCGC CATGCCGCG ATAATGGCCT
TACGCGTGGG CACCCCGCG GTACGGCCGC TATTACCGGA
3201 GCTTCTCGCC GAAACGTTTG GTGGCGGGAC CAGTGACGAA
CGAAGAGCGG CTTTGCAAAC CACCGCCCTG GTCACTGCTT
3241 GGCTTGAGCG AGGGCGTGCA AGATTCCGAA TACCGCAAGC
CCGAACTCGC TCCCGCACGT TCTAAGGCTT ATGGCGTTCCG
3281 GACAGGCCGA TCATCGTCGC GCTCCAGCGA AAGCGGTCCCT
CTGTCCGGCT AGTAGCAGCG CGAGGTCGCT TTCGCCAGGA
3321 CGCCGAAAAT GACCCAGAGC GCTGCCGGCA CCTGTCCTAC
GCGGCTTTTA CTGGGTCTCG CGACGGCCGT GGACAGGATG
3361 GAGTTGCATG ATAAAGAAGA CAGTCATAAG TGCGGCGACG
CTCAACGTAC TATTTCTTCT GTCAGTATTC ACGCCGCTGC
3401 ATAGTCATGC CCCGCGCCCA CCGGAAGGAG CTGACTGGGT
TATCAGTACG GGGCGCGGGT GGCCTTCCTC GACTGACCCA
3441 TGAAGGCTCT CAAGGGCATC GGTGAGATC CCGGTGCCTA
ACTTCCGAGA GTTCCCGTAG CCAGCTCTAG GGCCACGGAT
3481 ATGAGTGAGC TAACTTACAT TAATTGCGTT GCGCTCACTG
TACTCACTCG ATTGAATGTA ATTAACGCAA CGCGAGTGAC
3521 CCCGCTTTCC AGTCGGGAAA CCTGTCGTGC CAGCTGCATT
GGCGGAAAGG TCAGCCCTTT GGACAGCACG GTCGACGTAA
3561 AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT
TTACTTAGCC GGTTGCGCGC CCCTCTCCGC CAAACGCATA
3601 TGGGCGCCAG GGTGGTTTTT CTTTTACCA GTGAGACGGG
ACCCGCGGTC CCACCAAAAA GAAAAGTGGT CACTCTGCCC
3641 CAACAGCTGA TTGCCCTTCA CCGCCTGGCC CTGAGAGAGT
GTTGTCGACT AACGGGAAGT GGCGGACCGG GACTCTCTCA

```

