(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 764 457**

(51) Int. CI.:

C12N 15/09 (2006.01)
C12N 15/85 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **30.03.2016 PCT/EP2016/056926**

(87) Fecha y número de publicación internacional: **06.10.2016 WO16156404**

(96) Fecha de presentación y número de la solicitud europea: **30.03.2016 E 16715272 (7)**

(97) Fecha y número de publicación de la concesión europea: **11.12.2019 EP 3277807**

(54) Título: **Vectores de expresión eucariotas que incluyen elementos reglamentarios de los grupos de genes de globina**

(30) Prioridad:

31.03.2015 LU 92686
31.03.2015 EP 15161922

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
03.06.2020

(73) Titular/es:

GLYCOTOPE GMBH (100.0%)
Robert-Rössle-Strasse 10
13125 Berlin, DE

(72) Inventor/es:

GOLETZ, STEFFEN;
JAHN, DOREEN y
DANIELCZYK, ANTJE

(74) Agente/Representante:

SÁEZ MAESO, Ana

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

Vectores de expresión eucariotas que incluyen elementos reglamentarios de los grupos de genes de globina

Campo de la invención

La presente invención se refiere a nuevos casetes de expresión que pueden usarse para mejorar el rendimiento de producción de una proteína de interés. Los casetes de expresión comprenden elementos de regulación de la expresión de los grupos de genes de la globina humana, en particular el promotor de la $^{A}\gamma$ globina humana y la región de control del locus del grupo de genes de la β-globina o α-globina humana. La presente invención en particular proporciona un casete de expresión que comprende dichos elementos de regulación de la expresión de globina.

Antecedentes de la invención

La producción de proteínas recombinantes es un aspecto importante de la industria biotécnica de hoy. Está ganando cada vez más importancia a medida que aumenta el número de aplicaciones que requieren altas cantidades de proteínas de alta calidad en el mercado. La producción de alimentos y, en particular, la farmacología son dos áreas principales en las que aumenta constantemente la necesidad de proteínas recombinantes. Se necesitan mayores eficiencias de producción y, en consecuencia, menores costos del producto final para obtener un proceso comercialmente viable.

Sin embargo, al mismo tiempo, es esencial una alta calidad del producto y compatibilidad con aplicaciones humanas. Cada vez más y más aplicaciones requerían la producción recombinante de las proteínas en las células eucariotas, en particular en las células eucariotas superiores. Especialmente las proteínas que portan modificaciones postraduccionales como la glucosilación (glucoproteínas) difieren significativamente cuando se expresan en sistemas de células procariotas como *E. coli* o sistemas de células eucariotas como en líneas celulares humanas en particular. Estas diferencias en muchos casos afectan notablemente la actividad biológica, así como la inmunogenicidad de las proteínas producidas. Sin embargo, muchos sistemas de expresión que utilizan líneas celulares eucariotas superiores sufren de una tasa de expresión bastante baja de la proteína deseada, lo que da como resultado bajos rendimientos y altos costos de la proteína recombinante.

Por lo tanto, existe la necesidad en la técnica de proporcionar medios y métodos novedosos para aumentar el rendimiento de la producción de proteínas recombinantes, especialmente cuando se usan líneas celulares de expresión eucariotas.

Sumario de la invención

Como se demuestra por la presente invención, ciertos elementos de los grupos de genes de la globina humana pueden combinarse para proporcionar un casete de expresión que permite la expresión estable y alta de polipéptidos de interés en células eucariotas. En particular, la combinación de partes específicas de la región de control del locus del grupo de genes de β-globina o el grupo de genes de α-globina con el promotor de la $^{A}\gamma$ globina y opcionalmente también el potenciador de la $^{A}\gamma$ globina 3' forma un casete de expresión con tasas de expresión sorprendentemente altas y estables.

Por lo tanto, la presente invención proporciona en un primer aspecto un método *in vitro* para producir recombinantemente un polipéptido de interés, que comprende las etapas de

(a) proporcionar una célula huésped que comprende un casete de expresión que comprende, funcionalmente unidos entre sí,

(i) una región de control del locus que comprende al menos el elemento central del sitio 2 de hipersensibilidad a la ADNasa I (HS2) del grupo de genes de la β-globina humana;

(ii) una región promotora que comprende al menos una parte funcional del promotor del gen de la $^{A}\gamma$ globina humana; y

(iii) una región de codificación que comprende una secuencia de ácido nucleico que codifica el polipéptido de interés;

(b) cultivar la célula huésped en condiciones en las que la célula huésped expresa el polipéptido de interés; y

(c) aislar el polipéptido de interés.

En un segundo aspecto, la presente invención proporciona un casete de expresión que comprende, funcionalmente unidos entre sí,

(i) una región de control del locus que comprende al menos el elemento central del sitio 2 de hipersensibilidad a la ADNasa I (HS2) del grupo de genes de la β-globina humana;

(ii) una región promotora que comprende al menos una parte funcional del promotor del gen de la $^{A}\gamma$ globina humana

que incluye un sitio de inicio de la transcripción y opcionalmente al menos parte de la región 5' no traducida del gen de la $^{A}\gamma$ globina humana;

(iii) opcionalmente una región de codificación;

(iv) una región de terminación de la transcripción; y

(v) una región potenciadora que comprende al menos una parte funcional del potenciador 3' del gen de la $^{A}\gamma$ globina humana;

en el que el casete de expresión no comprende una secuencia de ácido nucleico que codifica la $^{A}\gamma$ globina humana completa.

En un tercer aspecto, la presente invención proporciona un vector que comprende el casete de expresión de acuerdo con el segundo aspecto y una célula huésped que comprende dicho casete de expresión o dicho vector.

Otros objetivos, características, ventajas y aspectos de la presente invención resultarán evidentes para aquellos expertos en la materia a partir de la siguiente descripción y las reivindicaciones adjuntas.

Definiciones

Como se usa en el presente documento, las siguientes expresiones generalmente pretenden tener preferiblemente los significados que se exponen a continuación, excepto en la medida en que el contexto en el que se usan indique lo contrario.

La expresión "comprende", como se usa en el presente documento, además de su significado literal también incluye y se refiere específicamente a las expresiones "consiste esencialmente en" y "consiste en". Por lo tanto, la expresión "comprende" se refiere a realizaciones en las que la materia que "comprende" elementos específicamente enumerados no comprende elementos adicionales, así como realizaciones en las que la materia que "comprende" elementos específicamente enumerados puede y/o de hecho abarca otros elementos. Asimismo, la expresión "tiene" debe entenderse como la expresión "comprende", que también incluye y se refiere específicamente a las expresiones "consiste esencialmente en" y "consiste en".

El término "ácido nucleico" como se usa en el presente documento se refiere a un polímero de ribonucleótido o desoxirribonucleótido. Un ácido nucleico puede ser ARN o ADN. Puede estar compuesto de una cadena sencilla de polímero o puede ser una cadena doble. El ácido nucleico puede ser de origen natural, recombinante o sintético. En realizaciones preferidas, un ácido nucleico es un ADN bicatenario.

Un "casete de expresión" es un constructo de ácido nucleico, generado o sintetizado, con elementos de ácido nucleico que son capaces de efectuar la expresión de un gen estructural en huéspedes que son compatibles con tales secuencias. Los casetes de expresión incluyen al menos un promotor y, opcionalmente, señales de terminación de la transcripción. Típicamente, el casete de expresión incluye un ácido nucleico a transcribir y un promotor. Los factores adicionales útiles para efectuar la expresión también pueden usarse como se describe en el presente documento. Por ejemplo, un casete de expresión también puede incluir secuencias de nucleótidos que codifican una secuencia señal que dirige la secreción de una proteína expresada desde la célula huésped. Un casete de expresión es preferiblemente parte de un vector de expresión. Las células huésped que se usarán para la expresión del ácido nucleico a transcribir se transforman o transfectan con el vector de expresión. Para permitir la selección de células transformadas que comprenden los constructos, se puede incluir convenientemente un gen marcador seleccionable en los vectores de expresión. Una persona experta en la materia reconocerá que este componente vectorial puede modificarse sin afectar sustancialmente su función.

La expresión "funcionalmente unido" significa que dos o más elementos de un casete de expresión están unidos entre sí de tal manera que su función está coordinada y permite la expresión de la secuencia de codificación (por ejemplo, la región de codificación). A modo de ejemplo, un promotor está funcionalmente unido a una secuencia de codificación cuando es capaz de asegurar la expresión de dicha secuencia de codificación. La construcción de un casete de expresión de acuerdo con la invención y el ensamblaje de sus diversos elementos se pueden llevar a cabo utilizando técnicas bien conocidas por los expertos en la materia, en particular las descritas en Sambrook et al., (1989, Molecular Cloning: A Laboratory Manual, Nolan C. ed., Nueva York: Cold Spring Harbor Laboratory Press).

Los términos "secuencia arriba" y "secuencia abajo", como se usan en el presente documento, se refieren a la posición de un elemento o secuencia de ácido nucleico en una molécula de ácido nucleico con respecto a otro elemento o secuencia de ácido nucleico en dicha molécula de ácido nucleico. "Secuencia arriba" se refiere a una posición que está más cerca del extremo 5' de la molécula de ácido nucleico y "secuencia abajo" se refiere a una posición que está más cerca del extremo 3' de la molécula de ácido nucleico. En el caso de los ácidos nucleicos bicatenarios, en particular el ADN, esa cadena del ácido nucleico que se usa como matriz para la transcripción de un ARN tal como un ARNm, es decir, la cadena sentido, se usa para determinar el extremo 5' y el 3' del ácido nucleico. Por lo tanto, "secuencia arriba" está en la dirección del extremo 5' de la cadena sentido mientras que "secuencia abajo" está en la dirección del extremo 3' de la cadena sentido.

Un "homólogo" de una secuencia de ácido nucleico o secuencia de aminoácidos objetivo comparte una homología o identidad de al menos 75%, más preferiblemente al menos 80%, al menos 85%, al menos 90%, al menos 93%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% con dicha secuencia de ácido nucleico o secuencia de aminoácidos objetivo. Una "homología" o "identidad" de una secuencia de aminoácidos o secuencia de nucleótidos se determina preferiblemente de acuerdo con la invención en toda la longitud de la secuencia objetivo o toda la longitud de la parte indicada de la secuencia objetivo. Cuando se refiere a una secuencia de ácido nucleico o secuencia de aminoácidos específica, la presente invención generalmente también abarca homólogos de dicha secuencia de ácido nucleico o secuencia de aminoácidos, respectivamente. Un homólogo de una secuencia de ácido nucleico o secuencia de aminoácidos objetivo en particular es un homólogo funcional que tiene las mismas o sustancialmente las mismas funciones y actividades de la secuencia de ácido nucleico o secuencia de aminoácidos objetivo de la que se deriva.

Un "promotor" es una secuencia de ácido nucleico que permite y controla la transcripción de una secuencia de ácido nucleico unida funcionalmente a la misma, en particular una secuencia de codificación. Un promotor contiene una secuencia de reconocimiento para unir la ARN polimerasa e incluye o está funcionalmente unido a un sitio de inicio de la transcripción. El promotor puede ser un promotor inducible que solo es activo en presencia (o ausencia) de una señal específica, o puede ser constitutivamente activo. La actividad del promotor puede regularse adicionalmente mediante elementos reguladores tales como regiones de control del locus y elementos potenciadores.

Una "secuencia codificadora" es una secuencia de ácido nucleico que codifica un producto génico tal como un polipéptido o ARN.

Una "parte" de un elemento de ácido nucleico en particular comprende al menos 5 ácidos nucleicos, preferiblemente al menos 10, al menos 15, al menos 20, al menos 30 o al menos 50 ácidos nucleicos de dicho elemento de ácido nucleico. Una "parte" de un elemento de ácido nucleico en particular comprende al menos 5 nucleótidos consecutivos, preferiblemente al menos 10, al menos 15, al menos 20, al menos 30 o al menos 50 nucleótidos consecutivos de dicho elemento de ácido nucleico. En particular, comprende al menos 1%, preferiblemente al menos 2%, al menos 3%, al menos 5%, al menos 7,5%, al menos 10%, al menos 15%, al menos 20% o al menos 25% de dicho elemento de ácido nucleico. Una "parte funcional" de un elemento de ácido nucleico es una parte de dicho elemento que es capaz de realizar la función prevista del elemento. Por ejemplo, una parte funcional de una región de control del locus, promotor o potenciador 3' es capaz de modular, en particular potenciar la expresión de una región de codificación a la que está unida funcionalmente. Una parte de un elemento de ácido nucleico en particular se refiere a una parte funcional de dicho elemento de ácido nucleico.

Un "péptido" o "polipéptido" como se usa en el presente documento se refiere a una cadena polipeptídica que comprende al menos 5 aminoácidos. Un péptido o polipéptido comprende preferiblemente al menos 10, al menos 15, al menos 20, al menos 25, al menos 30 o al menos 35 aminoácidos. El término "péptido" o "polipéptido" como se usa en el presente documento también se refiere a proteínas, que incluyen péptidos, polipéptidos y proteínas que se modificaron postraduccionalmente. En particular, el término péptido o polipéptido incluye péptidos y glicoproteínas glicosiladas. Los términos "péptido" y "polipéptido" se usan indistintamente en el presente documento.

