



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 309 168**

51 Int. Cl.:
C07K 14/535 (2006.01)
C12N 15/27 (2006.01)
A61K 38/18 (2006.01)
C07K 7/06 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **02727320 .0**
96 Fecha de presentación : **05.02.2002**
97 Número de publicación de la solicitud: **1392731**
97 Fecha de publicación de la solicitud: **03.03.2004**

54 Título: **Factor modificado estimulante de colonias de granulocitos (G-CSF) con inmunogenicidad reducida.**

30 Prioridad: **06.02.2001 EP 01102617**
19.02.2001 EP 01103954

45 Fecha de publicación de la mención BOPI:
16.12.2008

45 Fecha de la publicación del folleto de la patente:
16.12.2008

73 Titular/es: **Merk Patent GmbH**
Frankfurter Strasse 250
64293 Darmstadt, DE

72 Inventor/es: **Carr, Francis J.;**
Carter, Graham;
Jones, Tim y
Williams, Stephen

74 Agente: **Carvajal y Urquijo, Isabel**

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

Factor modificado estimulante de colonias de granulocitos (G-CSF) con inmunogenicidad reducida.

5 **Campo de la invención**

La presente invención se refiere a polipéptidos que se administran especialmente a seres humanos y, en particular, para uso terapéutico. Los polipéptidos son polipéptidos modificados por lo cual la modificación produce una propensión reducida del polipéptido para inducir una respuesta inmune después de su administración al sujeto humano. La invención, en particular, se refiere a la modificación del factor estimulante de colonias de granulocitos (G-CSF) humano para producir variantes de la proteína G-CSF que sean sustancialmente no inmunogénicas o menos inmunogénicas que cualquier polipéptido equivalente no modificado cuando se utiliza *in vivo*. La invención se refiere adicionalmente a péptidos epitópicos de células T derivados de dicha proteína no modificada mediante los cuales es posible crear variantes modificadas del factor estimulante de colonias de granulocitos con inmunogenicidad reducida.

15 **Antecedentes de la invención**

Existen muchas circunstancias en las que la eficacia de una proteína terapéutica se ve limitada por una reacción inmune indeseada contra dicha proteína terapéutica. Varios anticuerpos monoclonales de ratón se han mostrado prometedores como terapias en varios entornos de la enfermedad humana, pero en ciertos casos han fallado debido a la inducción de grados significativos de una respuesta de formación de anticuerpos humanos frente a anticuerpos murinos (HAMA) [Schroff, R. W. y col. (1985) *Cancer Res.* 45: 879-885; Shawler, D.L. y col. (1985) *J. Immunol.* 135: 1530-1535]. En el caso de anticuerpos monoclonales, se han desarrollado varias técnicas en un intento por reducir la respuesta HAMA [documentos WO 89/09622, EP 0239400, EP 0438310 y WO 91/06667]. Estos enfoques de ADN recombinante han reducido generalmente la información genética murina en la construcción del anticuerpo final aumentando al mismo tiempo la información genética humana en dicha construcción final. No obstante, los anticuerpos “humanizados” resultantes, en varios casos, han seguido induciendo una respuesta inmune en los pacientes [Issacs J.D. (1990) *Sem. Immunol.* 2: 449, 456; Rebello, P.R. y col. (1999) *Transplantation* 68: 1417-1420].

Los anticuerpos no son la única clase de molécula polipeptídica administrada como agente terapéutico frente a la que puede organizarse una respuesta inmune. Incluso proteínas de origen humano y con las mismas secuencias de aminoácidos que aparecen en humanos pueden también inducir una respuesta inmune en los mismos. Ejemplos notables incluyen el uso terapéutico del factor estimulante de colonias de macrófagos-granulocitos [Wadhwa, M. y col. (1999) *Clin. Cancer Res.* 5:1353-1361] e interferón alfa 2 [Russo, D. y col. (1996) *Bri. J. Haem.* 94:300-305; Stein, R. y col. (1988) *New Engl. J. Med.* 318:1409-1413] entre otros.

Un factor principal en la inducción de una respuesta inmune es la presencia dentro de la proteína de péptidos que puedan estimular la actividad de células T mediante su presentación asociada a moléculas MHC de clase II, llamados “epítomos de células T”. Estos posibles epítomos de células T se definen normalmente como cualquier secuencia de restos aminoácídicos con la capacidad de unirse a moléculas MHC de clase II. Dichos epítomos de células T pueden medirse para establecer la unión a MHC. De forma implícita, un “epítomo de células T” significa un epítomo que cuando se une a moléculas MHC puede ser reconocido por el receptor de células T (TCR), y que puede, al menos en principio, causar la activación de estas células T acoplándose al TCR y promover una respuesta de dichas células T. Sin embargo, habitualmente se entiende que ciertos péptidos en los que se ha establecido su unión a moléculas MHC de clase II pueden retenerse en una secuencia proteica porque dichos péptidos se reconoce como “propios” dentro del organismo en que se administra la proteína final.

Es sabido que algunos de estos péptidos epitópicos de células T pueden liberarse durante la degradación de péptidos, polipéptidos o proteínas dentro de las células y, posteriormente, ser presentados por moléculas del complejo principal de histocompatibilidad (MHC) para desencadenar la activación de células T. En el caso de péptidos presentados por MHC de clase II, dicha activación de las células T puede dar lugar a continuación, por ejemplo, a una respuesta de anticuerpos mediante estimulación directa de células B para producir dichos anticuerpos.

Las moléculas MHC de clase II son un grupo de proteínas altamente polimórficas que desempeñan una tarea central en la selección y activación de células T cooperativas.

El grupo de antígenos de leucocitos humanos DR (HLA-DR) es el isotipo predominante de este grupo de proteínas y es el objeto principal de la presente invención. Sin embargo, los isotipos HLA-DQ y HLA-DP realizan funciones similares, por tanto la presente invención es igualmente aplicable a estos últimos. La molécula MHC de clase II DR está compuesta por una cadena alfa y una beta que se insertan en sus extremos C-terminales a través de la membrana celular. Cada heterodímero posee un dominio de unión al ligando que se une a péptidos cuya longitud varía entre 9 y 20 aminoácidos de longitud, aunque el surco de unión puede alojar un máximo de 11 aminoácidos. El dominio de unión a ligando está compuesto por los aminoácidos 1 a 85 de la cadena alfa, y los aminoácidos 1 a 94 de la cadena beta. Recientemente se ha demostrado que las moléculas DQ tienen una estructura, homóloga y también se espera que las proteínas de la familia DP sean muy similares. En humanos se conocen aproximadamente 70 alotipos diferentes del isotipo DR, para DQ hay 30 alotipos diferentes y para DP se conocen 47 alotipos diferentes. Cada individuo alberga de dos a cuatro alelos DR, dos alelos DQ y dos DP. Se ha resuelto la estructura de varias moléculas DR y dichas estructuras señalan a un surco de unión a péptidos de extremo abierto con varios bolsillo hidrófobos que acoplan

restos hidrófobos (restos de bolsillo) del péptido [Brown y col. *Nature* (1993) 364: 33; Stem y col. (1994) *Nature* 368: 215]. El polimorfismo que identifica los diferentes alotipos de la molécula de clase II contribuye a una amplia diversidad de diferentes superficies de unión para péptidos dentro del surco de unión a péptidos y a nivel poblacional asegura una flexibilidad máxima con respecto a la capacidad de reconocer proteínas foráneas y organizar una respuesta inmune contra organismos patogénicos. Hay una cantidad considerable de polimorfismo dentro del dominio de unión al ligando con distintas “familias” dentro de diferentes poblaciones geográficas y grupos étnicos. Este polimorfismo afecta a las características de unión del dominio de unión a péptidos, por tanto diferentes “familias” de moléculas DR tendrán especificidades por péptidos con diferentes propiedades de secuencia, aunque puede haber cierto solapamiento. Esta especificidad determina el reconocimiento de epítomos de células Th (respuesta de células T de clase II) que finalmente son responsables de dirigir la respuesta de anticuerpos frente a epítomos de células B presentes en la misma proteína de la que deriva el epítomo de células Th. Por tanto, la respuesta inmune contra una proteína en un individuo está fuertemente influenciada por el reconocimiento del epítomo de células T que es una función de la especificidad de unión a péptidos del alotipo HLA-DR del individuo. Por lo tanto, para identificar epítomos de células T dentro de una proteína o péptido en el contexto de una población global, es deseable considerar las propiedades de unión de un conjunto de alotipos HLA-DR lo más diverso posible, cubriendo de este modo el porcentaje de población mundial más elevado posible.