```

3681 TGCAGCAAGC GGTCCACGCT GGTTTGCCCC AGCAGGCGAA
      ACGTCGTTCG CCAGGTGCGA CCAAACGGGG TCGTCCGCTT
3721 AATCCTGTTT GATGGTGgTT AACGGCGGGA TATAACATGA
      TTAGGACAAA CTACCACCAA TTGCCGCCCT ATATTGTACT
3761 GCTGTCTTCG GTATCGTCGT ATCCCACTAC CGAGATGTCC
      CGACAGAAGC CATAGCAGCA TAGGGTGATG GCTCTACAGG
3801 GCACCAACGC GCAGCCCGGA CTCGGTAATG GCGCGCATTG
      CGTGGTTGCG CGTCGGGCCT GAGCCATTAC CGCGCGTAAC
3841 CGCCCAGCGC CATCTGATCG TTGGCAACCA GCATCGCAGT
      GCGGGTCGCG GTAGACTAGC AACCGTTGGT CGTAGCGTCA
3881 GGAACGATG CCCTCATTCa GCATTTGCAT GGTTTGTTGA
      CCCTTGCTAC GGGAGTAAGT CGTAAACGTA CCAAACAACt
3921 AAACCGGACA TGGCACTCCA GTCGCCTTCC CGTTCGCTA
      TTTGGCCTGT ACCGTGAGGT CAGCGGAAGG GCAAGGCGAT
3961 TCGGCTGAAT TTGATTGCGA GTGAGATATT TATGCCAGCC
      AGCCGACTTA AACTAACGCT CACTCTATAA ATACGGTCGG
4001 AGCCAGACGC AGACGCGCCG AGACAGAAct TAATGGGCCC
      TCGGTCTGCG TCTGCGCGGC TCTGTCTTGA ATTACCCGGG
4041 GCTAACAGCG CGATTTGCTG GTGACCCAAT GCGACCAGAT
      CGATTGTGCG GCTAAACGAC CACTGGGTTA CGCTGGTCTA
4081 GCTCCACGCC CAGTCGCGTA CCGTCTTCAT GGGAGAAAAT
      CGAGGTGCGG GTCAGCGCAT GGCAGAAGTA CCCTCTTTTA
4121 AATACTGTTG ATGGGTGTCT GGTCAGAGAC ATCAAGAAAT
      TTATGACAAC TACCCACAGA CCAGTCTCTG TAGTTCTTTA
4161 AACGCCGGAA CATTAGTGCA GGCAGCTTCC ACAGCAATGG
      TTGCGGCCTT GTAATCACGT CCGTCGAAGG TGTCGTTACC
4201 CATCCTGGTC ATCCAGCGGA TAGTTAATGA TCAGCCCACT
      GTAGGACCAG TAGGTCGCCT ATCAATTACT AGTCGGGTGA
4241 GACGCGTTGC GCGAGAAGAT TGTGCACCGC CGCTTTACAG
      CTGCGCAACG CGCTCTTCTA ACACGTGGCG GCGAAATGTC
4281 GCTTCGACGC CGCTTCGTTT TACCATCGAC ACCACCACGC
      CGAAGCTGCG GCGAAGCAAG ATGGTAGCTG TGGTGGTGCG
4321 TGGCACCCAG TTGATCGGCG CGAGATTTAA TCGCCGCGAC
      ACCGTGGGTC AACTAGCCGC GCTCTAAATT AGCGGCGCTG
4361 AATTTGCGAC GCGCGGTGCA GGGCCAGACT GGAGGTGGCA
      TTAAACGCTG CCGCGCACGT CCCGGTCTGA CCTCCACCGT
4401 ACGCCAATCA GCAACGACTG TTTGCCCGCC AGTTGTTGTG
      TGCGGTTAGT CGTTGCTGAC AAACGGGCGG TCAACAACAC
4441 CCACGCGGTT GGAATGTAA TTCAGCTCCG CCATCGCCGC
      GGTGCGCCAA CCCTTACATT AAGTCGAGGC GGTAGCGGCG
4481 TTCCACTTTT TCCCGCGTTT TCGCAGAAAC GTGGCTGGCC
      AAGGTGAAAA AGGGCGCAAa AGCGTCTTTG CACCGACCGG
4521 TGTTTACCA CGCGGGAAAC GGTCTGATAA GAGACACCGG
      ACCAAGTGGT GCGCCCTTTG CCAGACTATT CTCTGTGGCC
4561 CATACTCTGC GACATCGTAT AACGTTACTG GTTTCACATT
      GTATGAGACG CTGTAGCATA TTGCAATGAC CAAAGTGTAA
4601 CACCACCCTG AATTGACTCT CTTCCGGGCG CTATCATGCC

```

