

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 440 777**

51 Int. Cl.:

C07K 16/00 (2006.01)

C07K 16/24 (2006.01)

C12N 15/63 (2006.01)

C12N 15/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **21.12.2005** **E 05855447 (8)**

97 Fecha y número de publicación de la concesión europea: **23.10.2013** **EP 1839120**

54 Título: **Vectores basados en anticuerpo anti-il-12, células huéspedes y métodos de producción y usos**

30 Prioridad:

21.12.2004 US 637936 P

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

30.01.2014

73 Titular/es:

JANSSEN BIOTECH, INC. (100.0%)
800/850 Ridgeview Drive
Horsham, PA 19044 , US

72 Inventor/es:

LU, JIN;
NESSPOR, THOMAS;
SCALLON, BERNARD y
SNYDER, LINDA

74 Agente/Representante:

IZQUIERDO FACES, José

ES 2 440 777 T3

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

DESCRIPCIÓN

Vectores basados en anticuerpo anti-il-12, células huéspedes y métodos de producción y usos

5 Campo de la invención

La presente invención se refiere a vectores y plásmidos que dirigen expresiones de un anticuerpo, células huéspedes y métodos para hacer uso de los mismos, incluyendo secuencias potenciadoras y promotoras del vector específico y su interacción con los factores de transcripción de célula huésped.

10 Antecedentes

Las moléculas de anticuerpo consisten en una combinación de dos polipéptidos de cadena pesada (P) y dos de cadena ligera (L). cada cadena pesada y ligeras comprende una región constante que contiene las regiones CL, CH1, región bisagra, CH2 y CH3, y una región variable que contiene las regiones hipervariables (regiones determinante del complemento (RDCs)); las RDCs controlan las características de enlace de antígeno del anticuerpo. Las dos cadenas pesadas se unen y las cadenas ligeras en una estructura en forma de Y por medio de puentes disulfuros de tal manera que las regiones variables de las cadenas ligeras (V.sub L.) y cadenas pesadas (V.sub P.) están situadas unas junto a otras.

Para generar anticuerpos, se han usado técnicas convencionales de hibridoma en las que los clones de las células híbridas que expresan genes que codifican las cadenas ligeras y pesadas de una molécula de anticuerpo se obtienen mediante inmunización con una molécula de anticuerpo. Esta técnica necesita la fusión de células de origen linfocítico, que contienen los genes para la formación de anticuerpos y células que forman líneas inmortales. Las células que tienen los genes en cuestión se obtienen generalmente mediante creación arbitraria de bibliotecas de células circulantes, y la filtración de los hibridomas con una reacción antígeno-anticuerpo después de que los clones del hibridoma se hayan multiplicado y cultivado. Esta técnica puede ser incierta y laboriosa con una producción limitada de anticuerpos, y se limita a aplicaciones para la producción de anticuerpos no humanos (por ejemplo, ratón).

Además, los anticuerpos monoclonales y sus fragmentos pueden expresarse en varios sistemas huéspedes, tales como levadura *E. coli*, y células huéspedes de mamíferos. En general, un vector de expresión de mamífero contendrá (1) elementos reguladores, normalmente en forma de secuencias promotoras y potenciadoras virales y caracterizado por un amplio rango de huésped y tejido; (2) una secuencia "policlonaje", que facilita la inserción de un fragmento de ADN en el vector plásmido; y (3) las secuencias responsables para el empalme de intrón y la poliadenilación de transcripciones de ARN. Esta región continua del sitio promotor-policlonaje-poliadenilación es referida comúnmente como la unidad de transcripción. El vector probablemente también tendrá (4) un gen o genes marcadores seleccionables (por ejemplo, el gen beta-lactamasa), que a menudo confieren resistencia a un antibiótico (tal como ampicilina), que permite la selección de transformadores positivos iniciales en *E. coli*; y (5) secuencias que facilitan la réplica del vector en huéspedes bacterianos y mamíferos.

A diferencia de la mayoría de los genes que se transcriben a partir de secuencias continuas genómicas de ADN, los genes de anticuerpo se montan a partir de segmentos de gen que pueden separarse comúnmente en la línea del germen. En particular, los genes de cadena pesada se forman mediante recombinación de tres segmentos genómicos que codifican las regiones variables (V), de diversidad (D) y unión (U)/constante (C) del anticuerpo. Los genes de cadena ligera funcional se forman uniendo dos segmentos de gen; uno codifica la región V y el otro codifica la región U/C. Los locus de la cadena pesada y la cadena ligera .kappa. contienen muchos segmentos de gen V (los cálculos aproximados varían entre 100s y 1000s) estimados para abarcar más de 1000 kb. El locus .lambda. es, en cambio, mucho más pequeño y ha demostrado abarcar aproximadamente 300 kb en el cromosoma 16 en el ratón. Consiste en cuatro segmentos de gen de unión/constantes y dos segmentos de gen variables. La recombinación que da como resultado los genes funcionales ocurre predominantemente entre elementos V.sub.1 y U.sub.1/C.sub.1 o U.sub.3/C.sub.3 o entre elementos V.sub.2 y U.sub.2/C.sub.2 (U.sub.4 / C.sub.4 es un pseudogen), aunque las recombinaciones entre V.sub.2 y U.sub.3/C.sub.3 o U.sub.1/C.sub.1 se ven muy raramente.

Un ejemplo de un vector de expresión de mamífero es CDM8. La unidad de transcripción de CDM8 está compuesta por un promotor quimérico (el promotor constitutivo AD169 de citomegalovirus humano fusionado con el promotor de polimerasa T7 ARN), una región de policlonaje y el empalme de antígeno de tumor pequeño (t) SV40 y las señales de poliadenilación de región temprana derivadas de pSV2. El promotor de citomegalovirus humano (HCMV) se expresa en una variedad de tipos de células de mamíferos, mientras que el promotor de polimerasa de ARN dependiente de ADN bacteriófago T7 puede moverse a la transcripción/traslación in vitro libre de células de inserciones clonadas. Esta fusión promotora particular permite que los experimentos iniciales se realicen dentro de los confines del tipo de célula de mamífero huésped, mientras que el análisis y utilización adicional de la inserción clonada puede realizarse potencialmente en un sistema de transcripción/traslación in vitro libre de células. El promotor HCMV constitutivamente expresado también se ha utilizado en otros vectores de expresión mamífera además de CDM8. Los orígenes de réplica en CDM8 incluyen (1).pi.VX (permitiendo, por ejemplo, réplica en *E. coli*) (2) origen SV40 (por ejemplo, permitiendo la réplica en una variedad de tipos de célula COS) (3) origen de polioma

(por ejemplo, permitiendo réplica en fibroblastos de ratón transformados por virus de polioma) y (4) el origen de bacteriófago M13 (por ejemplo, permitiendo la generación de plantilla de hebra sencilla para análisis de secuencia de ADN y/o mutagénesis de oligonucleótido dirigida al sitio).

Además, CDM8 tiene el gen *supF* para la selección en *E. coli*. En este sistema de selección antibiótica, una construcción de plásmido basada en CDM8 se transforma a cepa *E. coli* especializada que contiene un episoma que transporta genes que codifican la resistencia a los antibióticos, ampicilina y tetraciclina. Sin embargo, ambos genes contienen mutaciones de punto de terminación de cadena (codón "sin sentido") que inactivan el fenotipo de resistencia. El producto de gen *supF*, un tARN supresor sin sentido, restaura el fenotipo resistente para cada antibiótico. Por lo tanto, la selección se basa en el crecimiento de la cepa *E. coli* que transporta el episoma especializado en el medio que contiene ampicilina y tetraciclina. Las colonias que muestran este fenotipo se transforman supuestamente con la construcción de plásmido basada en CDM8.

El vector CDM8 es compatible con líneas celulares COS así como líneas celulares transformadas con el virus polioma. Las líneas celulares COS son células CV1 de mono verde africano transformadas con un virus mutante SV40 defectuoso de origen. Las células COS producen el antígeno T grande, que se requiere en trans para promover la réplica de SV40 o construcciones de plásmido, tales como CDM8, que contienen las respectivas secuencias cis-actoras que inician la réplica viral. Por lo tanto, las células COS transfectadas con una construcción basada en CDM8 soportará la réplica del plásmido, dando como resultado un mayor número de copia de plásmido y un sobreexpresión temporal del gen de interés.

El principal uso de CDM8 es la clonación de expresión de cADN y la sobreproducción de proteínas específicas en un sistema de expresión in vitro de mamíferos. La clonación de expresión toma varias formas dependiendo del modo de detección utilizado para identificar el cADN de interés; sin embargo, la etapa inicial consiste en aislar mRNA y sintetizar copias de ácido deoxirribonucleico de doble hebra de la población de mRNA (cADNs). Estos cADNs deben estar eficientemente ligados a un plásmido o vector de clonación de ADN bacteriófago y transferirse al huésped apropiado antes de la filtración y el análisis de la biblioteca. El vector CDM8 contiene dos sitios de restricción BstXI, haciéndolo tratable para el procedimiento del enlace "adaptador" de ligar cADNs al vector, es decir, el uso de fragmentos de ADN con extremo desafilado en un extremo (y por lo tanto compatibles con la ligadura con el cADN de extremo desafilado) pero que contienen un saliente no palindrómico (extremo engomado) en el otro extremo (por ejemplo, compatible con ligadura con vector de ADN digerido con BstXI, pero no con otros cADNs).

Otro ejemplo de vector de expresión de mamífero es pCMX. Este vector contiene: (1) el promotor temprano inmediato de HCMV, (2) una secuencia empalme/poliadenilación SV40 ARN, (3) una resistencia a un antibiótico, tal como el gen beta-lactamasa que confiere resistencia a la ampicilina antibiótica. El vector pCMX también puede usarse para la expresión temporal de una secuencia de ADN clonada en células COS transfectadas.

El control de la transcripción de ambos genes de cadena pesada y ligera .kappa. reorganizados depende de la actividad del promotor específico del tejido corriente arriba de la región V y del potenciador específico de tejido situado en el intrón J-C. Estos elementos actúan sinérgicamente. También, se ha identificado un segundo potenciador específico de célula B en el locus de cadena ligera .kappa. Este potenciador adicional está situado 9kb corriente debajo de C.sub..kappa.

Tal sistema de huésped de mamífero usado para producir anticuerpos es una célula huésped de mieloma de ratón que se ha transfectado con ADN clonado que codifica el anticuerpo deseado. Tales "anticuerpos monoclonales recombinantes" a menudo son distintos de los anticuerpos monoclonales derivados de hibridoma para los que el ADN no se ha clonado y para los que las células que producen el anticuerpo monoclonal se derivan inmortalizando una célula que produce un anticuerpo monoclonal natural aislada de un animal. Los genes de inmunoglobulina de cadena pesada y ligera (Ig) que se expresan en células de hibridoma están bajo el control del promotor endógeno natural que siempre se ha unido con la secuencia de la región variable particular que se expresa en oposición al promotor contenido en el vector recombinante.

En producción recombinante, la secuencia de anticuerpo monoclonal que se clonará debe ligarse en un vector apropiado después del tratamiento de enzima de restricción del vector. Esta tarea puede ser difícil e imprecisa como el proceso de incorporar la secuencia o secuencias de nucleótido de anticuerpo en un vector de expresión o plásmido es complejo.

Sin embargo, al clonar las secuencias de ADN de anticuerpo monoclonal antes de preparar los anticuerpos monoclonales derivados de célula transfectada, pueden usarse métodos de ADN recombinante para sustituir el promotor endógeno natural para un gen Ig por cualquier promotor preferido. Una razón principal para cambiar un promotor es realizar niveles más altos de producción de anticuerpo monoclonal.

Las secuencias promotoras, en conjunto con las secuencia potenciadoras corriente abajo, son responsables de dirigir la transcripción (es decir, síntesis de ARN) de los genes de cadena pesada y ligera en las células transfectadas al enlazarse con proteínas nucleares especializadas llamadas factores de transcripción. Es

evidente que hay menos sitios para el enlace del factor de transcripción en un promotor Ig que en un potenciador Ig; sin embargo, el hecho de que hay variabilidad de secuencia entre promotores pero solamente una única copia de la secuencia promotora hace que sea muy probable que haya variabilidad funcional entre promotores Ig. Un promotor puede ser “fuerte”, es decir, eficiente en enlazar una combinación favorable de factores de transcripción y lleva a altos niveles de síntesis de ARN de anticuerpo monoclonal, mientras que otro promotor puede ser “débil”, debido a que tiene una secuencia diferente de ADN. Ya que cada uno de los más de 200 genes HC de región variable y los más de 200 genes de LC de región variable en un repertorio Ig tiene su propio promotor naturalmente unido, y es probable que dos promotores no tengan secuencias idénticas, es probable que los muchos promotores diferentes Ig varíen significativamente con respecto a cómo de bien llevan la transcripción.

Los promotores Ig son solamente funcionales en células huésped de tipo linfóide, tales como células T y B (y células de mieloma), debido a su requisito para factores de transcripción específicos de gen Ig (por ejemplo, Oct-2 y OBF-1) no expresados en otros tipos de células. Además, incluso los factores de transcripción específicos de células pueden expresarse solamente en fases particulares de diferenciación celular de tal manera que la expresión óptima puede depender de unir el estado de diferenciación de la línea celular huésped con los motivos apropiados de secuencia en los promotores de gen Ig. Aunque la especificidad de célula huésped de promotores Ig puede verse como una desventaja menor para la expresión del anticuerpo monoclonal en una célula huésped no linfóide, la gran selección de promotores HC y LC permite una posibilidad de identificar y quizás optimizar más los promotores fuertes que pueden incorporarse a los vectores específicos de célula linfóide.

La expresión de anticuerpos monoclonales detrás de un promotor fuerte aumenta las posibilidades de identificar líneas celulares altamente productoras y obtener mayores producciones de anticuerpos monoclonales. Como consecuencia, los vectores Ig con promotores fuertes son muy deseables para expresar cualquier anticuerpo monoclonal de interés. Además, los vectores con sitios de clonación únicos de ADN corriente debajo de los promotores fuertes tendrían una conveniencia añadida.

Por consiguiente, hay una necesidad de nuevos vectores y plásmidos útiles para la expresión de anticuerpos que simplifiquen las técnicas de ligadura y permitan la customización de secuencias promotoras y potenciadoras con el fin de aumentar la producción de anticuerpos.

Resumen de la invención

La presente invención se refiere a vectores y plásmidos de expresión recombinante que comprenden sitios de restricción para la clonación de varios anticuerpos. En otra realización, la divulgación proporciona la expresión de secuencias de control en el vector, tales como secuencias potenciadoras y promotoras, que pueden customizarse en relación con el gen de anticuerpo que se clonará y transcribirá y el tipo de célula huésped que se usará, con el fin de llevar la transcripción eficientemente. La presente invención también comprende células huésped aisladas, por ejemplo, células de mamíferos y no mamíferos, que contienen tal vector o plásmido. La divulgación proporciona además métodos para producir un anticuerpo mediante cultivo, en un medio adecuado, una célula huésped que contiene un vector de expresión recombinante de la invención de tal manera que se produzca el anticuerpo.

En otra realización, la divulgación comprende un método para identificar, modular y/o determinar la interacción entre los factores de transcripción de célula huésped y las secuencias promotoras y potenciadoras de un vector de expresión. Esta interacción impulsa el proceso de transcripción. Los factores de transcripción y las secuencias promotoras y potenciadoras pueden customizarse para mejorar su afinidad para enlazarse entre sí, lo que puede aumentar la producción y eficiencia del proceso de transcripción.

Descripción de las figuras

La Fig. 1A es una representación esquemática del mapa vector para el plásmido p1560.

La Fig. 1B es una representación esquemática del mapa vector para el plásmido p1558.

Descripción de la invención

Las siguientes definiciones se exponen para ilustrar y definir el significado y alcance de varios términos usados para describir la invención en el presente documento.

Una “actividad” o “actividad biológica”, y una actividad funcional de un polipéptido se refieren a una actividad ejercida por una proteína o polipéptido in vivo, in situ o in vitro, de acuerdo con técnicas estándares. Tales actividades pueden ser una actividad directa, tal como una asociación con una actividad enzimática en una segunda proteína, o una actividad indirecta, tal como un proceso celular mediado por la interacción de la proteína con una segunda proteína o una serie de interacciones como en la señalización intracelular o la cascada de coagulación.

Un “anticuerpo” incluye cualquier polipéptido o péptido que contiene molécula que comprende al menos una parte de una molécula de inmunoglobulina, tal como aunque sin limitar, al menos una región determinante de la

complementariedad (RDC) de una cadena pesada o ligera o una parte de enlace de ligando de la misma, una región variables de cadena pesada o cadena ligera, una región constante de cadena pesada o cadena ligera, una región marco, o una parte, fragmento o variante de la misma. El término "anticuerpo" pretende además abarcar anticuerpos, fragmentos de digestión, partes especificadas y variantes de los mismos, incluyendo miméticos de anticuerpo o que comprende partes de anticuerpos que imitan la estructura y/o función de un anticuerpo o fragmento especificado o parte del mismo, incluyendo anticuerpos de cadena sencilla y fragmentos de los mismos. Por ejemplo, los fragmentos de anticuerpo incluyen, aunque no se limitan a, Fab (por ejemplo, por digestión de papaína), Fab' (por ejemplo, por digestión de pepsina y reducción parcial) y F(ab')₂ (por ejemplo, por digestión de pepsina, reducción parcial y reagregación), fragmentos de Fv o scFv (por ejemplo, mediante técnicas de biología molecular), están incluidas en la invención (véase, por ejemplo, Colligan, et al., eds., Current Protocols in Immunology, John Wiley & Sons, Inc., NY (1994-2001); Colligan et al., Current Protocols in Polypeptide Science, John Wiley & Sons, NY, NY (1997-2001).

Moléculas "quiméricas" o de "fusión" son ácidos nucleicos o polipéptidos que se crean combinando uno o más polinucleótidos (o sus partes) con secuencias adicionales de ácido nucleico. Tales secuencias combinadas pueden introducirse en un vector apropiado y expresarse para dar lugar a un polipéptido quimérico o de fusión.

"Complemento de" o "complementario a" una secuencia de ácido nucleico de la invención se refiere a una molécula de polinucleótido que tiene una secuencia base complementaria y una orientación inversa en comparación con un primer polinucleótido.

"Fragmento" es un polipéptido variante que tiene una secuencia de aminoácido que es completamente la misma como parte pero no toda de cualquier secuencia de aminoácido de un polipéptido o un polinucleótido variante que tiene una secuencia de ácido nucleico que es completamente la misma como parte pero no toda de cualquier secuencia de ácido nucleico de cualquier polipéptido. Los fragmentos pueden incluir, por ejemplo, polipéptidos de truncamiento que tienen una parte de una secuencia de aminoácido, o de variantes de la misma, tales como una serie continua de residuos que incluyen una secuencia de aminoácido heteróloga amino- y/o carboxi-terminal. Las formas de degradación de los polipéptidos producidos por o en una célula huésped también se incluyen. Otros fragmentos ejemplares se caracterizan por los atributos estructurales o funcionales, tales como fragmentos que comprenden hélice alfa o regiones que forman hélice alfa, lámina beta o regiones que forman lámina beta, giro o regiones que forman giro, bobina o regiones que forman bobina, regiones hidrofílicas, regiones hidrofóbicas, regiones alfa-anfipáticas, regiones beta-anfipáticas, regiones flexibles, regiones que forman superficie, regiones que se enlazan con sustrato, regiones extracelulares, y regiones de alto índice antigénico.

Más fragmentos ejemplares incluyen un polipéptido aislado que comprende una secuencia de aminoácido que tiene al menos 15, 20, 30, 40, 50 o 100 aminoácidos contiguos de una secuencia de aminoácido de longitud completa, o un polipéptido aislado que comprende una secuencia de aminoácido que tiene al menos 15, 20, 30, 40, 50 o 100 aminoácidos contiguos truncados o eliminados de la secuencia de aminoácido de longitud completa. Los fragmentos también incluyen polinucleótidos aislados que tiene tamaños o características similares.

"Identidad", como se conoce en la técnica, es una relación entre dos o más secuencias de polipéptido o dos o más secuencias de polinucleótido, como se determina comparando las secuencias. En la técnica, "identidad" también significa el grado de relación secuencial entre secuencias e polipéptido o polinucleótido, como lo determina la coincidencia entre series de tales secuencias. "Identidad" y "similitud" pueden calcularse fácilmente mediante métodos conocidos, incluyendo, aunque sin limitar, a aquellos descritos en Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, Nueva York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, Nueva York, 1993; Computer Analysis of Sequence Data, Parte I, Griffin, A. M., y Griffin, H. G., eds., Humana Press, Nueva Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; y Sequence Analysis Primer, Gribskov, M., y Devereux, J. Eds., M Stockton Press, Nueva York, 1991; y Carillo, H., y Lipman, D., Siam J. Applied Math., 48:1073 81988). Además, los valores para la identidad porcentual pueden obtenerse a partir de alineaciones de aminoácidos y secuencias nucleótidos generadas usando los ajustes estándares para el componente AlignX de Vector NTI Suite 8.0 (Informax, Frederick, MD).

Los métodos preferentes para determinar la identidad están diseñados para dar la mayor coincidencia entre las secuencias analizadas. Los métodos para determinar identidad y similitud se codifican en programas de ordenador públicamente disponibles. Los métodos para programas de ordenador preferentes para determinar identidad y similitud entre dos secuencias incluyen, aunque no se limitan a, el paquete de programa GCG (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984)); BLASTN y FASTA (Altschul, S. F. et al., J. Molec. Biol. 215:403-410 (1990)). El programa BLAST X está públicamente disponible en NCBI y otras fuentes (Manual BLAST, Altschul, S., et al., NCBINLM NIH Bethesda, Md. 20894; Altschul, S. et al., J. Mol. Biol. 215:403-410 (1990)). El algoritmo bien conocido de Smith Waterman también puede usarse para determinar identidad.

Los parámetros preferentes para la comparación de secuencia de polipéptidos incluyen los siguientes:

(1) Algoritmo: Needleman y Wunsch, J. Mol Biol. 48:443-453 (1970)
Matriz de comparación: BLOSSUM62 de Hentikoff y Hentikoff, Proc. Natl. Acad. Sci, USA. 89:10915-10919 (1992)
Penalización en la puntuación del alineamiento por cada hueco: 12

Penalización en la puntuación del alineamiento por cada hueco en longitud: 4

Un programa útil con estos parámetros está públicamente disponible como el programa "gap" de Genetics Computer Group, Madison Wis. Los parámetros anteriormente mencionados son los parámetros estándares para las comparaciones de secuencia de péptido (junto con las faltas de penalización para los huecos finales).

Los parámetros preferentes para la comparación de polinucleótidos incluyen los siguientes:

(1) Algoritmo: Needleman y Wunsch, J. Mol Biol. 48:443-453 (1970)

Matriz de comparación: coincidencias=+10, sin coincidencia=0

Penalización en la puntuación del alineamiento por cada hueco: 50

Penalización en la puntuación del alineamiento por cada hueco en longitud: 3

Disponible como: el programa "gap" de Genetics Computer Group, Madison Wis. Estos son parámetros estándares para comparaciones de secuencia de ácido nucleico.

A modo de ejemplo, una secuencia de polinucleótido puede ser idéntica a otra secuencia, que es 100% idéntica, o puede incluir hasta ciertos números enteros de alteraciones de nucleótidos en comparación con la secuencia referencia. Tales alteraciones se seleccionan del grupo consistente en al menos una eliminación de nucleótido, sustitución, incluyendo transición o transversión, o inserción, y donde las alteraciones ocurren en las posiciones terminales 5' o 3' de la secuencia de nucleótido referencia o en cualquier sitio entre esas posiciones terminales, intercaladas individualmente entre los nucleótidos en la secuencia referencia o en uno o más grupos contiguos en la secuencia referencia. El número de alteraciones de nucleótidos se determina multiplicando el número total de nucleótidos en la secuencia por el porcentaje numérico de la respectiva identidad porcentual (dividida entre 100) y restando el producto del número total de nucleótidos en la secuencia, o:

$n.\text{sub}.n.l\text{torsim}.x.\text{sub}.n-(x.\text{sub}.n.y)$,

donde $n.\text{sub}.n$ es el número de alteraciones de nucleótido, $x.\text{sub}.n$ es el número total de nucleótidos en la secuencia, e y es, por ejemplo 0,70 para 70%, 0,80 para 80%, 0,85 para 85%, 0,90 para 90%, 0,95 para 95%, etc., donde cualquier producto no entero de $x.\text{sub}.n$ e y se redondea hacia abajo al número entero más cercano para restar de $x.\text{sub}.n$.