Una respuesta inmune contra una proteína terapéutica, tal como la proteína que es el objeto de esta invención, procede mediante la ruta de presentación peptídica en el contexto del MHC de clase II. Aquí proteínas exógenas se internalizan y procesan para su presentación en asociación con moléculas MHC de clase II del tipo DR, DQ o DP. Las moléculas MHC de clase II se expresan en las células presentadoras de antígeno (APC) profesionales, como macrófagos y células dendríticas entre otras. El acoplamiento de un complejo peptídico del MHC de clase II con un receptor de células T afín sobre la superficie de la célula T, junto con la unión cruzada de ciertos correceptores diferentes como la molécula CD4, puede inducir un estado activado dentro de la célula T. La activación conduce a la liberación de citocinas que activan adicionalmente a otros linfocitos, como las células B, para producir anticuerpos o a células T asesinas como una respuesta inmune celular completa.

La capacidad de un péptido para unirse a una molécula MHC de clase II dada para la presentación sobre la superficie de una APC depende de varios factores, siendo los más destacable su secuencia primaria. Esto influirá tanto en su propensión a la escisión proteolítica así como también en su afinidad de unión dentro de la hendidura de unión a péptidos de la molécula MHC de clase II. El complejo MHC de clase II/péptido sobre la superficie de la APC presenta una cara de unión con el receptor de células T (TCR) en particular, capaz de reconocer determinantes proporcionados tanto por los restos expuestos del péptido como por la molécula MHC de clase II.

En la técnica existen procedimientos para identificar péptidos sintéticos capaces de unirse a moléculas MHC de clase II (por ejemplo, en los documentos WO98/52976 y WO00/34317). Es posible que dichos péptidos no funcionen como epítomos de células T en todas las situaciones, particularmente, *in vivo* debido a las rutas de procesamiento u otros fenómenos. La identificación de epítomos de células T es la primera etapa para la eliminación de epítomos. Se ha descrito previamente la identificación y eliminación de posibles epítomos de células T a partir de proteínas. En la técnica se han proporcionado métodos para posibilitar la detección de epítomos de células T habitualmente por medios informáticos que analizan motivos de secuencia reconocidos en epítomos de células T determinados experimentalmente o, alternativamente, usando técnicas por ordenador para predecir los péptidos de unión a MHC de clase II y en particular péptidos de unión a DR.

Los documentos WO98/52976 y WO00/34317 muestran enfoques de reconocimiento de plegamiento por ordenador para identificar secuencias polipeptídicas con el potencial para unirse a un subconjunto de alotipos MHC de clase II DR humanos. En estos contenidos, se eliminan epítomos de células T predichos usando una sustitución de aminoácidos lógica dentro de la secuencia primaria del anticuerpo terapéutico o de otra proteína no anticuerpo derivada tanto de humanos como de otras especies.

Se han usado otras técnicas que utilizan los complejos solubles de moléculas MHC recombinantes en combinación con péptidos sintéticos y son capaces de unirse a clones de células T de muestras de sangre periférica de humanos o animales de experimentación en la técnica [Kern, F. y col. (1998) *Nature Medicine* 4:975-978; Kwok, W.W. y col. (2001) *TRENDS in Immunology* 22: 583-588] y también pueden utilizarse en una estrategia de identificación de epítomos.

Como se ha descrito anteriormente y como consecuencia de ello, sería deseable identificar y retirar o al menos reducir los epítomos de células T de un péptido, polipéptido o proteína determinada en principio terapéuticamente valioso pero originalmente inmunogénico.

El G-CSF es una citocina hematopoyética importante actualmente utilizada en el tratamiento de indicaciones donde un aumento en el número de neutrófilos de la sangre proporcionará beneficios. Estas incluyen tratamiento contra el cáncer, diversas enfermedades infecciosas y afecciones relacionadas, como sepsis. El G-CSF también se usa solo o en combinación con otros compuestos y citocinas en la expansión *ex vivo* de células hematopoyéticas para el trasplante de médula ósea.

Normalmente se reconocen dos formas de G-CSF humano para esta citocina. Una es una proteína de 177 aminoácidos, la otra una proteína de 174 aminoácidos [Nagata y col. (1986), *EMBO J.* 5: 575-581], se ha encontrado que la

forma de 174 aminoácidos tiene la mayor actividad biológica específica *in vivo*. La tecnología de ADN recombinante han posibilitado la producción de G-CSF en cantidades a escala comercial utilizando sistemas de expresión de células hospedadoras tanto eucariotas como procariotas.

- 5 La secuencia de aminoácidos del factor estimulante de colonias de granulocitos (G-CSF) humano (representada según el código de una letra) es la siguiente:

10 TPLGPASSLPQSFLKCLEQVRKIQGDGAALQEKLCA₁TYKLCHPEELVLLGHSLGIPWAPLSSCPSQ
ALQLAGCLSQLHSGFLYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQMEELGMAPALQPT
QGAMPAFASAFQRRAGGVLVASHLQSFLEVS₁YRVLRLHLAQP.

- 15 Previamente se han descrito otros análogos polipeptídicos y fragmentos peptídicos de G-CSF, incluyendo formas modificadas por sustituciones de aminoácidos específicas de sitio y/o modificación por productos de adición química. Por tanto el documento US 4.810.643 describe análogos con restos Cys en particular sustituidos por otro aminoácido, y G-CSF con un resto de Ala en la primera posición (N-terminal). El documento EP 0 335 423 describe la modificación de al menos un grupo amino en un polipéptido que tiene actividad G-CSF. El documento EP 0 272 703 describe derivados de G-CSF que tienen un aminoácido sustituido o delecionado cerca del extremo N-terminal. El documento
20 EP 0 459 630 describe derivados de G-CSF en los que la Cys 17 y el Asp 27 se han sustituido por restos de Ser. El documento EP 0 243 153 describe un G-CSF modificado por la inactivación de al menos un sitio de procesamiento de proteasas KEX2 de levaduras para aumentar el rendimiento de la producción recombinante y el documento US 4.904.584 describe proteínas con lisina alterada. El documento WO 90/12874 describe variantes con Cys alterada adicionales y el documento de patente australiana AU-A-10948/92 describe la adición de aminoácidos a cualquier
25 extremo de una molécula G-CSF con el fin de ayudar al plegamiento de la molécula después de la expresión procariota. El documento AU-76380/91, describe variantes de G-CSF en las posiciones 50-56 de la forma de 174 aminoácidos de G-CSF, y en las posiciones 53-59 de la forma de 177 aminoácidos. También se describieron cambios adicionales en restos de His particulares.