```

      GTGGTGGGAC TTAAGTGAAGA GAAGGCCCGC GATAGTACGG
4641 ATACCGCGAA AGGTTTTGCG CCATTCGATG GTGTCCGGGA
      TATGGCGCTT TCCAAAACGC GGTAAGCTAC CACAGGCCCT
4681 TCTCGACGCT CTCCCTTATG CCACTCCTGC ATTAGGAAGC
      AGAGCTGCGA GAGGGAATAC GCTGAGGACG TAATCCTTCG
4721 AGCCAGTAG TAGGTTGAGG CCGTTGAGCA CCGCCGCCGC
      TCGGGTCATC ATCCAACCTC GGCAACTCGT GGCAGCGGCG
4761 AAGGAATGGT GCATGCAAGG AGATGGCGCC CAACAGTCCC
      TTCCTTACCA CGTACGTTCC TCTACCGCGG GTTGTGAGGG
4801 CCGGCCACGG GGCCTGCCAC CATACCCACG CCGAAACAAG
      GGCCGGTGCC CCGGACGGTG GTATGGGTGC GGCTTTGTTC
4841 CGCTCATGAG CCCGAAGTGG CGAGCCCGAT CTTCCCCATC
      GCGAGTACTC GGGCTTCACC GTCGGGGCTA GAAGGGGTAG
4881 GGTGATGTCG GCGATATAGG CGCCAGCAAC CGCACCTGTG
      CCACTACAGC CGCTATATCC GCGGTCGTTG GCGTGGACAC
4921 GCGCCGGTGA TGCCGGCCAC GATGCGTCCG GCGTAGAGGA
      CGCGGCCACT ACGGCGGGTG CTACGCAGGC CGCATCTCCT
4961 TCGAGATCGA TCTCGATCCC GCGAAATTAA TACGACTCAC
      AGCTCTAGCT AGAGCTAGGG CGCTTTAATT ATGCTAGATG
5001 TATAGGGGAA TTGTGAGCGG ATAACAATTC CCCTCTAGAA
      ATATCCCCTT AACACTCGCC TATTGTTAAG GGGAGATCTT
                                     NdeI
                                     ~~~~~~
                                     BsaI
                                     ~~~
                                     MetVal.
5041 ATAATTTTGT TTAAGTTTAA GAAGGAGATA TACATATGGT
      TATTAAACA AATTGAAATT CTCCTCTAT ATGTATACCA
      BsaI
      ~~~
      ·SerGlyPro ArgHisGlyAsp GlySerPhe SerAspGlu
5081 CTCAGGTCCG CGTCATGGTG ACGGTTCTTT CTCTGACGAA
      GAGTCCAGGC GCAGTACCAC TGCCAAGAAA GAGACTGCTT
                                     PmlI
                                     ~~~~~~
      MetAsnThrIle LeuAspAsn LeuAlaAla ArgAspPheIle.
5121 ATGAACACCA TCCTGGACAA CCTGGCTGCA CGTGAATTCA
      TACTTGTGGT AGGACCTGTT GGACCGACGT GCACTGAAGT
                                     BsaI
                                     ~~~~~~
      ·IAsnTrpLeu IleGlnThr LysIleThrAsp ProValSer.
5161 TCAACTGGCT GATCCAGACC AAAATCACCG ACCCGGTCTC
      AGTTGACCGA CTAGGTCTGG TTTTAGTGGC TGGGCCAGAG
      ·GlyProArg HisGlyAspGly SerPheSer AspGluMet
5201 AGGTCCGCGT CATGGTGACG GTTCTTTCTC TGACGAAATG
      TCCAGGCGCA GTACCACTGC CAAGAAAGAG ACTGCTTTAC
                                     PmlI

```