Las alteraciones de una secuencia de polinucleótido que codifica un polipéptido pueden crear mutaciones sin sentido, de sentido erróneo o de cambio de estructura en esta secuencia codificadora y por lo tanto alterar el polipéptido codificado por el polinucleótido seguido de tales alteraciones. Similarmente, una secuencia de polipéptido puede ser idéntica a una secuencia de referencia, es decir 100% idéntica, o puede incluir hasta cierto número entero de alteraciones de aminoácido en comparación con la secuencia de referencia de tal manera que la identidad porcentual sea inferior al 100%. Tales alteraciones se seleccionan del grupo consistente en al menos una eliminación de aminoácido, sustitución, incluyendo sustitución conservadora y no conservadora, o inserción, y donde las alteraciones pueden ocurrir en la posición amino- o carboxi-terminal de la secuencia de polipéptido de referencia o en cualquier sitio entre esas posiciones terminales, intercaladas individualmente entre los aminoácidos en la secuencia referencia o en uno o más grupos contiguos en la secuencia referencia. El número de alteraciones de aminoácidos para una identidad% dada se determina multiplicando el número total de aminoácidos en la secuencia por el porcentaje numérico de la respectiva identidad porcentual (dividida entre 100) y después restando ese producto del número total de aminoácidos en la secuencia, o:

$n.\text{sub}.a.l\text{torsim}.x.\text{sub}.a-(x.\text{sub}.a.y)$,

donde $n.\text{sub}.a$ es el número de alteraciones de aminoácido, $x.\text{sub}.a$ es el número total de aminoácidos en la secuencia, e y es, por ejemplo 0,70 para 70%, 0,80 para 80%, 0,85 para 85%, etc., donde cualquier producto no entero de $x.\text{sub}.a$ e y se redondea hacia abajo al número entero más cercano para restar de $x.\text{sub}.a$.

"Ácidos nucleicos" son polímeros de nucleótidos, donde un nucleótido comprende una base unida a un azúcar cuyos azúcares a su vez se unen entre sí mediante al menos una molécula bivalente, tal como un ácido fosfórico. En ácidos nucleicos que ocurren de manera natural, el azúcar es 2'-deoxirribosa (ADN) o ribosa (ARN). Los poli- u oligonucleótidos no naturales contienen bases modificadas, azúcares, o molécula de enlace, pero se entiende que generalmente imitan la naturaleza complementaria de los ácidos nucleicos que ocurren de manera natural a los que se han diseñado. Un ejemplo de oligonucleótido no natural es una composición de molécula anti-sentido que tiene una red central de fosforotiorato. Un "oligonucleótido" generalmente se refiere a ácidos nucleicos que tienen menos de 30 nucleótidos.

Un "polipéptido" es un polímero de residuos de aminoácidos unidos por enlaces peptídicos, y un péptido se refiere generalmente a polímeros de aminoácido de 12 o menos residuos. Los enlaces peptídicos pueden producirse naturalmente como los dirigidos por la plantilla de ácido nucleico o sintéticamente por métodos bien conocidos en la técnica.

Una "proteína" es una macromolécula que comprende una o más cadenas de polipéptido. Una proteína puede además comprender grupos sustituyentes unidos a los grupos laterales de los aminoácidos implicados en la formación de los enlaces peptídicos. Típicamente, las proteínas formadas por expresión de célula eucariótica también contienen carbohidratos. Las proteínas se definen en el presente documento en términos de su secuencia o red central de aminoácido y los sustituyentes no se especifican, sean conocidos o no.

El término “receptor” denota una molécula que tiene actividad biológica resultante de la interacción con un ligando específico o pareja de enlace. Los receptores unidos a la membrana celular se caracterizan por un dominio de enlace de ligando extracelular, uno o más dominios que cubren la membrana o transmembrana, y un dominio efector intracelular que está típicamente implicado en transducción de señal. En enlace de ligando a los receptores de la membrana celular provoca cambios en el dominio extracelular que se comunican con la membrana celular, interacción directa o indirecta con una o más proteínas intracelulares, y altera las proteínas celulares, tales como actividad enzimática, forma celular o perfil de expresión de gen. Los receptores también pueden estar liberados a la superficie celular y pueden ser citosólicos, nucleares o liberarse de la célula juntos. Los receptores asociados no celulares se llaman receptores solubles.

Todas las publicaciones o patentes citadas en el presente documento desvelan el estado de la técnica en el momento de la presente invención y/o proporcionan la descripción y capacidad de la presente invención. Las publicaciones se refieren a cualquier publicación científica o de patente, o cualquier otra información disponible en cualquier formato de medio, incluyendo formatos grabados, electrónicos o impresos. Se citan las siguientes referencias: Ausubel et al., ed., Current Protocols in Molecular Biology, John Wiley & Sons, Inc., NY, NY (1987-2001); Sambrook, et al., Molecular Cloning: A Laboratory Manual, 2ª Edición, Cold Spring Harbor, NY (1989); Harlow y Lane, antibodies, a Laboratory Manual, Cold Spring Harbor, NY (1989); Colligan, et al., eds., Current Protocols in Immunology, John Wiley & Sons, Inc., NY (1994-2001); Colligan et al., Current Protocols in Protein Science, John Wiley & Sons, NY, NY, (1997-2001).

Vectores de Expresión Recombinante y Células Huéspedes

La invención proporciona vectores, preferentemente, vectores de expresión, que contienen un ácido nucleico que codifica un polipéptido específicos, por ejemplo, el anticuerpo anti-IL-12, o puede usarse para obtener plásmidos que contienen varios genes de anticuerpo HC o LC o partes de los mismos. Como se usa en el presente documento, el término “vector” se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico con el que se ha unido. Un tipo de vector es un “plásmido”, que se refiere a un bucle de ADN circular de doble hebra al que pueden ligarse segmentos adicionales de ADN. Otro tipo de vector es un vector viral, donde segmentos adicionales de ADN pueden ligarse al genoma viral.

Ciertos vectores son capaces de réplica autónoma en una célula huésped en la que se introducen (por ejemplo, vectores bacterianos que tiene un origen bacteriano de réplica y vectores episómico de mamífero). Otros vectores (por ejemplo, vectores no episómico de mamífero) están integrados en el genoma de una célula huésped tras la introducción en la célula huésped, y de este modo se replican junto con el genoma huésped. Además, ciertos vectores, por ejemplo, vectores de expresión, son capaces de dirigir la expresión de genes a los que operativamente se unen. En general, los vectores de expresión de utilidad en técnicas de ADN recombinante tienen a menudo forma de plásmidos (vectores). Sin embargo, la invención pretende incluir otras formas de vectores de expresión, tales como vectores virales (por ejemplo, retrovirus defectuosos de réplica, adenovirus y virus adeno-asociados), que sirven funciones equivalentes.

Los vectores de expresión recombinantes de la invención comprenden un ácido nucleico en una forma adecuada para la expresión del ácido nucleico en una célula huésped. Esto significa que los vectores de expresión recombinantes incluyen una o más secuencias reguladoras, seleccionadas en base a las células huéspedes que se usarán para la expresión, que están operativamente unidas con la secuencia de ácido nucleico que se expresará. En un vector de expresión recombinante, “operativamente unido” pretende significar que la secuencia de nucleótido de interés está unida a la secuencia o secuencias reguladoras de manera que permite la expresión de la secuencia de nucleótido (por ejemplo, en un sistema de transcripción/traslación in vitro o en una célula huésped cuando el vector se introduce en la célula huésped). Además, la secuencia reguladora se optimiza en base a las características de la célula huésped, es decir, factores de transcripción.

El término “secuencia reguladora” pretende incluir promotores, potenciadores y otros elementos de control de expresión (por ejemplo, señales de poliadenilación). Tales secuencias reguladoras se describen, por ejemplo, en Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Las secuencias reguladoras incluyen aquellas que dirigen expresión constitutiva de una secuencia de nucleótido en muchos tipos de células huéspedes y aquellas que dirigen expresión de la secuencia de nucleótido solamente en ciertas células huéspedes (por ejemplo, secuencias reguladoras específicas de tejido). Aquellos expertos en la técnica apreciarán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped que se transformará, el nivel de expresión de proteína deseada, y similares. Los vectores de expresión de la invención pueden introducirse en las células huéspedes para producir de ese modo proteínas o péptidos, incluyendo proteínas o péptidos de fusión o quiméricos, codificados por ácidos nucleicos como se describe en el presente documento.

Los vectores de expresión recombinantes de la invención pueden estar diseñados para la expresión de un polipéptido en células procarióticas (por ejemplo, *E. coli*) o eucarióticas (por ejemplo, células de insectos (usando vectores de expresión de baculovirus), células de levadura o células de mamíferos). Las células huéspedes adecuadas se analizan con más detalle en Goeddel, supra. Alternativamente, el vector de expresión puede

transcribirse y trasladarse in vitro, por ejemplo usando secuencias reguladoras de promotor T7 y polimerasa T7.

La expresión de proteínas en procariotas se realiza a menudo en *E. coli* con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de proteínas de fusión o de no fusión. Los vectores de fusión añaden un número de aminoácidos a una proteína codificada en los mismos, normalmente a la terminal amino de la proteína recombinante. Tales vectores de fusión típicamente sirven tres fines: (1) aumentar la expresión de proteína recombinante; (2) aumentar la solubilidad de la proteína recombinante; y (3) ayudar en la purificación de proteína recombinante actuando como un ligando en la purificación por afinidad. A menudo, en vectores de expresión de fusión, se introduce un sitio de segmentación proteolítica en la unión de la fracción de fusión y la proteína recombinante para permitir la separación de la proteína recombinante de la fracción de fusión después de la purificación de la proteína de fusión. Tales enzimas, y sus secuencias semejantes de reconocimiento, incluyen Factor Xa, trombina y enteroquinasa. Los vectores de expresión de fusión típicos incluyen pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) Gene 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) y pRIT5 (Pharmacia, Piscataway, N.J) que fusionan glutatona S-transferasa (GST), proteína de enlace maltosa E, o proteína A, respectivamente, con la proteína recombinante diana.

Para ayudar en la purificación por afinidad, varios polipéptidos con etiqueta y sus respectivos anticuerpos son bien conocidos en la técnica. Ejemplos incluyen etiquetas poli-histidina (poli-his) o poli-histidina-glicina (poli-his-gli); el polipéptido con etiqueta de gripe HA y su anticuerpo 12CA5 (Field et al., Mol. Cell. Biol. 8:2159-2165 (1988)); la etiqueta c-myc y los anticuerpos 8F9, 3C7, 6E10, G4, 7 y 9E10 en en los mismos (Evan et al., Molecular and Cellular Biology, 5:3610-3616 (1985)); y la etiqueta de glicoproteína D de virus Herpes Simplex (gD) y su anticuerpo (Paborsky et al., Protein Engineering 3(6):547-553 (1990)). Otros polipéptidos con etiqueta incluyen el péptido Flag (Hopp et al., Bio Technology, 6:1204-1210 (1988)); el péptido de epítipo KT3 (Skinner et al., J. Biol. Chem., 266:15163-15166 (1991)); y la etiqueta de péptido de proteína 10 de gen T7 (Lutz-Freyermuth et al., Proc. Natl. Acad. Sci. USA, 87:6393-6397 (1990)). Una etiqueta preferente es la etiqueta FLAG.

Ejemplos de vectores de expresión inducibles adecuados de *E. coli* de no fusión incluyen pTrc (Amann et al., (1988) Gene 69:301-315) y pET 11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89). La expresión de gen diana del vector pTrc depende de la transcripción de polimerasa de ARN huésped de un promotor de fusión del híbrido trp-lac. La expresión del gen diana del vector pET 11d depende de la transcripción de un promotor de fusión T7 gn10-lac mediado por la polimerasa de ARN viral co-expresada (T7 gn1). Esta polimerasa viral es suministrada por la cepa huésped BL21 (DE3) o HMS174(DE3) de un prófago residente λ que alberga un gen T7 gn1 bajo el control transcripcional del promotor lacUV 5.

Una estrategia para maximizar la expresión de proteína recombinante en *E. coli* es expresar la proteína en la bacteria huésped con una capacidad disminuida para partir proteolíticamente la proteína recombinante (Gottesman, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 119-128). Otra estrategia es alterar la secuencia de ácido nucleico para insertarla en un vector de expresión de manera que los codones individuales para cada aminoácido sean aquellos preferentemente utilizados en *E. coli* (Wada et al. (1992) Nucleic Acids Res. 20:2111-2118). Tal alteración de secuencias de ácido nucleico de la invención puede realizarse mediante técnicas estándares de síntesis de ADN.

En otra realización, el vector de expresión es un vector de expresión de levadura. Ejemplos de vectores para expresión en levadura *S. cerevisiae* incluyen pYepSec1 (Baldari et al. (1987) EMBO J. 6:229-234), pMFa (Kurjan y Herskowitz, (1982) Cell 30:933-943), pJRY88 (Schultz et al. (1987) Gene 54:114-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), y pPicZ (Invitrogen Corp, San Diego, Calif.).

Alternativamente, el vector de expresión es un vector de expresión de baculovirus. Los vectores de baculovirus disponibles para la expresión de proteínas en células de insectos cultivadas (por ejemplo, células Sf 9) incluyen la serie pAc (Smith et al., (1983) Mol. Cell Biol. 3:2156-2165) y la serie pVL (Lucklow y Summers (1989) Virology 170:31-39).

En otra realización más, un ácido nucleico de la invención se expresa en células mamíferas usando un vector de expresión de mamífero. Ejemplos de vectores de expresión de mamíferos incluyen pCDM8 (Seed (1987) Nature 329:840) y pMT2PC (Kaufman et al. (1987) EMBO J. 6:187:195). Cuando se usan en células mamíferas, las funciones de control del vector de expresión a menudo las proporcionan los elementos reguladores virales. Por ejemplo, los promotores comúnmente usados se derivan de poliovirus, Adenovirus 2, citomegalovirus y Virus del Simio 40. Para otros sistemas de expresión adecuados para célula procariótica y eucariótica, véase capítulos 16 y 17 de Sambrook et al., supra.

En otra realización, el vector de expresión de mamífero recombinante es capaz de dirigir la expresión del ácido nucleico, preferentemente en un tipo de célula particular, tales como células de linfoma (por ejemplo, células de mieloma de ratón). En tipos específicos de células, los elementos reguladores específicos de tejido se usan para expresar el ácido nucleico. Los elementos reguladores específicos de tejido son bien conocidos en la técnica. Ejemplos no limitativos de promotores adecuados específicos de tejido incluyen el promotor de albúmina (específico

del hígado; Pinkert et al. (1987) Genes Dev. 1:268-277), promotores específicos de linfocitos (Calame y Eaton (1988) Adv. Immunol. 43:235-275), en particular promotores de receptores de célula T (Winoto y Baltimore (1989) EMBO J. 8:729-733), e inmunoglobulinas (Banerji et al. (1983) Cell 33:729-740; Queen y Baltimore (1983) Cell 33:741-748), promotores neuro-específicos (por ejemplo, promotor de neurofilamento, Byrne y Ruddell (1989) Proc. Natl. Acad. Sci. USA 86:5473-5477), promotores específicos de páncreas (Edlund et al., (1985) Science 230:912-916), y promotores específicos de glándula mamaria (por ejemplo, promotor de suero de leche; Patente de Estados Unidos Nº 4.873.316 y Publicación de Solicitud Europea Nº 264.166). Los promotores regulados con desarrollo también se incluyen, por ejemplo, en los promotores hox murinos (Kessel y Gruss (1990) Science 249:374-379) y el promotor α -fetoproteína (Campes y Tilghman (1989) Genes Dev. 3:537-546).

La invención proporciona además un vector de expresión recombinante que comprende una molécula de ADN clonada en el vector de expresión en una orientación antisentido. Es decir, la molécula de ADN está operativamente unida a una secuencia reguladora de una manera que permite la expresión (mediante transcripción de la molécula de ADN) de una molécula de AR que es antisentido para el polipéptido que codifica mRNA. Las secuencias reguladoras operativamente unidas a un ácido nucleico clonado en la orientación antisentido pueden elegirse para que dirijan la expresión continua de la molécula de ARN antisentido en una variedad de tipos de célula. Por ejemplo, los promotores y/o potenciadores virales, o secuencias reguladoras pueden elegirse con expresión constitutiva directa, específica de tejido o específica de tipo célula de ARN antisentido. El vector de expresión antisentido pueden tener forma de un plásmido recombinante, fagémido, o virus atenuado en el que los ácidos nucleicos antisentido se producen bajo el control de una región reguladora muy eficiente, cuya actividad puede determinarse pro el tipo de célula en el que el vector se introduce. Para un análisis de la regulación de expresión genética usando genes antisentido, véase Weintraub et al. (Reviews—Trends in Genetics, Vol. 1(1) 1986).

Otro aspecto de la invención pertenece a células huéspedes en las que un vector de expresión recombinante de la invención se ha introducido. Los términos "célula huésped" y "célula huésped recombinante" se usan intercambiabilmente en el presente documento. Se entiende que tales términos no solamente se refieren a la célula sujeto particular, sino a la progenie o progenie potencial de tal célula. Debido a que pueden ocurrir ciertas modificaciones en las subsiguientes generaciones debido a mutación o influencias ambientales, tal progenie puede no ser, de hecho, idéntica a la célula original, pero sigue estando incluida en el alcance del término como se usa en el presente documento.

Una célula huésped puede ser cualquier célula procariótica (por ejemplo, *E. coli*) o eucariótica (por ejemplo, células de insecto, células de levadura o mamíferos). Un número de líneas de célula huésped de mamífero adecuadas capaces de expresar polipéptidos glicosilados intactos se han desarrollado en la técnica, e incluyen las líneas celulares COS-1 (por ejemplo, ATCC CRL 1650), COS-7 (por ejemplo, ATCC CRL-1651), HEK293, BHK21 (por ejemplo, ATCC CR-10), CHO (por ejemplo, ATCC CRL 1610) y BSC-1 (por ejemplo, ATCC CRL-26), células Cos-7, células COS, células hepG2, P3X63Ag8.653, SP2/0-Ag14, células 293, células HeLa y similares, que están fácilmente disponibles, por ejemplo, en la Colección Americana de Cultivos Tipo, Manassas, Va (www.atcc.org). En el presente documento, "célula huésped" no incluye células madre del embrión humano.

Los vectores de expresión para estas células pueden incluir uno o más de las siguientes secuencias de control de expresión, un promotor, un potenciador y/o procesar sitios de información, tales como sitios de enlace de ribosomas, sitios de segmentación de ARN, sitios de poliadenilación y secuencias terminadoras transcripcionales (véase, por ejemplo, Ausubel et al., supra; Sambrook, et al., supra).

El vector de ADN puede introducirse en células procarióticas y eucarióticas mediante técnicas convencionales de transformación o transfección. Como se usa en el presente documento, los términos "transformación" y transfección" pretenden referirse a una variedad de técnicas reconocidas en la técnica para introducir ácido nucleico externo en una célula huésped, incluyendo co-precipitación de fosfato cálcico o cloruro cálcico, transfección mediada por DEAE-dextrano, lipofección o electroporación. Los métodos adecuados para transformar o transfectar células huéspedes pueden encontrarse en Sambrook, et al. (supra) y otros manuales de laboratorio.

Para transfección estable de células mamíferas, es bien conocido que, dependiendo del vector de expresión y la técnica de transfección usada, solamente una pequeña fracción de células puede integrar el ADN externo en su genoma. Con el fin de identificar y seleccionar estos integrantes, un gen que codifica un marcador seleccionable (por ejemplo, para resistencia a antibióticos) se introduce generalmente en las células huéspedes junto con el gen de interés. Los marcadores seleccionables preferentes incluyen aquellos que confieren resistencia a los fármacos, tales como, cloranfenicol, tetraciclinas, gentamicina, kanamicina, ampicilina, G418, higromicina, metotrexato, etc. Las células establemente transfectadas con el ácido nucleico introducido pueden identificarse mediante selección de fármacos (por ejemplo, células que han incorporado el gen marcador seleccionable sobrevivirán, mientras que otras células morirán).

Una célula huésped de la invención, tal como una célula huésped procariótica o eucariótica, puede usarse para producir un polipéptido (por ejemplo, una proteína o anticuerpo). Por consiguiente, la divulgación proporciona además métodos para producir un polipéptido usando las células huéspedes de la invención. En una realización, el

método comprende cultivar la célula huésped de la invención (en la que se ha introducido un vector de expresión recombinante que codifica un polipéptido) en un medio adecuado de tal manera que se produzca el polipéptido. En otra realización, el método comprende además aislar el polipéptido del medio o la célula huésped.

Las células huéspedes de la invención también pueden usarse para producir animales transgénicos no humanos. Por ejemplo, en una realización, una célula huésped de la invención es un oocito fertilizado de una célula madre embrionaria en la que al menos se ha introducido una secuencia que codifica un polipéptido. Tales células huéspedes pueden después usarse para crear animales transgénicos no humanos en los que se han introducido secuencias exógenas que codifican un polipéptido en su genoma o animales recombinantes heterólogos en los que se han alterado secuencias endógenas que codifican un polipéptido. Tales animales son útiles para estudiar la función y/o actividad del polipéptido y para identificar y/o evaluar moduladores de la actividad del polipéptido. Como se usa en el presente documento, un "animal transgénico" es un animal no humano, preferentemente, un mamífero, más preferentemente un roedor, tal como una rata o ratón, en el que una o más de las células del animal incluye un transgen. Otros ejemplos de animales transgénicos incluyen primates no humanos, ovejas, perros, vacas, cabras, pollos, anfibios, etc. Un transgen es ADN exógeno que se integra en el genoma de una célula de la que se desarrolla un animal transgénico y que permanece en el genoma del animal maduro, dirigiendo de ese modo la expresión de un producto de gen codificado a uno o más tipos de célula o tejidos del animal transgénico. Como se usa en el presente documento, un "animal recombinante homólogo" es un animal no humano, preferentemente un mamífero, más preferentemente un ratón, en el que el gen endógeno se ha alterado mediante recombinación homóloga entre el gen endógeno y una molécula de ADN exógeno introducida en una célula del animal, por ejemplo, una célula embrionaria del animal, antes del desarrollo del animal.

Un animal transgénico de la invención puede crearse introduciendo ácido nucleico que codifica un polipéptido en los pro-núcleos machos de un oocito fertilizado, por ejemplo, mediante microinyección, infección retroviral, y permitiendo que el oocito se desarrolle en un animal de acogida hembra pseudoembarazado. Las secuencias intrónicas y las señales de poliadenilación también pueden incluirse en el transgen para aumentar la eficiencia de expresión de transgen. Una secuencia o secuencias reguladoras específicas de tejido pueden estar operativamente unida al transgen para dirigir la expresión del polipéptido a células particulares. Métodos para generar animales transgénicos mediante manipulación de embrión y microinyección, particularmente, animales tales como ratones, se han convertido en convencionales en la técnica y se describen, por ejemplo, en Patentes de Estados Unidos Números 4.736.866, 4.870.009 y 4.873.191 y Hogan, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 1986). Se usan métodos similares para la producción de otros animales transgénicos no humanos. Un animal fundador transgénico puede identificarse en base a la presencia del transgen en su genoma y/o expresión de mRNA que codifica el transgen en tejidos o células de los animales. Un animal fundador transgénico puede después usarse para engendrar animales adicionales que lleven el transgen. Además, los animales transgénicos no humanos que llevan el transgen pueden además engendrar otros animales transgénicos no humano que lleven otros transgenes.

Para crear un animal recombinante homólogo, se prepara un vector que contiene al menos una parte de un gen que codifica un polipéptido en el que se ha introducido una eliminación, adición o sustitución para alterar de ese modo, por ejemplo, interrumpir funcionalmente, el gen. En una realización preferente, el vector está diseñado de tal manera que, después de la recombinación homóloga, el gen endógeno se interrumpe funcionalmente (es decir, deja de codificar una proteína funcional; también referido como un vector "noquear"). Alternativamente, el vector puede estar diseñado de tal manera que, después de la recombinación homóloga, el gen endógeno se mute o de otra manera se altere pero que aún codifique la proteína funcional (por ejemplo, la región reguladora corriente arriba puede alterarse para alterar de ese modo la expresión de la proteína endógena). En el vector de recombinación homólogo, la parte alterada del gen está flanqueada en sus extremos 5' y 3' por ácido nucleico adicional del gen para permitir que ocurra la recombinación homóloga entre el gen exógeno transportado por el vector y un gen endógeno en una célula madre embrionaria. Las secuencias adicionales de ácido nucleico flanqueantes tienen la longitud suficiente para la recombinación homóloga buena con el gen endógeno. Típicamente, se incluyen varias kilobases de ADN flanqueante (ambas en los extremos 5' y 3') en el vector (véase, por ejemplo, Thoma y Capecchi (1987) 51:503 para una descripción de vectores de recombinación homólogos). El vector se introduce en una línea de célula madre embrionaria (por ejemplo, mediante electroporación) y se seleccionan las células en las que el gen introducido se ha recombinado homológamente con el gen endógeno (véase, por ejemplo, Li et al. (1992) *Cell* 69:915). Las células seleccionadas se inyectan después en un blastocisto de un animal no humano (por ejemplo, un ratón) para formar quimeras de agregación (IRL, Oxford, 1987), págs. 113-152). Un embrión quimérico no humano puede después implantarse en un animal de acogida hembra pseudoembarazado y el embrión se lleva a término. La progenie que alberga el ADN homológamente recombinado en sus células germinales puede usarse para engendrar animales en los que todas las células del animal contengan el ADN homológamente recombinado por la transmisión de la línea celular del transgen. Los métodos para construir vectores de recombinación homólogos y animales recombinantes homólogos se describen con más detalle en Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 y en Publicaciones PCT Números WO 90/11354, WO 91/01140, WO 92/0968 y WO 93/04169.