- 30 Se entiende que muchos de los enfoques anteriores se han dirigido hacia mejoras en la producción comercial de G-CSF, por ejemplo, la mejora de la estabilidad *in vitro*. Ninguno de estos contenidos reconoce la importancia de los epítomos de células T para las propiedades inmunogénicas de la proteína ni se han concebido para influir directamente sobre dichas propiedades de un modo específico y controlado de acuerdo con el esquema de la presente invención.

- 35 Sin embargo, sigue existiendo la necesidad de análogos del factor estimulante de colonias de granulocitos (G-CSF) con propiedades mejoradas. Las mejoras deseadas incluyen esquemas y modalidades alternativos para la expresión y purificación de dicho agente terapéutico, pero también y especialmente, mejoras en las propiedades biológicas de la proteína. Existe una necesidad especial de mejora de las características *in vivo* cuando se administra a humanos. A este respecto, es muy deseable proporcionar un factor de estimulación de colonias de granulocitos (G-CSF) con potencial
40 reducido o ausente para inducir una respuesta inmune en humanos.

Resumen y descripción de la invención

- 45 La presente invención proporciona formas modificadas del “factor estimulante de colonias de granulocitos (G-CSF)”, en las que la característica inmune está modificada mediante la reducción en el número y la eliminación de posibles epítomos de células T.

- La presente invención proporciona formas modificadas de G-CSF humano con uno o más epítomos de células T eliminados.

- 50 La invención describe secuencias identificadas dentro de la secuencia primaria del G-CSF que son posibles epítomos de células T en virtud de su potencial de unión a MHC de clase II. Esta descripción se aplica específicamente a ambas formas reconocidas de la proteína G-CSF humana que son las especies de 177 y 174 aminoácidos.

- 55 La invención puede aplicarse a cualquier especie de molécula G-CSF con, sustancialmente, las mismas secuencias primarias de aminoácidos que las descritas en este documento e incluirían por lo tanto moléculas G-CSF obtenidas por medios de ingeniería genética u otros procesos y pueden no contener 177 ó 174 restos aminoácidos.

- 60 Las proteínas G-CSF, como las identificadas de fuentes murinas, bovinas, caninas y otros mamíferos, tienen en común muchas de las secuencias peptídicas de la presente descripción y tienen en común muchas secuencias peptídicas con sustancialmente la misma secuencia que las de la lista descrita.

- 65 La invención también describe posiciones específicas dentro de la secuencia primaria de la molécula de acuerdo con la invención que han de ser alteradas por sustitución, adición o deleción de aminoácidos específicos sin que se vea afectada básicamente la actividad biológica. En caso de que puede conseguirse la pérdida de inmunogenicidad solamente mediante una pérdida simultánea de actividad biológica es posible recuperar dicha actividad por alteraciones adicionales dentro de la secuencia de aminoácidos de la proteína.

La invención describe además métodos para producir dichas moléculas modificadas, principalmente, métodos para identificar dichos epítomos de células T que han de ser modificadas para eliminar o reducir el número de sitios inmunogénicos.

Cabría esperar que la proteína de acuerdo con esta invención presentara un aumento del tiempo en circulación en humanos, lo que sería especialmente beneficioso en entornos de enfermedad crónica o recurrente tal como es el caso de varias indicaciones del factor estimulante de colonias de granulocitos (G-CSF). La presente invención proporciona formas modificadas de proteínas G-CSF que se espera muestren propiedades potenciadas *in vivo*. Estas moléculas G-CSF modificadas pueden usarse en composiciones farmacéuticas.

En resumen, la invención se refiere a las siguientes cuestiones:

- una molécula modificada que tiene la actividad biológica del factor estimulante de colonias de granulocitos (G-CSF) humano y que es sustancialmente no inmunogénica o menos inmunogénica que cualquier molécula no modificada que tiene la misma actividad biológica cuando se usa *in vivo*;
- una molécula especificada en consecuencia, en la que dicha pérdida de inmunogenicidad se consigue eliminando uno o más epítomos de células T derivados de la molécula originalmente no modificada;
- una molécula especificada en consecuencia, en la que dicha pérdida de inmunogenicidad se consigue mediante la reducción del número de alotipos MHC capaces de unirse a péptidos derivados de dicha molécula;
- una molécula especificada en consecuencia, en la que se ha eliminado un epítomo de células T;
- una molécula especificada en consecuencia, en la que dichos epítomos de células T originalmente presentes son ligandos de MHC de clase II o secuencias peptídicas que muestran la capacidad de estimular o unirse a células T mediante la presentación en el entorno de clase II;
- una molécula especificada en consecuencia, en la que dichas secuencias peptídicas se seleccionan entre el grupo mostrado en la Tabla 1;
- una molécula especificada en consecuencia, donde se han alterado de 1 a 9 restos aminoacídicos, preferiblemente un resto aminoacídico en cualquiera de los epítomos de células T originalmente presentes;
- una molécula especificada en consecuencia, en la que la alteración de los restos aminoacídicos supone una sustitución, adición o delección de resto(s) aminoacídico(s) originalmente presente(s) por otro(s) resto(s) aminoacídico(s) en la(s) posición(es) específica(s);
- una molécula especificada en consecuencia, en la que una o más de las sustituciones de restos aminoacídicos se realizan como se indica en la Tabla 2;
- una molécula especificada en consecuencia, en la que (adicionalmente) una o más de las sustituciones de restos aminoacídicos se realizan como se indica en la Tabla 3 para la reducción del número de alotipos MHC capaces de unirse a péptidos derivados de dicha molécula;
- una molécula especificada en consecuencia, en la que, si es necesario, adicionalmente se realiza una alteración adicional habitualmente por sustitución, adición o delección de aminoácidos(s) específico(s) para reestablecer la actividad biológica de dicha molécula;
- una secuencia o molécula de ADN que codifica cualquier de las moléculas modificadas especificadas anteriormente y a continuación;
- una composición farmacéutica que comprende una molécula modificada que tiene la actividad biológica del factor estimulante de colonias de granulocitos (G-CSF) como se ha definido anteriormente y/o en las reivindicaciones, opcionalmente junto con un vehículo, diluyente o excipiente farmacéuticamente aceptable;
- un método para obtener una molécula modificada que tiene la actividad biológica del factor estimulante de colonias de granulocitos (G-CSF) como se define en cualquiera de las reivindicaciones citadas anteriormente que comprende las siguientes etapas: (i) determinar la secuencia de aminoácidos del polipéptido o parte del mismo; (ii) identificar uno o más posibles epítomos de células T dentro de la secuencia de aminoácidos de la proteína por cualquier método que incluye la determinación de la unión de los péptidos a moléculas MHC usando técnicas *in vitro* o *in silico* o ensayos biológicos; (iii) diseñar nuevas variantes de secuencia con uno o más aminoácidos dentro de los posible epítomos de células T identificados y modificados de modo que se reduzca sustancialmente o elimine la actividad del epítomo de células T, según se determina mediante la unión de los péptidos a moléculas MHC usando técnicas *in vitro* o *in silico* o ensayos biológicos; (iv) construir dichas variantes de secuencia mediante tecnología de ADN recombinante y probar estas

variantes para identificar una o mas variantes con las propiedades deseadas y (v) opcionalmente, repetir las etapas (ii) y (iv);