```

~::~~::~~
AsnThrIleLeu AspAsnLeu AlaAlaArg AspPheIleAsn·
5241 AACACCATCC TGGACAACCT GGCTGCACGT GACTTCATCA
      TTGTGGTAGG ACCTGTTGGA CCGACGTGCA CTGAAGTAGT
                        BsaI
~::~~::~~
·ATrpLeuIle GlnThrLys IleThrAspPro ValSerGly·
5281 ACTGGCTGAT CCAGACCAAA ATCACCGACC CGGTCTCAGG
      TGACCGACTA GGTCTGGTTT TAGTGGCTGG GCCAGAGTCC
·ProArgHis GlyAspGlySer PheSerAsp GluMetAsn
5321 TCCGCGTCAT GGTGACGGTT CTTTCTCTGA CGAAATGAAC
      AGGCGCAGTA CCACTGCCAA GAAAGAGACT GCTTTACTTG
                        PmlI
~::~~::~~
ThrIleLeuAsp AsnLeuAla AlaArgAsp PheIleAsnTrp·
5361 ACCATCCTGG ACAACCTGGC TGCACGTGAC TTCATCAACT
      TGGTAGGACC TGTTGGACCG ACGTGCAC TG AAGTAGTTGA
                        BsaI
~::~~::~~
·TLeuIleGln ThrLysIle ThrAspProVal SerGlyPro·
5401 GGCTGATCCA GACCAAAATC ACCGACCCGG TCTCAGGTCC
      CCGACTAGGT CTGGTTTTAG TGGCTGGGCC AGAGTCCAGG
·ArgHisGly AspGlySerPhe SerAspGlu MetAsnThr
5441 GCGTCATGGT GACGGTTCTT TCTCTGACGA AATGAACACC
      CGCAGTACCA CTGCCAAGAA AGAGACTGCT TTA CTTGTGTTGG
                        PmlI
~::~~::~~
IleLeuAspAsn LeuAlaAla ArgAspPhe IleAsnTrpLeu·
5481 ATCCTGGACA ACCTGGCTGC ACGTGACTTC ATCAACTGGC
      TAGGACCTGT TGGACCGACG TGCACTGAAG TAGTTGACCG
                        BsaI
~::~~::~~
·LIleGlnThr LysIleThr AspProValSer GlyProArg·
5521 TGATCCAGAC CAAAATCACC GACCCGGTCT CAGGTCCGCG
      ACTAGGTCTG GTTTTAGTGG CTGGGCCAGA GTCCAGGCGC
·HisGlyAsp GlySerPheSer AspGluMet AsnThrIle
5561 TCATGGTGAC GGTTC TTTCT CTGACGAAAT GAACACCATC
      AGTACCACTG CCAAGAAAGA GACTGCTTTA CTTGTGTTAG
                        PmlI
~::~~::~~
LeuAspAsnLeu AlaAlaArg AspPheIle AsnTrpLeuIle·
5601 CTGGACAACC TGGCTGCACG TGACTTCATC AACTGGCTGA
      GACCTGTTGG ACCGACGTGC ACTGAAGTAG TTGACCGACT
                        BsaI
~::~~::~~
·IGlnThrLys IleThrAsp ProValSerGly ProArgHis·
5641 TCCAGACCAA AATCACCGAC CCGGTCTCAG GTCCGCGTCA

```

```

AGGTCTGGTT TTAGTGGCTG GGCCAGAGTC CAGGCGCAGT
·GlyAspGly SerPheSerAsp GluMetAsn ThrIleLeu
5681 TGGTGACGGT TCTTTCTCTG ACGAAATGAA CACCATCCTG
ACCACTGCCA AGAAAGAGAC TGCTTTACTT GTGGTAGGAC
PmlI
~~~~~
AspAsnLeuAla AlaArgAsp PheIleAsn TrpLeuIleGln·
5721 GACAACCTGG CTGCACGTGA CTTTCATCAAC TGGCTGATCC
CTGTTGGACC GACGTGCACT GAAGTAGTTG ACCGACTAGG
BsaI
~~~~~
·GThrLysIle ThrAspPro ValSerGlyPro Arg
5761 AGACCAAAAT CACCGACCCG GTCTCAGGTC CGCGCTAATG
TCTGGTTTTA GTGGCTGGGC CAGAGTCCAG GCGCGATTAC
BamHI
~~~~~
5801 AGGATCCGAA TTCGAGCTCC GTCGACAAGC TTGCGGCCGC
TCCTAGGCTT AAGCTCGAGG CAGCTGTTCG AACGCCGGCG
5841 ACTCGAGCAC CACCACCACC ACCACTGAGA TCCGGCTGCT
TGAGCTCGTG GTGGTGGTGG TGGTGA CTCT AGGCCGACGA
5881 AACAAAGCCC GAAAGGAAGC TGAGTTGGCT GCTGCCACCG
TTGTTTCGGG CTTTCCTTCG ACTCAACCGA CGACGGTGGC
5921 CTGAGCAATA ACTAGCATAA CCCCTTG GGG CCTCTAAACG
GACTCGTTAT TGATCGTATT GGGGAACCCC GGAGATTTGC
5961 GGTCTTGAGG GGTTTTTTGC TGAAAGGAGG AACTATATCC
CCAGAACTCC CCAAAAAACG ACTTTCCTCC TTGATATAGG
6001 GGAT
CCTA