En otra realización, pueden producirse animales transgénicos no humanos que contienen sistemas seleccionados que permite la expresión regulada del transgen. Un ejemplo de tal sistema es el sistema recombinasa cre/loxP de bacteriófago P1. Para una descripción del sistema recombinasa cre/loxP, véase, por ejemplo, Lakso et al.

la. (1992) Proc. Natl. Acad. Sci. USA 89:6232-6236. Otro ejemplo de un sistema de recombinasa es el sistema FLP recombinasa de *Saccharomyces cerevisiae* (O'Gorman et al. (1991) Science 251:1351-1355). Si un sistema recombinasas cre/loxP se usa para regular la expresión del transgen, se requieren animales que contengan transgenes que codifiquen Cre recombinasa y una proteína seleccionada. Tales animales pueden proporcionarse a través de la construcción de animales transgénicos "dobles", por ejemplo, emparejando dos animales transgénicos, uno que contenga un transgen que codifica una proteína seleccionada y el otro que contenga un transgen que codifique una recombinasa.

Los clones de los animales transgénicos no humanos descritos en el presente documento también pueden producirse de acuerdo con los métodos descritos en Wilmut et al. (1997) Natura 385:810-813 y Publicaciones PCT Números WO 97/07668 y WO 97/07669.

Anticuerpos

La presente invención incluye además, aunque no se limita a, métodos de uso de ácidos nucleicos y polipéptidos codificados para hacer de este modo anticuerpos y anticuerpos anti-idiotipo, que incluyen composiciones de diagnóstico y terapéuticas, métodos y dispositivos. Tales anticuerpos afectan además opcionalmente a un ligando específico, tal como pero sin limitar a, donde tal anticuerpo modula, disminuye, aumenta, antagoniza, agoniza, mitiga, alivia, bloquea, inhibe, abroga y/o interfiere con al menos una actividad o enlace de proteína, o con actividad o enlace del receptor, in vitro, in situ y/o in vivo. Como un ejemplo no limitativo, un anticuerpo adecuado, parte especificada o variante puede enlazarse al menos con una proteína, o partes especificadas, variantes o dominios de la misma. Un anticuerpo adecuado, parte especificada o variante puede también afectar opcionalmente al menos a una de la actividad o función de la proteína, tal como pero sin limitar a, síntesis de ARN, ADN o polipéptido, liberación de proteína, señalización del receptor, segmentación de membrana, actividad de proteína, producción y/o síntesis de proteína. Los anticuerpos útiles en los métodos y composiciones de la presente invención pueden estar opcionalmente caracterizados por un enlace de alta afinidad con sus antígenos y, opcionalmente y preferentemente, por tener baja toxicidad.

Como se usa en el presente documento, un "anticuerpo", y similares incluyen cualquier polipéptido o péptido que contiene una molécula que comprende al menos una parte de una molécula de inmunoglobulina, tal como pero sin limitar a, al menos una región determinante de la complementariedad (RDC) de una cadena pesada o ligera o una parte de enlace de ligando de la misma, una región variable de cadena pesada o cadena ligera, una región constante de cadena pesada o cadena ligera, una región marco, o cualquier parte, fragmento o variante de la misma, o al menos una parte de un receptor o polipéptido de enlace, que puede incorporarse en un anticuerpo.

Los anticuerpos pueden incluir uno o más de al menos un RDC, al menos una región variable, al menos una región constante, al menos una cadena pesada (por ejemplo, g1, g2, g3, g4, m, a1, a2, d, e), al menos una cadena ligera (por ejemplo, kappa y lambda), o cualquier parte o fragmento de la misma, y pueden además comprender enlaces de disulfuro inter-cadena o intra-cadena, regiones bisagra, sitios de glicosilación que pueden separarse por una región bisagra, así como cadenas pesadas y cadenas ligeras. Las cadenas ligeras tienen típicamente un peso molecular de aproximadamente 25Kd y las cadenas pesadas oscilan entre 50K y 77 Kd. Las cadenas ligeras existen en dos formas distintas de isotipos, kappa (k) y lambda (l), que pueden combinarse con cualquiera de los tipos de cadena pesada. Todas las cadenas ligeras tienen al menos una región variable y al menos una región constante. Se considera que el anticuerpo IgG es una estructura típica de anticuerpo y tiene dos enlaces de disulfuro intra-cadena en la cadena ligera (uno en la región variable y uno en la región constante), con cuatro en la cadena pesada, y tal enlace abarca un bucle de péptido de aproximadamente 60-70 aminoácidos que comprenden un "dominio" de aproximadamente 110 aminoácidos en la cadena. Los anticuerpos IgG pueden caracterizarse en cuatro clases, IgG1, IgG2, IgG3 e IgG4. Cada clase de inmunoglobulina tiene un conjunto diferente de funciones. La siguiente tabla resume las propiedades fisicoquímicas de cada una de las clases y subclases de inmunoglobulina.

Propiedad	IgG1	IgG2	IgG3	IgG4	IgM	IgA1	IgA2	SigA	IgD	IgE
Cadena pesada	$\gamma 1$	$\gamma 1$	$\gamma 1$	$\gamma 1$	μ	$\alpha 1$	$\alpha 2$	$\alpha 1 \alpha 2$	δ	ϵ
Conc. Media de suero (mg/ml)	9	3	1	0,5	1,5	3,0	0,5	0,05	0,03	0,00005
Constante de sedimentación	7s	7s	7s	7s	19s	7s	7s	11s	7s	8s
Peso Mol. ($\times 10^3$)	146	146	170	146	970	160	160	385	184	188
Vida media (días)	21	20	7	21	10	6	6	¿	3	2
% distribución intravascular	45	45	45	45	80	42	42	Rastro	75	50
Carbohidrato (%)	2-3	2-3	2-3	2-3	12	7-11	7-11	7-11	9-14	12

La siguiente tabla resume ejemplos no limitativos de funciones efectoras de anticuerpo para clases y subclases de anticuerpo humano.

5	Función efectora	IgG1	IgG2	IgG3	IgG4	IgM	IgA	IgD	IgE
	Fijación complemento	+	+/-	++	-	++	-	-	-
	Transferencia placentar	+	+/-	+	+	-	-	-	-
10	Enlace con Staph A	+++	+++	-	+++	-	-	-	-
	Enlace con Strep G	+++	+++	+++	+++	-	-	-	-

+++ = muy alto; ++ = alto; + = moderado; +/- = mínimo; - = ninguno; ¿ = cuestionable

15 Como se describe más abajo, existen varios métodos para producir anticuerpos. Una vez que un anticuerpo se produce mediante cualquiera de estos métodos, su aminoácido y correspondientes secuencias genéticas pueden identificarse y, opcionalmente, modificarse, (por ejemplo, optimizarse, humanizarse, etc.) de tal manera que el anticuerpo pueda después producirse recombinantemente.

20 Por ejemplo, un polipéptido especificado, o un fragmento del mismo, puede usarse como un inmunogen para generar anticuerpos usando técnicas estándares para la preparación de anticuerpos policlonales y monoclonales. Puede usarse el polipéptido de longitud completa o proteína o alternatively, la divulgación proporciona fragmentos peptídicos antigénicos para su uso como inmunógenos. El péptido antigénico comprende al menos 8 (preferentemente 10, 15, 20 o 30 o más) residuos de aminoácido de una secuencia de proteína y abarca un epítipo de la proteína de tal manera que un anticuerpo se levanta contra el péptido forme un complejo inmune específico con la proteína.

25 Un inmunógen típicamente se usa para preparar anticuerpos inmunizando un sujeto adecuado (es decir, inmunocompetente), tal como un conejo, cabra, ratón u otro mamífero o vertebrado. Una preparación inmunogénica apropiada puede contener, por ejemplo, polipéptido recombinantemente expresado o químicamente sintetizado. La preparación puede además incluir un adyuvante, tal como adyuvante completo o incompleto de Freund, o un agente inmunoestimulante similar.

30 Las células que producen anticuerpos pueden obtenerse a partir de sangre periférica o, preferentemente, los nodos del bazo o linfáticos de humanos u otros animales adecuados que se han inmunizado con el inmunógen de interés. Cualquier otra célula huésped adecuada también puede usarse para expresar ácido nucleico heterólogo o endógeno que codifica un anticuerpo, fragmentos especificados o variantes del mismo. Las células fusionadas (hibridomas) o células recombinantes pueden aislarse usando condiciones selectivas de cultivo u otros métodos conocidos, y clonarse limitando la dilución y la clasificación celular, u otros métodos conocidos. Las células que producen anticuerpos con la especificidad deseada pueden seleccionarse mediante un ensayo adecuado (por ejemplo, ELISA).

35 En una técnica, un hibridoma se produce fusionando una línea celular inmortal (por ejemplo, una línea celular de mieloma, tal como, aunque sin limitar a, Sp2/0, Sp2/0-AG14, NS0, NS1, NS2, AE-1, L5, <243, P3X63Ag8.653, Sp2 SA3, Sp2 MAI, Sp2 SS1, Sp2 SA5, U937, MLA 144, ACT IV, MOLT4, DA-1, JURKAT, WEHI, K-562, COS, RAJI, NIH 3T3, HL-60, MLA 144, NAMALWA, NEURO 2A, o similares), o heteromieloma, productos de fusión de los mismos, o célula o célula de fusión derivada de los mismos, o cualquier línea celular adecuada como las conocidas en la técnica (véase, por ejemplo, www.atcc.org, www.lifetech.com, y similares), con células que producen anticuerpo, tales como, aunque sin limitarse a, bazo aislado o clonado, sangre periférica, linfa, amígdala, y otras células inmunes o que contienen célula B, o cualquier otras células que expresen secuencias constantes o variables o RDC de cadena pesada o ligera, ya sea ácido nucleico endógeno o heterólogo, recombinante o endógeno, viral, bacteriano, algal, procariótico, anfibio, insecto, reptil, pez, mamífero, roedor, equino, ovino, de cabra, oveja, primate, eucariótico, ADN genómico, cADN, rADN, ADN o ARN mitocondrial, ADN o ARN cloroplasto, hnARN, mARN, tARN, de único, doble o triple hebra, hibridizado, y similares y cualquier combinación de los mismos. Véase, por ejemplo, Ausubel, supra, y Colligan, Immunology, supra, capítulo 2.

40 Otros métodos adecuados de producir o aislar anticuerpos de la especificidad requerida que pueden usarse incluyen, aunque no se limitan a, métodos que seleccionan anticuerpo recombinante de un péptido o biblioteca de péptidos (por ejemplo, aunque sin limitar a, un bacteriófago, ribosoma, oligonucleótido, ARN, cADN, o similares, biblioteca de exposición; por ejemplo, como la disponible en Cambridge antibody Technologies, Cambridgeshire, Reino Unido; MorphoSys, Martinsried/Planegg, Alemania; Biovation, Aberdeen, Escocia, Reino Unido; BioInvent, Lund, Suecia; Dyax, Corp., Enzon, Affymax/Biosite; Xoma, Berkeley, CA; Ixys. Véase, por ejemplo, EP 368.684, PCT/GB91/01134; PCT/GB92/01755; PCT/GB92/00883; PCT/GB93/00605; US 08/350260(5/12/94); PCT/GB94/01422; PCT/GB94/02662; PCT/GB97/01835; (CAT/MRC); WO90/14443; WO90/14424; WO90/14430; PCT/US94/1234; WO92/18619; WO96/07754; (Scripps); EP 614 989 (MorphoSys); WO95/16027 (BioInvent); WO88/06630; WO90/3809 (Dyax); US 4.704.692 (Enzon); PCT/US91/02989 (Affymax); WO89/06283; EP 371 998;

EP 550400; (Xoma); EP 229 046; PCT/US91/07149 (Ixys); o péptidos o polipéptidos estocásticamente generados-
US 5723323, 5763192, 5814476, 5817483, 5824514, 5976862, WO 86/05803, EP 590 689 (Ixsys, ahora Applied
Molecular Evolution (AME) o la que se basa en la inmunización de animales transgénicos (por ejemplo, ratones
SCID, Nguyen et al., Microbiol. Immunol. 41:901-907 (1997); Sandhu et al., Crit. Rev. Biotechnol. 16:95-118 (1996);
Eren et al., Immunol. 93:154-161 (1998), así como patentes y solicitudes relacionadas) que son capaces de producir
un repertorio de anticuerpos humanos, como los conocidos en la técnica y/o descritos en el presente documento.
Tales técnicas incluyen, aunque no se limitan a, exposición de ribosoma (Hanes et al., Proc. Natl. Acad. Sci. USA,
94:4937-4942 (Mayo 1997); Hanes et al., Proc. Natl. Acad. Sci. USA, 95:14130-14135 (Nov. 1998)); tecnologías de
producción de anticuerpo de única célula (por ejemplo, método de anticuerpo de linfocitos seleccionado ("SLAM")
(Patente de Estados Unidos N° 5.627.052), Wen et al., J. Immunol. 17:887-892 (1987); Babcook et al., Proc. Natl.
Acad. Sci. Usa 93:7843-7848 (1996); microgotita de gel y citometría de flujo (Powell et al., Biotechnol. 8:333-337
(1990); One Cell Systems, Cambridge, MA; Gray et al., J. Imm. Meth. 182:155-163 (1995); Kenny et al., Bio/Technol.
13:787-790 (1995)); Selección de célula B (Steenbakkers et al., Molec. Biol. Reports 19:125-134 (1994); Jonak et al.,
Progress Biotech, Vol. 5, Inmunización In Vitro en Tecnología de Hibridoma, Borrebaeck, ed., Elsevier Science
Publishers B. V., Amsterdam, Holanda (1988)).

Los métodos para diseñar o humanizar anticuerpos no humano o humanos también pueden usarse y son
bien conocidos en la técnica. Generalmente, un anticuerpo humanizado o diseñado tiene uno o más residuos de
aminoácido de una fuente que no es humana, por ejemplo, pero sin limitarse a, ratón, rata, conejo, primate no
humano y otro mamífero. Estos residuos de aminoácido humanos a menudo son referidos como residuos de
"importación", que típicamente se toman de un dominio variable, constante y otro dominio de "importación" de una
secuencia humana conocida. Las secuencias Ig humanas conocidas se desvelan, por ejemplo, en los siguientes
sitios web:

www.ncbi.nlm.nih.gov/entrez/queri.fcgi; www.atcc.org/phage/hadb.html; www.sciquest.com/; www.abcam.com/;
www.antibodyresource.com/onlinecomp.html; www.public.iastate.edu/~pedro/research_tools.html; www.mgen.uni-heidelberg.de/SD/IT/IT.html;
www.whfreeman.com/Immunology/CH05/kuby05.html;
www.library.thinkquest.org/12429/Immune/Antibody.html; www.hhmi.org/grnts/lectures/1996/vlab/;
www.path.cam.ac.uk/~mrc7/mikeimages.html; www.antibodyresource.com/;
mcb.harvard.edu/BioLinks/Immunology.html; www.immunologylink.com/; pathbox.wustl.edu/~hcenter/index.html;
www.biotech.ufl.edu/~hcl/; www.pebio.com/pa340913/340913.html; www.nal.usda.gov/awic/pubs/antibody/;
www.m.ehime-u.ac.jp/~yasuhito/Elisa.html; www.biodesign.com/table.asp; www.icnet.uk/axp/facs/davies/links.html;
www.biotech.ufl.edu/~fccl/protocol.html; www.isac-net.org/sites_geo.html; aximt1.imt.uni-marburg.de/~rek/AEPStart.html;
baserv.uci.kun.nl/~jraats/links.html; www.recab.uni-hd.de/immuno.bme.nwu.ed/;
www.mrc-cpe.cam.ac.uk/imt-doc/public/INTRO.html; www.ibt.unam.mx/vir/V_mice.html; imgt.cnusc.fr:8104/;
www.biochem.ucl.ac.uk/~martin/abs/index.html; antibody.batch.ac.uk/; abgen.cvm.tamu.edu/lab/wwwabgen.html;
www.unizh.ch/~honegger/AHOseminar/Slide01.html; www.cryst.bbk.ac.uk/~ubcg07s/;
www.nimr.mrc.ac.uk/CC/caewg/caewg.html; www.path.cam.ac.uk/~mrc7/humanisation/TAHHP.html;
www.ibt.unam.mx/vir/structure/stat_aim.html; www.biosci.missouri.edu/smithgp/index.html;
www.cryst.bioc.cam.ac.uk/~fmolina/Web-pages/Pept/spottech.html; www.jerini.de/fr_products.html;
www.patents.ibm.com/ibm.html. Kabat et al., Sequences of Polypeptides of Immunological Interest, US. Dept. Health
(1983).

Tales secuencias importadas pueden usarse para reducir la inmunogenicidad o reducir, aumentar o
modificar en enlace, afinidad, constante de asociación, constante de disociación, avidez, especificidad, vida
media, o cualquier otra característica adecuada, como las conocidas en la técnica. Generalmente, parte de todas las
secuencias RDC no humana o humanas se mantienen mientras que las secuencias no humanas de las regiones
variable y constante se sustituyen por humanos u otros aminoácidos. Los anticuerpos pueden opcionalmente
humanizarse con la retención de una elevada afinidad para el antígeno y otras propiedades biológicas favorables.
Para conseguir este objetivo, los anticuerpos humanizados pueden prepararse opcionalmente mediante un proceso
de análisis de las secuencias parentales y varios productos conceptuales humanizados que usan modelos
tridimensionales de las secuencias parentales y humanizadas. Los modelos tridimensionales de inmunoglobulina
están comúnmente disponibles y son familiares para aquellos expertos en la técnica. Hay disponibles programas de
ordenador que ilustran y exponen estructuras conformacionales tridimensionales probables de secuencias de
inmunoglobulina seleccionadas candidatas. La inspección de estas exposiciones permite el análisis del papel
probable de los residuos en el funcionamiento de la secuencia de inmunoglobulina candidata, es decir, el análisis de
residuos que influyen en la habilidad de la inmunoglobulina candidata para enlazarse con su antígeno. De esta
manera, los residuos FR pueden seleccionarse y combinarse de las secuencias de consenso e importación para que
se consigan las características deseadas del anticuerpo, tales como la mayor afinidad para el antígeno o antígenos
diana(s). En general, los residuos RDC están directamente y más sustancialmente implicados en la influencia del
enlace de antígeno. La humanización o diseño de anticuerpos de la presente invención puede realizarse usando
cualquier método descrito, tales como aunque sin limitar, aquellos descritos en, Winter (Jones, et al. Nature 321:522
(1986); Riechmann et al., Nature 332:3213 (1988); Verhoeyen et al., Science 239:1534 (1988)), Sims et al. Immunol.
151:2296 (1993); Chothia y Lesk, J. Mol. Biol. 196:901 (1987), Carter et al., Proc. Natl. Acad. Sci. U.S.A. 89:4285
(1992); Presta et al., J. Immunol. 151:2623 (1993); Patentes de Estados Unidos Números: 5723323; 5976862;
5824514; 5817483; 5814476; 5763192; 5723323; 5766886; 5714352; 6204023; 6180370; 5693762; 5530101;
5585089; 5225539 y 4816567; PCT: US98/16280; US96/18978; US91/09630; US94/01234; GB89/01334;

GB91/01134; GB92/01755; WO90/14443; WO90/14424 y WO90/14430; EP 229246.

Los anticuerpos también pueden generarse opcionalmente mediante inmunización de un animal transgénico (por ejemplo, ratón, rata, hámster, primate no humano, y similares) capaz de producir un repertorio de anticuerpos humano, como se describe en el presente documento y/o se conoce en la técnica. Las células que producen un anticuerpo pueden aislarse de tales animales e inmortalizarse usando métodos adecuados, tales como los métodos descritos en el presente documento.

Los ratones transgénicos que pueden producir un repertorio de anticuerpos humanos que se enlazan con antígenos humanos pueden producirse mediante métodos conocidos (por ejemplo, aunque sin limitar a, Patentes de Estados Unidos Números: 5.770.428, 5.569.825, 5.545.806, 5.625.126, 5.625.825, 5.633.425, 5.661.016 y 5.789.650 presentado por Lonberg et al; Jakobovits et al. WO 98/50433, Jakobovits et al. WO 98/24893, Lonberg et al. WO 98/24884, Lonberg et al. WO 97/13852, Lonberg et al. WO 94/25585, Kucherlapate et al. WO 96/34096, Kucherlapate et al. EP 0463 151 B1, Kucherlapate et al. EP 0710 719 A1, Surani et al. Patente de Estados Unidos N° 5.545.807, Bruggemann et al. WO 90/04036, Bruggemann et al. EP 0438 474 B1, Lonberg et al. EP 0814 259 A2, Lonberg et al. GB 2 272 440 A, Lonberg et al. Nature 368:856-859 (1994), Taylor et al., Nature Genetics 15:146-156 (1997), Taylor et al., Nucleic Acids Reserch 20(23):6287-6295 (1992), Tuailon et al., Proc Ntal Acad Sci USA 90(8):3720-3724 (1993), Lonberg et al., Int Rev Immunol 13(1):65-93 (1995) y Fishwald et al., Nat Biotechnol 14(7):845-851 (1996). Generalmente, estos ratones comprenden al menos un transgen que comprende ADN de al menos un locus de inmunoglobulina humana que se reorganiza funcionalmente, o que puede sufrir recolocación funcional. Los lugares de inmunoglobulina endógena en tales ratones pueden interrumpirse o borrarse para eliminar la capacidad del animal para producir anticuerpos codificados por genes endógenos.

Los anticuerpos también pueden prepararse en leche administrando al menos un anticuerpo que codifica ácido nucleico a animales transgénicos o mamíferos, tales como cabras, vacas, caballos, ovejas, y similares que produzcan anticuerpos en su leche. Tales animales pueden proporcionarse usando métodos conocidos, Véase, por ejemplo, aunque sin limitar las Patentes de Estados Unidos Números 5.827.690; 5.849.992, 4.873.316, 5.849.992, 5.994.616, 5.565.962, 5.304.489, y similares. Los anticuerpos pueden además prepararse usando al menos un anticuerpo que codifica un ácido nucleico para proporcionar plantas transgénicas y células de plantas cultivadas (por ejemplo, aunque sin limitar a, tabaco y maíz) que producen tales anticuerpos, partes especificadas o variantes en las partes de la planta y en las células cultivadas de las mismas.

Los anticuerpos pueden enlazarse con antígenos con un amplio rango de afinidades (K_D). En una realización preferente, al menos un anticuerpo monoclonal humano de la presente invención puede enlazarse opcionalmente con su antígeno con alta afinidad. Por ejemplo, un anticuerpo monoclonal humano puede enlazarse con un antígeno humano con un K_D igual o inferior a aproximadamente 10^{-7} M, tal como pero sin limitar a 0,1-9,9 (o cualquier rango o valor en el mismo) $\times 10^{-7}$, 10^{-8} , 10^{-9} , 10^{-10} , 10^{-11} , 10^{-12} , 10^{-13} o cualquier rango o valor en el mismo.

La afinidad o avidez de un anticuerpo para un antígeno puede determinarse experimentalmente usando cualquier método adecuado. (Véase, por ejemplo, Berzofsky, et al., "Antibody-Antigen Interactions". En Fundamental Immunology, Paul, W. E., Ed., Raven Press: Nueva York, NY (1984); Kuby, Janis Immunology, W. H. Freeman & Company: Nueva York, NY (1992); y métodos descritos en los mismos). La afinidad medida de una interacción particular anticuerpo-antígeno puede variar si se mide bajo diferentes condiciones (por ejemplo, concentración de sal, pH). De este modo, las mediciones de afinidad y otros parámetros de enlace con antígeno (por ejemplo, K_D , D_a , D_d) se hacen preferentemente con soluciones estandarizadas de anticuerpo y antígeno, y un tampón estandarizado, tal como el tampón descrito en el presente documento.