Una parte de un péptido, polipéptido o proteína comprende preferiblemente al menos 3 aminoácidos consecutivos de dicho péptido, polipéptido o proteína, preferiblemente al menos 5, al menos 10, al menos 15 o al menos 20 aminoácidos consecutivos de dicha proteína.

El término "composición farmacéutica" y términos similares se refieren particularmente a una composición adecuada para administrar a un ser humano, es decir, una composición que contiene componentes que son farmacéuticamente aceptables. Preferiblemente, una composición farmacéutica comprende un compuesto activo o una sal o profármaco del mismo junto con un vehículo, diluyente o excipiente farmacéutico tal como tampón, conservante y modificador de tonicidad.

Descripción detallada de la invención

La presente invención está dirigida a un casete de expresión que comprende, funcionalmente unidos entre sí,

(i) una región de control del locus que comprende al menos el elemento central del sitio 2 de hipersensibilidad a la ADNasa I (HS2) del grupo de genes de la β-globina humana;

(ii) una región promotora que comprende al menos una parte funcional del promotor del gen de la $^{A}\gamma$ globina humana que incluye un sitio de inicio de la transcripción y opcionalmente al menos parte de la región 5' no traducida del gen de la $^{A}\gamma$ globina humana;

(iii) opcionalmente una región de codificación;

(iv) una región de terminación de la transcripción; y

(v) una región potenciadora que comprende al menos una parte funcional del potenciador 3' del gen de la $^{A}\gamma$ globina humana;

en la que el casete de expresión no comprende una secuencia de ácido nucleico que codifica la $^{A}\gamma$ globina humana completa.

La hemoglobina es una metaloproteína en la sangre de humanos y animales que permite el transporte de oxígeno y dióxido de carbono. La hemoglobina es una proteína globular de múltiples subunidades que consta de cuatro subunidades (globinas), cada una de las cuales consiste en una cadena polipeptídica estrechamente asociada con un grupo hemo que transporta un ion de hierro. Existen varios tipos diferentes de subunidades de globina y la composición de la subunidad de la hemoglobina cambia a lo largo de la vida. Por ejemplo, el feto humano tiene hemoglobina F compuesta de dos globinas α y dos globinas γ ($\alpha_2\gamma_2$), mientras que en el adulto predomina la hemoglobina A con dos globinas α y dos globinas β ($\alpha_2\beta_2$). Los diferentes polipéptidos de globina son expresados por los grupos de genes de globina humana que son responsables de la expresión de las diferentes subunidades en los diferentes estados de desarrollo del ser humano. El grupo de genes de la β-globina humana comprende una región de control del locus y cinco genes de globina diferentes, es decir, el gen de la ε-globina, Gγ-globina, Aγ-globina, δ-globina y β-globina. Del mismo modo, el grupo de genes de α-globina también comprende una región de control del locus y los genes de la ζ-globina, α2-globina, α1-globina y θ-globina. La Figura 1 muestra la estructura de los grupos de genes de la α-globina y la β-globina humana. Cada gen de globina separado dentro del grupo de genes tiene sus propias secuencias promotora y potenciadora específicas que controlan la expresión de la secuencia de codificación de la subunidad de globina respectiva. Sin embargo, estos promotores están regulados por la región de control del locus. En particular, la región de control del locus es capaz de activar o inactivar los promotores y este patrón de activación cambia a lo largo de la vida. Ambas regiones de control del locus (una en el grupo de genes de α-globina y una en el grupo de genes de la β-globina) orquestan la expresión de los diferentes genes de globina durante el desarrollo, dando como resultado la composición de subunidades específicas de las diferentes proteínas de hemoglobina.

Las regiones de control del locus comprenden varios sitios de hipersensibilidad a la ADNasa I (HS) que son responsables de la activación e inactivación de los promotores. Se ha encontrado que, por ejemplo, HS2 del grupo de genes de la β-globina controla la actividad del promotor de la $^{A}\gamma$ globina. HS2 del grupo de genes de la β-globina puede subdividirse adicionalmente en una región central y subdominios moduladores. La región central es esa parte de HS2 que es principalmente responsable de la activación del promotor de la globina. Los subdominios moduladores M1 a M5 modulan el efecto del dominio central, ya sea positiva o negativamente. La región central se coloca entre los subdominios moduladores M1 y M2, con M3 a M5 a continuación. En particular, M1 y M2 potencian adicionalmente el efecto de activación de la región central sobre el promotor de la $^{A}\gamma$ globina.

Ahora se descubrió que el uso de elementos de expresión de los grupos de genes de la globina humana proporciona una expresión estable y alta de un producto objetivo tal como un polipéptido de interés en células eucariotas, en particular en células sanguíneas humanas o células derivadas de ellas. La presente divulgación proporciona un casete de expresión que comprende al menos una parte funcional del promotor de la $^{A}\gamma$ globina humana y al menos una parte funcional de una región de control del locus del grupo de genes de la α-globina o la β-globina humana. El casete de expresión comprende además una región de codificación que contiene una secuencia de ácido nucleico que codifica un polipéptido de interés y/o un sitio de clonación para la introducción de dicha región de codificación. El casete de expresión puede comprender además una región potenciadora que comprende al menos una parte funcional de un potenciador 3' del gen de la $^{A}\gamma$ globina humana. La región de control del locus, el promotor y el potenciador 3' se colocan todos en el casete de expresión para que puedan modular y, en particular, habilitar y mejorar la expresión de la región de codificación. El casete de expresión también puede comprender una región de terminación de la transcripción en la que se termina la transcripción de la región de codificación. Los elementos del casete de expresión en particular están funcionalmente unidos entre sí.

Los ejemplos específicos del casete de expresión comprenden o consisten en la secuencia de ácido nucleico de cualquiera de las SEQ ID NOs: 1 a 5. Los elementos de expresión respectivos también se muestran en la Figura 2 en la que se indican los diferentes elementos de los casetes de expresión. En particular, HS4, HS3, HS2 y HS40 son partes de la región de control del locus del grupo de genes de la α-globina o la β-globina humana. Después de esta región de control del locus, se indica la parte funcional del promotor de la $^{A}\gamma$ globina humana ($^{A}\gamma$-Prom). A continuación, sigue una secuencia de codificación o sitio de clonación para la introducción de una secuencia de codificación (CS) y una región de terminación de la transcripción que incluye una señal de poliadenilación (γpA). Al final del casete de expresión, se coloca la parte funcional de un potenciador 3' del gen de la $^{A}\gamma$ globina humana ($^{A}\gamma$-Enh).

En ciertas realizaciones, el casete de expresión no comprende un intrón 2 de la β-globina, en particular ningún intrón de un gen de la β-globina o cualquier otro gen de globina.

El promotor de la $^{A}\gamma$ globina humana

El casete de expresión utiliza una parte funcional del promotor de la $^{A}\gamma$ globina humana para permitir y controlar la expresión de la región de codificación, en particular el polipéptido de interés. La parte funcional del promotor de globina humana se coloca en particular secuencia arriba de la región de codificación. Está funcionalmente unida a la región de codificación y permite y controla la transcripción de la misma. En ciertas realizaciones, la parte funcional del promotor de la $^{A}\gamma$ globina humana abarca un sitio de inicio de la transcripción en el que comienza la transcripción del ARNm prematuro. Además, la parte funcional del promotor de la $^{A}\gamma$ globina humana también puede comprender al

menos parte de una región 5' no traducida (5' UTR), en particular al menos parte de la 5' UTR del gen de la $^{A}\gamma$ globina humana. En realizaciones específicas, la parte funcional del promotor de la $^{A}\gamma$ globina humana abarca al menos esa parte del gen de la $^{A}\gamma$ globina humana que permite la transcripción del ARNm de la $^{A}\gamma$ globina humana. En ciertas realizaciones, la parte funcional del promotor de la $^{A}\gamma$ globina humana comprende una caja CCAAT. En particular, la parte funcional del promotor de la $^{A}\gamma$ globina humana comprende o consiste en los nucleótidos -299 a -26, especialmente los nucleótidos -299 a +36, los nucleótidos -384 a -26 o los nucleótidos -384 a +36, con respecto al sitio de inicio de la transcripción, del gen de la $^{A}\gamma$ globina humana. En particular, la parte funcional del promotor de la $^{A}\gamma$ globina humana comprende y especialmente consiste en la secuencia de ácido nucleico de la posición 1123 a 1542 de la SEQ ID NO: 1. También es posible usar fragmentos más cortos del promotor de la $^{A}\gamma$ globina humana que todavía son funcionales. El experto en la materia es capaz de determinar partes funcionales adecuadas del promotor de la $^{A}\gamma$ globina humana. En particular, los métodos para determinar la actividad de una secuencia promotora se conocen en la técnica y se describen en los ejemplos a continuación.

Como alternativa, se puede usar un homólogo de dicha parte funcional del promotor de la $^{A}\gamma$ globina humana. Dicho homólogo preferiblemente es al menos 85%, preferiblemente al menos 90%, al menos 95%, al menos 97%, al menos 98% o al menos 99% idéntico con una de las secuencias de ácido nucleico definidas anteriormente en toda su longitud. En ciertas realizaciones, el homólogo tiene la misma o sustancialmente la misma función y/o actividad que la parte funcional del promotor de la $^{A}\gamma$ globina del que se deriva, y en particular proporciona una tasa de expresión de la región de codificación que alcanza al menos el 75%, preferiblemente al menos el 80%, al menos el 85% o al menos el 90% de la tasa de expresión alcanzada usando la parte funcional del promotor de la $^{A}\gamma$ globina del que se deriva en las mismas condiciones.

La región de control del locus

La región de control del locus está funcionalmente unida a la región promotora y es capaz de modular y, en particular, mejorar la actividad de la parte funcional del promotor de la $^{A}\gamma$ globina humana. La región de control del locus en particular está posicionada secuencia arriba de la región promotora.

En ciertas realizaciones preferidas, se usa la región de control del locus o una parte funcional de la misma del grupo de genes de la β-globina humana. De acuerdo con la técnica, la región de control del locus de β-globina humana abarca cuatro sitios de hipersensibilidad a la ADNasa I específicos de eritroides denominados HS1 a HS4 ubicados 6 a 20 kpb secuencia arriba del primer gen de globina de este grupo de genes, es decir, el gen de la ε-globina. En particular, HS2 es responsable de controlar la expresión de los genes de la globina y se considera que constituye un componente funcional principal de la región de control del locus. HS2 de la región de control del locus de la β-globina humana se subdivide en un elemento central y otros subdominios moduladores, en los que el elemento central se coloca entre los subdominios moduladores M1 y M2. El elemento central es la región mínima de la región de control del locus de la β-globina que es capaz de potenciar la actividad promotora del promotor de la $^{A}\gamma$ globina humana.

En una realización preferida, la región de control del locus comprende el elemento central del sitio 2 de hipersensibilidad a la ADNasa I (HS2) del grupo de genes de la β-globina humana. En particular, el elemento central de HS2 del grupo de genes de la β-globina humana tiene la secuencia de ácido nucleico de la posición 906 a 939 de la SEQ ID NO: 1. En ciertas realizaciones, la región de control del locus comprende o consiste en el elemento central de HS2 y los dos subdominios moduladores adyacentes M1 y M2, es decir, el elemento M1-núcleo-M2 de HS2 del grupo de genes de la β-globina humana. Dicho elemento M1-núcleo-M2 puede tener la secuencia de ácido nucleico de la posición 742 a 995 de la SEQ ID NO: 1. Por lo tanto, incluye la región central de HS2 que tiene la secuencia de ácido nucleico de la posición 906 a 939 de la SEQ ID NO: 1.

En una realización adicional, la región de control del locus comprende al menos una parte funcional del sitio 2 de hipersensibilidad (HS2) del grupo de genes de la β-globina humana que comprende la secuencia de ácido nucleico de la posición 741 a 1109 de la SEQ ID NO: 1. Esta parte funcional de HS2 comprende el elemento M1-núcleo-M2 y una secuencia adicional de ácido nucleico directamente secuencia abajo del mismo.

En ciertas realizaciones, la región de control del locus que comprende al menos una parte funcional de HS2 del grupo de genes de β-globina humana comprende además al menos una parte del sitio 3 de hipersensibilidad (HS3) del grupo de genes de la β-globina humana . En particular, el HS3 o parte del mismo se coloca secuencia arriba del HS2 o parte del mismo en la región de control del locus del casete de expresión. El HS3 o parte del mismo en particular comprende o consiste en la secuencia de ácido nucleico de la posición 310 a 735 de la SEQ ID NO: 1. Además, la región de control del locus también puede comprender el sitio 4 de hipersensibilidad (HS4) o una parte del mismo del grupo de genes de la β-globina humana, que en particular se coloca secuencia arriba de HS2 y HS3, si está presente. En ciertas realizaciones, HS4 o la parte del mismo comprende o consiste en la secuencia de ácido nucleico de la posición 13 a 294 de la SEQ ID NO: 1. Por lo tanto, la región de control del locus del casete de expresión de acuerdo con la invención puede comprender, en la dirección secuencia abajo, opcionalmente al menos una parte de HS4, opcionalmente al menos una parte de HS3 y al menos una parte funcional de HS2 del grupo de genes de la β-globina humana. En realizaciones alternativas, la región de control del locus no comprende el sitio 3 de hipersensibilidad (HS3) del grupo de genes de la β-globina humana.