- un método especificado en consecuencia, en el que la etapa (iii) se lleva a cabo mediante sustitución, adición o delección de 1 a 9 restos aminoacídicos en cualquiera de los epítomos de células T originalmente presentes;
- un método especificado en consecuencia, en el que la alteración se hace con referencia a una secuencia proteica homóloga y/o técnicas de modelado *in silico*;
- un método especificado en consecuencia, en el que la etapa (ii) anterior se lleva a cabo mediante las siguientes etapas: (a) seleccionar una región del péptido que tiene una secuencia de restos aminoacídicos conocida; (b) muestrear secuencialmente segmentos de restos aminoacídicos solapantes de tamaño uniforme predeterminado y constituidos por al menos tres restos aminoacídicos de la región seleccionada; (c) calcular la puntuación de la capacidad de unión a la molécula MHC de clase II para cada uno de dichos segmentos muestreados sumando los valores asignados para cada cadena lateral de resto aminoacídico hidrófobo presente en dicho segmento de restos aminoacídicos muestreado y (d) identificar al menos uno de dichos segmentos adecuado para la modificación, en función de la puntuación de la capacidad de unión a la molécula MHC de clase II calculado para ese segmento, cambiando el valor de unión a MHC de clase II global para el péptido sin reducir sustancialmente la utilidad terapéutica del mismo; la etapa (c) se lleva a cabo preferiblemente usando una función de puntuación de Böhm modificada para incluir el término de repulsión de energía ligando-proteína de van der Waal 12-6 y el término de energía conformacional (1) proporcionando una primera base de datos de modelos de moléculas MHC de clase II; (2) proporcionando una segunda base de datos de estructuras peptídicas permitidas para dichos modelos de moléculas MHC de clase II; (3) seleccionando un modelo a partir de dicha primera base de datos; (4) seleccionando una estructura peptídica permitida de dicha segunda base de datos; (5) identificando cadenas laterales de restos aminoacídicos presentes en cada segmento muestreado; (6) determinando el valor de afinidad de unión para todas las cadenas laterales presentes en cada segmento muestreado; y repitiendo las etapas (1) a (5) para cada uno de dichos modelos y cada una de dichas estructuras;
- un péptido epitópico de células T de 13 mer que tiene una posible actividad de unión a MHC de clase II y se obtiene a partir del factor estimulante de colonias de granulocitos (G-CSF) inmunogenéticamente no modificado, seleccionado entre el grupo representado en la Tabla 1 y su uso para la producción de G-CSF que tiene sustancialmente ninguna o menos inmunogenicidad que cualquier molécula no modificada con la misma actividad biológica cuando se usa *in vivo*;
- una secuencia peptídica que consta de al menos 9 restos aminoacídicos consecutivos de un péptido epitópico de células T de 13 mer especificado anteriormente y su uso para la producción de G-CSF que tiene sustancialmente ninguna o menos inmunogenicidad en comparación con cualquier molécula no modificada con la misma actividad biológica cuando se usa *in vivo*.

La expresión “epítomo de células T” significa, de acuerdo con la comprensión de esta invención, una secuencia de aminoácidos que es capaz de unirse a MHC II, capaz de estimular a las células T y/o también de unirse (sin una activación necesariamente apreciable) a células T formando complejo con MHC II. El término “péptido” como se usa en este documento y en las reivindicaciones adjuntas, es un compuesto que incluye dos o más aminoácidos. Los aminoácidos están unidos entre sí por un enlace peptídico (definido a continuación en este documento). Se conocen 20 aminoácidos diferentes de origen natural implicados en la producción biológica de péptidos, y pueden unirse en cualquier cantidad y en cualquier orden para formar una cadena o anillo peptídico. Los aminoácidos de origen natural empleados en la producción biológica de péptidos tiene todos la configuración L. Pueden prepararse péptidos sintéticos empleando métodos convencionales, utilizando aminoácidos L, aminoácidos D o diversas combinaciones de aminoácidos de las dos configuraciones diferentes. Algunos péptidos contienen solamente unas pocas unidades de aminoácidos. Los péptidos cortos, por ejemplo, que tienen menos de diez unidades de aminoácidos, se denominan en ocasiones “oligopéptidos”. Otros péptidos contiene una gran cantidad de restos aminoacídicos, por ejemplo hasta 100 o más, y se denominan “polipéptidos”. Por convencionalismos, un “polipéptido” puede considerarse como una cadena peptídica que contiene tres o más aminoácidos, mientras que un “oligopéptido” habitualmente se considera como un tipo particular de polipéptido “corto”. Por tanto, como se usa en este documento, se entiende que cualquier referencia a un “polipéptido” también incluye un oligopéptido. Además, cualquier referencia a un “péptido” incluye polipéptidos, oligopéptidos, y proteínas. Cada ordenamiento diferente de aminoácidos da lugar a polipéptidos o proteínas diferentes. La cantidad de polipéptidos (y por tanto la cantidad de diferentes proteínas) que puede formarse es prácticamente ilimitada. El “carbono alfa (C α)” es el átomo de carbono del componente carbono-hidrógeno (CH) que está en la cadena peptídica. Una “cadena lateral” es un grupo colgante del C α que puede comprender un grupo o resto simple o complejo, con dimensiones físicas que pueden variar significativamente en comparación con las dimensiones del péptido.

La invención puede aplicarse a cualquier especie de molécula G-CSF con sustancialmente con las mismas secuencias de aminoácidos primarias que las descritas en este documento e incluirían, por lo tanto, moléculas G-CSF obtenidas mediante tecnología de ingeniería genética u otros procesos y pueden no contener ni 177 ni 174 restos aminoacídicos.

Las proteínas del factor estimulante de colonias de granulocitos (G-CSF), como las identificadas a partir de otras fuentes de mamíferos tienen en común muchas de las secuencias peptídicas de la presente descripción así como muchas secuencias peptídicas sustancialmente con las mismas secuencias que las descritas en la lista.

5 La invención está concebida para superar la realidad práctica de que proteínas solubles introducidas en organismos autólogos pueden desencadenar una respuesta inmune que da lugar al desarrollo de anticuerpos en el hospedador que se unen a la proteína soluble. Un ejemplo entre otros, es el interferón alfa 2 contra el que una proporción de pacientes humanos sintetizan anticuerpos a pesar del hecho de que esta proteína se produce de forma endógena [Russo, D. y col. (1996) *ibid*; Stein, R. y col. (1988) *ibid*]. Es probable que la misma situación se aplique al uso terapéutico del factor
10 estimulante de colonias de granulocitos (G-CSF) y la presente invención busca abordar este hecho proporcionando proteínas con función de factor estimulante de colonias de granulocitos (G-CSF) con propensión alterada a inducir una respuesta inmune tras su administración al hospedador humano.

15 El método general de la presente invención que conduce al factor estimulante de colonias de granulocitos (G-CSF) modificado comprende las siguientes etapas:

(a) determinar la secuencia de aminoácidos del polipéptido o parte del mismo;

20 (b) identificar uno o más posibles epítomos de células T dentro de la secuencia de aminoácidos de la proteína por cualquier método que incluya la determinación de la unión de los péptidos a moléculas MHC usando técnicas *in vitro* o *in silico* o ensayos biológicos;

(c) diseñar nuevas variantes de secuencia con uno o más aminoácidos dentro de los posibles epítomos de células T
25 identificados modificados de tal modo que se reduzca sustancialmente o elimine la actividad del epítomo de células T determinada por la unión de los péptidos a moléculas MHC usando técnicas *in vitro* o *in silico* o ensayos biológicos. Dichas variantes de secuencia se obtienen modo que se evite la formación de nuevos posibles epítomos de células T por las variaciones de secuencia salvo que dichos nuevos epítomos potenciales de células T se modifiquen, a su vez, de modo que se reduzca sustancialmente o se elimine la actividad del epítomo de células T y
30

(d) construir dichas variantes de secuencia mediante tecnología de ADN recombinante y probar dichas variantes para identificar una o más variantes con las propiedades deseables de acuerdo con técnicas recombinantes bien conocidas.