Un anticuerpo dirigido contra un polipéptido (por ejemplo, anticuerpo monoclonal) puede usarse para aislar el polipéptido mediante técnicas estándares, tael como cromatografía por afinidad o inmunoprecipitación. Además, tal anticuerpo puede usarse para detectar la proteína (por ejemplo, en un lisado celular o sobrenadante celular) con el fin de evaluar la abundancia y patrón de expresión del polipéptido. Los anticuerpos pueden también usarse mediante diagnóstico para controlar los niveles de proteína en tejido como parte de un procedimiento clínico de análisis, por ejemplo, para determinar la eficacia de un régimen de tratamiento dado. La detección puede facilitarse uniando el anticuerpo a una sustancia detectable. Ejemplos de sustancias detectables incluyen varias enzimas, grupos prostéticos, materiales fluorescentes, materiales luminiscentes, materiales bioluminiscentes y materiales radioactivos. Ejemplos de enzimas adecuadas incluyen peroxidasa de rábano picante, fosfatasa alcalina, β -galactosidasa o acetilcolinesterasa; ejemplos de complejos adecuados de grupo prostético incluyen estreptavidina/biotina y avidina/biotina; ejemplos de materiales fluorescentes adecuados incluyen umbeliferona, fluoresceína isotiocianato, rodamina, diclorotriacilnamina fluoresceína, cloruro de dansilo o ficoeritrina; un ejemplo de material luminiscente incluye luminol; ejemplos de materiales bioluminiscentes incluyen luciferasa, luciferina y acuorina, y ejemplos de material radioactivo adecuado incluyen ^{125}I , ^{131}I , ^{135}S o ^3H .

Características de Vectores

Se investigaron los promotores inherentes de líneas celulares transfectadas con una fuerte capacidad de producción de anticuerpo monoclonal. Se descubrió inesperadamente que se obtuvieron buenos niveles de

producción, es decir, aproximadamente 300 mg/L, o 25-30 pg por célula por día sobre una base de productividad específica, incluso con promotores de gen humano que se reconocieron por factores de transcripción de ratón en la célula huésped de mieloma murino.

Se observó que las líneas celulares transfectadas C379B y C381B fueron fuertes productores de un anticuerpo monoclonal humano con la citoquina IL-12 (anticuerpo anti-IL-12 o IL-12 mAb) codificado por los genes IgHC y LC completamente humanos. La invención proporciona las secuencias completas de ADN de los plásmidos de expresión HC y LC usados para crear las líneas celulares C379B y C381 B y que contienen el gen de anticuerpo anti-IL-12. Los vectores y plásmidos se han diseñado para permitir una inserción conveniente con una única etapa de varios genes de región variable de anticuerpo. En una realización preferente, el vector se usa en un proceso de una única etapa para sustituir la secuencia de región variable anti-IL-12 por una secuencia de región variable que codifique otro anticuerpo monoclonal de interés, por ejemplo, un anticuerpo para una interleuquina humana, un factor de crecimiento, etc. Además, los vectores y plásmidos permiten la sustitución de regiones constantes de anticuerpo.

Los mapas de dos plásmidos de expresión de acuerdo con la invención, p1560 y p1558 se representan esquemáticamente en las Figs. 1A y 1B. Sus secuencias se desvelan en las SEQ ID NOS: 11 y 12. El plásmido HC específicamente descrito en el presente documento, p1560, codifica regiones constantes del isotipo IgG1 humano. El plásmido LC específicamente descrito en el presente documento, p1558, codifica una región constante del isotipo kappa humano. La secuencia completa de ADN para cada uno de los dos plásmidos tiene alrededor de 20 kb en longitud y sus características seleccionadas, incluyendo, aunque sin limitación, el posicionamiento de las varias secuencias de polinucleótido de control de expresión, se enumeran en la Tabla 1.

Los genes de anticuerpo en esto dos plásmidos están presentes en la orientación inversa/complementaria con el fin de que los ejes centrales del vector pSV2gpt se presenten en sus orientación convencional, como se representa en las Figuras 1A y 1B. Como consecuencia, los números de posición del nucleótido correspondientes al extremo de terminal carboxil de la secuencia codificadora de anticuerpo (secuencia codificadora cH3) son inferiores y los números de posición del nucleótido correspondientes al extremo de terminal amino de la secuencia codificadora de anticuerpo (sitio de clonación de región V) más altos. Además, todas las secuencias que codifican anticuerpos corresponden a la "hebra menos" como se presente en las secuencias mostradas aquí. Esta orientación es razonable para la descripción de los sitios de enlace del factor de transcripción ya que tales sitios de enlace pueden aplicarse a la "hebra más" o "hebra menos".

Tabla 1. Posiciones de nucleótido de características seleccionadas en p1560 y p1558

Característica	p1560	p1558
Secuencia flanqueante 5'	13200-14950	12614-14871
Región promotora de gen Ig	11209-13199	10623-12613
Secuencia de señal	11060-11070 y 11154-11199	10422-10432 y 10559-10612
Secuencia codificadora de región V	10702-11059	10100-10421
Potenciador de intrón	8449-9030	6765-7209
Secuencia codificadora de CH1	5002-2595	N.A
Secuencia codificadora bisagra	4566-4610	N.A
Secuencia codificadora de CH2	4118-4447	N.A
Secuencia codificadora de CH3	3701-4020	N.A
Secuencia codificadora de kappa C	N.A	5943-6262
Marcador seleccionable de gpt	16200-16655	16122-16577
Gen de resistencia a Amp	18410-19267	18332-19189

Los vectores de la presente invención pueden usarse para expresar las secuencias de región constante para IgG1, IgG2, IgG3, IgG4, IgD, IgE o IgM humano, IgG1, IgG2a, IgG2b, IgG3 de ratón o IgG1, IgG2a, IgG2b, IgG2c de rata. Alternativamente, los vectores de la presente invención pueden usarse para expresar las versiones ΔCH1 de las secuencias de región constante anteriormente enumeradas.

Cualquier anticuerpo, proteína derivada de inmunoglobulina, proteína de fusión, otra proteína o parte de la misma puede sustituirse por la secuencia codificadora de región V en los vectores/plásmidos de la presente invención, por ejemplo, dominio extracelular de un receptor TNF. Para tal proteína de fusión que incluye el dominio CH1, sería necesario alguna forma de LC para la secreción de la célula.

El plásmido de expresión p1560 HC como el mostrado esquemáticamente en el presente documento contiene el promotor HC y un potenciador de intrón; sin embargo, el tipo de promotor y potenciador puede variar. El plásmido de expresión p1560 HC contiene una única secuencia que codifica la región constante. Es una variante alotípica que codifica un residuo Arg en la posición 214 (G1m(1,3)alotipo) en lugar del residuo Lys (G1m(1,17)alotipo)

Análisis Secuencial del Promotor de Gen HC

Para definir los elementos en el promotor del gen HC de anticuerpo anti-IL-12 que podría impactar la transcripción del gen y ser parcialmente responsable de los altos niveles de expresión del anticuerpo anti-IL-12 observados en células transfectadas, se realizó un análisis bioinformático de la secuencia 2000-base corriente arriba del codón de inicio de traslación HC. Este análisis identificó motivos de secuencia reconocidos por los factores de transcripción relevantes (TFs). La base de datos TF más actualizada, TRANSFAC 7.2 (Matys et al., 2003), se usó para la comparación. Después, se aplicaron varios algoritmos de búsqueda de matriz y patrón para identificar secuencias relevantes. Algunos de los resultados se validaron posteriormente mediante la bibliografía relevante que sostiene la conclusión de que la sinergia y la combinación de estos TFs pueden impulsar la alta producción de anticuerpos.

Se construyeron varios modelos de factor de transcripción de alta calidad y específicos de ratón con diferentes parámetros para una búsqueda de matriz (Goessling et al., 2001). Un modelo específico de factor de transcripción de linfocito de ratón también se construyó para una búsqueda de matriz. Además, se realizaron varias búsquedas de ajustes para las diferentes longitudes posteriores (6 bps y superiores).

En base a estos resultados, se identificaron 21 sitios potenciales de enlace para TFs de ratón (Tabla 2). La mayoría de estos TFs fueron específicos de linfocito B. algunos de los TFs pueden activarse durante diferentes fases del desarrollo de célula B. Su activación también es dependiente de la presencia de otros sitios de enlace y la interacción de otros factores.

Los códigos de acceso de la base de datos TRANSFAC para estos TFs de ratón son: T00032 para Ap1; T01575 para STATx; T00111 para c-Ets-1; T01852 para HMG_IY; T00479 para Lyf-1; T00278 para YY1; T00613 para NF-Y; T00989 para CREB; T00402 para IRF; T02057 para IPF1; T00422 para IRF-1; T01432 para cMaf; T00814 para TFE3-S; T00017 para C/EBPBeta; T01114 para C/EBPdelta; T01864 para POU2F2/Oct-2; T00644 para POU2F2a/Oct-1; T00651 para POU5F1/Oct-3; T00702 para PU.1; T00169 para c-Rel; y T01159 para TFIID.

Se descubrió que seis TFs estaban localizados 500 bp corriente arriba del sitio de iniciación de la transcripción. Estos son POU2F2/Oct-2, POU2F2a/Oct-1, POU5F1/Oct-3, pU.1, c-Rel y TFIID de ratón. Se identificaron los siguientes TFs auxiliares o TFs en asociación con funciones potenciadoras: Ap1; STATx; c-Ets-1; HMG_IY; Lyf-1; YY1 y TFIID.

Tabla 2: Motivos secuenciales en promotor p1560 relevantes para transcripción

TF	Posición en p1560	Sito de enlace ID	Secuencia reconocida por TF (SEQ ID NO)	Puntuación Central	Matriz/Parche	Fuentes/Método de Búsqueda	Anotación
AP-1	12292-	V\$API_Q4	agTGACTgacg (13)	1	0,984	HQ_mus/Inmunológico/Parche	Interactivo con c-ETs-1 y NF-Atp
STATx	12643-	V\$STAT_01	TTCCCTgaa (14)	1	0,969	HQ_mus	Expresado en célula B, particularmente en centro germinal (Henderson, 1998)
C-Ets-1	12697-	V\$ETS1_B	gcAGGAAGtgaaagt (15)	1	0,964	HQ_mus/Inmunológico/Parche	Desarrollo acelerado de célula B de proB a célula de plasma
HMG_IY	12381-	V\$HMG_IY_06	GGAAAGt (16)	1	0,979	Inmunoespecífico	Factor auxiliar para otros factores de transcripción como NF-kB o ATF-2
LyF-1	12239-	V\$LYFI_01	TtTGGGAgg (17)	1	0,989	Inmunoespecífico	Expresado de proB a célula de plasma
YY1	11788-	V\$YY1_Q6	CaaaATGGC (18)	1	0,997	Inmunoespecífico	Represor ubicuo
NF-Y	12311	V\$NFY_Q6_01	AgtagCCAATGG (19)	1	0,974	HQ_mus	Se enlaza con elementos de caja Y de genes MHC clase II
CREB	12289-	V\$CREB_01	TGACGtag (20)	1	0,971	HQ_mus	Se enlaza en linfocitos; media la respuesta cAMP

TF	Posición en p1560	Sito de enlace ID	Secuencia reconocida por TF (SEQ ID NO)	Puntuación Central	Matriz/Parche	Fuentes/Método de Búsqueda	Anotación
IRF/ICSBP	11955-; 12694-	V\$IRF_Q6	aaaaaTGAAAGaact (21); ggaagTGAAAGtaat (22)	1	0,967	HQ_mus	Expresado en bazo y timo. Inducido por IFN-gama
IPF1	12058-	V\$IPFI_Q4	gtgcTAATGaa a (23)	1	0,965	HQ_mus	Células beta de páncreas; promotor de insulina factor 1
IRF-1/IRF-2	11960+	RATONE S COX2_02	TTTCATTTTT (24)	0	100	Parche	Requerido para linfopoyesis B; interactivo con ICSBP
C-Maf	12833+; 11862	RATONE S IL4_02	TCAGCA (25)	0	100	Parche	Expresado en célula Th2, implicado en activación de gen específico de Th2
TFE3-S	13105+	RATONE S IGH_10	CATGTG (26)	0	100	Parche	Expresado de proB a células de plasma
C/EBP Beta	12868+	RATONE S INOS_06	TGATGTAAT (27)	0	100	Parche	Expresado de células GC B a células de plasma; sinergiza con Nf-kB/rel
C/EBP Delta	11886-	V\$CEBP DELTA_Q6	ggtgcaGCAATg (28)	1	0,965	HQ_mus/inmuno	Expresado ubicuamente; sinergiza con Nf-kB
PU.1	11506+; 12280-	RATONE S GSHPX1_01	CTTCTC (29)	0	100	Parche	Expresado de progenitor a GC B; sin células B si PU.1 es deficiente
POU2 F2/Oct-2	11304+	RATONE S IGH_44	ATTTGCAT (30)	0	100	Parche	Expresado de proB a células de plasma; defecto en secreción si es deficiente
TFIID	11601+	RATONE S MBP_04	TTCAAA (31)	0	100	Parche	Interactivo con PU.1,REL
C-Rel	11580-	V\$CREL_01	TggccTTTCC (32)	1	0,968	HQ_mus	Prevalece en células B maduras, activador potente
POU5 F1/Oct-3	11304+	RATONE S IGH_45	ATTTGCAT (33)	0	100	Parche	Pueden tener función de represión de potenciador de IgH
POU2 F1/Oct-1	11303-; 12598-	V\$OCTI_B	tATGCAaatg (34); CCGAAtatgcA ATTC (35)	1	1	HQ_mus/Parche	El enlace de ADN se reducir por GR de una manera dependiente de ligando
<p>- Posición en vector HC en p1560 corresponde al primer nucleótido; "+" indica más hebra; "-" indica menos hebra.</p> <p>- Sitio de enlace en la identificación de sitio de enlace TF asignado por la base de datos TRANSFAC.</p> <p>- Puntuación cenral indica la puntuación de enlace calculada por TRANSFAC para los cinco nucleótidos más conservados, consecutivos usados en una matriz.</p> <p>-Puntuación coincidencia/parche indica la similitud de una subsecuencia a la matriz de TRANSFAC. La puntuación es de 0 a 1, siendo 1 la mejor coincidencia.Si la puntuación es 100, fue una búsqueda Parche.</p> <p>- Fuente/método de búsqueda indica si se usó el modelo de matriz específico de inmunocélula, el modelo de matriz de alta calidad y/o el método de búsqueda de parche.</p> <p>- La subsecuencia Oct-2 se publicó por Henderson et al., 1995.</p>							

Análisis Secuencial del Promotor de Gen LC

La región del promotor de gen LC se analizó mediante los mismos métodos, que produjeron una lista de 20 factores potenciales de transcripción de ratón y sus sitios de enlace (mostrados en la Tabla 3). Los códigos de acceso de la base de datos TRANSFAC para estos TFs son: T02057 para IPF12, T01786p ara E12, T00169 para c-Rel, T00549 para NF-AT, T01675 para NKx2-5, T01575 para STATx, T01852 para HMG1Y; T00111 para c-Ets-1, T00702 para Pu.1, T00278 para YY1, T00032 para Ap1, T01429 para Sox-5, T00017 para C/EBPBeta, T00814 para TF3-S, T01554 para Mitf, T01159 para TFIID, T00613 para NF-Y, T01114 para C/EBPdelta, T01864 para

POU2F2/Oct-2 y T00644 para POU2F2a/Oct-1. Entre ellos, se descubrió que NF-AT, C/EBPdelta, POU2F2a/Oct-1, Sox-5, E12, Statx, NF-Y, TFE3-S, Mitf, C/EBPBeta, TFIID y POU2F estaban localizados 500 bps corriente arriba del sitio de inicio de transcripción. Los TFs auxiliares de cadena ligar fueron Ap1, STATx, c-ETS-1, HMGIIY, YY1 y TFIID.

5 Tabla 3. Motivos secuenciales en promotor p1558 relevantes para transcripción

	TF	Posición en p1560	Sito de enlace ID	Secuencia reconocida por TF (SEQ ID NO)	Puntuación Central	Matriz/Par che	Fuentes/Método de Búsqueda	Anotación
10	HMGIIY	11827-	V\$HMGIIY_Q6	aaTTTCC (36)	1	1,00	InmunoEspec	Factor auxiliar para otro TF tal como NF-kB o ATF-2
	NF-AT	12249-	V\$NFAT_06	gctTTTCCtaaa (37)	1	0,983	InmunoEspec	Células T activadas; compuestas de un componente AP-1 y de tipo Rel
	PU.1	11676-	V\$PUL_Q6	CTTCCcct (38)	1	0,969	InmunoEspec	Expresado de progenitor a células GC B. Sin células B es deficiente.
15	C/EBP delta	10913-	V\$CEBP DELTA_Q6	cATTGCtccatc (39)	1	0,966	HQ_mus/ InmunoEspec	Expresado de manera ubicua; actúa sinérgicamente con NF-kB
	POU2F 1a/Oct-1	10696-	V\$OCT1_Q6	ctcaTTTGCatgttc (40)	1	0,994	HQ_mus/ Parche	Enlace de ADN se reduce por GR de una manera dependiente de ligando.
20	Sox-5	11090	V\$SOX5_01	ggaACAATgt (41)	1	0,992	HQ_mus/ Parche	Células germinales post-meióticas; gen 5 de caja HMG relacionado con SRY
	Nkx2-5	11937-	V\$NKK25_01	TtAAGTG (42)	0	0,986	HQ_mus	Expresado en estroma de nodo linfático; activador transcripcional TSL-1
	C-Rel	12253-	V\$CREL_01	gggctTTTCC (43)	1	0,979	HQ_mus	Prevalece en células B maduras; activador potente
25	E12	10781-; 12363-	V\$E12_Q6	gaCAGGTgggg (44); ggCAGGTgggt (45)	1	0,976	HQ_mus/ Parche	Redundancia funcional de E2A y E2-2, excepto en células B maduras
	STATx	10857-	V\$STAT_01	TTCCcataa (46)	1	0,972	HQ_mus	Expresado en células B, particularmente en centros germinales (Henderson, 1998)
	IPFI	12530-	V\$IPFI_Q4	tatCATTaaggc (47)	1	0,960	Parche	Células Beta de páncreas; factor 1 de promotor de insulina
30	NF-Y	10929-	V\$NFY_Q6	catacaCCAATga (48)	1	0,966	HQ_mus	Se enlaza con elementos de caja Y de genes II de clase MHC
	TFE3-S	11133-	RATONE S IGH_10	CATGTG (49)	0	100	Parche	Expresados de proB a células de plasma
	Mitf	11133-	RATONE S TBX2_01	CATGTG (50)	0	100	Parche	Célula de melanocito; forma heterodímeros con proteínas relacionadas con TFE3
35	Ap-1	11163+; 12025+; 12092+	RATONE SIL 2_11	AGGTAAT (51)	0	100	Parche	Interactúa con c-Ets-1 o NF-Atp
	Nkx2-5	11937-	V\$NKK25_01	TtAAGTG (42)	0	0,986	HQ_mus	Expresado en estroma de nodo linfático; activador transcripcional TSL-1
	C-Rel	12253-	V\$CREL_01	gggctTTTCC (43)	1	0,979	HQ_mus	Prevalece en células B maduras; activador potente
40	C-EBP beta	11065-	V\$CEBP B_02	aaatgtTGCAAgta (52)	1	0,917	InmunoEspec	Expresado de células B de centro germinal para células de plasma
	YY1	11564-	RATONE S CMYC_05	ACATGG (53)	0	100	Parche	Represor ubicuo
45	c-Ets-1	11792-	RATONE S TIMP1_02	CAGGAAG (54)	0	100	Parche	Desarrollo acelerado de célula B de proB para células de plasma
	TFIID	11891-; 11015+	RATONE S MBP_04	TTCAAA (55)	0	100	Parche	Interactivo con PU.1 y REL
50	POU2F 2/Oct-1	10700-	RATONE S IGH_44	ATTGTCAT (56)	0	100	Parche	Expresado de proB a célula de plasma. Secreción defectuosa si es deficiente

Resumen de las Propiedades del Vector

55 Los únicos promotores HC y LC aquí descritos han demostrado ser capaces de impulsar altos niveles de expresión de gen, como en el caso de las líneas celulares transfectadas C379B y C381B. La identificación de motivos secuenciales que pueden determinar niveles de transcripción proporciona información para aumentar la transcripción de genes customizando las secuencias del promotor y potenciador y usando células huéspedes que expresan la mejor combinación de factores de transcripción para esos promotores. Esto podría incluir el uso de una

60 célula huésped que se ha modificado para sobreexpresar los factores de transcripción que pueden aumentar la expresión y/o una célula huésped que se ha modificado para expresar de manera mínima los factores de transcripción que pueden impedir expresión, entre otros. Por ejemplo, las células huéspedes pueden co-transfectarse con un gen que codifica el factor de transcripción OBF-1 para obtener niveles más altos de OBF-1, o usan ARN antisentido, interferente (por ejemplo, siARN o shARN), o técnicas con gen noqueado para reducir la

65 expresión de TFs que pueden regular de manera negativa la transcripción de gen Ig, por ejemplo, NF- μ NR.

La divulgación también comprende métodos para identificar la interacción entre las secuencias de las regiones promotoras y potenciadoras y factores de transcripción en vectores/plásmidos y huéspedes celulares, moderar la interacción alterando las secuencias para efectuar transcripción, traslación y niveles de expresión de genes, y determinar las modificaciones y ajustarlas con el fin de controlar los niveles. La alteración secuencial puede ser por mutaciones a las secuencias en la región o por una completa sustitución de las regiones compatibles con la línea celular usada.

EJEMPLOS

Los siguientes ejemplos específicos pretenden ilustrar la invención y no deberían construirse como limitadoras del alcance de la invención.

EJEMPLO 1: Clonación de genes de anticuerpo humano anti-humano interlequina-12 y preparación de líneas celulares muy productivas

Resumen

La línea celular de hibridoma C340A que secreta el mAb humano anti-humano IL-12 produce menos de 0,4 µg/ml de cultivo celular gastado. Este nivel de producción fue suficiente para generar las grandes cantidades de material necesario para futuros estudios. Para aumentar el nivel de producción de mAb IL-12 y para tener una línea celular bien caracterizada como la fuente, los genes de cadena pesada y ligera de mAb IL-12 se clonaron e introdujeron en célula de mieloma de ratón P3X63Ag8.653 (653) o SP2/0-Ag14 (SP2/0). Los genes de cadena pesada y ligera de mAb IL-12 se clonaron de una biblioteca de bacteriófago preparada a partir de ADN genómico de hibridoma C340A y transfirieron a vectores de plásmido. Los plásmidos de expresión resultantes de cadena pesada y ligera se secuenciaron después y las secuencias se compararon con secuencias de línea celular conocida. Estos plásmidos se introdujeron en células de mieloma de ratón y los clones transfectados se filtraron para expresión Ig humano mediante ELISA para identificar los productores más altos. Los clones parentales seleccionados se subclonaron después para identificar los clones con mayor producción y homogéneos. Los subclones de interés se caracterizaron después mediante análisis de curva de crecimiento y se analizaron para estabilidad de producción de mAb durante conforme avanzaba el tiempo. Un ensayo de enlace de antígeno realizado con sobrenadantes de estas líneas celulares mostró que el mAb recombinante se unió a interlequina 12 humana inmovilizada (huIL-12) con una afinidad indistinguible del anticuerpo purificado del hibridoma original C340, confirmando que los genes que se clonaron y expresaron fueron los genes que codifican IL-12 mAb. Los subclones con mayor producción en cultivos de matraz T gastado fueron transfectante C379B 653 que produjo 135 µg/ml de IL-12 mAb y transfectante C381B SP2/0 que produjo 150 µg/ml de IL-12 mAb.

Materiales

Trisol se obtuvo de Gibco BRL, Grand Island, NY. Proteinasa K fue de Sigma Chemical, St. Lois, MO. Las enzimas de restricción se compraron en New England Biolabs, Beverly, MA. El Kit de Clonación Lambda EMBL3/Gigapack II Gold fue de Stratagene, La Jolla, CA. Las membranas Protran fueron suministradas por Schleicher & Schuell, Keen, NH. ARNasa y el Kit de Etiquetado Random Prime fueron de Boehringer Mannheim, Alemania. ADNasa fue de Pharmacia, Uppsala, Suecia. α -³²P-dCTP se compró en DuPont NEN, Boston, MA. IL-12 humano se compró en RDI Immunochemicals, Flanders, NJ. Los oligonucleótidos a la medida se compraron en Biosource International, Camarillo, CA. Los nombres, números de identificación de secuencia, y secuencias de los oligonucleótidos en este trabajo fueron los siguientes:

Nombre	SEQ ID	Secuencia
5'-10	1	CCCAGGTGCAGCTGGTG
5'-46	2	CTCAGGTGCAGCTGGTGG
5'-63	3	CCCAGGTGCAGCTACAG
5'-73	4	CCGAGGTGCAGCTGGTG
huJH4	5	AACCTCGAGTTAACGGAGG
huJH6	6	GCAGGAAACCCACAGG
5'VK1	7	ATCCAGATGACCCAGTCT
5'VK2	8	ATCGTGTTGACACAGTCTCCA
huJK3	9	AATATGCACAAAACCTTGCAC
huJK5	10	ATTTGAGCCTCTAAAGGTC

Clonación de los Genes IL-12 mAb de Cadena Pesada y Ligera

Para clonar y aislar los genes IL-12 mAb de cadena pesada y ligera de ADN de hibridoma C3430, se preparó primero una biblioteca genómica. El ADN genómico total se aisló de 5×10^7 células mediante digestión con 0,1 mg/ml proteinasa K en un tampón que contenía 100 mM NaCl, 10 mM Tris-Cl (pH 8,0), 25 mM EDTA (pH 8,0) y 0,5% SDS, y una posterior extracción de fenol/cloroformo y precipitación de etanol. Después se digirieron doscientos µg de ADN genómico C340 con 8 unidades/ml enzima de restricción Sau3A y las alícuotas se tomaron cada 5 minutos. En base al análisis de tamaños de fragmento mediante electroforesis de gel de agarosa, se acumularon las

alícuotas de 5, 10 y 15 minutos que mostraron la digestión parcial deseada. Estas alícuotas se cargaron en un gradiente de sacarosa de 10% a 40% y giraron a 22.000 RPM durante 42 horas en un ultracentrifugadora Beckman (rotor SW41Ti)O para fraccionar los fragmentos por tamaño. Se recogieron fracciones de quinientos μ l y después se analizaron sobre un gel de agarosa para visualizar los fragmentos de ADN. El ADN se aisló de las fracciones que

contenían fragmentos en el rango de tamaño 15-23 kba. Los fragmentos de ADN genómico purificado se ligaron al vector bacteriófago Lambda EMBL3 y se empaquetaron en partículas bacteriófagas usando el Kit de Clonación Lambda EMBL3/Gigapack III Gold de acuerdo con el protocolo del fabricante. La biblioteca genómica final consistió en aproximadamente 1 millón de clones.