En realizaciones particulares, la región de control del locus tiene una secuencia de ácido nucleico seleccionada del grupo que consiste en la posición 13 a 1109 de la SEQ ID NO: 1, posición 20 a 819 de la SEQ ID NO: 2, posición 18 a 386 de la SEQ ID NO: 3 y posiciones 13 a 266 de la SEQ ID NO: 4.

También se describen casetes de expresión en los que la región de control del locus comprende al menos una parte funcional del sitio 40 de hipersensibilidad (HS40) del grupo de genes de la α-globina humana. La parte de HS40 en particular comprende o consiste en el elemento central de HS40 que puede tener la secuencia de ácido nucleico de la posición 24 a 278 de la SEQ ID NO: 5. En particular, HS40 o la parte del mismo comprende o consiste en la secuencia de ácido nucleico de la posición 7 a 372 de la SEQ ID NO: 5.

Además, en ciertas realizaciones se puede usar una región de control del locus que comprende o que consiste en una secuencia de ácido nucleico que es un homólogo de una de las regiones de control del locus anteriores. En particular, dicho homólogo tiene una identidad de secuencia de al menos 90%, preferiblemente al menos 95%, al menos 97%, al menos 98% o al menos 99% con una de las regiones de control del locus anteriores en toda la longitud, y/o tiene la misma o sustancialmente la misma función que la región de control del locus de la que se deriva. En realizaciones preferidas, la región de control del locus homólogo proporciona una tasa de expresión de la región de codificación que alcanza al menos 75%, preferiblemente al menos 80%, al menos 85% o al menos 90% de la tasa de expresión alcanzada usando la región de control del locus de la que se deriva bajo las mismas condiciones.

La región de codificación

La región de codificación del casete de expresión comprende una secuencia de ácido nucleico que codifica un producto de interés, en particular un polipéptido de interés, que debe expresarse mediante el casete de expresión. La expresión de las secuencias de ácido nucleico de la región de codificación está regulada por la región promotora y, por lo tanto, está funcionalmente unida a la misma. Cuando el casete de expresión, opcionalmente presente en un vector, se introduce en una célula huésped adecuada, dicha célula huésped produce un producto, en particular un polipéptido, codificado por las secuencias de ácido nucleico de la región de codificación.

La región de codificación del casete de expresión en particular contiene o consiste en una secuencia de ácido nucleico que codifica un polipéptido de interés.

El polipéptido de interés puede ser cualquier polipéptido, incluidas las proteínas. El polipéptido puede ser de cualquier origen, incluidos polipéptidos derivados de mamíferos y humanos, así como polipéptidos artificiales. En ciertas realizaciones, el polipéptido comprende uno o más sitios de glicosilación y en particular es un polipéptido glicosilado tal como una glicoproteína o una parte de la misma, anticuerpos o derivados o partes de los mismos; hormonas peptídicas, gonadotropinas tales como FSH (hormona estimulante de folículos), CG (gonadotropina coriónica), LH (hormona luteinizante) y TSH (hormona estimulante de la tiroides) incluyendo todas las isoformas y variantes de las mismas; eritropoyetina; factores de coagulación sanguínea tales como el factor VII, VIII, IX o el factor de von Willebrand; enzimas lisosomales y citoquinas. Además, el polipéptido de interés puede seleccionarse del grupo que consiste en cualquiera de las moléculas de proteína del grupo de citoquinas y sus receptores, por ejemplo los factores de necrosis tumoral TNF-alfa y TNF-beta; renina hormona de crecimiento humano y hormona de crecimiento bovina; factor de liberación de la hormona del crecimiento; hormona paratiroidea; hormona estimulante de la tiroides; lipoproteínas; alfa-1-antitripsina; cadena A y cadena B de insulina; gonadotropinas, por ejemplo, hormona estimulante del folículo (FSH), hormona luteinizante (LH), tirotropina y gonadotropina coriónica humana (hCG); calcitonina; glucagón; factores de coagulación tales como el factor VIIIC, el factor IX, el factor VII, el factor tisular y el factor von Willebrands; factores anticoagulantes como la proteína C; factor natriurético auricular; surfactante pulmonar; activadores de plasminógeno, tales como uroquinasa, orina humana y activador de plasminógeno de tipo tisular; bombesina; trombina factor de crecimiento hemopoyético; encefalinasa; proteína inflamatoria de macrófagos humanos; una albúmina sérica tal como la albúmina sérica humana; sustancia inhibidora mulleriana; cadena A y cadena B de relaxina; prorrelaxina; péptido asociado a gonadotropina de ratón; factor de crecimiento vascular endotelial; receptores para hormonas o factores de crecimiento; integrina; proteína A y D; factores reumatoides; factores neurotróficos tales como factor neurotrófico derivado de hueso, neurotrofina-3, neurotrofina-4, neurotrofina-5, neurotrofina-6 y factor beta de crecimiento nervioso; factor de crecimiento derivado de plaquetas; factores de crecimiento de fibroblastos; factor de crecimiento epidérmico; factor de crecimiento transformante tal como TGF-alfa y TGF-beta; factor I y II de crecimiento similar a la insulina; proteínas de unión al factor de crecimiento similar a insulina; proteínas CD tales como CD-3, CD-4, CD-8 y CD-19; eritropoyetina (EPO); factores osteoinductores; inmunotoxinas; una proteína morfogenética ósea; un interferón tal como interferón-alfa, interferón-beta, e interferón-gamma; factores estimulantes de colonias (CSF), por ejemplo, M-CSF, GM-CSF y G-CSF; interleuquinas (IL), por ejemplo, IL-1 a IL-12; superóxido dismutasa; receptores de células T; proteínas de membrana superficial; factor de aceleración de la descomposición; anticuerpos e inmunoadhesinas; glicoforina A; y proteínas de mucina tales como MUC1.

En ciertas realizaciones, el polipéptido de interés es un anticuerpo o una parte o derivado del mismo. En particular, el polipéptido de interés puede ser la cadena pesada o la cadena ligera de un anticuerpo o una parte del mismo. Además, el polipéptido de interés puede ser una parte o derivado de un anticuerpo seleccionado del grupo que consiste en (i) fragmentos Fab, fragmentos monovalentes que consisten en la región variable y el primer dominio constante de cada cadena pesada y ligera; (ii) fragmentos $F(ab)_2$, fragmentos bivalentes que comprenden dos fragmentos Fab unidos por un puente disulfuro en la región bisagra; (iii) fragmentos Fd que consisten en la región variable y el primer dominio

constante CH1 de la cadena pesada; (iv) fragmentos Fv que consisten en la región variable de la cadena pesada y la cadena ligera de un solo brazo de un anticuerpo; (v) fragmentos scFv, fragmentos Fv que consisten en una sola cadena polipeptídica; (vi) fragmentos $(Fv)_2$ que consisten en dos fragmentos Fv unidos covalentemente entre sí; (vii) un dominio variable de cadena pesada; y (viii) multicuerpos que consisten en una región variable de la cadena pesada y una región variable de la cadena ligera unidas covalentemente entre sí de tal manera que la asociación de las regiones variables de la cadena pesada y la cadena ligera solo pueden ocurrir de forma intermolecular pero no intramolecular. En realizaciones en las que el casete de expresión incluye una región de codificación que codifica un anticuerpo o una parte o derivado del mismo, la región de control del locus en particular comprende al menos una parte de HS2 del grupo de genes de la β-globina humana, preferiblemente al menos una parte de HS2 y al menos una parte de HS3 del grupo de genes de la β-globina humana o al menos una parte de HS2 y al menos una parte de HS4 del grupo de genes de la β-globina humana, especialmente al menos una parte de HS2 y al menos una parte de HS3 y al menos una parte de HS4 del grupo de genes de la β-globina humana.

En ciertas realizaciones, la región de codificación codifica más de un polipéptido, en particular dos polipéptidos. En estas realizaciones, la región de codificación puede contener dos o más secuencias de ácido nucleico separadas que se transcriben en ARNm separados, cada uno con su propio sitio de inicio de la transcripción, sitio de terminación de la transcripción y señal de poliadenilación. Alternativamente, la región de codificación puede contener una secuencia de ácido nucleico que se transcribe en un ARNm que comprende dos o más secuencias separadas de ácido nucleico codificante que codifican cada una un polipéptido separado. En estas realizaciones, la región de codificación puede comprender uno o más sitios internos de entrada al ribosoma, uno para cada secuencia de ácido nucleico codificante además del primero. Estos sitios internos de entrada al ribosoma permiten la traducción de más de un polipéptido de una sola transcripción. En ciertas realizaciones, la región de codificación codifica dos polipéptidos, en particular la cadena pesada y la cadena ligera de un anticuerpo.

En ciertas realizaciones, la región de codificación comprende una secuencia de ácido nucleico que codifica un péptido señal que comprende en particular una señal de localización extracelular. La secuencia de ácido nucleico que codifica un péptido señal puede ser la única secuencia de codificación de la región de codificación o puede estar presente además de otras secuencias de codificación, tales como aquellas secuencias de ácido nucleico que codifican un polipéptido como se describió anteriormente. El péptido señal en particular induce una expresión secretora del polipéptido de interés codificado por la región de codificación. El péptido señal puede separarse del polipéptido restante durante la expresión. El péptido señal en particular se coloca secuencia arriba de las otras secuencias de ácido nucleico codificantes comprendidas en la región de codificación o secuencia arriba del sitio de clonación, en particular al comienzo de la región de codificación. Además, la secuencia de ácido nucleico que codifica el péptido señal se posiciona en el marco con las otras secuencias de ácido nucleico codificantes comprendidas en la región de codificación.

En realizaciones particulares, la región de codificación comprende la secuencia de ácido nucleico que codifica un polipéptido de interés.

En ciertas realizaciones, la región de codificación no comprende un gen informador o un gen marcador seleccionable. En realizaciones adicionales, la región de codificación no comprende una secuencia de ácido nucleico que codifica una proteína globina o una parte de la misma que comprende al menos 20 aminoácidos consecutivos de una proteína globina.

El sitio de clonación

En ciertas realizaciones, el casete de expresión comprende un sitio de clonación para integrar una secuencia de ácido nucleico. El sitio de clonación puede estar presente en el casete de expresión en lugar de la región de codificación y puede servir para introducir dicha secuencia de ácido nucleico en el casete de expresión. Además, el sitio de clonación puede estar presente en el casete de expresión además de la región de codificación, en particular en realizaciones en las que la región de codificación solo comprende una secuencia de ácido nucleico que codifica un péptido señal.

El sitio de clonación presente en el casete de expresión es adecuado para introducir una región de codificación, en particular un ácido nucleico que codifica un polipéptido de interés, en el casete de expresión. Los sitios y métodos de clonación adecuados para introducir fragmentos de ácido nucleico en otras moléculas de ácido nucleico tales como casetes de expresión o vectores son comúnmente conocidos en la técnica. En ciertas realizaciones, el sitio de clonación comprende al menos una, en particular al menos dos, al menos tres, al menos cuatro o al menos cinco secuencias de reconocimiento de enzimas de restricción. Las enzimas de restricción adecuadas y sus secuencias de reconocimiento son conocidas en la técnica. Ejemplos de enzimas de restricción son EcoRI, EcoRV, HindIII, BamHI, XbaI, PvuI, KpnI, BstXI, XmaI, SmaI, NotI, XhoI y ClaI. Un ejemplo de secuencia de ácido nucleico de un sitio de clonación múltiple está representada por la secuencia de ácido nucleico de la posición 1559 a 1664 de la SEQ ID NO: 1.

La región de terminación de la transcripción

En ciertas realizaciones, el casete de expresión comprende una región de terminación de la transcripción. La región de terminación de la transcripción está funcionalmente unida a la región promotora y termina la transcripción de la

región de codificación. Se coloca secuencia abajo de la región de codificación y/o el sitio de clonación.

En realizaciones específicas, la región de terminación de la transcripción comprende un sitio de terminación de la transcripción y/o una señal de poliadenilación. La señal de poliadenilación puede ser cualquier señal de poliadenilación que sea capaz de inducir la poliadenilación del ARNm prematuro en células eucariotas, en particular en células humanas. Puede comprender o consistir en la secuencia de ácido nucleico de la posición 1725 a 1730 de la SEQ ID NO: 1 o un homólogo de la misma.

La región potenciadora

En ciertas realizaciones, el casete de expresión comprende una región potenciadora, en particular una región potenciadora 3'. La región potenciadora 3' se posiciona secuencia abajo de la región de codificación y/o sitio de clonación y secuencia abajo de la región de terminación de la transcripción, si está presente. Está funcionalmente unida con la región promotora y mejora la expresión de la región de codificación. La región potenciadora en particular comprende o consiste en al menos una parte funcional del potenciador 3' del gen de la $^{A}\gamma$ globina humana. En ciertas realizaciones, la región potenciadora comprende o consiste en la secuencia de ácido nucleico de la posición 2136 a 2881 de SEQ ID NO: 1 o un homólogo de la misma. En ciertas realizaciones, la región potenciadora homóloga proporciona una tasa de expresión de la región de codificación que alcanza al menos 75%, preferiblemente al menos 80%, al menos 85% o al menos 90% de la tasa de expresión alcanzada usando la región potenciadora a partir de la cual se deriva bajo las mismas condiciones.