35 La identificación de posibles epítomos de células T de acuerdo con la etapa (b) puede realizarse según métodos descritos previamente en la técnica anterior. Los métodos adecuados se describen en los documentos WO 98/59244, WO 98/52976 y WO 00/34317, y pueden usarse preferiblemente para identificar la propensión de los péptidos derivados del factor estimulante de colonias de granulocitos (G-CSF) para unirse a una molécula MHC de clase II.

40 Otro método muy eficaz para identificar epítomos de células T mediante ordenador se describe en el Ejemplo que es una realización preferida de acuerdo con esta invención.

45 En la práctica se obtendrán varias proteínas variantes del factor estimulante de colonias de granulocitos (G-CSF) y se comprobará si presentan las características inmunes y funcionales deseadas. Las proteínas variantes se producirán más preferiblemente mediante tecnología de ADN recombinante aunque pueden contemplarse otros procedimientos que incluyen la síntesis química de fragmentos del factor estimulante de colonias de granulocitos (G-CSF).

50 En la Tabla 1 se presentan los resultados de un análisis según la etapa (b) del esquema anterior y que se corresponden con las secuencias de la proteína G-CSF humana completa de las formas tanto 174 como 177.

55

60

65

TABLA 1

Secuencias peptídicas en el factor estimulante de colonias de granulocitos humano (G-CSF) con posible actividad de unión a MHC de clase II humano

5	TPLGPASSLPQSF,	SSLPQSFLKCLE,	QSFLKCLEQVRK,	SFLKCLEQVRKI,
	FLLKCLEQVRKIQ,	KCLEQVRKIQGDG,	EQVRKIQGDGAAL,	RKIQGDGAALQEK,
10	AALQEKLVSSECAT,	EKLVSSECATYKLC,	KLVSECATYKLCH,	AALQEKLVSSECATYKL,
	EKLVSSECATYKLCHPE,	ATYKLCHPEELVL,	YKLCHPEELVLLG,	EELVLLGHSLGIP,
	ELVLLGHSLGIPW,	HSLGIPWAPLSSC,	IPWAPLSSCPSQA,	APLSSCPSQALQL,
15	QALQLAGCLSQLH,	GCLSQLHSGFLY,	SQLHSGFLYQGL,	SGLFLYQGLLQAL,
	GLFLYQGLLQALE,	LFLYQGLLQALEG,	FLYQGLLQALEGI,	QGLLQALEGISPE,
	GLLQALEGISPEL,	QALEGISPELGPT,	EGISPELGPTLDT,	PTLDTLQLDVADF,
20	DTLQLDVADFATT,	LQLDVADFATTIW,	LDVADFATTIWQQ,	TTIWQQMEELGMA,
	TIWQQMEELGMAP,	QQMEELGMAPALQ,	EELGMAPALQPTQ,	LGMALQPTQGA,
	PALQPTQGAMPAF,	GAMPAFASAFQRR,	PAFASAFQRRAGG,	SAFQRRAGGVLA,
25	GGVLVASHLQSF,	GVLVASHLQSFLE,	VLVASHLQSFLEV,	SHLQSFLEVSYRV,
	QSFLEVSYRVLRH,	SFLEVSYRVLRHL,	LEVSYRVLRHLAQ	

30 Las secuencias peptídicas incluyen péptidos identificados en las formas tanto de 177 como de 174 aminoácidos de G-CSF. Los péptidos son de 13 mer, los aminoácidos se identifican usando el código de una letra.

35 En las Tablas 2 y 3 se presentan los resultados del diseño y las construcciones según las etapas (c) y (d) del esquema anterior y que se corresponden con la molécula modificada de esta invención.

40 (Tabla pasa a página siguiente)

45

50

55

60

65

ES 2 309 168 T3

TABLA 2

Sustituciones que llevan a la eliminación de posibles epítomos de células T del factor estimulante de colonias de granulocitos (G-CSF) (TS = tipo salvaje)

5

10

15

20

25

30

35

40

45

50

55

60

65

Nº de resto	Resto TS	Sustitución											
3	L	A	C	D	E	G	H	K	N	P	O	R	S
9	L	A	C	D	E	G	H	K	N	P	O	R	S
14	L	A	C	D	E	G	H	K	N	P	O	R	S
15	L	A	C	D	E	G	H	K	N	P	O	R	S
18	L	A	C	D	E	G	H	K	N	P	O	R	S
21	V	A	C	D	E	G	H	K	N	P	O	R	S
24	I	A	C	D	E	G	H	K	N	P	O	R	S
31	L	A	C	D	E	G	H	K	N	P	O	R	S
35	L	A	C	D	E	G	H	K	N	P	O	R	S
39	Y	A	C	D	E	G	H	K	N	P	O	R	S
41	L	A	C	D	E	G	H	K	N	P	O	R	S
47	L	A	C	D	E	G	H	K	N	P	O	R	S
48	V	A	C	D	E	G	H	K	N	P	O	R	S
49	L	A	C	D	E	G	H	K	N	P	O	R	S
50	L	A	C	D	E	G	H	K	N	P	O	R	S
54	L	A	C	D	E	G	H	K	N	P	O	R	S
56	I	A	C	D	E	G	H	K	N	P	O	R	S
58	W	A	C	D	E	G	H	K	N	P	O	R	S
61	L	A	C	D	E	G	H	K	N	P	O	R	S
69	L	A	C	D	E	G	H	K	N	P	O	R	S
71	L	A	C	D	E	G	H	K	N	P	O	R	S
75	L	A	C	D	E	G	H	K	N	P	O	R	S
78	L	A	C	D	E	G	H	K	N	P	O	R	S
82	L	A	C	D	E	G	H	K	N	P	O	R	S
83	F	A	C	D	E	G	H	K	N	P	O	R	S
84	L	A	C	D	E	G	H	K	N	P	O	R	S
85	Y	A	C	D	E	G	H	K	N	P	O	R	S
88	L	A	C	D	E	G	H	K	N	P	O	R	S
89	L	A	C	D	E	G	H	K	N	P	O	R	S
92	L	A	C	D	E	G	H	K	N	P	O	R	S
95	I	A	C	D	E	G	H	K	N	P	O	R	S
99	L	A	C	D	E	G	H	K	N	P	O	R	S
103	L	A	C	D	E	G	H	K	N	P	O	R	S
106	L	A	C	D	E	G	H	K	N	P	O	R	S
108	L	A	C	D	E	G	H	K	N	P	O	R	S
110	V	A	C	D	E	G	H	K	N	P	O	R	S
113	F	A	C	D	E	G	H	K	N	P	O	R	S
117	I	A	C	D	E	G	H	K	N	P	O	R	S
118	W	A	C	D	E	G	H	K	N	P	O	R	S
121	M	A	C	D	E	G	H	K	N	P	O	R	S
124	L	A	C	D	E	G	H	K	N	P	O	R	S
130	L	A	C	D	E	G	H	K	N	P	O	R	S
137	M	A	C	D	E	G	H	K	N	P	O	R	S
140	F	A	C	D	E	G	H	K	N	P	O	R	S
144	F	A	C	D	E	G	H	K	N	P	O	R	S
151	V	A	C	D	E	G	H	K	N	P	O	R	S
152	L	A	C	D	E	G	H	K	N	P	O	R	S
153	V	A	C	D	E	G	H	K	N	P	O	R	S
157	L	A	C	D	E	G	H	K	N	P	O	R	S
160	F	A	C	D	E	G	H	K	N	P	O	R	S
161	L	A	C	D	E	G	H	K	N	P	O	R	S
163	V	A	C	D	E	G	H	K	N	P	O	R	S