Para filtrar la biblioteca genómica para los genes IL-12 mAb de cadena pesada y ligera, se mezclaron 640.000 bacteriófagos de la biblioteca Sau3A con cepa E.coli Y1090r y se colocaron en dieciséis placas de agar de 150 mm en una densidad de 40,000 clones por placa. El ADN bacteriófago de las 16 placas se transfirió a dos conjuntos de membranas de nitrocelulosa Protran. En resumen, las membranas de nitrocelulosa se colocaron en las placas, se dejaron 2 minutos, se retiraron y posteriormente se trataron con solución de desnaturalización, solución de neutralización, y tampón Tris como se describe en el manual de instrucciones del Kit de Clonación Lambda EMBL3/Gigapack II Gold. Se usó un entrecruzador Stragene UV para fijar ADN desnaturalizado a las membranas.

Las sondas para las cadenas pesada y ligera se prepararon usando fragmentos de ADN genómico que contenía secuencias de región constante de cadena pesada de IgG1 humano (2,8 kb fragmento EcoR1/HindIII de p747) o secuencias de región constante de cadena ligera kappa humana (fragmento 2,4 kb NcoI/Xba de p95). Ambas sondas se etiquetaron con α -³²P-dCTP usando un Kit de Etiquetado de ADN Random Primed.

Las membranas de nitrocelulosa se prehibridizaron durante una hora en un tampón de hibridización (50% formamida, 5X SSC, 3,75 mM Tris pH 7,8, 1X Solución Denhardtts, 25 μ g/ml ADN cortado y 5% sulfato de dextrano). Las sondas con un CPM de aproximadamente 1×10^8 se desnaturalizaron a 100 °C y después se añadieron al tampón de hibridización y se dejaron hibridizar a 42 °C durante la noche. Los filtros se lavaron dos veces en 2X SSC, 0,1% SDS durante 10 minutos a TA, después dos veces en 0,2X SSC, 0,1% SDS a 65 °C mientras se agitaba. Los filtros se secaron y se expusieron a película de rayos x.

Preparación de Plásmidos de Expresión IL-12 mAb

Se detectaron señales autorradiográficas correspondientes a treces clones bacteriófagos de cadena pesada y nueve de cadena ligera. Para aislar estos clones bacteriófagos, se recogió un tapón de agarosa del área correspondiente a cada señal positivo usando el extremo grande una pipeta Pasteur y se transfirió a un tubo que contenía 500 μ l de tampón de dilución de lamba y 20 μ l de cloroformo. Para purificar los clones bacteriófagos para homogeneidad, los aislados bacteriófagos se volvieron a colocar en placas en 1000-300 pfu/placa y se realizó una segunda ronda de filtrado. Del segundo filtrado, se eligieron los cuatro positivos más distintos para la cadena pesada y ligera para una tercera ronda de filtrado. Para la tercera y última ronda de filtrado los aislados fagos se titularon como antes y después se colocaron en placas a 100 pfu/placa. Las placas se recogieron para cada clon positivo y los clones bacteriófagos purificados se almacenaron en tampón de dilución lambda con cloroformo a 4 °C.

Con el fin de caracterizar los insertos de ADN clonado, se purificó ADN de los clones bacteriófagos positivos usando un método de cultivo líquido. Doscientos μ l de cultivo bacteriano Y1090r se combinaron con 100 μ l de bacteriófago en tampón de dilución lambda 1X y la mezcla se incubó a 37 °C durante 15 minutos para permitir que el bacteriófago se uniera a la bacteria. Después, se añadieron 10 ml de medio LB que contenía 10 mM MgSO₄ y los cultivos se incubaron durante la noche a 37 °C con agitación. El siguiente día se añadieron 100 μ l de cloroformo a cada tubo para lisar por completo la bacteria. Después de clarificación, se añadieron 20 μ g de ARNasa y 20 μ g de ADNasa y las muestras se incubaron a 37 °C durante 30 minutos. Para precipitar las partículas fago, después se añadieron 20 ml de una solución que contenía 20% PEG y 2,5 M NaCl y las muestras se incubaron en hielo durante 1,5 horas. Las partículas bacteriófagas se transformaron en bolas y después se lavaron una vez con tampón de dilución de lambda 1X. La bola se volvió a suspender después en una solución de 0,2% SDS y 20 mM EDTA y se incubó en 68EC durante 15 minutos. El ADN se extrajo de las partículas bacteriófagas usando fenol y cloroformo, y después el ADN se precipitó con isopropanol y se lavó con 70% etanol. El ADN purificado se volvió a suspender en 50 μ l de tampón TE.

Se usó PCR para determinar cuál, si alguna, de los clones contenía regiones variables humanas además de la región constante conocida por que estuvieran presentes en base a la homología para las sondas radioetiquetadas. Debido a que la secuencia de las regiones variables en los extremo 5' era desconocida, los cebadores se acumularon y usaron para preparar las reacciones de secuencias. Para la reacción HC, se usó una mezcla de cuatro cebadores 5'-10, 5'-46, 5'-63 y 5'-73, que representan cada una de las posibles secuencias. En el extremo 3', se usaron los cebadores huJH4 y huJH6 que son homólogos a las regiones JH4 y JH6. Para LC, en el extremo 5', se usó una mezcla de dos cebadores de consenso 5'VK1 y 5'VK2. En el extremo 3', se usaron los cebadores huJK3 o huJK5 homólogos a las regiones JK3 o JK5. Las reacciones estándares PCR 100 μ l se realizaron usando 0,5 nanogramos de ADN fago como plantilla. Los cebadores se templaron a 48 °C y se realizaron 25 ciclos de PCR. Las reacciones se fraccionaron sobre un gel de agarosa para su visualización.

Para transferir los insertos del vector bacteriófago a un plásmido que contiene el gen marcador seleccionable *gpt*, ADN bacteriófago que contenía regiones variables y constantes se digirió con *Sa*1I. Los fragmentos resultantes se purificaron y clonaron entre los sitios *Xho*I y *Sall* de plásmido p1351. Debido a que había un sitio interno *Sa*1I en la secuencia codificadora de la región variable de cadena pesada, dos fragmentos *Sa*1I que tenían un tamaño de 13 kb y 4 kb tuvieron que transferirse al vector de expresión p1351 secuencialmente. Primero el fragmento *Sall* de 13 kb se clonó en los sitios *Xho*I y *Sall* de p1351 y el plásmido resultante designó p1557. Posteriormente, el fragmento de 4 kb se transfirió a p1557 y el plásmido de expresión final resultante se llamó p1560. El inserto *Sall* de cadena ligera de 15 kb fue capaz de transferirse en una etapa y el plásmido de expresión final se designó p1558.

Transfecciones celulares, Filtrado para Productores Altos y Expresión de los Genes Clonados

El plásmido p1560 de cadena pesada se linealizó mediante digestión con enzima de restricción *Pvu*II y el plásmido p1558 de cadena ligera se linealizó usando enzima de restricción *Sall*. Las células 653 y SP2/o se transfectaron por separado con los plásmidos linealizados pre-mezclados mediante electroporación y las células cultivadas y transfectantes se seleccionaron usando ácido micofenólico como el descrito. Los sobrenadantes celulares de las colonias resistentes al ácido micofenólico se sometieron a ensayo aproximadamente dos semanas más tarde para IgG humano. Para este experimento, los sobrenadantes celulares se incubaron en placas ELISA con 96 pozos que habían sido previamente cubiertas con anticuerpos específicos Fc de IgG de cabra anti-humano. IgG humano unido se detectó usando anticuerpo IgG (H+L) anti-humano de cabra conjugado con alcalina fosfatasa y sustratos de alcalina fosfatasa como se ha descrito. Las células de los clones más productores se transfirieron a platos de cultivo con 24 pozos en medio estándar, IMDM, 5% FBS, 2 mM glutamina, mezcla de selección de ácido micofenólico (5g/L Xantana, 250 mg/L Hipoxantana, 50 mg/L ácido micofenólico, 50 mM NaOH): más tarde los sobrenadantes recogidos de los cultivos gastados se cuantificaron cuidadosamente con ELISA haciendo diluciones dobles en serie de cada muestra y comparando los valores O.D con la curva estándar preparada usando mAb purificado de células C340. Los clones seleccionados se transfirieron después a matraces T75 y la producción de IgG humano se cuantificó de manera similar mediante ELISA. En base a estos valores, los transfectantes con mayor producción se subclonaron sembrando una media de una célula por pozo y realizando ensayos ELISA de sobrenadantes celulares a partir de colonias individuales de subclones.

Los subclones seleccionados para los ensayos finales fueron el transfectante C379B de 653 y los transfectantes C381B de SP2/O. Los niveles máximos de producción para C318B fueron 135 µg/ml y 150 µg/ml, respectivamente. Las densidades celulares pico para C379B y C381B fueron 1×10^6 células/ml y $1,45 \times 10^6$ células/ml, respectivamente.

Ensayo para Enlace de IL-12

Los sobrenadantes de las tres líneas parentales (clon 2 y clon 18 de transfectantes 653 y clon 1 de transfectante SP2/O) se usaron para analizar las características de enlace de anti-huIL-12 mAb recombinante. Las concentraciones de mAg en las tres muestras de sobrenadante se determinaron primero mediante ELISA. Las cantidades de titulación de las muestras sobrenadantes, o mAb purificado de células C340 como un control positivo, se incubaron después en placas de 96 pozos que previamente se habían cubierto con 2 µg/ml de huIL-12. El mAb unido se detectó después con anticuerpo IgG (H+L) anti-humano de cabra conjugado con alcalina fosfatasa y los apropiados sustratos de alcalina fosfatasa.

Las características de enlace de antígeno de sobrenadantes de tres líneas celulares parentales separadas se compararon con mAb purificado de células C340. Se ha demostrado que mAb de sobrenadantes de célula transfectada se unió específicamente a huIL-12 inmovilizado de una manera indistinguible de mAb purificado de la línea celular original C340. Este resultado confirmó que se clonaron y expresaron los genes correctos.

Secuenciación

Las secuencias de los insertos de plásmido de expresión de cadena ligera y pesada se determinaron usando química terminadora BigDye y un secuenciador de ADN automatizado ABI377. Los cebadores se diseñaron para secuenciar corriente arriba o corriente debajo de secuencias conocidas en la región variable, región constante y eje central del vector. Cada reacción de secuenciación dio aproximadamente 500 bp de secuencia adicional que se alineó usando el programa AssemblyLIGN. Después de cada ronda de secuenciación, los cebadores se diseñaron para continuar las secuencias corriente arriba y corriente debajo de la secuencia conocida. Las secuencias de p1558 y p1560 se muestran en la SEQ ID NO: 11 y 12.

El inserto clonado que contiene las secuencias codificadoras de cadena ligera de IL-12 mAb tiene aproximadamente 15 kb de longitud. El codón de inicio está situado 4310 pares base del extremo 5' del inserto. Hay un intrón 3838 bp J-C entre las secuencias que codifican la región variable y constante y aproximadamente 6 kb de secuencia no codificadora corriente debajo de la región constante. La región constante de la cadena ligera es el alotipo Km(3), el mismo que cA2 y 7E3.

El inserto clonado que contiene las secuencias codificadoras de cadena pesada de IL-12 mAb tiene aproximadamente 16 kb de longitud. El codón de inicio está situado 4802 bp del extremo 5' del inserto. Hay un intrón 5406 y aproximadamente 4,8 kb de secuencia no codificadora corriente debajo de la región constante. El alotipo IgG es G1m(f), definido por la arginina (arg) en la posición 214. Éste es diferente que cA2 que tiene una lisina (lis) en la posición 214 y por lo tanto del alotipo G1m(z).

Aunque se ha ilustrado y descrito anteriormente con referencia a ciertas realizaciones específicas, la presente invención no se limita sin embargo a los detalles mostrados. Más bien, la presente invención está dirigida a un vector de expresión que tiene varias secuencias, células huéspedes y kits desvelados en el presente documento, cuyo alcance se define en las reivindicaciones adjuntas.

Lista de Referencia

- (1) V Matys, e Fricke, R Geffers, E Gößling, M Haubrock, R Hehl, K Hornischer, D Karas, A Kel, O Kel-Margoulis, D-U Klosos, S Land, B Lewicki-Potapov, H Michael, R Münch, I Reuter, S Rotert, H Saxel, M Scheer, S Thiel y E Wingender. TRANSFAC®: regulación transcripcional, de patrones y perfiles Nucleic Acids Res. 31, 374-378 (2003).
- (2) E Goessling, O Kel-Margoulis, A Kel, E Wingender. MATCH™- una herramienta para buscar sitios de enlace del factor de transcripción de secuencias de ADN: Aplicación para el análisis de cromosomas humanos. Conferencia Alemana sobre Bioinformática 2001. <http://www.bioinfo.de/isb/gcb01/poster/index.html>
- (3) A Henderson y K Calame, Regulación transcripcional durante el desarrollo de célula B. Annu. Rev. Immunol. 16:163-200 (1998).
- (4) A Henderson y K Calame, Lecciones en regulación transcripcional aprendidas de estudios sobre genes de inmunoglobulina. Critical Rev en Eucaryotic Gene Expr 5:255-280 (1995).

LISTADO DE SECUENCIA

<110> CENTOCOR, INC.

<120> ANTI-IL-12 ANTIBODY BASED VECTORS, HOST CELLS, AND METHODS OF PRODUCTION AND USES

<130> CEN5089 PCT

<140> TO BE ASSIGNED

<141> 2005-12-21

<150> 60/637,936

<151> 2004-12-21

<160> 56

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 1

cccaggtgca gctggtg 17

<210> 2

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 2
 ctcaggtgca gctggtgg 18

 5 <210> 3
 <211> 17
 <212> DNA
 <213> Artificial Sequence

 10 <220>
 <223> Primer

 <400> 3
 cccaggtgca gctacag 17

 15 <210> 4
 <211> 17
 <212> DNA
 <213> Artificial Sequence

 20 <220>
 <223> Primer

 <400> 4
 ccgaggtgca gctggtg 17
 25 <210> 5
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 30 <220>
 <223> Primer

 <400> 5
 aacctcgagt taacggagg 19
 35 <210> 6
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 40 <220>
 <223> Primer

 <400> 6
 gcaggaaacc ccacagg 17
 45 <210> 7
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 50 <220>
 <223> Primer

 <400> 7
 atccagatga cccagtct 18
 55 <210> 8
 <211> 21
 <212> DNA
 <213> Artificial Sequence

 <220>
 60 <223> Primer

 <400> 7
 atccagatga cccagtct 18
 65 <210> 8
 <211> 21
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Primer

<400> 8
 atcgtgttga cacagtctcc a 21
 5 <210> 9
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 10 <220>
 <223> Primer
 <400> 9
 aatatgcaca aaactgtcac 20
 15 <210> 10
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 20 <220>
 <223> Primer
 <400> 10
 atttgagcct ctaaaggctc 19
 25 <210> 11
 <211> 19490
 <212> DNA
 <213> Artificial Sequence
 30 <220>
 <221> misc_feature
 <222> (809)
 <223> Expression plasmid where n can equal a, c, g, or t
 35 <220>
 <221> misc_feature
 <222> (995)
 <223> Expression plasmid where n can equal a, c, g, or t
 40 <220>
 <221> misc_feature
 <222> (1046)
 <223> Expression plasmid where n can equal a, c, g, or t,
 45 <220>
 <221> misc_feature
 <222> (8220)
 <223> Expression plasmid where n can equal a, c, g, or t
 50 <220>
 <221> misc_feature
 <222> (8283)
 <223> Expression plasmid where n can equal a, c, g, or t
 55 <220>
 <221> misc_feature
 <222> (13101)
 <223> Expression plasmid where n can equal a, c, g, or t
 60 <400> 11
 gctcgacctg caggtcaacg gatccacttc acctgtaggc aaggcacagc acaggggtga 60
 gcgaggccac agccctgccc ccgagcccca cccacccctc agggcactga ggccacctct 120
 65 ctgcccccaa ggcccaccca cccgtcagtc cacgcaggcc acagccctgc ccctgaggcc 180

	catccggccc	ctcatgccac	ccaggtgccca	tggcctcacc	actgcctgct	ctgaggett	240
5	gtgatagaga	gcagagccag	ggcaccaaca	gcatgtggac	agcacagaag	acagcgtcag	300
	ggacaggtgg	ggacagcgtg	ggggacagt	tcagacacag	gtgaggacag	tgtgggggac	360
	agtgtcaggg	acaggtggag	acagagtgtg	gaagagtgtt	ggggacagtt	gaggacagca	420
	tggaggagag	tgttggggag	agatggggac	agtgtcaggg	agaggtgggg	actgtgtgga	480
	ggacagcatc	agggacaggt	ggggacagca	tgggggacag	tgggtgcatac	aggaggggac	540
10	ggtgtggggg	acagtgtcag	ggattatagg	ggacagagt	ggagacagt	tcagggaccg	600
	gtggagacca	tgtgggggac	aggtggggac	agcatggggg	acaggtgggg	acagcatggg	660
	ggacagtgtc	agggatagga	ggggacagga	aacagtgggg	acattgtcag	ggacagggga	720
	gatagcatgg	gggacagtgt	tggggacatg	ggggacagca	tggaggacag	tgttggggac	780
	tgggtggggac	agcatggggg	acagtgtang	agacaggttg	ggacaggatg	gaggatagt	840
15	ttggggacag	gtggggacag	tgtgggggac	agtgtcaggg	acaggagggg	agatcgtgga	900
	gaacagtgtc	cgggacaggt	ggggacagca	tgggggacag	tatcagggac	aggtggggag	960
	agtgtggggg	acagtgttgg	actggtaggt	acagnccggg	ggacagcatc	ggggacaggt	1020
	ggggactgca	tgggggacaa	tatcanggac	aggtggggac	atggaggaga	gtgttggggg	1080
20	caggtgggga	cagcatgggg	gactgtgttg	aggacaggtg	gggacagctt	gggggacagt	1140
	ggtgcatacg	ggaggggacg	gcgtkgggga	cagtgtcagg	gattataggg	gacagagtgg	1200
	gagacagtgt	cagggacagg	tggggacagc	atggcggaca	gtgtcagggg	taggagggaa	1260
	caggagga	cagtggggac	attgtcgggg	acagggggga	tagcgtgggg	gacagtgttg	1320
	gggacaggtg	gggacagtgt	ggaggaaagt	tttggggact	ggtggggaca	gcatggggga	1380
25	cagtgtaggg	gacaggtggg	gacaggaggg	gacagcatgg	aggatagtgt	tggggacagg	1440
	tggggacagt	gtgggggaca	gtgtcgggga	caggagggga	cagcgtgggg	gacagtgtca	1500
	gggacaggtg	gggacagcat	gggggacagt	gttgtactgg	taggtacagc	ctgggggaca	1560
	gcattgggga	caggtggaga	ctgcatgggg	gacaatatca	gggacaggtg	gggacagcat	1620
	ggaggagagt	gttgggggaca	ggtgggggaca	gcatggggga	cagtgttgag	gacaggtggg	1680
30	gacaaacgtg	gggtacagt	tccggagatg	gtggggacag	catggaggaa	agtgtcgggt	1740
	ttaggtaaga	acaacgtgga	ggagagtgtc	ggggacaggt	ggggacagt	tgggggat	1800
	tgtcacggac	aggtggggac	agcatgtggg	acaggggttc	atacaggagg	ggacagcatg	1860
	ggggacagt	ccaggtactg	taggggacag	cgtgggggac	agggccagga	actatagggg	1920
35	acagagtggg	ggatgggtgtc	agggacaggt	ggggaaagca	tgggggacag	tgtcagggac	1980
	aggtggaaac	tgtgggggac	agtgttgggg	acaaaggggg	acaggtggg	ggacagtgtt	2040
	ggggacagat	ggggacagca	tgggggacag	tgtcggggac	atgtggggag	agcctggggg	2100
	acactgttgg	acaggtgggg	gcagcattgg	ggacaatgtc	agggacagtt	tgtgagagca	2160
	tgggggacag	cgtcagggac	aggtggggac	agcctgtggg	acagtgtcag	agacagtttg	2220
40	tgacagcatg	ggggacaatg	tcaaggacag	ctgggggaca	cgtgcggccg	accttgaaga	2280
	aggtgacggg	ggcactgtag	cacacgctta	acaggaagag	tgtgatgaag	atggtgatgg	2340
	tccgtccacag	cccgtccagc	tccccgtcct	gcgcctccgc	acagctctcc	tccagtgtga	2400
	gctctggaca	ggaaaggggt	ggtcagtgct	gtgtccccct	gggcttgggc	ctctgggggt	2460
	gattccctct	gtggcggggc	ccaggatgta	gggcccggcc	gggatgggcc	aacaatgtcc	2520
45	tgaggtcagc	tccccacagc	tggccgccc	gggaccagc	tttggccccg	ggactcagcc	2580
	agacacccgg	ccctagatag	cgcctggcc	ctcagcagga	cccgtcccc	gtctcccgtg	2640
	tccctccctg	agggccagag	ggcaggagga	tggtagagcc	cacacctcat	gtgacccag	2700
	ctgcagggaa	gggcggcatt	gggaagtggg	ccagtgccag	ggacgcgacg	tggcgtgtgt	2760
	tccctgtgtg	gtgggggccc	gtgtgtgtgt	ggcggtgca	ggggcacctt	gtgagaggag	2820
50	ggctgggttt	gtctgagctg	gtcagcatgt	ggagaagctg	ccgagcggct	cgtgggcctt	2880
	gaggtgccgc	gtggggctcg	tgggggccc	tgtccgagga	gtgttcacgt	gtgcgaggac	2940
	cttctgtctg	tctgggtgct	gtgcggttcg	ccgggtgag	gctccgtgtg	tgaggcgtgc	3000
	acgtgtgtgt	gtgggtggccg	tgtggccggc	caacctcagt	gcgggggttg	tgaacgggt	3060
	ctgggctgag	tgtgtgtgtg	ggcatctgca	ccagtctctc	cacaggggcc	gagagtgc	3120
55	gtccccagga	gtcgggtgtg	tccccatgtg	gggtgcgaggc	tgggcagggc	tggcaggggg	3180
	tagtgccgtg	ggggtagatg	gggtgaggag	ggcctgtccc	tacgcgcag	gactaggcat	3240
	gcccccgagt	gggcatcggg	gtcggaggac	agggcgctca	caggacagga	cagtctccta	3300
	cagaggcagg	ggctgtgtgt	ctgtccccag	gggctcctag	ggcttctcgt	ggcccagccc	3360
	agggcagctg	ctgctggagg	gagggccacg	ctggcaaatc	ccccaccctg	ccgagggcag	3420
60	cccctggctg	agccccaccc	tagggcgccc	agggcacacct	gcacagcctg	ggccagtgtg	3480
	gggacagtgg	gacccgctct	gcctccctca	tgcactcag	gcctcagact	cggcctgacc	3540
	cgtggaaaga	accatcacag	tctcgcaggg	gcccagggca	gcgctgggtg	ctttatttcc	3600