```

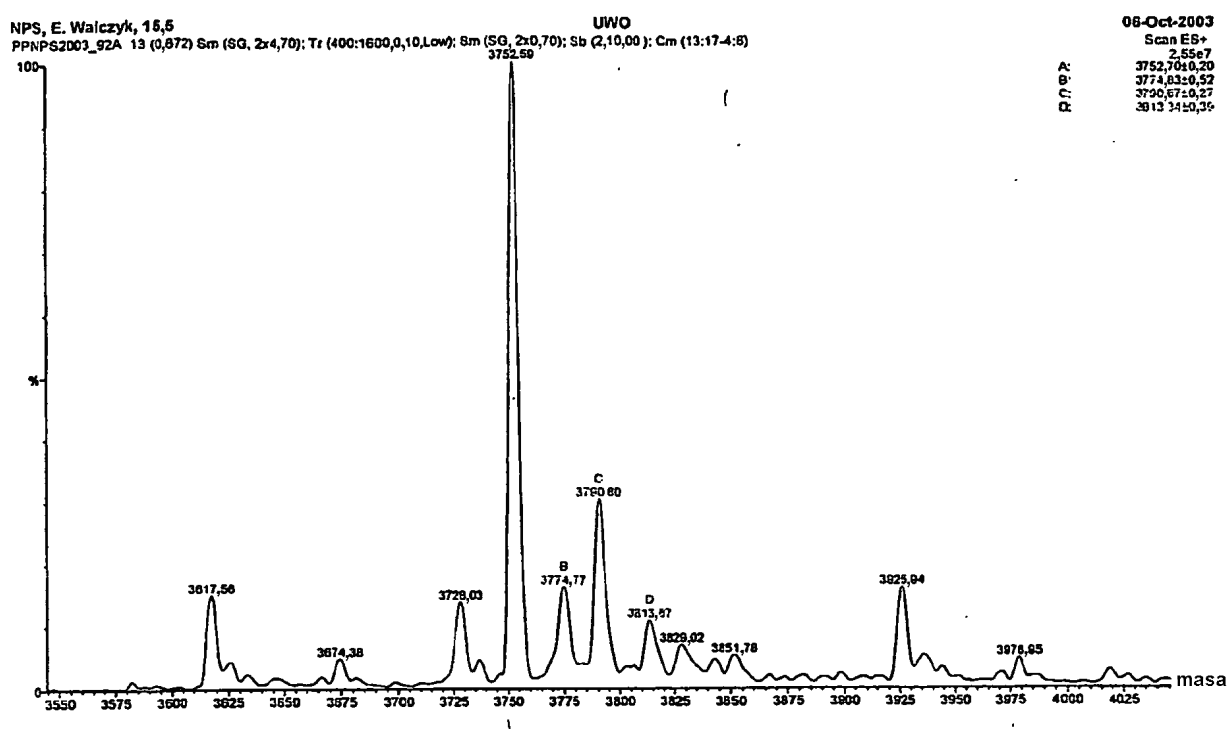


Figura 5