El vector que comprende el casete de expresión

En un aspecto, la presente invención se refiere a un vector que comprende el casete de expresión de acuerdo con la invención. El vector puede ser cualquier vector adecuado para transferir el casete de expresión a una célula huésped. Los vectores respectivos son conocidos en la técnica. En particular, el vector está adaptado para su transferencia a células eucariotas, tales como células de mamífero, en particular células humanas.

Además del casete de expresión, el vector puede comprender elementos adicionales. Por ejemplo, el vector puede comprender uno o más marcadores de selección. En ciertas realizaciones, al menos uno de los marcadores de selección es adecuado para seleccionar células huésped que comprenden el vector, en particular células huésped eucariotas, tales como células huésped de mamífero, en particular células huésped humanas, contra células huésped que no comprenden el vector. Ejemplos adecuados de los marcadores de selección son genes que proporcionan resistencia contra un compuesto antibiótico. Además, el vector puede comprender elementos adecuados para amplificarlo en una célula huésped procariota tal como células de *E. coli*. Tales elementos, por ejemplo, incluyen un origen de replicación, tal como Col E1 Ori, y un marcador de selección procariota, tal como un gen que proporciona resistencia contra un bactericida, por ejemplo, ampicilina

En ciertas realizaciones, el vector es un ADN bicatenario circular o lineal, en particular un ADN bicatenario circular.

En ciertas realizaciones, el vector comprende el casete de expresión con la región de codificación que comprende una secuencia de ácido nucleico que codifica un polipéptido de interés.

El gen marcador seleccionable

En ciertas realizaciones, el vector comprende además un gen marcador seleccionable. El gen marcador seleccionable no necesita estar funcionalmente unido con los elementos del casete de expresión. El gen marcador seleccionable permite la selección de células huésped que comprenden el vector. Las células que contienen el vector se cultivan preferiblemente en presencia de un agente de selección adecuado que reduce o inhibe la proliferación de células que no comprenden el gen marcador seleccionable.

En realizaciones específicas, el gen marcador seleccionable es un gen marcador seleccionable amplificable que permite la amplificación del gen marcador y la amplificación conjunta del casete de expresión que está presente en el mismo vector. Cuando se usa un gen marcador seleccionable amplificable, la amplificación del casete de expresión en células transfectadas, en particular, se logra mediante cultivo gradual de las células en presencia de concentraciones crecientes del agente de selección. En ciertas realizaciones, el gen marcador seleccionable codifica una dihidrofolato reductasa (DHFR), tal como una variante de DHFR resistente a antifolato, y el agente de selección correspondiente es un antifolato, tal como metotrexato.

Otros ejemplos de genes marcadores seleccionables amplificables adecuados y sus agentes de selección correspondientes son el gen de resistencia a la neomicina (por ejemplo, aminoglicósido fosfotransferasa) y la geneticina (G418); puromicina N-acetil-transferasa y puromicina; metalotioneína y cadmio; CAD (carbamoil-fosfato sintetasa: aspartato transcarbamilasa: dihidroorotasa) y N-fosfoacetil-L-aspartato; adenosina-desaminasa y Xyl-A- o adenosina, 2'-desoxicoformicina; AMP (adenilato)-desaminasa y adenina, azaserina, coformicina; UMP-sintasa y 6-azauridina, pirazofurano; IMP 5'-deshidrogenasa y ácido micofenólico; xantina-guanina-fosforribosil transferasa y ácido micofenólico con xantina limitante; HGPRTasa mutante o timidina quinasa mutante e hipoxantina, aminopterina y timidina (HAT); timidilato sintetasa y 5-fluorodeoxiuridina; P-glicoproteína 170 (MDR1) y adriamicina, vincristina, colchicina; ribonucleótido reductasa y afidicolina; glutamina sintetasa y metionina sulfoximina (MSX); asparagina

sintetasa y β-aspartilhidroxamato, albizziina, 5'azacitidina; argininosuccinato sintetasa y canavanina; ornitina descarboxilasa y α-difluorometil-ornitina; HMG-CoA-reductasa y compactina; N-acetilglucosaminil transferasa y tunicamicina; treonil-tARN sintetasa y borrelidina; y Na^+K^+-ATPasa y ouabaína.

La célula huésped que comprende el casete de expresión o el vector

En otro aspecto, la presente invención proporciona una célula huésped que comprende el casete de expresión de acuerdo con la invención o el vector de acuerdo con la invención. La célula huésped puede ser cualquier célula adecuada para la transfección con el casete o vector de expresión y, en particular, adecuada para la producción del polipéptido de interés. En ciertas realizaciones, la célula huésped se deriva de una línea celular de expresión establecida. La célula huésped en particular es una célula eucariota, tal como una célula de mamífero, en particular una célula humana, o una célula derivada de la misma. En particular, la célula huésped es una célula sanguínea, tal como un glóbulo blanco, una célula precursora sanguínea o una célula leucémica, o una célula derivada de las mismas. En ciertas realizaciones, la célula huésped es una célula de origen leucocitario.

En realizaciones específicas, la célula huésped se deriva de células de leucemia mieloide humana. Ejemplos específicos de células huésped son K562, NM-F9, NM-D4, NM-H9D8, NM-H9D8-E6, NM-H9D8-E6Q12, GT-2X, GT-5s y células derivadas de cualquiera de dichas células huésped. K562 es una línea celular de leucemia mieloide humana presente en la American Type Culture Collection (ATCC CCL-243). Las líneas celulares restantes se derivan de las células K562 y se han seleccionado por características de glicosilación específicas. Las líneas celulares derivadas de K562 pueden cultivarse y mantenerse en las condiciones bien conocidas adecuadas para K562. Todas estas líneas celulares, excepto las células K562, se depositaron de acuerdo con el tratado de Budapest. La información sobre la deposición se puede encontrar al final de la memoria descriptiva.

Los ejemplos de células huésped también se describen, por ejemplo, en el documento WO 2008/028686. En ciertas realizaciones, la célula huésped está optimizada para la expresión de glicoproteínas que tienen un patrón de glicosilación específico. En particular, el uso de codones en la región de codificación y/o el promotor y los elementos adicionales del casete o vector de expresión son compatibles y en particular optimizados para el tipo de célula huésped utilizada.

En ciertas realizaciones, la célula huésped es una célula huésped aislada. En realizaciones específicas, la célula huésped no está presente en el cuerpo humano o animal.

La célula huésped puede transfectarse de forma transitoria o estable con el casete o vector de expresión de acuerdo con la invención. Se prefiere la transfección estable, en particular mediante la integración del casete de expresión en el genoma de la célula huésped. Los métodos de transfección para la transfección estable o transitoria son comúnmente conocidos en la técnica. En ciertas realizaciones, la célula huésped se transfecta con el vector que comprende el casete de expresión con la región de codificación que comprende una secuencia de ácido nucleico que codifica un polipéptido de interés.

El método de producción

De acuerdo con otro aspecto, la presente invención proporciona un método *in vitro* para producir de forma recombinante un polipéptido de interés, que comprende las etapas de

(a) proporcionar una célula huésped que comprende un casete de expresión que comprende, funcionalmente unidos entre sí,

(i) una región de control del locus que comprende al menos el elemento central del sitio 2 de hipersensibilidad a la ADNasa I (HS2) del grupo de genes de β-globina humana;

(ii) una región promotora que comprende al menos una parte funcional del promotor del gen de la $^A\gamma$ globina humana; y

(iii) una región de codificación que comprende una secuencia de ácido nucleico que codifica el polipéptido de interés;

(b) cultivar la célula huésped en condiciones en las que la célula huésped expresa el polipéptido de interés; y

(c) aislar el polipéptido de interés.

La célula huésped en particular comprende un casete de expresión como se define en el presente documento, que tiene uno o más de los elementos definidos en este documento.

Las condiciones adecuadas para cultivar las células huésped y expresar el polipéptido de interés dependen de la célula huésped específica, el vector y el casete de expresión utilizados en el método. La persona experta puede determinar fácilmente las condiciones adecuadas y también son conocidas en la técnica para una pluralidad de células huésped. En ciertas realizaciones, la célula huésped se transfecta con un vector que comprende el casete de expresión y que además comprende un gen marcador seleccionable. En estas realizaciones, las condiciones de cultivo en la etapa (b) pueden incluir la presencia de un agente de selección correspondiente en el medio de cultivo celular.

El aislamiento del polipéptido de interés en particular se refiere a la separación del polipéptido de interés de los componentes restantes del cultivo celular. En ciertas realizaciones, la región de codificación del casete de expresión comprende además una secuencia de ácido nucleico que codifica un péptido señal para la expresión secretora, y en la etapa (b) el polipéptido de interés es secretado por la célula huésped. En estas realizaciones, la etapa (c) en particular comprende separar el medio de cultivo celular que comprende el polipéptido de interés de las células huésped, por ejemplo por centrifugación, y separar el polipéptido de interés de algunos o la mayoría de los componentes del medio de cultivo celular, por ejemplo por métodos cromatográficos. Los métodos y medios adecuados para aislar el polipéptido de interés son conocidos en la técnica y la persona experta puede aplicarlos fácilmente.

En ciertas realizaciones, el método para producir un polipéptido de interés comprende además después de la etapa (c) la etapa de

(d) formular el polipéptido de interés como una composición farmacéutica. La formulación del polipéptido de interés como una composición farmacéutica en particular comprende el intercambio de la solución tampón o los componentes de la solución tampón de la composición que comprende el polipéptido de interés. Además, la etapa de formulación puede incluir liofilización del polipéptido de interés. En particular, el polipéptido de interés se transfiere a una composición que solo comprende ingredientes farmacéuticamente aceptables.

Los intervalos numéricos descritos en este documento incluyen los números que definen el intervalo. Los títulos proporcionados en el presente documento no son limitaciones de los diversos aspectos o realizaciones de esta invención que pueden leerse con referencia a la especificación en su conjunto. De acuerdo con una realización, la materia descrita en el presente documento que comprende ciertas etapas en el caso de métodos o que comprende ciertos ingredientes en el caso de composiciones se refiere a la materia que consiste en las etapas o ingredientes respectivos. Se prefiere seleccionar y combinar aspectos y realizaciones específicos descritos en este documento y la materia específica que surge de una combinación respectiva de realizaciones específicas también pertenece a la presente descripción.

Figuras

La Figura 1 muestra la estructura de los grupos de genes de la globina humana, incluida la región de control del locus (LCR) con los diferentes sitios de hipersensibilidad a la ADNasa (HS) y los diferentes genes de la globina. A: grupo de genes de la β-globina humana en el cromosoma 11; B: grupo de genes de la α-globina humana en el cromosoma 16.

La Figura 2 muestra los elementos de ejemplos de casetes de expresión como los utilizados en los vectores pHBG1A-E. HS: sitio de hipersensibilidad a la ADNasa; $^{A}\gamma$-Prom: promotor del gen de la $^{A}\gamma$ globina; CS: secuencia de codificación/sitio de clonación; γpA: señal de poliadenilación del gen de la $^{A}\gamma$ globina; $^{A}\gamma$-ENh: potenciador 3' del gen de la $^{A}\gamma$ globina.

La Figura 3 muestra el rendimiento de proteína del factor VII obtenido después de la transfección transitoria de diferentes vectores que comprenden la secuencia de codificación del factor VII. Los vectores que comprenden los casetes de expresión mostrados en la Figura 2 con la secuencia de codificación del factor VII introducido en el sitio de clonación y un gen que codifica DHFR como marcador seleccionable amplificable se transfectaron transitoriamente en células NM-H9D8. El rendimiento total del factor VII se determinó después del cultivo. pEFdhfrmut(-): vector de control con secuencia de codificación del factor VII. Se muestran los resultados de tres experimentos independientes.

La Figura 4 muestra una comparación de la transfección estable de un vector de acuerdo con la invención y un vector de control que codifica un anticuerpo. Las células NM-H9D8-E6Q12 se transfectaron de manera estable con el vector de control pEF o el vector pHB de acuerdo con la invención. Ambos vectores comprenden una secuencia de codificación para un anticuerpo y un gen que codifica DHFR como marcador seleccionable amplificable. Para la amplificación del vector en las células, se aumentó gradualmente la presión de selección, es decir, la concentración del agente de selección metotrexato en el medio de cultivo. El gráfico muestra la presión de selección máxima que fue posible para el vector respectivo después de un tiempo de cultivo dado. Una mayor presión de selección posible (concentración de metotrexato) indica una amplificación más fuerte del vector en las células transfectadas, lo que debería dar como resultado una mayor producción de la proteína de interés.

La Figura 5 muestra la productividad del conjunto de las células transfectadas de forma estable de la Figura 4. La producción de anticuerpos en picogramos por célula por día se muestra para las diferentes presiones de selección para el vector de acuerdo con la invención y el vector de control.