5	atgctggg	cccgggaagt	atgtacacgg	ggtacgtgcc	aagcatcctc	gcgcgacccc	3660
	gagagcccg	ggagcagggg	cttgccggcc	ctggcactca	tttacccgga	gacagggaga	3720
	ggctcttctg	cgtgtagtgg	ttgtgcagag	cctcatgcat	cacggagcat	gagaagacgt	3780
	ccccctgctg	ccacctgctc	ttgtccacgg	tgagcttgct	gtagaggaag	aaggagccgt	3840
	cggagtccag	cacgggaggg	gtggctctgt	agttgttctc	cggctgcccc	ttgctctccc	3900
10	actccacggc	gatgtcgctg	ggatagaagc	ctttgaccag	gcagggtcagg	ctgacctggg	3960
	tcttggtcag	ctcatcccgg	gatgggggca	gggtgtacac	ctgtgggtct	cggggctgcc	4020
	ctgtagggac	agagggttgg	acagcgggtc	ctctcagggc	agaggggtgg	ccgagccggc	4080
	ctctgtccat	gtggccctcg	caccccacgg	gtcccacctt	tggttttggg	gatgggtttc	4140
	tcgatggggg	ctgggagggc	tttgttgagg	accttgccact	tgtactcctt	gccattcagc	4200
15	cagtctgggt	gcaggacggg	gaggacgctg	accacacggg	acgtgctggt	gtactgctcc	4260
	ccccgcggct	ttgtcttggc	attatgcacc	tccacgccgt	ccacgtacca	gttgaacttg	4320
	acctcagggg	cttcgtgggt	cacgtccacc	accacgcctg	tgacctcagg	ggtccgggag	4380
	atcatgaggg	tgtccttggg	ttttgggggg	aagaggaaga	ctgacgggtc	ccccagggag	4440
	tcagggtcct	aggaagagat	ggagggtggc	gtgtcagcac	cgggtggggg	cctgtccctg	4500
20	gatgcaggct	actctagggc	acctgtcccg	ccttgagctg	gagggcgagg	cctgggctgg	4560
	cttacctggg	cacggtgggc	atgtgtgagt	tttgtcaca	gatttgggct	ctgcagagag	4620
	aagattggga	gttactggaa	tctgggagga	gagaaggtgt	ccgagctgag	ggagtggaga	4680
	gtttggcctt	tggggtgggc	ttaggtcagg	ggcaggggtc	tcccggatat	ggctcttggc	4740
	aggtctgagc	ccagcacctg	cccctttgtg	tgcagggcct	gggttagggg	cacctagcct	4800
25	gtgcctgccc	agagcctggg	gaaaaagcca	gaagaccctc	tccctgagca	tgagtggggc	4860
	gggcagaggg	ctccgggtga	agaggcagac	ggggcctgcc	ttgctgcctt	ggactggggc	4920
	tgcataggcg	ggatgcgtcc	aggcaggagc	gctgagcctg	gcttccagca	gacaccctcc	4980
	ctccctgtgc	tggcctctca	ccaactctct	tgtccacctt	gggtgtgctg	ggcttgtgat	5040
	tcacgttgca	gatgtaggte	tgggtgcccc	agctgctgga	gggcacgggc	accacgctgc	5100
30	tgagggagta	gagtcctgag	gactgtagga	cagccgggaa	gggtgtgcac	ccgctggtca	5160
	gggcgcctga	gttccacgac	accgtcaccg	gttcggggaa	gtagtccctg	accaggcagc	5220
	ccagggccgc	tgtgccccca	gaggtgctct	tggaggaggg	tgccaggggg	aagaccgatg	5280
	ggcccttggg	ggaggctgca	agagaggtgg	tgccatgtga	ccgcgggtgt	ggacagagct	5340
35	gggcccaggg	cgcagagggc	cctgggttct	taactgtccg	cgagggttcag	cgtccagtgt	5400
	ctgggctcat	gggcattggg	tgtgcacctg	gctggcgcca	cctgcctcac	cttagccccc	5460
	tccctgcccc	aaagccaagg	tcaggcctgg	cctgccccag	aaagcttcag	ctgctcgacc	5520
	tgcaggcatg	caagcttgca	ggaccgggtg	ccctgtgggt	cccttctgca	ggcaccctct	5580
	cagctccagga	ggcggggctc	ggcagccagc	tcagcgctct	gtgctgtccg	ggagtcagca	5640
40	cagtcacagt	tctctagctt	ggcctcagct	ctggccatcg	gtgccacctc	agggacggct	5700
	catgcccatt	ggccccactc	cagcctttta	tgggtgctct	gcttgaccag	tggacactgt	5760
	tctcagatgg	cttctgggtg	gtcccccgag	ccccctgaag	ccccctgacc	tgcgctcca	5820
	gcatggccct	cccctagtga	gtgggcctga	cttgcccagg	gccctggtca	tagcctgccc	5880
	tctgccctcc	aaggcccttt	tcttctgtgc	agcagagggg	ccagacactg	catagggtcg	5940
45	gcgccttca	gccccagggg	cccggaaacc	cctgccttgg	aatatcgccc	cgggagcctc	6000
	ctcctcagcc	tctccccctc	tttccccctt	gccccagtgt	gcagcagccc	aggtcagggc	6060
	cctgagtgcc	tggatgcccc	ctgcctccca	gtgtcctgca	ttacttctgg	aggctcagtc	6120
	accacaacct	caccctccca	ggcctggcct	ggccttctcg	gccaccagcc	cacctcctcc	6180
	ctctctccag	agcttcccc	ggcaagggtc	ctgctggggt	caaccacagg	ccccagcac	6240
50	aggtaggagc	cttgccacctg	cccctggccc	tccccacctt	gcatgggtgc	aggaccccca	6300
	ggccacaggg	aggccccatt	tctctctgoc	gctggcccag	tggccctgga	gtcccactgc	6360
	aggtgggggt	tgcctctgac	ctctgaggag	gctaagtgcc	ctgcccctcag	ccaggccatc	6420
	ccctctgctc	agccccaggg	ccccgctcac	caccccttcc	cctcacctgc	accacaggct	6480
	ctggctgact	ctgcccaggc	cctgaatggg	ccccctctgg	tgccctctgc	tgtacactg	6540
55	tcctgcacca	cctccactca	gcttcattgt	gtgggtggcc	ctgggtcctg	gcagcccatc	6600
	ttgtccttcc	tggggcacca	gcctcagagg	ccttccctgc	caggggtccg	tgggaccagc	6660
	cgtgggaccc	tccctgggtc	aagcacacgt	tccccctgca	gccacacctg	ccccctgctg	6720
	agagcccgag	cccgagccct	ggaacgcctt	cccttctcca	tcccagctcg	cccttgccaa	6780
	ctgctcagtg	ggatggactc	acactccctt	ccgggcacca	ggaggctgca	ctgaccttcc	6840
60	accagccctc	agctgtctgt	tgccagcaac	taccagctc	ctgccaaaagt	ctagagagctg	6900
	agtgatgcct	cccaccagcc	gtgctcacct	gtgggtgctt	tgcctgagc	tctagtgcct	6960
	gtccccctgg	gcaacttagc	ccagctcagc	tcagctcagc	tcaaccaggt	tcaactcagc	7020

65

	ccagttcagc	tcagctcagc	ccagttcagc	cttggttagt	ctaggtcagt	ttaggtcagt	7080
5	tttgcccatc	tgagtcatt	tctgaaagct	ggatggagtt	gtcatggcca	gaaatggcca	7140
	gcccaccaga	cctgtttgtc	tcagctaaag	ccatctcatt	gccaggttcc	tgcacagcca	7200
	ggctggcttc	catcttttgt	ctccctctac	ttgatacccc	agttccctgc	agtcctgccc	7260
	cagcgcacc	tggtttttgg	ttccaaagca	ttaccaatca	ttaccacctc	ccactacctg	7320
	ggtggaatat	ttcttttgtg	cttttaaagtc	attaaaacat	cttgagaatg	agaccaagaa	7380
10	tttaggagcc	tgtgctgtga	taaaaatgag	caggtccctc	tgctctagaa	gtggcagcat	7440
	atcttctgca	ccaagaggag	ggtattgaga	tgctcagagc	ctccaccttc	ccggagcatc	7500
	ccctcccttc	tgagtctgca	gtaaacccct	gcctttaaat	tccctctaga	taacagtcac	7560
	cattggaaac	aaccaagaaa	tgcattttat	ctgaatttgc	cacttaaaat	tctgccattt	7620
	accataaatc	gctttggaag	gcatgggcta	ctttcaaggg	tgcgatgatg	acctacagtc	7680
15	aatgacttag	acaagggcga	tgccagtggt	gcttggtatg	ttctcaagca	tattaccca	7740
	tgccatcccc	attcagaggt	tgtggaacag	ctcgtgcgac	ctctccttca	aattggacttt	7800
	agggaaaagt	aaatgggagt	gacccagaca	atggtcactc	aaaagactca	cataaatgag	7860
	tctcctgctc	ttcatcaagc	aattaagacc	agttcccttc	ctagtggaaa	taagacgtta	7920
20	aatacaaatg	tttaagagaa	gcaaattgcag	cagcggcgcc	tgccctgtct	ttaccatgtc	7980
	gggcgcctgg	tactgcgag	ccttgcaaaag	ctttggcatg	gaatcattcc	tccaagtcca	8040
	ttaacaaggg	ctggggcctg	agcagccagt	cggcccgcca	gcagaagcca	cgcacccag	8100
	ctctgggtag	tccggggaga	cccaaagccc	aggccgggcc	tggcagccac	cctcccagag	8160
	cctccgctag	gccagtcctg	ctgacgcccgc	atcggtgatt	cggaaacagaa	tctgtccttn	8220
25	taaggtgtct	ccacagtcct	gtcttcagca	ctatctgatt	gagttttctc	ttatgccacc	8280
	aantaacatg	cttaactgaa	ataattcagg	ataatgatgc	acattttacc	taaaacttat	8340
	cctaaagtga	gtagttgaaa	agggctctga	aaaatactaa	aatgaaggcc	actctatcag	8400
	aatatcaaag	tgtttctcct	taatcacaaa	gagaaaacga	gttaacctaa	aaagattgtg	8460
	aacacagtc	ttatgaaaat	aatgctctga	ggtatcgaaa	aagtatttga	gattagttat	8520
30	cacatgaagg	gataacaagc	taatttataa	aactttttga	atacagtcac	aaactctccc	8580
	taagactgtt	taattttctt	aacatcttac	tttaaaaatg	aatgcagttt	agaagttgat	8640
	atgctgtttg	cacaaactag	cagttgataa	gctaagattg	gaaatgaaat	tcagatagtt	8700
	aaaaaaagcc	ttttcagttt	cggctcagct	cgccttattt	tagaaacgca	aattgtccag	8760
	gtgttggttt	gctcagtaga	gcactttcag	atctgggccc	gggcaaaaacc	acctcttcac	8820
35	aaccagaagt	gataaaattt	ccaattgtgt	ttttttgctt	cctaaaatag	actctcgccg	8880
	tgacctgctt	cctgccacct	gctgtgggtg	ccggagaccc	ccatgcagcc	atcttgactc	8940
	taatttcata	cctgtctcca	gcttcgctca	attaattaaa	aaaataaact	tgatttatga	9000
	tggtcaaaac	gcagtcctgc	atcggggccg	acagactgtt	gctagtattt	cttagctgag	9060
40	cttgcttttg	cctcaatttc	agacacatat	cactcatggg	tgtaaatcaa	atgataagaa	9120
	tttcaaatac	ttggacagtt	aaaaaaatta	atatacttga	aaatctctca	cattttttaag	9180
	tcataatttt	cttaaccatt	tttctcagaa	gccacttcaa	acatatcctg	tcttttaaca	9240
	gtaagcatgc	ctcctaagat	aaacaatcct	tttctcatgg	aaaccagctt	caaggcactg	9300
	aggtcctgga	gcctccctaa	gccccgtgca	ggacggcagc	cactgtttct	gggctacccc	9360
45	tgcccccaac	cctgctctca	tcaagaccgg	ggctacgcgt	ccctcctggc	tggtattcac	9420
	cactccgaca	gttctctttc	cagccaaaaa	agaatctaag	atgcaggttg	acacacagcg	9480
	cacctcataa	ttctaaagaa	aatattttcac	gattcgttgc	tgtgcagcga	tcttgagtc	9540
	ctacagacac	cgctcctgag	acacattcct	cagccatcac	aaagacccct	ggtttgttca	9600
	ggcattttgc	ccaaatgtgg	cgccccaaag	ccccaggctc	agtgactcca	tcagacgcac	9660
50	ccaacctgag	tccatttttc	caaaggcatc	ggaaaatcca	cagaggctcc	cagatcctca	9720
	aggcacccca	gtgcccgctc	cctcctggcc	agtcgcgcca	ggtccctctg	gaacatgccc	9780
	cgaggaccaa	cctgcaatga	tcaggaaacc	ccacaggcag	tagcagaaaa	caaaggccct	9840
	agaatggcaa	ctgactgtcc	gtggccctgt	cctgccttcc	tcattggaatc	ctctgttggc	9900
	ctcccacgta	ccccacattt	tggcctgacc	cctcagaagc	cagaccactg	tcggcctggg	9960
55	aagcccaact	gcaagcagac	ggcttctaag	tactcccag	gagtcacaaa	acccgggggg	10020
	gcacccgtcc	cagagagcgg	gtgccttgga	gcgggacaga	gtcccaccac	gcaatcatca	10080
	cgacagcccc	tgagaatgct	ccaggtgaag	cggagagagg	tctccccaga	ccagccgaag	10140
	gagcccccca	ctgcccagca	tctgtggccg	ggcttgggga	ggacaggctg	ggttcccat	10200
	cgacgggtcc	ctctccccgg	ctttctttcc	tgacctccaa	aatgactcca	agactctgac	10260
60	cctgagaccc	tggaagttg	agtctcccta	agtggacgca	gagagggggg	ggtgaggact	10320
	cacctgagga	gacggtgacc	agggttccct	ggccctaggg	gtcgaaccag	atgtcacatc	10380
	gtgacaacaa	agccaggacc	ccaggcaaga	attggcgccc	cgtatgtccc	tgggacccac	10440

65

5	t	cagattgaa	cccggggagg	gccgggggtt	ggtggggaact	ggacccagga	ggcctagggt	10500
	g	gccctggcc	acagagagag	gcgtcctatt	gggctcagga	ggaaggagaa	gggtggagccc	10560
	g	ctattgtct	cagggcaccc	tcctgagtc	cccagggtgc	tccggggctc	ccttggcagg	10620
	a	gacgcagca	ccctcatgtc	ccccaaaaa	tgcaacaaaa	tccttcagac	ttaaagtagg	10680
	a	gagagggtg	tgaggactca	catgaggaga	cggtgaccag	ggttccctgg	cccagaaagt	10740
	c	aaagtaacc	ctgtccagga	cgtctctctg	cacagtaata	catggcgggtg	tccgaggcct	10800
10	t	caggctgtt	ccactgcaag	taggcggtag	tgatggactt	gtcgactgac	atggtgacct	10860
	g	gccttggaa	ggacggacta	tatctgatat	cagagtcaac	aggagacatg	attcctatcc	10920
	a	gtccaggcc	tttcccgggc	atctggcgca	cccagccgag	ccagtagggtg	gtaaagctgt	10980
	a	gccagaacc	cttacaggag	atcttcagag	actccccggg	ctttttcacc	tctgctccag	11040
	g	ctgcaccag	ctgcacctcg	gcacagactc	ctgtgtgggga	gacacaaaat	ttgaatcagg	11100
15	g	ggtcctttc	cacccgttct	cctctgtgac	ctcaagacct	ggcgaggact	gacctggag	11160
	a	acagccagg	aggaggccga	ggatggcggt	tgaccccatg	atgggtggagg	acagaaaatg	11220
	a	agccctgag	atcccagctg	ggcagtgagg	gagactcaact	gtggaggggga	gccctgggtt	11280
	t	aagtgggga	ggcccccact	tgcatttgca	tagttgtcac	cctgccctga	agggaggagt	11340
20	c	tacagcgtt	tataaccag	aacctcaaat	gcagaaaaag	gctggcctga	gcctcctggg	11400
	a	ggggcagag	taaggtctca	ataattcctt	acaccctgct	attgtccctc	tccactcttt	11460
	t	acatgtttc	ttcgagaccc	ataccagggc	ctcccttctt	ccacccttct	ctgtgacctt	11520
	g	tgaaggtga	taaatctagt	tgaaactatg	tgtgttaaaa	aaaatgagaa	atagtgccag	11580
	g	aaaggccat	gaagagaaaa	ttcaaataga	cttatgcctg	ataacaagaa	ctactaaaaa	11640
25	a	actactgtc	tattccctgg	caatttcgtg	tgagttggag	tacggcctgg	gcatacagca	11700
	a	gggcaggag	cacctcagga	cctcaacagt	cttcaagatg	attaacttgc	cagaccttca	11760
	c	cccatgcaa	atcgcacatt	tttctggcc	attttgtctt	ctagattttt	acactctgcc	11820
	a	attcaacat	gaataggga	tatttgttta	ggtctctgac	ttgctgatgg	acctgaaggg	11880
	a	tgcccatgt	ctgcacccaa	ctcctggtag	tgctgtttta	agtcccttgt	gtcaacccca	11940
30	g	caccttctt	gtttagttct	ttcatttttt	aacatttatt	ttatgatatc	cacattgctt	12000
	g	gaggaggcc	cttaactatc	ccctttgtct	gccccatttt	tttgtgaagc	actctcattt	12060
	c	attagcacc	taaaagatcc	agagaaaaag	ccaaagcaac	acaaactaca	caagtctcta	12120
	a	actatacac	gatctgatat	ccagtgcacc	agggaaagctg	tctcatcaag	aacatcctac	12180
35	t	tgaggttgg	gatcagctac	gtctgtcata	tgtctgttgt	gagctgtggg	tccagcatcc	12240
	t	cccaaaccat	ctctctctca	tctgcagcac	acgaagccag	agaagctgct	acgtcagtca	12300
	c	ctctcatagt	ccattggcta	actttgttca	caatccgatg	atataaatta	atgatatcaa	12360
	t	catatcaca	aaatgatgcc	actttccatg	ctgtcctctc	tgttatagtt	gtctagtgga	12420
	t	ctctctccc	cagcccaaac	tctgggctac	cttctgtgtc	actcttatgt	gttgtgtgtt	12480
40	g	ttttctggac	aactgtaccc	tttgggtggtg	aatgggagtc	tagcagcctc	ggctcacagg	12540
	g	cacactgca	gggctggatg	tgcatctttt	tactattatt	ttagctatca	cagatatgaa	12600
	t	tgcatattc	gggttgttcc	tcattatttt	tacttttctt	ttttcaggga	acaactccta	12660
	g	gaagctgta	tctgttgttt	ctaaattcca	tggattactt	tcacttctctg	ctccttattg	12720
	g	agtgtacat	gccatctcat	gccatgatgt	tctgtagcca	tctgggcatt	gaatgtcttt	12780
45	t	atcccgcac	tttctcatcc	tttgatgtta	taaaaaatcc	tctagccatg	cgtcagcaat	12840
	c	aggaacttt	ctagcattct	ccaactctga	tgtaattgtg	ttgagtgtcc	caagaccatc	12900
	c	tccagggccc	ttaatcagta	caaataaagg	acacaagaaa	agctgttatt	ctcatgggtg	12960
	c	agcttttatt	atagcgaattg	aatatgaatt	aaaatgagca	aaggcaccag	ggagaaggcc	13020
	c	tggagaatcc	aggcatgagc	tcccagggtg	tctttccctg	aggagtctct	tgccccagt	13080
50	t	ctccagca	gtgatgcattg	ncaacatgtg	tgaagcattg	tccaccaggg	aagctcacct	13140
	g	agtgtctgt	gcccagggtc	gtttattggg	cccatcacag	atgtgtgtga	cacctgcaca	13200
	a	ttgacctcc	agtgtctcaga	cgccggcctc	ttgagcacta	ataggcattc	accataagtc	13260
	a	ttatgaaaa	cagctagtat	agcgtgcacc	caggctacac	acacagagac	acagacaaac	13320
	a	caaacaaaa	atacatttta	actaataaaa	taatcacaat	actaagagag	ataagaggaa	13380
55	t	gttttggaa	gtgatttcta	tggttatgac	cttgatgatg			