TABLA 3

Sustituciones adicionales que llevan a la eliminación de un posible epítipo de células T para 1 o más alotipos MHC

	Nº de resto	Resto TS		Sustitución															
	6	A	H	P	O	R	S	T											
	8	L	H	L															
10	14	L	F	I	M	V	W	Y											
	16	K	A	C	F	G	I	L	M	P	V	W	Y						
	17	C	D	E	H	K	N	P	O	R	S	T							
	18	L	F	I	M	V	W	Y											
	19	E	A	C	G	L	M	V	W	Y									
15	21	V	M	W	Y														
	22	R	H	P	O	S	T												
	24	I	W	Y															
	29	A	D	E	G	H	K	N	P	O	R	T							
	30	A	H	K	P	O	R	S	T										
20	31	L	F	I	M	V	W	Y											
	32	O	A	C	G	I	L	M	P	V	W	Y							
	33	E	A	C	G	H	P												
	34	K	H	P	T	W													
	35	L	I	M	W	Y													
25	39	Y	M	W															
	40	K	A	C	F	G	I	L	M	P	V	W	Y						
	41	L	M	W	Y														
	42	C	H	P	T														
	43	H	A	G	P														
30	45	E	P	T															
	47	L	F	I	M	V	W	Y											
	48	V	M	W	Y														
	49	L	F	I	W														
	50	L	F	M															
35	51	G	P	T	W														
	53	S	P	Y															
	54	L	F	I	M	V	W	Y											
	61	L	F	I	M	W	Y												
40	62	S	A	C	F	G	I	L	M	P	V	W	Y						
	63	S	A	C	G	P													
	64	C	D	E	H	I	K	N	P	O	R	S	T	W	Y				
	66	S	D	H	P														
	67	O	P	T															
45	69	L	W																
	71	L	F	I	M	V	W	Y											
	74	L	H	P															
	75	L	F	I	M	V	W	Y											
	76	S	A	C	G	P													
50	77	O	A	C	F	G	L	P	M	P	V	W	Y						
	79	H	A	C	G	P	W	Y											
	80	S	F	H	P	W	Y												
	81	G	E	H	P	O	R	S	T										
	82	L	F	I	M	V	W	Y											
55	86	O	A	C	G	P													
	87	G	F	I	L	P	W												
	90	O	F	I	P	T	W	Y											
	92	L	W	Y															
	98	E	P	T															
60	99	L	I	M	V	W	Y												
	100	G	H	P	T														
	103	L	M	W	Y														
	106	L	F	I															
	108	L	F	I	M	V	W												
65	110	V	M	W															
	112	D	A	C	G	P													
	113	F	M	W															
	117	I	M	W															

	Nº de resto	Resto TS	Sustitución													
5	119	O	A	C	F	G	I	L	M	P	V	W	Y			
	120	O	I	P	T											
	122	E	P													
	123	E	H	P												
	124	L	F	I	M	W	Y									
10	127	A	P													
	129	A	D	E	H	N	P	Q	R	S	T	W				
	130	L	F	I	M	V	W	Y								
	131	O	A	C	G	P										
15	134	O	A	C	G	P										
	135	G	D	H	N	P	Q	R	S	T						
	136	A	P													
	142	S	A	C	P											
	149	G	D	H	P											
20	150	G	P													
	151	V	I	M	W	Y										
	152	L	I	M	V	W	Y									
	153	V	F	M	W	Y										
	154	A	P													
25	155	S	P													
	156	H	P													
	159	S	F	H	I	P	T	V	W	Y						
	161	L	F	I	M	V	W	Y								
30	162	E	A	C	G	P										
	163	V	F	I	M	W	Y									
	164	S	A	C	G	M	P	T								
	166	R	F	P	T	Y										
	167	V	P													
35	168	L	F	H	I	K	M	N	P	Q	R	S	T	V	W	Y
	169	R	P	T												
	170	H	A	C	F	G	I	L	M	P	V	W	Y			
	171	I	A	C	D	E	F	G	H	I	K	L	M	N	P	Q
40	171ctd		R	S	T	V	W	Y								

La invención se refiere a análogos del factor estimulante de colonias de granulocitos (G-CSF) humano en los que se han hecho sustituciones de al menos un resto aminoacídico en la posición 153 que da lugar a una reducción sustancial en la actividad o a la eliminación de uno o más posibles epítomos de células T de la proteína. Una o más sustituciones de aminoácidos en puntos particulares dentro de cualquier posible ligando de MHC de clase II identificados en la Tabla 1 puede dar lugar a una molécula del factor estimulante de colonias de granulocitos (G-CSF) con un potencial inmunogénico reducido cuando se administra como compuesto terapéutico al hospedador humano. Preferiblemente, las sustituciones de aminoácidos se hacen en puntos apropiados dentro de la secuencia peptídica predicha para conseguir una reducción sustancial o la eliminación de la actividad del epítomo de células T. En la práctica un punto apropiado se equiparará preferiblemente con un resto aminoacídico que se une dentro de uno de los bolsillos hidrófobos que se encuentran dentro del surco de unión de MHC de clase II.

Es más preferido alterar la unión dentro del primer bolsillo de la hendidura en la llamada P1 o de anclaje del péptido P1. La calidad de la interacción de unión entre el resto de anclaje P1 del péptido y el primer bolsillo del surco de unión de MHC de clase II está reconocida como un determinante principal de la afinidad de unión global para el péptido completo. Una sustitución apropiada en esta posición del péptido será aquella que se realice por un resto que se acomoda menos fácilmente dentro del bolsillo, por ejemplo, la sustitución por un resto más hidrófilo. También se consideran restos aminoacídicos del péptido en posiciones que equiparan la unión dentro de otras regiones del bolsillo de la hendidura de unión de MHC y pertenecen al alcance de la presente invención.

Se entiende que sustituciones de un único aminoácido dentro de un posible epítomo de células T determinado son la vía más preferida por la que puede eliminarse dicho epítomo. Pueden contemplarse combinaciones de sustituciones dentro de un único epítomo y, por ejemplo, puede ser particularmente apropiado cuando epítomos definidos individualmente se solapan entre sí. Además, pueden hacerse sustituciones de aminoácidos tanto individuales dentro de un epítomo dado como en combinación dentro de un único epítomo en posiciones no equiparables a los "restos de bolsillo" con respecto al surco de unión de MHC de clase II, pero en cualquier punto dentro de la secuencia peptídica. Pueden hacerse sustituciones con referencia a una estructura homóloga o método estructural producido usando técnicas *in*

silico conocidas en la técnica que pueden basarse en características estructurales conocidas de la molécula de acuerdo con esta invención. Todas estas sustituciones están dentro del alcance de la presente invención.

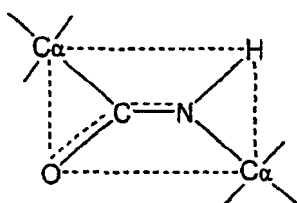
Pueden contemplarse sustituciones de aminoácidos diferentes a las de los péptidos identificados anteriormente particularmente cuando se hacen en combinación con sustituciones hechas dentro de un péptido enumerado. Por ejemplo puede contemplarse un cambio que reestablece la estructura o actividad biológica de la molécula variante. Dichos cambios compensatorios y cambios para incluir una delección o adición de restos aminoácidos en particular del polipéptido del factor estimulante de colonias de granulocitos (G-CSF) que produce una variante con actividad deseada y en combinación con cambios en cualquiera de los péptidos descritos pertenecen al alcance de la presente invención.