La Figura 6 muestra el aumento de la productividad de las células transfectadas de forma estable de la Figura 4 por amplificación del vector debido a la presión de selección. La producción de anticuerpos en picogramos por célula por día se muestra para las diferentes presiones de selección para el grupo celular inicial, para el grupo celular después de la amplificación y para los clones de células individuales después de la amplificación. A: vector de control pEF; B: vector pHB de acuerdo con la invención.

Ejemplos

Ejemplo 1: Construcción de vectores que comprenden el promotor de la $^{A}\gamma$ globina humana y elementos de las regiones de control del locus de los grupos de genes de la globina humana

Para la construcción de los vectores de globina, se eliminaron las regiones potenciadoras y promotoras de un vector parental (por ejemplo, pEF que tiene un gen de resistencia a la puromicina o neomicina o un gen dhfr como marcador seleccionable). El promotor de la $^{A}\gamma$ globina humana, la señal de poliadenilación y las regiones potenciadoras 3' y diferentes constructos de las regiones de control del locus de los grupos de genes de la globina humana se sintetizaron y clonaron en el vector en los sitios apropiados. La Figura 2 muestra ejemplos de constructos de los casetes de expresión de los vectores construidos. Luego, se introdujo una secuencia de ácido nucleico que codifica un polipéptido de interés en el sitio de clonación.

Ejemplo 2: transfección transitoria de los vectores de globina

La transfección transitoria se realizó con Lipofectamina® LTX y el reactivo Plus^{MR} de acuerdo con las instrucciones del fabricante. Brevemente, se sembraron 2 x 10^5 células en placas de 6 pozos durante su fase de crecimiento logarítmico. El ADN plasmídico se diluyó en medio de suero reducido Opti-MEM I y reactivo Plus^{MR}. Después de un tiempo de incubación (15 min), se añadió Lipofectamina® LTX a la solución. Después de una incubación adicional (30 min), la mezcla se añadió gota a gota en la suspensión celular. La expresión se analizó después de 72 horas por ELISA. Se obtuvieron títulos de proteína más altos para los vectores pHBG1Cdhfr, pHBG1Ddhfr y pHBG1Edhfr en comparación con el vector pEFdhfrmut(-) (Figura 3).

Ejemplo 3: transfección estable de los vectores de globina

La transfección de la línea celular NM-H9D8 se realizó por nucleofección (Tecnología Nucleofector^{MR}, Amaxa) usando ADN plasmídico de los dos plásmidos de expresión que codifican la cadena pesada y ligera del anticuerpo, respectivamente (ambas linealizadas) de acuerdo con las instrucciones del fabricante. Para la selección y amplificación de los grupos productores de anticuerpos, se agregaron metotrexato y puromicina a concentraciones crecientes y los grupos se seleccionaron por la secreción de moléculas de anticuerpos activos.

Los grupos transfectados con plásmidos pHB podrían amplificarse en un período de tiempo más corto (Figura 4) y conducir a niveles de proteína más altos (Figura 5) que podrían confirmarse para los clones de células individuales resultantes, respectivamente (Figura 6).

Identificación del material biológico depositado.

Las líneas celulares DSM ACC 2606 y DSM ACC 2605 se depositaron en la DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Mascheroder Weg 1b, 38124 Braunschweig (DE) por Nemod Biotherapeutics GmbH & Co. KG, Robert-Rössle-Str. 10, 13125 Berlín (DE) el 14 de agosto de 2003. Glycotope tiene derecho a referirse a estos materiales biológicos ya que mientras tanto fueron asignados por Nemod Biotherapeutics GmbH & Co. KG a Glycotope GmbH.

Las líneas celulares DSM ACC 2806, DSM ACC 2807, DSM ACC 2856, DSM ACC 2858 y DSM ACC 3078 se depositaron en la DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, InhoffenstraBe 7B, 38124 Braunschweig (DE) por Glycotope GmbH Robert-Rössle-Str. 10, 13125 Berlín (DE) en las fechas indicadas en la siguiente tabla.

Nombre de la línea celular	Número de Acceso	Depositante	Fecha de depósito
NM-F9	DSM ACC 2606	Nemod Biotherapeutics	14 de agosto de 2003
NM-D4	DSM ACC 2605	Nemod Biotherapeutics	14 de agosto de 2003
NM-H9D8	DSM ACC 2806	Glycotope GmbH	15 de septiembre de 2006
NM-H9D8-E6	DSM ACC 2807	Glycotope GmbH	5 de octubre de 2006
NM-H9D8-E6Q12	DSM ACC 2856	Glycotope GmbH	8 de agosto de 2007
GT-2x	DSM ACC 2858	Glycotope GmbH	7 de septiembre de 2007
GT-5s	DSM ACC 3078	Glycotope GmbH	28 de julio de 2010

Listado de secuencias

<110> Glycotope GmbH

<120> Vectores de expresión eucariotas que incluyen elementos reglamentarios del grupo de genes de globina

<130> 56 981 K

<150> LU92686

<151> 2015-03-31

<150> EP15161922.8
<151> 2015-03-31

<160> 5

<170> BiSSAP 1.0

<210> 1

<211> 2899

<212> ADN

<213> secuencias artificiales

<220>
<221> fuente

<222> 1..2899

<223> /tipo de molécula = "ADN" / nota = " casete de expresión HBG1A" / organismo = "secuencias artificiales"

<400> 1

tcgcgatgat caacttttag agagctcttg gggaccccag tacacaagag gggacgcagg 60

gtatatgtag acatctcatt ctttttctta gtgtgagaat aagaatagcc atgacctgag 120

tttatagaca atgagccctt ttctctctcc cactcagcag ctatgagatg gcttgccctg 180

cctctctact aggctgactc actccaaggc ccagcaatgg gcagggctct gtcagggctt 240

tgatagcact atctgcagag ccagggccga gaaggggtgg actccagaga ctctcctgat 300

cattaattaa gctactgctc atgggccctg tgctgcactg atgaggagga tcagatggat 360

ggggcaatga agcaaaggaa tcattctgtg gataaaggag acagccatga agaagtctat 420

gactgtaaat ttgggagcag gagtctctaa ggacttggat ttcaaggaat tttgactcag 480

caaacacaag accctcacgg tgactttgcg agctggtgtg ccagatgtgt ctatcagagg 540

ttccagggag ggtggggtgg ggtcagggct ggccaccagc tatcagggcc cagatgggtt 600

ataggctggc aggctcagat aggtggttag gtcaggttgg tggtgctggg tggagtccat 660

gactcccagg agccaggaga gatagaccat gagtagaggg cagacatggg aaaggtgggg 720

gaggcacagc atagcttaat taagccagtt tttccttagt tcctgttaca tttctgtgtg 780

tctccattag tgacctccca tagtccaagc atgagcagtt ctggccaggc ccctgtcggg 840

gtcagtgccc caccccgcc ttctggttct gtgtaacctt ctaagcaaac cttctggctc 900

aagcacagca atgctgagtc atgatgagtc atgctgaggc ttagggtgtg tgcccagatg 960

ttctcagcct agagtgatga ctcctatctg ggtccccagc aggatgctta cagggcagat 1020

ggcaaaaaaa aggagaagct gaccacctga ctaaaactcc acctcaaacg gcatcataaa 1080

gaaaatggat gcctgagaca gaatgtgacg catttaaatg atcctcactg gagctacaga 1140

caagaaggta aaaaacggct gacaaaagaa gtcctggtat cctctatgat gggagaagga 1200

aactagctaa agggaagaat aaattagaga aaaactggaa tgactgaatc ggaacaaggc 1260

aaaggctata aaaaaaatta agcagcagta tcctcttggg ggccccttcc ccacactatc 1320

tcaatgcaaa tatctgtctg aaacggtccc tggctaaact ccacccatgg gttggccagc 1380

cttgccttga ccaatagcct tgacaaggca aacttgacca atagtcttag agtatccagt 1440

gaggccaggg gccggcggct ggctagggat gaagaataaa aggaagcacc cttcagcagt 1500

tccacacact cgcttctgga acgtctgaga ttatcaataa gctcctagtc cagacgccaa 1560

gcttggtacc gagctcggat ccactagtaa cggccggcca gtgtgctgga attctgcaga 1620

tatccatcac actgcccggg cggccgctcg agcatcgatc tagagcctct tgcccatgat 1680

tcagagcttt caaggatagg ctttattctg caagcaatac aaataataaa tctattctgc 1740

tgagagatca cacatgattt tcttcagctc tttttttac atctttttaa atatatgagc 1800

cacaaagggt ttatattgag ggaagtgtgt atgtgtattt ctgcatgcct gtttgtgttt 1860

gtggtgtgtg catgctcctc atttatttt atatgagatg tgcattttga tgagcaaata 1920

aaagcagtaa agacacttgt acacgggagt tctgcaagtg ggagtaaatg gtgttggaga 1980

aatccggtgg gaagaaagac ctctatagga caggacttct cagaaacaga tgttttggaa 2040

gagatgggaa aaggttcagt gaagacctgg gggctggatt gattgcagct gagtagcaag 2100

gatggttctt aatgaaggga aagtgttcca agctggaatt caaggtttag tcaggtgtag 2160

caattctatt ttattaggag gaatactatt tctaatggca cttagctttt cacagccctt 2220

gtggatgcct aagaaagtga aattaatccc atgccctcaa gtgtgcagat tggtcacagc 2280

atttcaaggg agagacctca ttgtaagact ctgggggagg tggggactta ggtgtaagaa 2340

atgaatcagc agaggctcac aagtcagcat gagcatgtta tgtctgagaa acagaccagc 2400

actgtgagat caaaatgtag tgggaagaat ttgtacaaca ttaattggaa ggtttactta 2460

atggaatttt tgtatagttg gatgttagtg catctctata agtaagagtt taatatgatg 2520

gtgttacgga cctggtgttt gtgtctcctc aaaattcaca tgctgaatcc ccaactccca 2580

actgacctta tctgtggggg aggcttttga aaagtaatta ggtttagctg agctcataag 2640

agcagatccc catcataaaa ttattttcct tatcagaagc agagagacaa gccatttctc 2700

tttcctcccg gtgaggacac agtgagaagt ccgccatctg caatccagga agagaaccct 2760

gaccacgagt cagccttcag aaatgtgaga aaaaactctg ttgttgaagc cacccagtct 2820

tttgtatttt gttatagcac cttacactga gtaaggcaga tgaagaagga gaaaaaaata 2880

acagctggtt aactacgta 2899

<210> 2

<211> 2609

<212> ADN

<213> secuencias artificiales

<220>
<221> fuente

<222> 1..2609

<223> /tipo de molécula = "ADN" / nota = " casete de expresión HBG1B" / organismo = "secuencias artificiales"