5	agccaaagca	gatgtttcac	ggagagtgtg	ggcaggggca	tgcctgcagg	tcgagtgcctt	13920
	tctacataag	ttatgggacc	atagcagagt	ttcccaactc	ttctgagtca	ctttctgata	13980
	ctttagtgtc	ttgccacact	ctggtcagta	actccttcga	tctcccctta	attaattaaa	14040
	ctctgttgtc	agcccaattg	ccaccagcag	ctagttttcc	ttaaaggggtg	tagtgtgtgt	14100
	tggatgggc	agagcattcc	actagagcca	tgagaaaatc	ggagttctac	aaggacacag	14160
10	ccatggataa	agacaattcc	tactcaagct	agctgaagaa	ctgtaaattt	gaaagccagt	14220
	gtctgagcgc	cctctactgg	aagcaatggc	acattactcc	cagcatcatt	ttccagcaca	14280
	aaagacctaa	attgttgaga	tgagacattg	gcttagcatt	ttcttgactc	tcactttcag	14340
	tttcttcag	aactcctgat	gttgacacaa	gacatgacat	atgaagggtg	aagttaatta	14400
	aaagacgctt	tccagcatcc	aagatgaggg	aaagggtgaga	aatactgaga	ggctgtgtaa	14460
15	aatggagtcg	tttcttggtg	catgctaagg	ataaatagca	tccttgccga	ttcttgccag	14520
	cctgcccaga	aaaaatgtgag	ttctagcagc	agtttctcag	ttatcactga	cagaacctga	14580
	ttgtctaatg	atttatagct	ttattacata	gtggccaaga	gcaagcccct	ctggtaagtc	14640
	acttcttcaa	tttggtgaca	agaaaccagc	catcaggaat	gatgtatttc	cttgtaagtt	14700
	ccttgagaaa	tacgagattt	ttcctctgtg	tgcctgtgtg	tgggtgtttat	gcccgtgtgt	14760
20	gtgtgtgtgt	gtgtgtgtgt	gtatgtgtgt	atgtgtatgt	gtgtgtgtgt	ttgtatgtgt	14820
	gtgtgtatgt	atgtgtgtgt	gtgtatctct	agtcctgtct	ttttctttct	actgactttg	14880
	cagttgtcct	gattatattgt	tggtttacta	cagtactggg	gatccgttga	cctgcaggtc	14940
	gaccggatcc	agacatgata	agatacattg	atgagtttgg	acaaaccaca	actagaatgc	15000
	agtgaaaaaa	atgctttatt	tgtgaaattt	gtgatgctat	tgccttattt	gtaaccatta	15060
25	taagctgcaa	taaaacaagt	aacaacaaca	attgcattca	ttttatgttt	caggttcagg	15120
	gggaggtgtg	ggaggttttt	taaagcaagt	aaaacctcta	caaagtgtgt	atggctgatt	15180
	atgatctcta	gtcaaggcac	tatacatcaa	atattcctta	ttaaccctta	tacaaaattaa	15240
	aaagctaaag	gtacacaatt	tttgagcata	gttattaata	gcagacactc	tatgcctgtg	15300
	tggagtaaga	aaaaacagta	tgttatgatt	ataactgtta	tgcctactta	taaagggtac	15360
30	agaatatatt	tccataattt	tcttgtatag	cagtgcagct	ttttcctttg	tgggtgtaaat	15420
	agcaaagcaa	gcaagagtcc	tattactaaa	cacagcatga	ctcaaaaaac	ttagcaattc	15480
	tgaaggaaaag	tccttggggg	cttctacctt	tctcttcttt	tttgaggagg	tagaatgttg	15540
	agagtcagca	gtagcctcat	catcactaga	tggcatttct	tctgagcaaa	acaggttttc	15600
35	ctcattaaag	gcattccacc	actgctccca	ttcatcagtt	ccatagggtt	gaatctaaaa	15660
	tacacaaaca	attagaatca	gtagttaaac	acattataca	cttaaaaaatt	ttatatattac	15720
	cttagagctt	taaatctctg	taggtagttt	gtccaattat	gtcacaccac	agaagtaagg	15780
	ttccttcaca	aagatccggg	gccactcat	aaatccagtt	gccgccacgg	tagccaatca	15840
	ccgtatcgta	taaatcatcg	tccgtacgtt	cggcatcgct	catcacataa	cgtgcctgga	15900
40	cgtcgaggat	ttcgcgtggg	tcaatgccgc	gccagatcca	catcagacgg	ttaatcatgc	15960
	gataccagtg	agggatgggt	ttaccatcaa	gggccgactg	cacagggcgt	tgtgcgccgt	16020
	gattaaagcg	gcggactagc	gtcggagttt	caggatgttt	aaagcggggg	ttgaacaggg	16080
	tttcgctcag	gtttgcctgt	gtcatggatg	cagcctccag	aatacttact	ggaaactatt	16140
	gtaaccgcgc	tgaagttaaa	aagaacaacg	cccggcagtg	ccaggcgttg	aaaagattag	16200
45	cgaccggaga	ttggcggggc	gaatacgacg	cccatatccc	acggctgttc	aatccaggta	16260
	tcttgccggga	tatcaacaac	atagtcacat	accagcggac	gaccagccgg	ttttgcgaag	16320
	atgggtgaaa	agtgcgcttt	tggatacatt	tcacgaatcg	caaccgcagt	accaccggta	16380
	tccaccaggt	catcaataac	gatgaagcct	tcgccatcgc	cttctgcgcg	tttcagcact	16440
	ttaagctcgc	gctggttgtc	gtgatcgtag	ctggaaatac	aaacgggtatc	gacatgacga	16500
50	ataccagttt	cacgcgccag	taacgcaccc	ggtaccagac	cgccacggct	tacggcaata	16560
	atgcctttcc	attgttcaga	aggcatcagt	cggcttgcca	gtttacgtgc	atggatctgc	16620
	aacatgtccc	aggtgacgat	gtatttttcc	ctcatgtgaa	gtgtcccagc	ctgtttatct	16680
	acggcttaaa	aagtgttcga	ggggaaaata	ggttgcgcga	gattatagag	atctggcgca	16740
	ctaaaaacca	gtatttcaca	tgagtccgcg	tctttttacg	cactgcctct	ccctgacgcg	16800
55	ggataaaagt	gtattctcaa	acatatctcg	caagcctgtc	tttgtgtccaa	gctagctttt	16860
	tgcaaaagcc	taggctccca	aaaaagcctc	ctcactactt	ctggaatagc	tcagaggccg	16920
	aggcggcctc	ggcctctgca	taaataaaaa	aaattagtc	gccatggggc	ggagaatggg	16980
	cggaaactggg	cggagttagg	ggcgggatgg	cggagtttag	gggcgggact	atggttgctg	17040
	actaattgag	atgcatgctt	tgcatacttc	tgcctgctgg	ggagcctggg	tgtgactaa	17100
60	ttgagatgca	tgccttgcat	acttctgcct	gctggggagc	ctggggactt	tccacaccct	17160
	aactgacaca	cattccacag	ctgcctcgcg	cgtttcgggtg	atgacgggtga	aaacctctga	17220
	cacatgcagc	tcccggagac	ggtcacagct	tgtctgtaag	cggatgccgg	gagcagacaa	17280

```

5      gccggtcagg gcgcgtcage ggggtgttggc ggggtgtcggg gcgcagccat gacccagtca 17340
      cgtagcgata gcggagtgtat tactggctta actatgcggc atcagagcag attgtactga 17400
      gaggatgacca tatgcggtgt gaaataccgc acagatgcgt aaggagaaaa taccgcatca 17460
      ggcgctcttc cgcttctctc ctcactgact cgctgcgctc ggtcgttcgg ctgcggcgag 17520
      cggatcacgc tcaactcaaa gcggtaatat gggtatccac agaatacagg gataacgcag 17580
      gaaagaacat gtgagcaaaa ggccagcaaa agggcaggaa ccgtaaaaaag gccgcgttgc 17640
      tggcggtttt ccataggctc cgccccctcg acgagcatca caaaaatcga cgctcaagtc 17700
      agagggtggcg aaacccgaca ggactataaa gataccaggc gtttccccct ggaagctccc 17760
10     tcgtgcgctc tcctgttccg acctgcccgc ttaccggata cctgtccgcc tttctccctt 17820
      cgggaagcgt gcgcttttct catagctcac gctgtaggtg tctcagttcg gtgtaggtcg 17880
      ttcgctccaa gctgggctgt gtgcacgaac cccccgttca gcccgaccgc tgcgccttat 17940
      ccggttaacta tcgtcttgag tccaacccgg taagacacga cttatcgcca ctggcagcag 18000
      ccactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt 18060
15     ggtggccta ctacggctac actagaagga cagtatttgg tatctgcgct ctgtcgaagc 18120
      cagttacctt cggaaaaaga gttggtagct cttgatccgg caaaciaaac accgctggta 18180
      gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag 18240
      atcctttgat cttttctacg gggctctgacg ctcaagtggaa cgaaaactca cgttaaggga 18300
      ttttggatcat gagattatca aaaaggatct tcacctagat ctttttaaat taaaaatgaa 18360
20     gttttaaatc aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa 18420
      tcagtgggc acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc 18480
      ccgtcgtgta gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga 18540
      taccgcgaga cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa 18600
25     gggccgagcg cagaagtggc cctgcaactt tatccgctc catccagtct attaatgtt 18660
      gccgggaagc tagagtaagt agttcgccag ttaatagttt gcgcaacgtt gttgccattg 18720
      ctgcaggcat cgtggtgtca cgctcgtcgt ttggtatggc ttcattcagc tccggttccc 18780
      aacgatcaag gcgagttaca tgatccccca tgttgtgcaa aaaagcgggt agctccttcg 18840
      gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt atcactcatg gttatggcag 18900
30     cactgcataa ttctcttact gtcatgccat ccgtaagatg cttttctgtg actggtgagt 18960
      actcaaccaa gtcatcttga gaatagtgtg tgccggcgacc gagttgctct tgcccgcgct 19020
      caacacggga taataccgcg ccacatagca gaactttaa agtgctcatc attggaaaac 19080
      gttcttcggg gcgaaaactc tcaaggatct taccgctgtt gagatccagt tcgatgtaac 19140
      ccactcgtgc acccaactga tcttcagcat cttttacttt caccagcgtt tctgggtgag 19200
35     caaaaacagg aaggcaaaat gccgcaaaa aggggaataag ggcgacacgg aaatgttgaa 19260
      tactcatact cttccttttt caatattatt gaagcattta tcagggttat tgtctcatga 19320
      gcggatacat atttgaatgt atttagaaaa ataaacaaat aggggttccg cgcacatttc 19380
      cccgaaaagt gccacctgac gtctaagaaa ccattattat catgacatta acctataaaa 19440
40     ataggcgat caccaggccc tttcgtcttc aagaagaatt ccaagctatt 19490

```

<210> 12
 <211> 19412
 <212> DNA
 <213> Artificial Sequence

<220>
 <221> misc_feature
 <222> (790)
 <223> Expression plasmid where n can equal a, c, g, or t

<220>
 <221> misc_feature
 <222> (13735)
 <223> Expression plasmid where n can equal a, c, g, or t

<400> 12

```

      ttccaagcta tgctcgacct gcagggtcaac ggatccacat atataattta gaagttttca 60
      ctagccatgt tataaaagta aaaagaaatc tgtgcaatca attttaataa tatttttgtt 120

```

	tgcacacaag	atatacaggg	taatctcaac	atgtaatcat	ataaaatatt	aatattttac	180
5	acgtattttt	tgtacaaaat	gttaaaaatt	ggatgtgttt	ttacacgtat	agcccatctt	240
	agtatggact	ggccttatct	ccagagctca	ataactctgt	gtggctagt	gcaggatatt	300
	ggatatcaag	tctagaacat	acgtgaatta	attcaagggt	caagcacatt	aaggctgaaa	360
	gagttatgac	tgaagaacaa	gttcaaggag	gaaatcaaac	aatagccaaa	tttcaactgag	420
	ctactccttc	cattactaat	gataacaaat	atztatagat	cccttagtat	atactaggcc	480
10	tgatgcttag	aacttttatat	gaactgtttc	atttaatcct	cacagcaatc	ttacaagtaa	540
	gacctccacc	ccctttcaga	tgagaaaact	gaagcacaaa	agtgcagtaa	cttgctatta	600
	tgactttgtga	gcatattttc	ccaggcaatt	atatgcacag	gatgatgtag	gtaaagcagc	660
	ccaataattt	attcactata	tttctccatg	ctcatggtaa	ggactgactt	gatccctggt	720
	ggtcacccac	ttgtgctccc	tcagttaggg	aagggctgca	gagcccctgg	cccggtgtcct	780
15	agtgaactcan	gcttgggac	tccagcagct	gctatgggtg	acaatttagt	ttagtataat	840
	gttagatcct	accctgataa	ctgttataaa	aaatcctcct	tccccctccc	cagcacaggt	900
	acccagccca	aggggcactt	tctatggaaa	gatgaaaaga	actgacatgg	acttgagctc	960
	actgggggtg	caggtttcaa	gaaagaacaa	agagagctgt	taggttgaag	agacgcagtt	1020
	gaagagttaa	gtcttggaat	tggttaatacc	tggaacaagg	cacgggtgtga	aaagacaaaag	1080
20	aaacacacaa	acatcaatct	gtcctcctct	tttctctttc	tgtttcattt	cttatcctcc	1140
	atcagtcaga	ctttgcttcc	tctttctcct	ccttcacctc	ctcctcctcg	cccttcccaa	1200
	tttctttcta	tcttccattt	attgttcttc	caaaattcgt	atactcatcc	tctcttctcc	1260
	ctttctcttt	tctccaaaat	aatatgttca	acttactgac	ttgctcaggg	gattgtgggg	1320
25	gcgaatgtat	tagcaccaaa	cctaggaaac	atggcgaatt	ctaaaaactt	gttctcatag	1380
	aaaaaagagt	taaaaataga	aagctacagt	ccagttttga	agccacagaa	gacctcactc	1440
	tgatgtatag	acattagagt	cagaaaaaga	ctttgaaaag	accagtgaag	ctccttgtcc	1500
	cacagacagt	gtgagggagg	gacaagggga	attgtgtgcc	caagcgtact	tgggatgggtg	1560
	taggtggctg	ataagagtgt	cagtggaggt	gggtggttagc	atgttgagtg	ccccaatctt	1620
30	gtgacactaa	acattttattg	ttataactaac	cccactgcc	actaccagga	agttttcatt	1680
	tttcttactg	tgtcttagtc	tccatggcat	ctataaaactg	ttttccccc	atgtccaggc	1740
	atcattcctc	tcccacagtt	gatgtctggc	tctgtgccat	gggtgtgggc	ctgcctcttc	1800
	tctctcagaa	tgactgctgt	ttcatgcttg	gtgagccacg	cagtatgcc	agctgctgtc	1860
	tctgacctgg	agaaaaggata	cccacccatc	cacaggtggg	gcatatcatg	ggatggcctt	1920
35	cctggttttc	cttggcagac	acattttttct	cacttttttt	tttctctcct	cccttctctt	1980
	tttccctccc	tctttgattt	cttctctttt	cctttttctt	ctcttgcttc	acttctgccc	2040
	tagatctacc	atgggaccca	gggagctttg	gagccatgca	acccacagt	tgctaagctt	2100
	tttactcatg	ctatttcttt	taatcttcac	aattctttac	ctatcaggca	ttaccatcag	2160
	atctgcacct	tacaggtgag	aaagggaggg	ccacaccgat	taagaaatgt	gctcagggcc	2220
40	acacagcctg	tacatgtcag	agtaagggca	ttctcatgcc	aagggttcatt	ttcagagcca	2280
	aaagcatacc	acgtctcagg	agaagcaggc	tagatctctt	gtctgcttgg	cttttaactc	2340
	atgctgacct	gaactttctg	gatttgagta	actctcagac	ctgtggagcc	cgatttctgt	2400
	gagtggattt	tgaggccac	ctggggaggt	gaacccttgc	tggggtatgg	ctcattagt	2460
45	tagaagggag	ttccttcaga	ggtacgtgt	ctctcggcc	atccccattg	tattagtttc	2520
	ttgggtctgc	cctaacaaag	taccacaaac	taagagccgt	aaaacaacag	aagttttattc	2580
	tgtcataggt	ctggagacag	aaatctgaat	tcaaggcaca	ggcagagtta	gtgtcttctg	2640
	caggttctcg	gggagaaacc	atcccacacc	tctccagctt	ctgggtggctc	ccagcaatcc	2700
	ttgggtgctct	ttgggtctca	gctgcacccc	tctgggtctct	gcctccttgg	tcatgtggct	2760
50	ttcttcctat	gtgtcactgc	atctccaatc	tcttctctct	tataaggaca	ccagtcattg	2820
	gatttagggc	ccactcggat	ccagcataaa	aaaaaagttt	aaaaaaatta	tttggaata	2880
	attatagatt	cacagaaaag	agcaaagaca	ggatataaag	cacagaggtc	tgatgtgtcc	2940
	tgacccag	ttctcctggt	ggttacttct	tacgtaatta	cagtacgata	ttaaaacca	3000
	gaatttgaca	ttgataaaaat	aagtgtgtac	agttctatgt	cattttatca	catgtctaga	3060
55	tgcttgcaac	taccactgca	atcaagatac	agattttatcc	catgatcaca	aagatctccc	3120
	tttatagttt	cgctactct	tctttctcca	ccaccctaac	ccctggcaac	cagaatctgt	3180
	tctctatccc	tataattttg	tcacttcaac	aatgttacgt	aaaaggaatc	acatggtag	3240
	tgaccttttc	agattggctt	ttttactca	gtagaaagtc	cttcagaatc	actccagttc	3300
	tctgtatcaa	cacttcgttc	ctttttattg	cctcatagca	ttctatggca	tggatatatt	3360
60	acaattttgtt	aggtcagttg	cctattgagg	gatgggttgg	ctctttctag	ctttggctac	3420
	aacaaataaaa	gctgctgtga	acaatcatgt	atgggttctg	tgtggacata	agtttttatt	3480
	tcctgggggt	aaattcccaa	ggagtacaat	tcctgggtca	tatggtaggt	gtcttttgtt	3540

	tagtttttca	agaaactgac	aactatatttc	tagagtagct	ataccctttt	atattccac	3600
	cagcaatata	tgaatgatcc	attttcacccg	catectcaca	aacatttcag	tatgacctaa	3660
5	ttttgacttg	attacctctg	caaagacctt	atttccaaat	aaagtcacat	ttatgtgacc	3720
	catttatgag	taccaagagc	aaggagttga	gcataccttt	ttgaaggaca	caattccacc	3780
	tgcagcacct	gtcttcacett	gcttaggtta	tgggtgggtt	tcacttcaga	gtgttatctt	3840
	ctgagaagaa	ggtttgaatt	cataactgga	ttatgcctgc	ctttgattgt	tcaatttggc	3900
10	ccccagggat	gagtgagcaa	gttcatgtca	ttggaaggga	cattactgag	tctcaccctt	3960
	actcaaacct	ccaaagtgtg	aaagatttga	tactcatcaa	cattctcaag	aggcagaagg	4020
	gcaaggtgct	tccttatgta	gtttctccat	gactgtcctt	ccccactttc	ccagtactga	4080
	ctggaaagaa	aaacaaagca	ggacaaaggg	attgtcagtc	atltgaggga	agggatgctg	4140
	tgtaggttac	tcctgaaccc	cttcagagtg	taggcagaga	gaggtactta	gaaaggtgtg	4200
15	ttggatggat	gagtagctga	gaggggcagg	gccagagtgc	cccatagaat	gaggaagagt	4260
	cctcatttac	atgtgccctt	cccccttccc	aaacatgtga	gtgttcatac	atgggtttcc	4320
	agtgtctggg	caaggttaat	tctgtcctct	gtccccctgc	tgaaggctctg	tttcatttaa	4380
	ttctctatga	tgtgtgctcc	tagctagcca	aacagggcac	agatgccagg	aaaatagcag	4440
	gctctgtgtt	ttgttgagca	ccttgagccc	tggagggcag	actcattcag	ggctcaagag	4500
20	ggcaggatta	ggcaggtagc	tatttgatcc	caggccccc	cacttaccag	ctgctgtggtc	4560
	ttgacctatt	acttatctct	cttctatcta	ttactgggt	ataagatgtt	ctttatgtga	4620
	agttattttg	aagcataatg	tgtatacaga	acagtgcaca	gatcataact	gcacaacccc	4680
	acacgaggaa	cagaacatta	ccaagacccc	agaagcccct	tcttgtcttc	tcccaggcac	4740
	taactcgccc	acccccagct	aagagtaatt	tccatccaaa	tttctaactt	tatttttacc	4800
25	tggttttgaa	ctttatatgg	atgaaaacat	acagaacata	tcttttgcat	tgagtttctc	4860
	ttatgtaaca	ttatttttat	gagattcctc	tgtgtgagta	ggtatagtgt	tagcttgtta	4920
	attctcctca	ttgggtattc	taatgtataa	acataccaca	gtttaataac	tcattctatt	4980
	gttggacaat	cgggctattt	ccaatttttg	gcttttgtaa	ataaagctgc	tttgaacatt	5040
	cttgcatagt	tttttggtgg	gtggggacac	tcagtcactc	agaagtggga	ttgctgagtt	5100
30	atagaatata	tttgtctctc	agctttatag	aacctttgca	acattttaat	gaatgttttt	5160
	aatctcacag	tatagtgtgg	attaaattag	gttgtggaaa	gtactcaata	cggcccttgg	5220
	cagacaggag	tcctcagta	cctgtttaaa	aatgtaaata	aggcatgaca	ttccattaag	5280
	tctaagtacc	atgatgaact	taaacatttc	cctattgttt	gtttatttaa	ttgcttttat	5340
35	tgtgttggtta	ttttatatca	tgttgcaatg	aatgattctc	tttatagcag	gtttcttctt	5400
	ttgaaaaatt	tgccttggtt	gattctcagg	gaattgaatc	aaaggatata	tgactgtaag	5460
	acctgtcacc	cttaaaaagg	actatgaggg	cttgctgagg	aggggaaaac	aagggaagcaa	5520
	gtctctccta	ccatggccca	ggggactgtg	aggacagaag	gcttgtgggt	ttgaggggagg	5580
	actgtcttgc	agaggatgat	agggtaaaa	agaattgaag	atgattttta	taaatgggta	5640
40	tgcgccttag	gatgactaca	tatttagtcc	cttataagag	aaattgagta	gttggtaaaa	5700
	caacagataa	taattattaa	atgaggaaag	agagaaacca	cagggtgcaa	gattcacttt	5760
	atlttattcat	tctcctccaa	cattagcata	attaaagcca	aggaggagga	gggggggtgag	5820
	gtgaaagatg	agctggagga	ccgcaatagg	ggtagggtccc	ctgtggaaaa	agggtcagag	5880
	gccaaaggat	gggagggggg	caggctggaa	ctgaggagca	ggtgggggca	cttctccctc	5940
45	taacactctc	ccctgttgaa	gctctttgtg	acgggcgagc	tcaggccctg	atgggtgact	6000
	tcgcaggcgt	agactttgtg	tttctcgtag	tctgctttgc	tcagcgtcag	ggtgctgctg	6060
	aggctgtagg	tgctgtcctt	gctgtcctgc	tctgtgacac	tctcctggga	gttaccgat	6120
	tggagggcgt	tatccacctt	ccactgtact	ttggcctctc	tgggatagaa	gttattcagc	6180
	aggcacacaa	cagaggcagt	tccagatttc	aactgctcat	cagatggcgg	gaagatgaag	6240
50	acagatgggtg	cagccacagt	tcctgaggaa	agaagcaaac	aggatgggtg	ttaaagtaaca	6300
	aagttctgcc	cttgggtgtg	ttgtttgcgg	ataatcacag	ggcatgttag	ggacagacag	6360
	aaaacagcat	gcttatccca	gataattata	gcaaggagac	caagaagcgt	atttaaaatc	6420
	ttgatgtttt	gagtttcttc	ctagcttccc	cctattcctt	aataaagttc	taaattgttt	6480
55	tgttgagct	ctttgcagcc	attctgaggg	ctttgcatgc	ttttctgacc	ttgcagtaaa	6540
	ctcaatgctt	taggcaaaga	atggccacgt	catccgacct	cctcagagtt	tagaattcag	6600
	aacaggtctg	aagaagacca	ggcagcggct	gagtcaaggga	aagcctccgt	ccgcttttat	6660
	ttccctctgt	cctcttccag	gactgtgctg	ggataacagg	ctcccggggg	ttactttggc	6720
	tgggctgggc	taaaacctcc	ctgcagagca	ggcctgagc	cctgcctctg	cgcctgggtg	6780
	gtgtcagccc	ctccaccttc	tgactgttcc	agcaactctc	taagccctcc	caaaggcctc	6840
60	aaggcctgta	accatctgca	gcaattttca	gccataccag	gagaggtcaa	ctgtaattctt	6900
	ggccacctgc	ctaagaggaa	gtggctagct	tcacttctga	ccctcagcaa	ctgccagggtg	6960

65

	gcctcttggg	aatccccctc	tgggggatcc	caccgcgttg	gtgggagagc	agtagttaa	7020
	atgtaaaata	agaatctttt	gctgggagaa	gtcaacagat	agggagaagt	cagctgata	7080
5	cagaaatagt	tttttttttt	tttaaaacta	acttcactgt	taaccaagca	gttcaacatg	7140
	aaagactgaa	tctcttatgt	ttaatatatt	cttctctttt	aatcttcata	actaattttt	7200
	ttcagataat	tgtataaaat	aaccatggta	gcaaaataat	gtgatcactg	gaaaataagc	7260
	agggaaaaac	atgctatgaa	gatactccta	tctgggtgaa	ttcttgatag	ctttacattt	7320
10	ttcatctggc	atttaaacat	taaacagtta	atgtatttga	catgaaaatt	atttcaagtt	7380
	atcttattag	ttttaataga	gtttaaaaag	tgtttaaaag	agttttcaaa	aggctctaaa	7440
	atcattttga	aatagtttaa	aacagttttg	aatcgttgta	agttagtttt	aatagagctt	7500
	taaaaaggcc	ctaaaatagt	cctatcaagt	tgttcagac	caaaataatc	tccttaata	7560
	tcacttttga	gatcagctgg	ggtaaaccgac	agcaacacaa	tgacaaatca	ttaaactatt	7620
15	ttagagatta	tgaattaaa	atactcagat	taaaattttc	ctatcacaga	attaaggtac	7680
	tggaaaatat	gttttaagttt	ttattaatcc	attgctatag	gttttagatat	tttgtaaac	7740
	tgaataaaaa	tcacacactg	gcagctacat	ttttgaaagt	taaaaacatg	gtcacgaata	7800
	tatcttattt	taaaatcagt	taatatacct	taatggattt	taatgccaaa	ttcaaagtga	7860
20	attgatcaag	ccctcagtg	ccaggctcatg	gggtgtgattt	ttactctgaa	agaattacat	7920
	atctctttct	ttttgggttga	gctttttgtta	tttaaataca	tttgatgaga	ggatattgaa	7980
	ataattaaat	agcactgaaa	aaaaaaagct	ttaaattatt	tacaatcccc	taattggaat	8040
	tttcaacta	gagatatcat	aatgaatgtg	aattttattt	ctgaaatctc	taataaatca	8100
	gtcttctccc	tggttttccc	agctcagcgc	ccattacgtt	tctgttctct	ttcccttagt	8160
25	ggcattattt	gtatcactgt	gcatcaggaa	agctggctac	ggcagcatca	atcgggcaga	8220
	cacagggtgg	ccacggccac	tagcggcaag	gcggtgccc	caagagcgcg	gtggcatggc	8280
	caccaaagcc	actcaatcga	gaaagaccgc	ggctctgtct	acagctcgcg	gtgccacggc	8340
	cttcttggca	gaataaaaa	gtagacaagt	aataacagag	gataatgaaa	gaacatactc	8400
	tttaaaatat	ttcctatttt	tttcacagac	ccacggctcat	taaaaaatgc	aattatttac	8460
30	tttttttcat	ttaaacacat	ttctttgaga	ttgagctttt	gggaataacc	acctttccac	8520
	cattacaata	agagataa	tcacgttttag	tctaattgtac	aaattggatt	tttaaaaaat	8580
	gagctctatc	tgtgaagccc	ttattcctat	agaatgtgtc	tttttgagtt	tattacttat	8640
	tacagactct	aaaaacaaca	ttgctgctga	ttttcaagta	agctgcctct	tctacatagc	8700
	aaataggtag	acttcacttt	tccctgattt	ttcttagggc	gtgctattga	tttttattgt	8760
35	tgtctgacaa	aataatttat	caaacaagag	ggagaaagac	taaaaaatgt	atttttccac	8820
	ttttctgtat	catgcataat	cagcaacaac	caatacaata	tttggaaga	gtgaacaaaa	8880
	ataaatttac	ttttgtctct	tagaaataca	aggggttctt	tttagttaca	cttttttttt	8940
	ttactttgtg	tcattcagtt	tagagcaatt	taattctttt	ttctccaaat	ccatttttga	9000
	agctgagttt	aacttttgca	acccatggca	aatcttaaat	gccctcattt	accaatcttt	9060
40	accaaactcc	tatttaagcc	tctaaaagtc	aatactggcc	atcagaccca	aatttcagaa	9120
	gacaatagtg	aaaaattact	tacgtttaat	ctccagtcgt	gtcccttggc	cgaagggatc	9180
	cacagtgtta	acttaattac	tttcccctta	acaaaaatct	cttttcgctg	ttaatatcac	9240
	taacctgacc	gatgcagaga	aaatcttgca	atbgagatgc	ctcacttaac	tggctagcgc	9300
	ttggctgttc	cttaagatga	actaattttc	tatcccttac	tcactctgact	ttttgaaaga	9360
45	atctggtagt	ctttggaatt	gacctgagct	aatatctcaa	acacaaaaac	gtcccaaat	9420
	taaaacctta	taagaaaaag	cattagggaa	gtgcacttac	gtttgatctc	caccttgggtc	9480
	cctccgccga	aagttagcca	cagttagggg	tctcaccctt	tcccctcaac	aaaaacctct	9540
	cttgaagcca	atcatttgag	ataggtgct	tgttcagaga	aaaatctagc	tatttcttcc	9600
50	ccatttcccc	catgaatcct	attctcctct	caaaccctaat	gattcgtcta	tttgctcagc	9660
	tttttaagtt	cattttctgg	tgtcctgcta	tttacttctg	ggtcaccagg	tttattcaac	9720
	caaaatatca	caaaacttgc	acaaatgata	caatggcact	aaaatctcac	gaataattga	9780
	gacagatgta	cttacgtttg	atatccactt	tggctccagg	gccgaaagtg	aatcacagtg	9840
	attcgtctta	acttttccct	ttacaaaaac	ctccctgaaa	gctcagcaag	cctctttccc	9900
55	ccaatgaagt	tattttgatt	tagaaatctt	aaaaattagc	cacaagctag	cgctcctgtg	9960
	aacaatttcc	cctcctctgt	acctaacctg	ggaatgaagt	ttgttagatc	cctggcatcc	10020
	gactaatgaa	aatccacaca	aaggaacaca	aagtaaacta	attagcaaca	gtgaagaatc	10080
	agtggaaaaa	agtacttacg	tttgatctcc	agcttgggtc	cctggccaaa	agtgtacggg	10140
	taaatattat	actgttggca	gtaataagtt	gcaaaatctt	caggctgcag	gctgctgatg	10200
60	gtgagagtga	aatctgtccc	agatccactg	ccgctgaacc	ttgatgggac	cccactttgc	10260
	aaactggatg	cagcatagat	cagggactta	ggggctttct	ctggtttctg	ctgataccag	10320
	gctaaccagc	tgctaatace	ctgactcgcc	cgacaagtga	tgggtgactct	gtctcctaca	10380