En cuanto a lo que esta invención se refiere a composiciones del factor estimulante de colonias de granulocitos (G-CSF) modificado, las composiciones que contienen dichas proteínas G-CSF modificadas deben considerarse dentro del alcance de la invención. En otro aspecto, la presente invención se refiere a ácidos nucleicos que codifican entidades del factor estimulante de colonias de granulocitos (G-CSF) modificado. En un aspecto adicional, la presente invención se refiere a métodos para el tratamiento terapéutico de seres humanos usando las proteínas G-CSF modificadas.

Ejemplo

Existen varios factores que desempeñan tareas importantes en la determinación de la estructura total de una proteína o polipéptido. En primer lugar, el enlace peptídico, es decir, ese enlace que une los aminoácidos en una cadena, es un enlace covalente. Este enlace es de estructura plana, esencialmente una amida sustituida. Una “amida” es cualquier compuesto de un grupo de compuestos orgánicos que contienen el grupo -CONH-.

El enlace peptídico plano que une $C\alpha$ de aminoácidos adyacentes puede representarse como se muestra a continuación:



Como los átomos O=C y C-N están en un plano relativamente rígido, no existe rotación libre alrededor de estos ejes. Por tanto, el plano representado esquemáticamente por la línea discontinua se denomina en ocasiones como un plano “amida” o “plano peptídico” donde están los átomos de oxígeno (O), carbono (C), nitrógeno (N) e hidrógeno (H) de la estructura peptídica. En esquina opuestas de este plano amida se localizan los átomos $C\alpha$. Como no hay rotación sustancialmente alrededor de los átomos O=C y C-N en el plano peptídico o amida, una cadena polipeptídica comprende, por tanto, una serie de enlaces peptídicos planos que unen los átomos $C\alpha$.

Un segundo factor que desempeña una tarea importante en la definición de la estructura total o conformación de un polipéptido o proteína es el ángulo de rotación de cada plano amida alrededor del enlace común $C\alpha$. Las expresiones “ángulo de rotación” y “ángulo de torsión” se consideran, a partir de ahora en este documento, expresiones equivalentes. Suponiendo que los átomos O, C, N, y H permanecen en el plano amida (que habitualmente es una suposición válida, aunque puede haber ligeras desviaciones de la planaridad de estos átomos en algunas conformaciones), estos ángulos de rotación definen la conformación estructural del polipéptido N y R, es decir, la estructura existente entre restos adyacentes. Estos dos ángulos se conocen como ϕ y ψ . El conjunto de los ángulo ϕ_1 y ψ_1 , donde el subíndice 1 representa un resto en particular de una cadena polipeptídica, define por tanto, de forma eficaz, la estructura secundaria del polipéptido. En la bibliografía se recogen convencionalismos usados para definir los ángulo ϕ y ψ , es decir, los puntos de referencia en que los planos amida forman un ángulo de cero grados, y la definición de qué ángulo es ϕ y cual es ψ , para un polipéptido dado. Véase, por ejemplo, Ramachandran y col. *Adv. Prot. Chem.* 23:283-437 (1968), en las páginas 285-94, dichas páginas se incorporan en este documento por referencia. El presente método puede aplicarse a cualquier proteína, y se basa en parte en el descubrimiento de que en seres humanos la posición de anclaje del Bolsillo 1 principal de los surcos de unión de la molécula MHC de clase II tiene una especificidad bien diseñada para cadenas laterales de aminoácidos particulares. La especificidad de este bolsillo está determinada por la identidad de los aminoácidos en la posición 86 de la cadena beta de la molécula MHC de clase II. Este sitio está localizado en la parte inferior del Bolsillo 1 y determina el tamaño de la cadena lateral que puede acomodarse en este bolsillo. Marshall, K. W., *J. Immunol.*, 152: 4946-4956 (1994). Si este resto es una glicina, entonces todos los aminoácidos hidrófobos alifáticos y aromáticos (siendo los hidrófobos alifáticos: valina, leucina, isoleucina, metionina y siendo los aromáticos: fenilalanina, tirosina y triptófano) pueden acomodarse en el bolsillo, prefiriéndose las cadenas laterales aromáticas. Si este resto de bolsillo es una valina, entonces la cadena lateral de este aminoácido sobresale del bolsillo y restringe el tamaño de las cadenas laterales del péptido que pueden acomodarse de modo que sólo pueden hacerlo cadenas laterales hidrófobas alifáticas. Por lo tanto, en una secuencia de restos aminoácidos, independientemente

de si se encuentra un aminoácido con una cadena lateral hidrófoba alifática o aromática, existe la posibilidad de que esté presente un epítipo de células T restringido a un MHC de clase II. Si la cadena lateral es hidrófoba alifática, sin embargo, es aproximadamente dos veces más probable que esté asociado con un epítipo de células T que si fuera una cadena lateral aromática (suponiendo una distribución aproximadamente uniforme de los tipos de Bolsillo 1 en toda la población global).

Un método por ordenador que abarca la presente invención esboza la probabilidad de que regiones peptídicas contengan epítopos de células T del siguiente modo:

(1) Se analiza la secuencia primaria de un segmento peptídico de longitud predeterminada, y se identifican todas las cadenas laterales hidrófobas alifáticas y aromáticas presentes. (2) Se asigna a las cadenas laterales hidrófobas alifáticas un valor mayor que a las cadenas laterales aromáticas; preferiblemente aproximadamente dos veces el valor asignado a las cadenas laterales aromáticas, por ejemplo, un valor de 2 para una cadena lateral hidrófoba alifática y un valor de 1 para una cadena lateral aromática. (3) Se suman los valores determinados presentes en cada segmento de restos aminoácido solapantes (ventana) de longitud uniforme predeterminada dentro del péptido, y se asigna el valor total para un segmento particular (ventana) a un único resto aminoácidos en una posición intermedia del segmento (ventana), preferiblemente a un resto situado aproximadamente en el punto medio del segmento muestreado (ventana). Este procedimiento se repite para cada segmento de restos aminoácidos solapantes muestreado (ventana). Por tanto, a cada resto aminoácido del péptido se le asigna un valor que hace referencia a la probabilidad de que esté presente un epítipo de células T en ese segmento particular (ventana). (4) Los valores calculados y asignados como se ha descrito en la Etapa 3, anteriormente, pueden representarse frente a las coordenadas de los aminoácidos de la secuencia de restos aminoácidos completa que se está evaluando. (5) Se considera que todas las partes de la secuencia que tienen una puntuación de un valor predeterminado, por ejemplo, un valor de 1, contienen probablemente un epítipo de células T y puede modificarse, si se desea. Este aspecto particular de la presente invención proporciona un método general por el que pueden describirse las regiones peptídicas que probablemente contienen epítopos de células T. Las modificaciones del péptido en estas regiones tienen la posibilidad de modificar las características de unión a MHC de clase II.

De acuerdo con otro aspecto de la presente invención, pueden predecirse epítopos de células T con mayor exactitud usando un método por ordenador más sofisticado que tiene en cuenta las interacciones de los péptidos con modelos de alelos MHC de clase II. La predicción por ordenador de epítopos de células T presentes en un péptido de acuerdo con este aspecto particular contempla la construcción de modelos de al menos 42 alelos MHC de clase II en base a las estructuras de todas las moléculas MHC de clase II conocidas y un método para el uso de estos modelos en la identificación por ordenador de epítopos de células T, la construcción de bibliotecas de estructuras peptídicas para cada modelo que permitan obtener la variabilidad conocida en las posiciones relativas del carbono alfa (α) de la estructura peptídica, la construcción de bibliotecas de conformaciones de cadenas laterales de aminoácidos para cada acoplamiento estructural con cada modelo para cada uno de los 20 aminoácidos alternativos en las posiciones críticas de interacción entre el péptido y la molécula MHC de clase II y el uso de estas bibliotecas de estructuras y conformaciones de cadenas laterales en conjunto con una función de valoración para seleccionar la estructura óptima y la conformación de cadena lateral de un péptido particular acoplado con una particular molécula MHC de clase II en particular y la obtención de un valor de unión a partir de esta interacción.