<400> 2

tcgcgatgat cattaattaa gctactgctc atgggccctg tgctgcactg atgaggagga 60

tcagatggat ggggcaatga agcaaaggaa tcattctgtg gataaaggag acagccatga 120

agaagtctat gactgtaaat ttgggagcag gagtctctaa ggacttggat ttcaaggaat 180

tttgactcag caaacacaag accctcacgg tgactttgcg agctggtgtg ccagatgtgt 240

ctatcagagg ttccagggag ggtggggtgg ggtcagggct ggccaccagc tatcagggcc 300

cagatgggtt ataggctggc aggctcagat aggtggttag gtcaggttgg tggtgctggg 360

tggagtccat gactcccagg agccaggaga gatagaccat gagtagaggg cagacatggg 420

aaaggtgggg gaggcacagc atagcttaat taagccagtt tttccttagt tcctgttaca 480

tttctgtgtg tctccattag tgacctccca tagtccaagc atgagcagtt ctggccaggc 540

ccctgtcggg gtcagtgccc caccccgcc ttctggttct gtgtaacctt ctaagcaaac 600

cttctggctc aagcacagca atgctgagtc atgatgagtc atgctgaggc ttagggtgtg 660

tgcccagatg ttctcagcct agagtgatga ctcctatctg ggtccccagc aggatgctta 720

cagggcagat ggcaaaaaaa aggagaagct gaccacctga ctaaaactcc acctcaaacg 780

gcatcataaa gaaaatggat gcctgagaca gaatgtgacg catttaaatg atcctcactg 840

gagctacaga caagaaggta aaaaacggct gacaaaagaa gtcctggtat cctctatgat 900

gggagaagga aactagctaa agggaagaat aaattagaga aaaactggaa tgactgaatc 960

ggaacaaggc aaaggctata aaaaaaatta agcagcagta tcctcttggg ggccccttcc 1020

ccacactatc tcaatgcaaa tatctgtctg aaacggtccc tggctaaact ccacccatgg 1080

gttggccagc cttgccttga ccaatagcct tgacaaggca aacttgacca atagtcttag 1140

agtatccagt gaggccaggg gccggcggct ggctagggat gaagaataaa aggaagcacc 1200

cttcagcagt tccacacact cgcttctgga acgtctgaga ttatcaataa gctcctagtc 1260

cagacgccaa gcttggtacc gagctcggat ccactagtaa cggccggcca gtgtgctgga 1320

attctgcaga tatccatcac actgcccggg cggccgctcg agcatcgatc tagagcctct 1380

tgcccatgat tcagagcttt caaggatagg ctttattctg caagcaatac aaataataaa 1440

tctattctgc tgagagatca cacatgattt tcttcagctc ttttttttac atcttttaa 1500

atatatgagc cacaaagggt ttatattgag ggaagtgtgt atgtgtattt ctgcatgcct 1560

gtttgtgttt gtggtgtgtg catgctcctc atttattttt atatgagatg tgcattttga 1620

tgagcaaata aaagcagtaa agacacttgt acacgggagt tctgcaagtg ggagtaaatg 1680

gtgttggaga aatccggtgg gaagaaagac ctctatagga caggacttct cagaaacaga 1740

tgttttggaa gagatgggaa aaggttcagt gaagacctgg gggctggatt gattgcagct 1800

gagtagcaag gatggttctt aatgaaggga aagtgttcca agctggaatt caaggtttag 1860

tcaggtgtag caattctatt ttattaggag gaatactatt tctaatggca cttagctttt 1920

cacagccctt gtggatgcct aagaaagtga aattaatccc atgccctcaa gtgtgcagat 1980

tggtcacagc atttcaaggg agagacctca ttgtaagact ctgggggagg tggggactta 2040

ggtgtaagaa atgaatcagc agaggctcac aagtcagcat gagcatgtta tgtctgagaa 2100

acagaccagc actgtgagat caaaatgtag tgggaagaat ttgtacaaca ttaattggaa 2160

ggtttactta atggaatttt tgtatagttg gatgttagtg catctctata agtaagagtt 2220

taatatgatg gtgttacgga cctggtgttt gtgtctcctc aaaattcaca tgctgaatcc 2280

ccaactccca actgacctta tctgtggggg aggcttttga aaagtaatta ggtttagctg 2340

agctcataag agcagatccc catcataaaa ttattttcct tatcagaagc agagagacaa 2400

gccatttctc tttcctcccg gtgaggacac agtgagaagt ccgccatctg caatccagga 2460

agagaaccct gaccacgagt cagccttcag aaatgtgaga aaaaactctg ttgttgaagc 2520

cacccagtct tttgtatttt gttatagcac cttacactga gtaaggcaga tgaagaagga 2580

gaaaaaaata acagctggtt aactacgta 2609

<210> 3

<211> 2176

<212> ADN

<213> secuencias artificiales

<220>
<221> fuente

<222> 1..2176

<223> /tipo de molécula = "ADN" / nota = " casete de expresión HBG1C" / organismo = "secuencias artificiales"

<400> 3

tcgcgatgat cattaattaa gccagttttt ccttagttcc tgttacattt ctgtgtgtct 60

ccattagtga cctcccatag tccaagcatg agcagttctg gccaggcccc tgtcggggtc 120

agtgccccac ccccgccttc tggttctgtg taaccttcta agcaaacctt ctggctcaag 180

cacagcaatg ctgagtcatg atgagtcatg ctgaggctta gggtgtgtgc ccagatgttc 240

tcagcctaga gtgatgactc ctatctgggt ccccagcagg atgcttacag ggcagatggc 300

aaaaaaaagg agaagctgac cacctgacta aaactccacc tcaaacggca tcataaagaa 360

aatggatgcc tgagacagaa tgtgacgcat ttaaatgatc ctcactggag ctacagacaa 420

gaaggtaaaa aacggctgac aaaagaagtc ctggtatcct ctatgatggg agaaggaaac 480

tagctaaagg gaagaataaa ttagagaaaa actggaatga ctgaatcgga acaaggcaaa 540

ggctataaaa aaaattaagc agcagtatcc tcttgggggc cccttcccca cactatctca 600

atgcaaatat ctgtctgaaa cggtccctgg ctaaactcca cccatgggtt ggccagcctt 660

gccttgacca atagccttga caaggcaaac ttgaccaata gtcttagagt atccagtgag 720

gccaggggcc ggcggctggc tagggatgaa gaataaaagg aagcaccctt cagcagttcc 780

acacactcgc ttctggaacg tctgagatta tcaataagct cctagtccag acgccaagct 840

tggtaccgag ctcggatcca ctagtaacgg ccggccagtg tgctggaatt ctgcagatat 900

ccatcacact gcccgggcgg ccgctcgagc atcgatctag agcctcttgc ccatgattca 960

gagctttcaa ggataggctt tattctgcaa gcaatacaaa taataaatct attctgctga 1020

gagatcacac atgattttct tcagctcttt tttttacatc tttttaaata tatgagccac 1080

aaagggttta tattgaggga agtgtgtatg tgtatttctg catgcctgtt tgtgtttgtg 1140

gtgtgtgcat gctcctcatt tatttttata tgagatgtgc attttgatga gcaaataaaa 1200

gcagtaaaga cacttgtaca cgggagttct gcaagtggga gtaaatggtg ttggagaaat 1260

ccggtgggaa gaaagacctc tataggacag gacttctcag aaacagatgt tttggaagag 1320

atgggaaaag gttcagtgaa gacctggggg ctggattgat tgcagctgag tagcaaggat 1380

ggttcttaat gaagggaaag tgttccaagc tggaattcaa ggtttagtca ggtgtagcaa 1440

ttctatttta ttaggaggaa tactatttct aatggcactt agcttttcac agcccttgtg 1500

gatgcctaag aaagtgaaat taatcccatg ccctcaagtg tgcagattgg tcacagcatt 1560

tcaagggaga gacctcattg taagactctg ggggaggtgg ggacttaggt gtaagaaatg 1620

aatcagcaga ggctcacaag tcagcatgag catgttatgt ctgagaaaca gaccagcact 1680

gtgagatcaa aatgtagtgg gaagaatttg tacaacatta attggaaggt ttacttaatg 1740

gaatttttgt atagttggat gttagtgcat ctctataagt aagagtttaa tatgatggtg 1800

ttacggacct ggtgtttgtg tctcctcaaa attcacatgc tgaatcccca actcccaact 1860

gaccttatct gtgggggagg cttttgaaaa gtaattaggt ttagctgagc tcataagagc 1920

agatccccat cataaaatta ttttccttat cagaagcaga gagacaagcc atttctcttt 1980

cctcccggtg aggacacagt gagaagtccg ccatctgcaa tccaggaaga gaaccctgac 2040

cacgagtcag ccttcagaaa tgtgagaaaa aactctgttg ttgaagccac ccagtctttt 2100

gtattttgtt atagcacctt acactgagta aggcagatga agaaggagaa aaaaataaca 2160

gctggttaac tacgta 2176

<210> 4

<211> 2054

<212> ADN

<213> secuencias artificiales

<220>
<221> fuente

<222> 1..2054

<223> /tipo de molécula = "ADN" / nota = "casete de expresión HBG1D" / organismo = "secuencias artificiales"

<400> 4

tcgcgatgat caaagccagt ttttccttag ttcctgttac atttctgtgt gtctccatta 60

gtgacctccc atagtccaag catgagcagt tctggccagg cccctgtcgg ggtcagtgcc 120

ccacccccgc cttctggttc tgtgtaacct tctaagcaaa ccttctggct caagcacagc 180

aatgctgagt catgatgagt catgctgagg cttagggtgt gtgcccagat gttctcagcc 240

tagagtgatg actcctatct gggtccattt aaatgatcct cactggagct acagacaaga 300

aggtaaaaaa cggctgacaa aagaagtcct ggtatcctct atgatgggag aaggaaacta 360

gctaaaggga agaataaatt agagaaaaac tggaatgact gaatcggaac aaggcaaagg 420

ctataaaaaa aattaagcag cagtatcctc ttgggggccc cttccccaca ctatctcaat 480

gcaaatatct gtctgaaacg gtccctggct aaactccacc catgggttgg ccagccttgc 540

cttgaccaat agccttgaca aggcaaactt gaccaatagt cttagagtat ccagtgaggc 600

caggggccgg cggctggcta gggatgaaga ataaaaggaa gcacccttca gcagttccac 660

acactcgctt ctggaacgtc tgagattatc aataagctcc tagtccagac gccaagcttg 720

gtaccgagct cggatccact agtaacggcc ggccagtgtg ctggaattct gcagatatcc 780

atcacactgc ccgggcggcc gctcgagcat cgatctagag cctcttgccc atgattcaga 840

gctttcaagg ataggcttta ttctgcaagc aatacaaata ataaatctat tctgctgaga 900

gatcacacat gattttcttc agctcttttt tttacatctt tttaaatata tgagccacaa 960

agggtttata ttgagggaag tgtgtatgtg tatttctgca tgcctgtttg tgtttgtggt 1020

gtgtgcatgc tcctcattta tttttatatg agatgtgcat tttgatgagc aaataaaagc 1080

agtaaagaca cttgtacacg ggagttctgc aagtgggagt aaatggtgtt ggagaaatcc 1140

ggtgggaaga aagacctcta taggacagga cttctcagaa acagatgttt tggaagagat 1200

gggaaaaggt tcagtgaaga cctgggggct ggattgattg cagctgagta gcaaggatgg 1260

ttcttaatga agggaaagtg ttccaagctg gaattcaagg tttagtcagg tgtagcaatt 1320

ctattttatt aggaggaata ctatttctaa tggcacttag cttttcacag cccttgtgga 1380

tgcctaagaa agtgaaatta atcccatgcc ctcaagtgtg cagattggtc acagcatttc 1440

aagggagaga cctcattgta agactctggg ggaggtgggg acttaggtgt aagaaatgaa 1500

tcagcagagg ctcacaagtc agcatgagca tgttatgtct gagaaacaga ccagcactgt 1560

gagatcaaaa tgtagtggga agaatttgta caacattaat tggaaggttt acttaatgga 1620

atttttgtat agttggatgt tagtgcatct ctataagtaa gagtttaata tgatggtgtt 1680

acggacctgg tgtttgtgtc tcctcaaaat tcacatgctg aatccccaac tcccaactga 1740

ccttatctgt gggggaggct tttgaaaagt aattaggttt agctgagctc ataagagcag 1800

atccccatca taaaattatt ttccttatca gaagcagaga gacaagccat ttctctttcc 1860

tcccggtgag gacacagtga gaagtccgcc atctgcaatc caggaagaga accctgacca 1920

cgagtcagcc ttcagaaatg tgagaaaaaa ctctgttgtt gaagccaccc agtcttttgt 1980

attttgttat agcaccttac actgagtaag gcagatgaag aaggagaaaa aaataacagc 2040

tggttaacta cgta 2054

<210> 5

<211> 2160

<212> ADN

<213> secuencias artificiales

<220>
<221> fuente

<222> 1..2160

<223> /tipo de molécula = "ADN" / nota = "casete de expresión HBG1E" / organismo = "secuencias artificiales"

<400> 5

tcgcgatgct ctcaggaaga ccctctggaa cctatcaggg accacagtca gccaggcaag 60

cacatctgcc caagccaagg gtggaggcat gcagctgtgg gggtctgtga aaacacttga 120

gggagcagat aactgggcca accatgactc agtgcttctg gaggccaaca ggactgctga 180

gtcatcctgt gggggtggag gtgggacaag ggaaaggggt gaatggtact gctgattaca 240

acctctggtg ctgcctcccc ctcctgttta tctgagaggg aaggccatgc ccaaagtgtt 300

cacagccagg cttcaggggc aaagcctgac ccagacagta aatacgttct tcatctggag 360

ctgaagaaat tcatttaaat gatcctcact ggagctacag acaagaaggt aaaaaacggc 420

tgacaaaaga agtcctggta tcctctatga tgggagaagg aaactagcta aagggaagaa 480

taaattagag aaaaactgga atgactgaat cggaacaagg caaaggctat aaaaaaaatt 540

aagcagcagt atcctcttgg gggccccttc cccacactat ctcaatgcaa atatctgtct 600

gaaacggtcc ctggctaaac tccacccatg ggttggccag ccttgccttg accaatagcc 660

ttgacaaggc aaacttgacc aatagtctta gagtatccag tgaggccagg ggccggcggc 720

tggctaggga tgaagaataa aaggaagcac ccttcagcag ttccacacac tcgcttctgg 780

aacgtctgag attatcaata agctcctagt ccagacgcca agcttggtac cgagctcgga 840

tccactagta acggccggcc agtgtgctgg aattctgcag atatccatca cactgcccgg 900

gcggccgctc gagcatcgat ctagagcctc ttgcccatga ttcagagctt tcaaggatag 960

gctttattct gcaagcaata caaataataa atctattctg ctgagagatc acacatgatt 1020

ttcttcagct ctttttttta catcttttta aatatatgag ccacaaaggg tttatattga 1080

gggaagtgtg tatgtgtatt tctgcatgcc tgtttgtgtt tgtggtgtgt gcatgctcct 1140

catttatttt tatatgagat gtgcattttg atgagcaaat aaaagcagta aagacacttg 1200

tacacgggag ttctgcaagt gggagtaaat ggtgttggag aaatccggtg ggaagaaaga 1260

cctctatagg acaggacttc tcagaaacag atgttttgga agagatggga aaaggttcag 1320

tgaagacctg ggggctggat tgattgcagc tgagtagcaa ggatggttct taatgaaggg 1380

aaagtgttcc aagctggaat tcaaggttta gtcaggtgta gcaattctat tttattagga 1440

ggaatactat ttctaatggc acttagcttt tcacagccct tgtggatgcc taagaaagtg 1500

aaattaatcc catgccctca agtgtgcaga ttggtcacag catttcaagg gagagacctc 1560

attgtaagac tctgggggag gtggggactt aggtgtaaga aatgaatcag cagaggctca 1620

caagtcagca tgagcatgtt atgtctgaga aacagaccag cactgtgaga tcaaaatgta 1680

gtgggaagaa tttgtacaac attaattgga aggtttactt aatggaattt ttgtatagtt 1740

ggatgttagt gcatctctat aagtaagagt ttaatatgat ggtgttacgg acctggtgtt 1800

tgtgtctcct caaaattcac atgctgaatc cccaactccc aactgacctt atctgtgggg 1860

gaggcttttg aaaagtaatt aggtttagct gagctcataa gagcagatcc ccatcataaa 1920

attattttcc ttatcagaag cagagagaca agccatttct ctttcctccc ggtgaggaca 1980

cagtgagaag tccgccatct gcaatccagg aagagaaccc tgaccacgag tcagccttca 2040

gaaatgtgag aaaaaactct gttgttgaag ccacccagtc ttttgtattt tgttatagca 2100

ccttacactg agtaaggcag atgaagaagg agaaaaaaat aacagctggt taactacgta 2160

REIVINDICACIONES

1. Un método *in vitro* para producir recombinantemente un polipéptido de interés, que comprende las etapas de

(a) proporcionar una célula huésped que comprende un casete de expresión que comprende, funcionalmente unidos entre sí,