65

	gatgcagaca	gtgaggatgg	agactgggtc	atctggatgt	cacatctggc	acctgagatt	10440
5	ggaaacataa	aaacaaatgt	ccacacaatt	aatcatgttg	taagagaatt	tccttgaata	10500
	gtaaagcagt	actgagcacg	ctgggctgag	taaactgcta	gtgttctcca	tccttacctg	10560
	ggaaacagag	cagcaggagc	cccaggagct	gagcgaggac	cctcatgtcc	atgctgtgtc	10620
	ctgactgggt	ctgattcctg	cacaaagtct	gaccagccta	ttaataaggc	ttcagggcag	10680
	gaggttgtgc	tctgggaaca	tgcaaattgag	caggggatgg	ggcaggctgg	gcacagctgc	10740
10	agagctggcg	catctgagta	actcagcacc	agctcagtg	ccccacctgt	cccaggtaag	10800
	atcaaggtag	ctcaaatttg	tctgcagaga	atgtgtttct	actggggact	attttgttat	10860
	gggaaacatt	ttatggtttc	tttttgacaa	tttgaaatat	tccttgggag	tcgatggagc	10920
	aatgtatttc	attgggtgat	ggggattatt	taggagaata	ttcttttttg	taggaaacac	10980
	atagtaaaat	tttagaccct	acaattttca	ggtcttcaaa	agactctcat	gtgatttctg	11040
15	ttaggggaagg	tggtacttat	catatacttg	caacatttct	gtgagtttaa	cattgttctt	11100
	ttctaaaaaa	aaaattaaaa	ataaaattta	ttcacatgat	gctacatata	tttgtaaattg	11160
	ttaggtaatg	gtgttatgcc	attgttctta	ccactgtaag	atcaagcaat	ttacttcaga	11220
	tacactaagt	tgataccgtg	tttcctcaat	gcatgcagca	attacagatc	caccattatc	11280
	aagagctcta	ggtctcttta	atacccagag	actaaatggg	ctgcacctta	ttcctgtttt	11340
20	gggcaccttc	atagtctacc	ttcttttctg	ccattaagta	ttatttccca	acattcatct	11400
	ctcttagtga	gggtgatcat	cgcagggagc	atgtccctgc	cacgcaccat	aggtgacact	11460
	ttcttctttt	actttttatc	agggacatca	tcctgaccca	gacaccaaga	tccttgttaa	11520
	catcttttgg	aaagaggcac	tcaatctttt	atcaagcaac	tgcccatgta	catggagaaa	11580
25	tcaagtgggt	tccagatgaa	actggacagg	gatttgcact	cgttatatct	catatctcta	11640
	atgtgcccga	aaatgtcccg	gcctggttcc	gtgggagggg	aagtggatcc	aactacatca	11700
	gcatcagtgg	gctgcagcct	aggactccat	aaaattttac	tgatgcctga	ctaggagagc	11760
	caaatcacag	tgctgtagcc	tctgcacaaa	ccttcctgct	gctttataag	ctgcctgaat	11820
	tttaagggaa	attgcttata	ttggaagaaa	ggaagaaagc	tccatttgtc	ctctaaatgt	11880
30	ttgctgaaaa	taaactgaca	aaagggaagat	taataagaga	aaaggcaaac	aaaattcact	11940
	taaagtgcag	caggatatca	tagcaagggtg	attaccacga	taactcaatg	ggatccaggt	12000
	gtttatgttt	cctttctagg	gaagaggtaa	ttgggaaatg	taggcaacct	ggagagaata	12060
	gatgcgaata	gaaatgcata	ctcaaaaagaa	caggtaatca	cctccctagg	taaagtatca	12120
	actttgagtc	tcttccattt	ttgattcctc	ttttgtgtta	atcttccctg	acataaaaaat	12180
35	tctcaggaag	agttttctta	aaaattgggt	tccttctgaa	gaatttgctt	ttaggcaggt	12240
	aggggatgtt	taggaaaagc	cccctgtgca	tttctgtctt	tctaaatgcc	tttggtttta	12300
	tataatcatc	atacaaatgc	agcacagttt	gagatgttat	tttctggatt	cctttacttg	12360
	caaccacact	gccaaagatc	tgttccagag	agatgtggct	acagactgaa	agagcagttt	12420
	tcccctcaac	aacgtggagg	ttggggcact	gaccactggt	gcagatgaaa	acctgtgtag	12480
40	aacttttgca	tctccaactt	atgtagtaat	agattacttt	tgactggaag	ccttaatgat	12540
	aaaataaata	gttgattaac	ccttttttat	gttatacata	ttatatactc	tattctttcca	12600
	ataaagtatg	ctaaagaaaa	aatgttattg	agaaaacctt	aaaaatgaga	aaatatattt	12660
	agtacttatt	aagcgcttgc	tcacagggtga	cacacagaag	aaaatataag	tggaatctgca	12720
	aatttcaaac	ccaagttatt	caaggggttaa	ctgtaccatg	atgaatgtag	cagccccat	12780
45	ctatagtcta	gggcttttcc	ctttgtctga	tctctgtctc	tttccaatgg	ctatatatat	12840
	gtcttatgtg	cttcatgatc	ttgggcagag	aggtctgctt	gcatgccttg	ctggccagat	12900
	ggcctgcaat	gtgtatctct	taggaaagtg	ctatggtttg	actgtgtcct	ccagaatcca	12960
	tcgttataaa	gttaaatctc	agtgaatgg	tattgccagg	tggggcctat	tgagatgtgt	13020
50	tagatcatga	gggtggagcc	ctctagtggg	ataccttaat	gccactataa	acagggttta	13080
	tggggctgga	atctctctgt	ctcttctgct	ggtctgtcat	gttaagacat	ggccttcggt	13140
	cctttgaagg	actgaaagct	ccaggcgtca	tcttgaaagc	agagaaagag	gagcttaacc	13200
	tgcccatgcc	ttgatcttaa	atttcccat	ctccagaaat	gtgagaaaat	aaatttctgt	13260
	tctttatgaa	ttacacagtg	tcaagggaacc	taacctgtta	tagcagcttg	aaagagaaca	13320
55	agagagacag	ctcaaatca	gtgaggacag	gatgagggtg	atacatacct	cagcttcctc	13380
	gtctctcagg	tggaatagcc	cagaggaatt	taagtccatg	tttccacatg	tggttgatct	13440
	tcagttatcc	tgagttaggt	gggttggtga	tggtgtcttt	accattcatc	ttctgttccc	13500
	tccctcactt	tcttcttttc	ctcccagtg	aaatttgctg	cctaaacagg	aatcctcatt	13560
	gctgggtggc	ccaaactaag	atagtaaata	aaatcattaa	tcttttgtat	gaggggtatt	13620
60	tctcatctga	attcagatcg	atgtggtacc	tggatcctcg	acctgcaggc	atgcaagctt	13680
	aggaagaaaa	atctaaaaaa	atgaaagtca	cagaaatagc	aagtagaatg	ctggttacca	13740
	gagactgcat	tgntaaaggg	tacaaaatat	cagctctgata	gggggactaa	gttttggtga	13800

5	tctgttgcat	agcatggcga	cattattata	tagatgtatt	tctggcaggc	ttgattgcct	13860
	attacatggg	aagtcaacat	acactaagac	actgggggct	gctgcagaga	aagagattta	13920
	atcccaaggc	aattaaatga	gaagacagga	ggaagcctca	tatctacctc	cccaagcact	13980
	ttgggggttaa	agactttaag	tggttttgga	tggagaggca	gattgggtga	agagtgaagg	14040
	atgaagtcac	gggactggga	ggtgaagaaa	ctgctttctt	atgttgactc	ggttctttgg	14100
	gaggggggtg	gtcttttagac	aagttggtgt	cagctattct	actggaattc	aggatctggg	14160
10	aaatatctca	aagattaggt	tttatgctca	taatggtgaa	ggtgttatat	ttgggaacaa	14220
	tggggacggt	aatggtctgt	atttattgca	acttgacttt	tattagttag	aagctaaggg	14280
	aagtggctca	gaatgttgct	tgattaatgc	ttaactatat	ttcagtctgg	aacctggaat	14340
	actgttcttg	ttaaccttat	gacagtgggt	ttgtagttaa	taatactata	ttctgtattt	14400
	caaaataaca	aagtatatct	aatgttctca	gtacaaaaaa	tgataaatat	tggaggtgag	14460
15	gactatgtta	attaacctga	tttgattatt	cgacaatgta	tacatgtatg	catctcattg	14520
	taccccatat	gtatatacat	tactatttgt	caattaaaaat	caaaataaaa	cttttaaaaa	14580
	ttcatgtagc	cattgcctgg	aataaataaa	gaaaaaatatg	tgtgcacatg	cttaactata	14640
	cccctgtaaa	aattgaagct	actaagtcca	gtaaaataact	gaatcacaca	tatttctgtc	14700
20	tttagcttta	gtggaaatag	tcaaaatact	gggtcccaac	aatacttata	ttagtctttt	14760
	cttcccatat	ccaccacctt	gttatttttt	aattcattttg	taatacatatt	agtttcattt	14820
	actcttggtt	gtaaccacatt	tggatccggt	gacctgcagg	tcgaccggat	ccagacatga	14880
	taagatacat	tgatgagttt	ggacaaacca	caactagaat	gcagtgaata	aatgcttta	14940
	tttgtaaaat	ttgtgatgct	attgctttat	ttgtaaccat	tataagctgc	aataaacaag	15000
25	ttaacaacaa	caattgcatt	cattttatgt	ttcagggtca	gggggagggtg	tgggaggttt	15060
	tttaagcaa	gtaaaacctc	tacaaatgtg	gtatggctga	ttatgatctc	tagtcaaggc	15120
	actatacatc	aaatattcct	tattaacccc	tttacaaatt	aaaaagctaa	aggtacacaa	15180
	tttttgagca	tagttattaa	tagcagacac	tctatgcctg	tgtggagtaa	gaaaaaacag	15240
	tatgttatga	ttataactgt	tatgcctact	tataaagggt	acagaatatt	ttcccataat	15300
30	tttcttgat	agcagtgcag	ctttttcctt	tgtggtgtaa	atagcaaagc	aagcaagagt	15360
	tctattacta	aacacagcat	gactcaaaaa	acttagcaat	tctgaaggaa	agtccttggg	15420
	gtcttctacc	tttctcttct	tttttgagg	agtagaatgt	tgagagtcag	cagtagcctc	15480
	atcatcacta	gatggcattt	cttctgagca	aaacagggtt	tcctcattaa	aggcattcca	15540
	ccactgctcc	cattcatcag	ttccataggt	tggaaatctaa	aatacacaaa	caattagaat	15600
35	cagtagttta	acacattata	cacttaaaaa	ttttatatatt	accttagagc	tttaaatctc	15660
	tgtaggtagt	ttgtccaatt	atgtcacacc	acagaagtaa	ggttccttca	caaagatccg	15720
	gggcccactc	ataaatccag	ttgccgccac	ggtagccaat	caccgtatcg	tataaatcat	15780
	cgtcggtacg	ttcggcatcg	ctcatcacaa	tacgtgcctg	gacgtcgagg	atttcgcgtg	15840
40	ggtcaatgcc	gcgccagatc	cacatcagac	ggttaaatcat	gcgataccag	tgagggatgg	15900
	ttttaccatc	aagggccgac	tgcacaggcg	ggttgcgccc	gtgattaaag	cggcgagacta	15960
	gcgtcgagg	ttcaggatgt	ttaaagcggg	gtttgaacag	ggtttcgctc	aggtttgcct	16020
	gtgtcatgga	tgcagcctcc	agaatactta	ctggaaacta	ttgtaaccgg	cctgaagtta	16080
	aaaagaacaa	cgcccggcag	tgccaggcgt	tgaaaagatt	agcgaccgga	gattggcgga	16140
45	acgaatatca	cgcccatatc	ccacggctgt	tcaatccagg	tatcttgccg	gatataca	16200
	acatagtcac	caaccagcgg	acgaccagcc	ggttttgcga	agatggtgac	aaagtgcgct	16260
	tttgataaca	tttcacgaat	cgcaaccgca	gtaccaccgg	tatccaccag	gtcatcaata	16320
	acgatgaagc	cttcgccatc	gccttctgct	cgtttcagca	ctttaagctc	gcgctgggtg	16380
	tcgtgatcgt	agctggaaat	acaaacggta	tcgacatgac	gaatacccag	ttcacgcgcc	16440
50	agtaacgcac	ccggtaccag	accgccacgg	cttacggcaa	taatgccttt	ccattgttca	16500
	gaaggcatca	gtcggcttgc	gagtttacgt	gcacggatct	gcaacatgtc	ccaggtgacg	16560
	atgtattttt	cgtcatgtg	aagtgtccca	gcctgtttat	ctacggctta	aaaagtgttc	16620
	gaggggaaaa	taggttgctc	gagattatag	agatctggcg	cactaaaaac	cagtatttca	16680
	catgagtcgg	cgtcttttta	cgcactgcct	ctccctgacg	cgggataaag	tggatttctc	16740
55	aaacatatct	cgcaagcctg	tcttgtgtcc	aagctagctt	tttgcaaaag	cctaggcctc	16800
	caaaaaagcc	tcctcactac	ttctggaata	gctcagaggc	cgaggcggcc	tcggcctctg	16860
	cataaataaa	aaaaattagt	cagccatggg	gcggagaatg	ggcggaaactg	ggcggagttta	16920
	ggggcgggat	gggcggagtt	aggggcggga	ctatggttgc	tgactaattg	agatgcatgc	16980
	tttgcatact	tctgcctgct	ggggagcctg	ggtgctgact	aattgagatg	catgctttgc	17040
60	atacttctgc	ctgctgggga	gcctggggac	tttccacacc	ctaactgaca	cacattccac	17100
	agctgcctcg	cgcgtttcgg	tgatgacggt	gaaaacctct	gacacatgca	gctcccggag	17160
	acggtcacag	cttgtctgta	agcggatgcc	gggagcagac	aagcccgta	gggcgcgtca	17220

65

[illegible]

<210> 13
<211> 11
<212> DNA
<213> Mus musculus

<400> 13
agtgactgac g 11

<210> 14
<211> 9
<212> DNA
<213> Mus musculus

<400> 14
ttccctgaa 9

<210> 15
<211> 15
<212> DNA
<213> Mus musculus

<400> 15
gcaggaagtg aaagt 15

$\langle 210 \rangle$ 16
 $\langle 211 \rangle$ 7

<212> DNA
 <213> Mus musculus

 5 <400> 16
 ggaaagt 7

 10 <210> 17
 <211> 9
 <212> DNA
 <213> Mus musculus

 15 <400> 17
 tttgggagg 9

 20 <210> 18
 <211> 9
 <212> DNA
 <213> Mus musculus

 25 <400> 18
 caaaatggc 9

 30 <210> 19
 <211> 13
 <212> DNA
 <213> Mus musculus

 35 <400> 19
 agttagccaa tgg 13

 40 <210> 20
 <211> 8
 <212> DNA
 <213> Mus musculus

 45 <400> 20
 tgacgtag 8

 50 <210> 21
 <211> 15
 <212> DNA
 <213> Mus musculus

 55 <400> 21
 aaaaatgaaa gaact 15

 60 <210> 22
 <211> 15
 <212> DNA
 <213> Mus musculus

 65 <400> 22
 ggaagtgaaa gtaat 15

 70 <210> 23
 <211> 12
 <212> DNA
 <213> Mus musculus

 75 <400> 23
 gtgctaataa aa 12

 80 <210> 24
 <211> 10
 <212> DNA
 <213> Mus musculus

<400> 24
 tttcatttt 10

 5 <210> 25
 <211> 6
 <212> DNA
 <213> Mus musculus

 10 <400> 25
 tcagca 6

 <210> 26
 <211> 6
 <212> DNA
 15 <213> Mus musculus

 <400> 26
 catgtg 6

 20 <210> 27
 <211> 9
 <212> DNA
 <213> Mus musculus

 25 <400> 27
 tgatgtaat 9

 <210> 28
 <211> 12
 30 <212> DNA
 <213> Mus musculus

 <400> 28
 ggtgcagcaa tg 12
 35
 <210> 29
 <211> 6
 <212> DNA
 <213> Mus musculus
 40
 <400> 29
 cttctc 6

 <210> 30
 45 <211> 8
 <212> DNA
 <213> Mus musculus .

 <400> 30
 50 atttgcatt 8

 <210> 31
 <211> 6
 <212> DNA
 55 <213> Mus musculus

 <400> 31
 ttcaaa 6

 60 <210> 32
 <211> 10
 <212> DNA
 <213> Mus musculus

 65 <400> 32
 tggcctttcc 10

5 <210> 33
 <211> 8
 <212> DNA
 <213> Mus musculus

 <400> 33
 atttgcac 8

 10 <210> 34
 <211> 10
 <212> DNA
 <213> Mus musculus

 15 <400> 34
 tatgcaaatg 10
 <210> 35
 <211> 15
 <212> DNA
 <213> Mus musculus

 20 <400> 35
 ccgaatatgc aattc 15

 25 <210> 36
 <211> 7
 <212> DNA
 <213> Mus musculus

 30 <400> 36
 aatttcc 7

 <210> 37
 <211> 12
 <212> DNA
 35 <213> Mus musculus

 <400> 37
 gcttttctca aa 12

 40 <210> 38
 <211> 8
 <212> DNA
 <213> Mus musculus

 45 <400> 38
 ctccccct 8

 <210> 39
 <211> 12
 <212> DNA
 50 <213> Mus musculus

 <400> 39
 cattgctcca tc 12

 55 <210> 40
 <211> 15
 <212> DNA
 <213> Mus, musculus

 60 <400> 40
 ctcatgtgca tggtc 15

 65 <210> 41
 <211> 10
 <212> DNA

<213> Mus musculus
 <400> 41
 ggaacaatgt 10
 5
 <210> 42
 <211> 7
 <212> DNA
 <213> Mus musculus
 10
 <400> 42
 ttaagtg 7
 <210> 43
 <211> 10
 <212> DNA
 <213> Mus musculus
 15
 <400> 43
 gggctttcc 10
 20
 <210> 44
 <211> 11
 <212> DNA
 <213> Mus musculus
 25
 <400> 44
 gacaggtggg g 11
 30
 <210> 45
 <211> 11
 <212> DNA
 <213> Mus musculus
 35
 <400> 45
 ggcaggtggg t 11
 40
 <210> 46
 <211> 9
 <212> DNA
 <213> Mus musculus
 45
 <400> 46
 ttcccataa 9
 50
 <210> 47
 <211> 12
 <212> DNA
 <213> Mus musculus
 55
 <400> 47
 tatcattaag gc 12
 60
 <210> 48
 <211> 13
 <212> DNA
 <213> Mus musculus
 65
 <400> 48
 catacaccaa tga 13
 70
 <210> 49
 <211> 6
 <212> DNA
 <213> Mus musculus
 75
 <400> 49

catgtg 6

<210> 50
<211> 6
5 <212> DNA
<213> Mus musculus

<400> 50
catgtg 6

10 <210> 51
<211> 7
<212> DNA
<213> Mus musculus

15 <400> 51
aggtaat 7

20 <210> 52
<211> 14
<212> DNA
<213> Mus musculus

25 <400> 52
aaatgttgca agta 14

<210> 53
<211> 6
30 <212> DNA
<213> Mus musculus

<400> 53
acatgg 6

35 <210> 54
<211> 7
<212> DNA
<213> Mus musculus

40 <400> 54
caggaag 7

<210> 55
<211> 6
45 <212> DNA
<213> Mus musculus

<400> 55
ttcaaa 6

50 <210> 56
<211> 8
<212> DNA
<213> Mus musculus

55 <400> 56
atttgcac 8

REIVINDICACIONES

1. Un vector de expresión que comprende la secuencia de ácido nucleico de una cualquiera de SEQ ID NOs: 11-12.
- 5 2. Una célula huésped que comprende los vectores de expresión de acuerdo con la reivindicación 1.
3. La célula huésped de la reivindicación 2, donde la célula huésped es una célula huésped de mamífero.
- 10 4. La célula huésped de la reivindicación 3 donde la célula huésped es una célula huésped de mieloma murino.
5. Un kit que comprende el vector de la reivindicación 1.

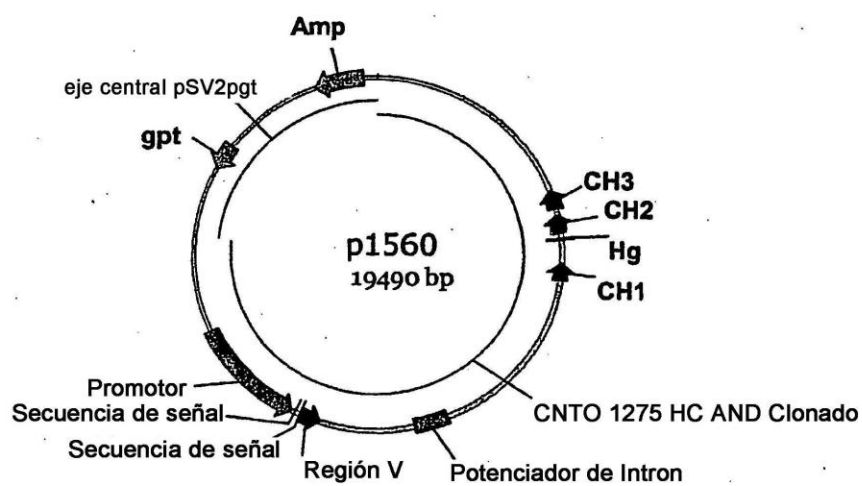


FIG. 1A

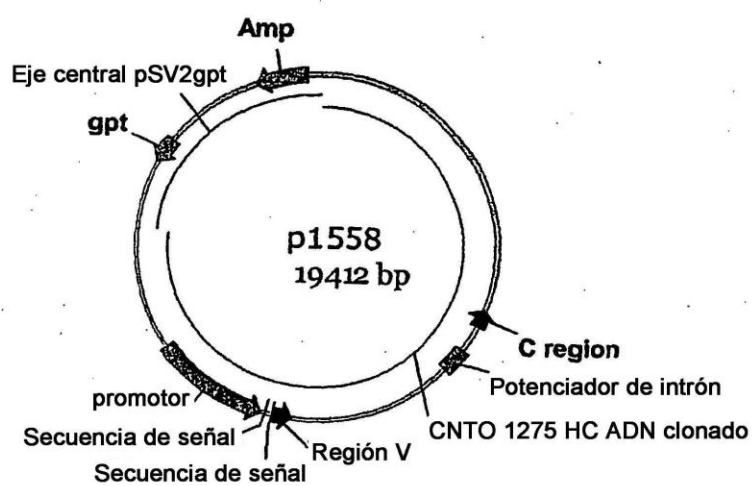


FIG. 1B