(i) una región de control del locus que comprende al menos el elemento central del sitio 2 de hipersensibilidad a la ADNasa I (HS2) del grupo de genes de la β-globina humana;

(ii) una región promotora que comprende al menos una parte funcional del promotor del gen de la $^{A}\gamma$ globina humana; y

(iii) una región de codificación que comprende una secuencia de ácido nucleico que codifica el polipéptido de interés;

(b) cultivar la célula huésped en condiciones en las que la célula huésped expresa el polipéptido de interés; y

(c) aislar el polipéptido de interés.

2. El método de acuerdo con la reivindicación 1, que comprende además después de la etapa (c) la etapa de

(d) formular el polipéptido de interés como una composición farmacéutica.

3. El método de acuerdo con la reivindicación 1 o 2, en el que el elemento central del sitio 2 de hipersensibilidad a la ADNasa I (HS2) del grupo de genes de la β-globina humana tiene la secuencia de ácido nucleico de la posición 906 a 939 de la SEQ ID NO: 1.

4. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en el que la región de control del locus del casete de expresión comprende

(i) el elemento M1-central-M2 del sitio 2 de hipersensibilidad a la ADNasa I (HS2) del grupo de genes de la β-globina humana, que tiene preferiblemente la secuencia de ácido nucleico de la posición 742 a 995 de la SEQ ID NO: 1; y

(ii) opcionalmente al menos una parte del sitio 2 de hipersensibilidad (HS2) del grupo de genes de la β-globina humana que comprende la secuencia de ácido nucleico de la posición 741 a 1109 de la SEQ ID NO: 1; y

(iii) opcionalmente, el sitio 3 de hipersensibilidad (HS3) o una parte del mismo del grupo de genes de la β-globina humana, que tiene la secuencia de ácido nucleico de la posición 310 a 735 de la SEQ ID NO: 1, y/o el sitio 4 de hipersensibilidad (HS4) o una parte del mismo del grupo de genes de la β-globina humana, que tiene la secuencia de ácido nucleico de la posición 13 a 294 de la SEQ ID NO: 1.

5. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en el que el casete de expresión comprende uno o más de los siguientes:

(i) en la región promotora, un sitio de inicio de la transcripción y, opcionalmente, al menos parte de la región 5' no traducida del gen de la $^{A}\gamma$ globina humana; y/o

(ii) en los nucleótidos de la región promotora -384 a +36, con respecto al sitio de inicio de la transcripción, del gen de la $^{A}\gamma$ globina humana, y/o la secuencia de ácido nucleico de la posición 1123 a 1542 de la SEQ ID NO: 1 ; y/o

(iii) una región de terminación de la transcripción, en la que la región de terminación de la transcripción comprende opcionalmente una señal de poliadenilación, que comprende preferiblemente la secuencia de ácido nucleico de la posición 1725 a 1730 de la SEQ ID NO: 1, y un sitio de terminación de la transcripción; y/o

(iv) una región potenciadora que comprende al menos una parte funcional del potenciador 3' del gen de la $^{A}\gamma$ globina humana, unida funcionalmente a los otros elementos del casete de expresión, en la que la región potenciadora del casete de expresión comprende la secuencia de ácido nucleico de la posición 2136 a 2881 de SEQ ID NO: 1.

6. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que el polipéptido de interés es una glicoproteína o una parte de la misma.

7. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 6, en el que el casete de expresión se transfecta de forma estable en la célula huésped.

8. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 7, en el que la célula huésped es una célula sanguínea, preferiblemente un glóbulo blanco, una célula precursora sanguínea o célula leucémica, o una célula derivada de la misma.

9. Un casete de expresión que comprende, funcionalmente unidos entre sí,

(i) una región de control del locus que comprende al menos el elemento central del sitio 2 de hipersensibilidad a la ADNasa I (HS2) del grupo de genes de la β-globina humana;

(ii) una región promotora que comprende al menos una parte funcional del promotor del gen de la $^{A}\gamma$ globina humana que incluye un sitio de inicio de la transcripción y opcionalmente al menos parte de la región 5' no traducida del gen de la $^{A}\gamma$ globina humana;

(iii) opcionalmente una región de codificación;

(iv) una región de terminación de la transcripción; y

(v) una región potenciadora que comprende al menos una parte funcional del potenciador 3' del gen de la $^{A}\gamma$ globina humana;

en el que el casete de expresión no comprende una secuencia de ácido nucleico que codifica la $^{A}\gamma$ globina humana completa.

10. El casete de expresión de acuerdo con la reivindicación 9, en el que el elemento central del sitio 2 de hipersensibilidad a la ADNasa I (HS2) del grupo de genes de la β-globina humana tiene la secuencia de ácido nucleico de la posición 906 a 939 de la SEQ ID NO: 1.

11. El casete de expresión de acuerdo con la reivindicación 10, en el que la región de control del locus comprende

(i) el elemento M1-central-M2 del sitio 2 de hipersensibilidad a la ADNasa I (HS2) del grupo de genes de la β-globina humana, que tiene preferiblemente la secuencia de ácido nucleico de la posición 742 a 995 de la SEQ ID NO: 1; y

(ii) opcionalmente al menos una parte del sitio 2 de hipersensibilidad (HS2) del grupo de genes de la β-globina humana que comprende la secuencia de ácido nucleico de la posición 741 a 1109 de la SEQ ID NO: 1; y

(iii) opcionalmente, el sitio 3 de hipersensibilidad (HS3) o una parte del mismo del grupo de genes de la β-globina humana, que tiene la secuencia de ácido nucleico de la posición 310 a 735 de la SEQ ID NO: 1, y/o el sitio 4 de hipersensibilidad (HS4) o una parte del mismo del grupo de genes de la β-globina humana, que tiene la secuencia de ácido nucleico de la posición 13 a 294 de la SEQ ID NO: 1.

12. El casete de expresión de acuerdo con una cualquiera de las reivindicaciones 9 a 11, en el que la región promotora comprende los nucleótidos -384 a +36, con respecto al sitio de inicio de la transcripción, del gen de la $^{A}\gamma$ globina humana, y/o la secuencia de ácido nucleico de la posición 1123 a 1542 de SEQ ID NO: 1.

13. El casete de expresión de acuerdo con una cualquiera de las reivindicaciones 9 a 12, en el que la región de terminación de la transcripción comprende una señal de poliadenilación, que comprende preferiblemente la secuencia de ácido nucleico de la posición 1725 a 1730 de la SEQ ID NO: 1, y un sitio de terminación de la transcripción.

14. El casete de expresión de acuerdo con una cualquiera de las reivindicaciones 9 a 13, en el que la región potenciadora comprende la secuencia de ácido nucleico de la posición 2136 a 2881 de la SEQ ID NO: 1.

15. El casete de expresión de acuerdo con una cualquiera de las reivindicaciones 9 a 14, que comprende además un sitio de clonación que comprende al menos una secuencia de reconocimiento de una enzima de restricción.

16. El casete de expresión de acuerdo con una cualquiera de las reivindicaciones 9 a 15, en el que la región de codificación comprende una secuencia de ácido nucleico que codifica un polipéptido de interés, que preferiblemente es una glicoproteína o una parte de la misma, tal como un anticuerpo, la cadena pesada o la cadena ligera de un anticuerpo, o una parte del mismo.

17. Un vector que comprende el casete de expresión de acuerdo con una cualquiera de las reivindicaciones 9 a 16 y opcionalmente uno o más genes marcadores seleccionables.

18. Una célula huésped que comprende el casete de expresión de acuerdo con una cualquiera de las reivindicaciones 9 a 16 o el vector de acuerdo con la reivindicación 17.

19. La célula huésped de acuerdo con la reivindicación 18, que es un célula sanguínea, preferiblemente un glóbulo blanco, una célula precursora sanguínea o una célula leucémica, o una célula derivada de las mismas.

20. Uso del casete de expresión de acuerdo con una cualquiera de las reivindicaciones 9 a 16 o la célula huésped de acuerdo con las reivindicaciones 18 o 19 en el método de acuerdo con una cualquiera de las reivindicaciones 1 a 8.

21. El método de acuerdo con una cualquiera de las reivindicaciones 1 a 8, en el que el casete de expresión es un casete de expresión de acuerdo con una cualquiera de las reivindicaciones 9 a 16 y/o la célula huésped es una célula huésped de acuerdo con la reivindicación 18 o 19.

Figura 1

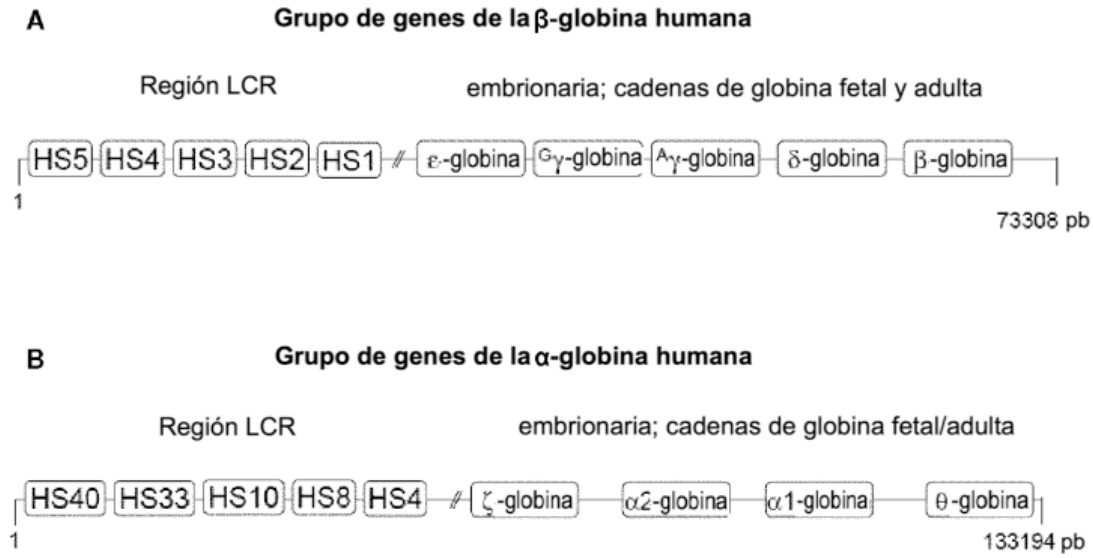


Figura 2

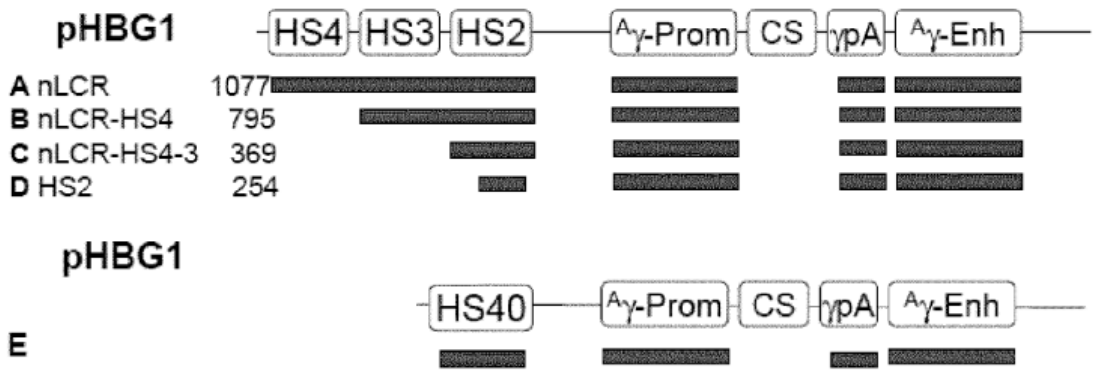


Figura 3

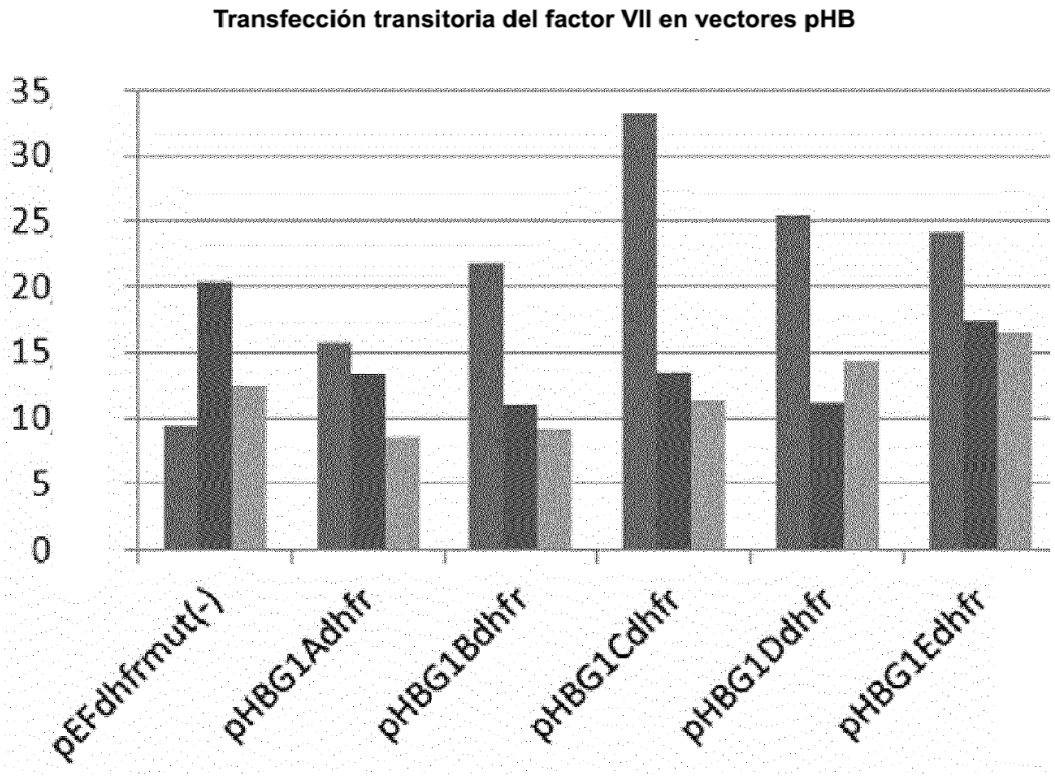

Transfección transitoria del factor VII en vectores pHB
35
30
25
20
15
10
5
0
pEFdhfrmut(-)
pHBG1Adhfr
pHBG1Bdhfr
pHBG1Cdhfr
pHBG1Ddhfr
pHBG1Edhfr

Figura 4

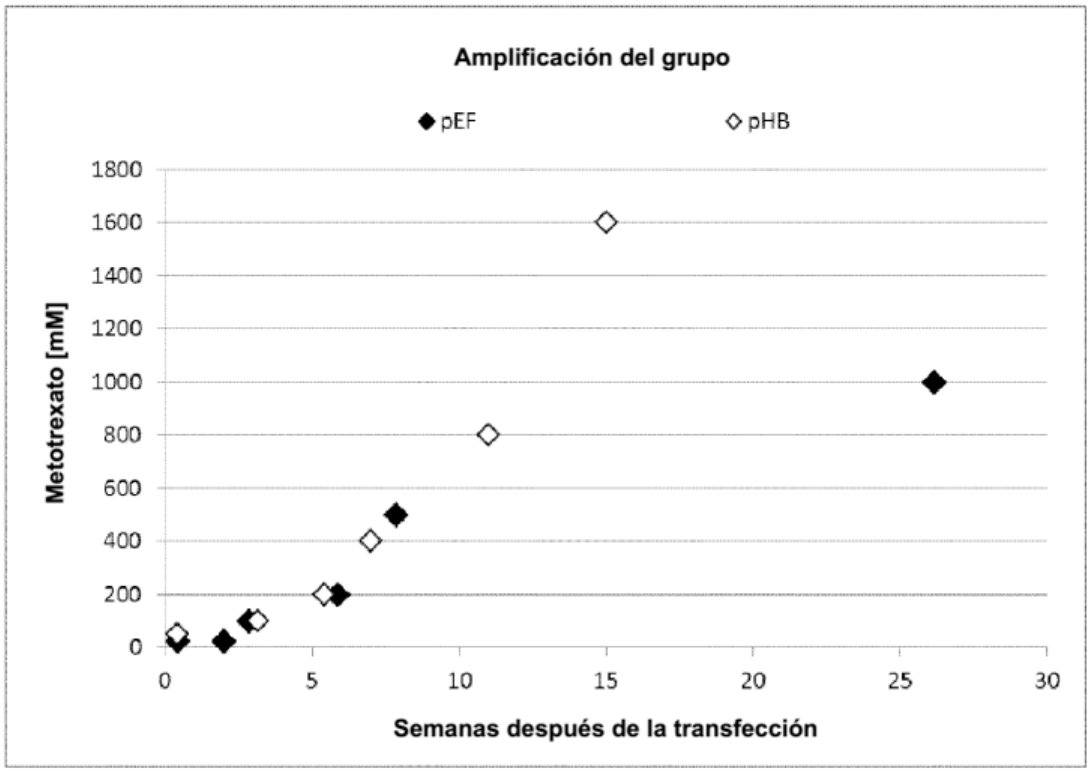

Amplificación del grupo
pEF
pHB
Metotrexato [mM]
1800
1600
1400
1200
1000
800
600
400
200
0
0
5
10
15
20
25
30
Semanas después de la transfección

Figura 5

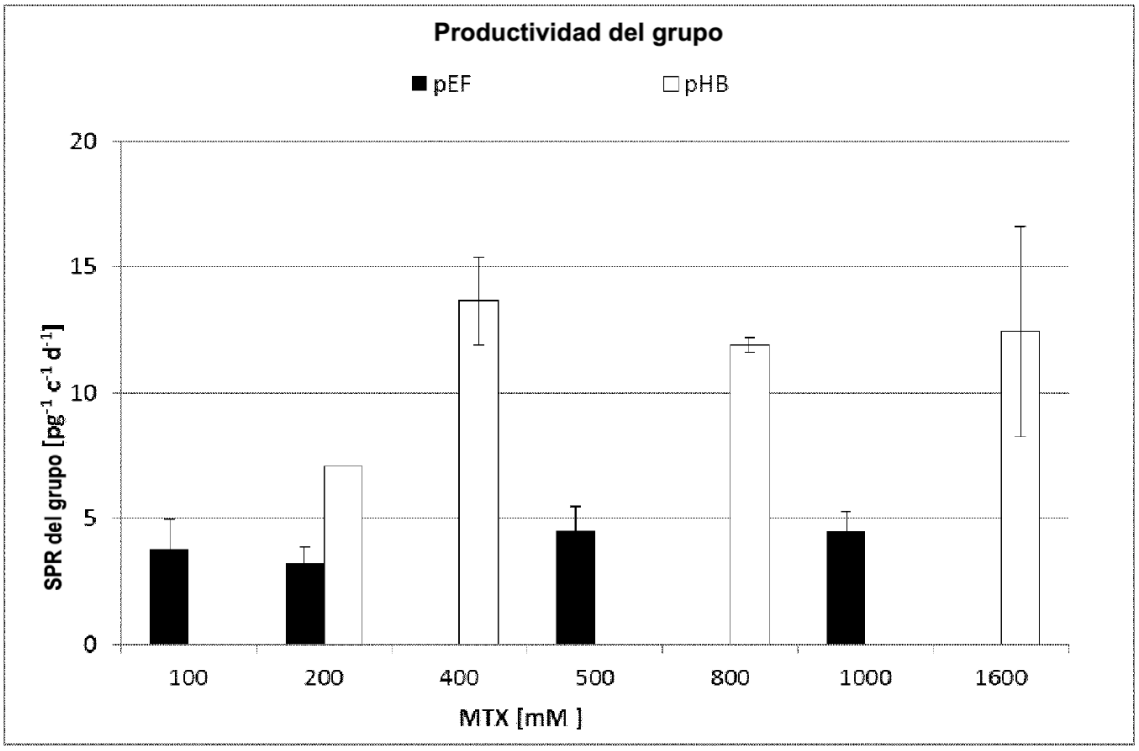
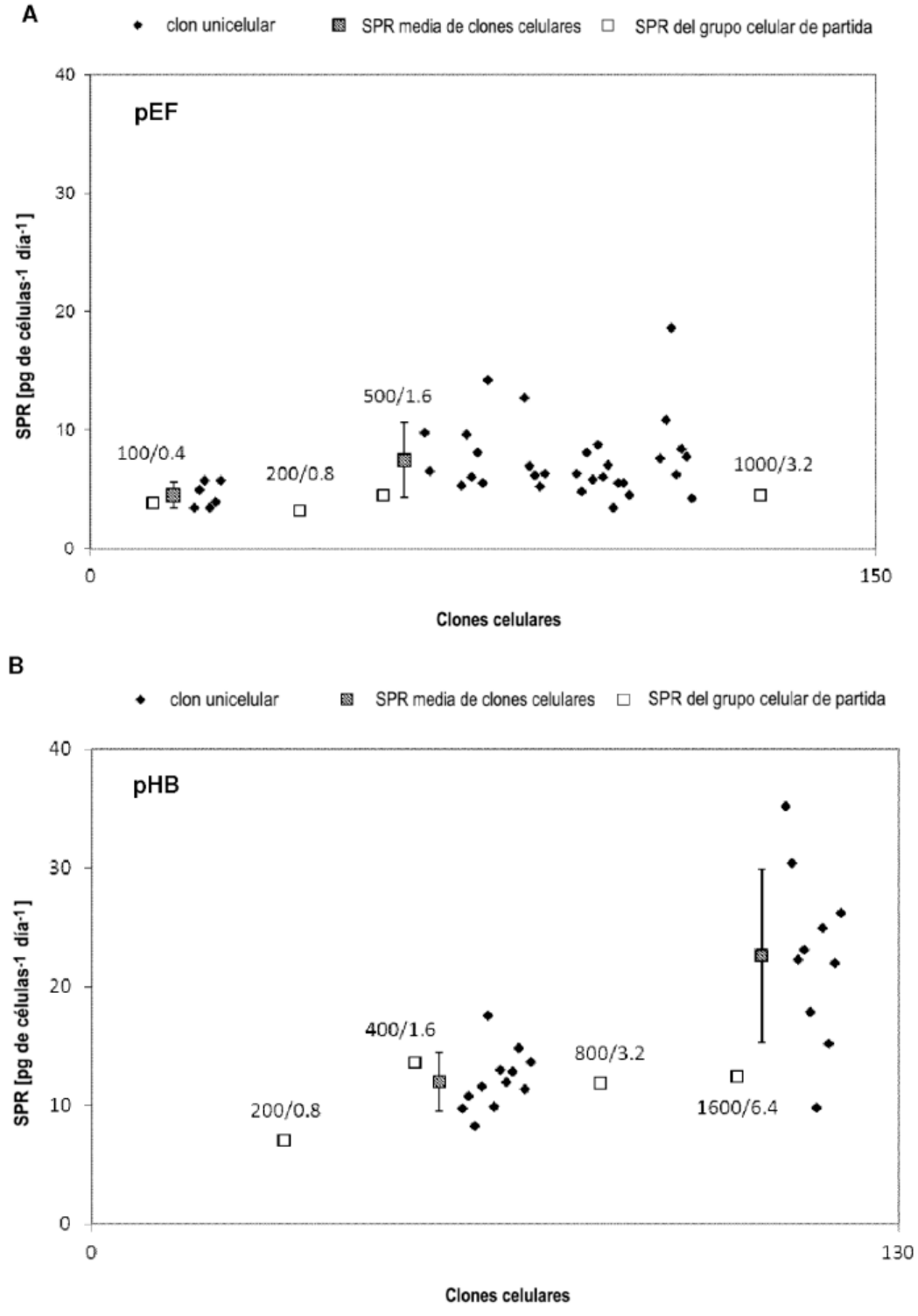

Productividad del grupo
pEF
pHB
20
15
10
5
0
SPR del grupo [pg-1 c-1 d-1]
100
200
400
500
800
1000
1600
MTX [mM]

Figura 6


A
clon unicelular
SPR media de clones celulares
SPR del grupo celular de partida
pEF
SPR [pg de células-1 día-1]
40
30
20
10
0
100/0.4
200/0.8
500/1.6
1000/3.2
0
150
Clones celulares
B
clon unicelular
SPR media de clones celulares
SPR del grupo celular de partida
pHB
SPR [pg de células-1 día-1]
40
30
20
10
0
200/0.8
400/1.6
800/3.2
1600/6.4
0
130
Clones celulares