Los modelos de moléculas MHC de clase II pueden obtenerse mediante modelado de homología a partir de varias estructuras similares encontradas en el Banco de datos de proteínas de Brookhaven ("BDP"). Estos pueden hacerse usando un software de modelado de homología semiautomático (Modeller, Sali A. & Blundell TL., 1993, *J. Mol Biol* 234: 779-815) que incorpora una función de hibridación simulada, en conjunto con el campo de fuerzas CHARMM para la minimización de la energía (disponible en Molecular Simulations Inc., San Diego, Ca.). También pueden utilizarse métodos de modelado alternativos.

El presente método difiere significativamente de otros métodos por ordenador que utilizan bibliotecas de datos de unión obtenidos experimentalmente de cada aminoácido alternativo en cada posición en el surco de unión para un pequeño conjunto de moléculas MHC de clase II (Marshall, K.W., y col., *Biomed. Pept. Proteins Nucleic Acids*, 1(3): 157-162) (1995) o incluso otros métodos por ordenador que utilizan datos de unión experimentales similares para definir las características de unión de tipos particulares de bolsillos de unión dentro del surco, usando de nuevo un subconjunto relativamente pequeño de moléculas MHC de clase II, y después "mezclando y ajustando" los tipos de bolsillo de esta biblioteca de bolsillo para crear artificialmente moléculas MHC de clase II "virtuales" adicionales (Sturniolo T., y col., *Nat. Biotech.*, 17(6): 555-561 (1999)). Ambos métodos anteriores presentan la desventaja principal de que, debido a la complejidad de los ensayos y a la necesidad de sintetizar grandes cantidades de variantes peptídicas, solamente puede analizarse experimentalmente una pequeña cantidad de moléculas MHC de clase II. Por lo tanto, el primer método anterior solamente puede hacer predicciones para una pequeña cantidad de moléculas MHC de clase II. El segundo método anterior también parte de la suposición de que un bolsillo revestido con aminoácidos similares en una molécula tendrá las mismas características de unión cuando esté en el contexto de un alelo de Clase II diferente y presentan desventajas adicionales porque solamente pueden crearse "virtualmente" aquellas moléculas MHC de clase II que contienen bolsillos incluidos en la biblioteca de bolsillos. Usando el enfoque de modelado descrito en este documento, puede deducirse la estructura de cualquier número y tipo de moléculas MHC de clase II, por lo tanto, pueden seleccionarse específicamente alelos que sean representativos de la población global. Además, puede aumentarse la cantidad de moléculas MHC de clase II analizadas creando modelos adicionales en lugar de tener que generar datos adicionales mediante experimentación compleja.

El uso de una biblioteca estructural permite la variación en las posiciones de los átomos $C\alpha$ de los diversos péptidos que se están analizando cuando se acoplan con moléculas MHC de clase II particulares. Esto está de nuevo en contraste con los métodos por ordenador previos alternativos descritos anteriormente que dependen del uso de estructuras peptídicas simplificadas para el análisis de la unión de aminoácidos en bolsillos particulares. Estas estructuras simplificadas probablemente no son representativas de conformaciones estructurales encontradas en péptidos “reales”, lo que conduce a imprecisiones en la predicción de la unión del péptido. La presente biblioteca estructural se crea superponiendo las estructuras de todos los péptidos unidos a moléculas MHC de clase II encontrados en el Banco de datos de proteínas y anotando la desviación de la media cuadrática (RMS, por sus siglas en inglés) entre los átomos $C\alpha$ de cada uno de los once aminoácidos localizados dentro del surco de unión. Aunque esta biblioteca puede obtenerse a partir de una pequeña cantidad de estructuras murinas y humanas disponibles adecuadas (actualmente 13), para permitir la posibilidad de una variabilidad incluso mayor, la cifra RMS para cada posición $C\alpha$ se aumenta en un 50%. Después se determina la posición $C\alpha$ promedio para cada aminoácido y se traza una esfera alrededor de este punto cuyo radio es igual a la desviación RMS en esa posición más el 50%. Esta esfera representa todas las posiciones $C\alpha$ permitidas. Trabajando a partir del $C\alpha$ con la desviación RMS mínima (la del aminoácido en el Bolsillo 1 mencionado anteriormente, equivalente a la Posición 2 de los 11 restos en el surco de unión), la esfera se cuadrícula tridimensionalmente, y cada vértice dentro de la cuadrícula se usa a continuación como una posible localización para un $C\alpha$ de ese aminoácido. El plano amida posterior, correspondiente al enlace peptídico con el aminoácido siguiente se inserta en cada uno de estos $C\alpha$ y se rotan los ángulos ϕ y ψ por etapas a intervalos establecidos para colocar el $C\alpha$ posterior. Si el $C\alpha$ posterior está dentro de la “esfera de posiciones permitidas” para este $C\alpha$ entonces se acepta la orientación del dipéptido, mientras que si está fuera de la esfera, entonces se rechaza el dipéptido. A continuación se repite este proceso para cada uno de las posiciones $C\alpha$ posteriores, de modo que el péptido crece desde la “semilla” del $C\alpha$ del Bolsillo 1, hasta que se han colocado los nueve $C\alpha$ posteriores en todas las posibles permutaciones de los $C\alpha$ precedentes. Después, se repite el proceso una vez más para el único $C\alpha$ que precede al bolsillo 1 para crear una biblioteca de posiciones $C\alpha$ estructurales localizadas dentro del surco de unión. La cantidad de estructuras generada depende de varios factores: el tamaño de las “esferas de posiciones permitidas”; la finura del cuadrículado de la “esfera principal” en la posición del Bolsillo 1, la finura de la rotación por etapas de los ángulos ϕ y ψ usados para colocar los $C\alpha$ posteriores. Usando este proceso, puede crearse una gran biblioteca de estructuras. Cuanto mayor es la biblioteca estructural, más probable será que se encuentre la adaptación óptima para un péptido particular dentro del surco de unión de una molécula MHC de clase II. Puesto que todas las estructuras no serán adecuadas para acoplarse a todos los modelos de moléculas MHC de clase II debido a choques con aminoácidos de los dominios de unión, se crean subconjuntos de la biblioteca para cada alelo que comprenden estructuras que pueden acomodar a ese alelo. El uso de la biblioteca estructural, junto con los modelos de moléculas MHC de clase II crea una base de datos exhaustiva que consta de conformaciones de cadena lateral permitidas para cada aminoácido en cada posición del surco de unión para cada molécula MHC de clase II acoplada con cada estructura permitida. Este conjunto de datos se genera usando una función de solapamiento estérico simple donde se acopla una molécula MHC de clase II con una estructura y se inserta una cadena lateral de aminoácidos en la estructura en la posición deseada. Cada uno de los enlaces rotatorios de la cadena lateral se rota por etapas a intervalos establecidos y se anotan las posiciones resultantes de los átomos dependientes de ese enlace. Se anota la interacción del átomo con átomos de las cadenas laterales del surco de unión y se aceptan o rechazan las posiciones de acuerdo con los siguientes criterios: la suma total del solapamiento de todos los átomos colocados hasta el momento no debe exceder de un valor predeterminado. Por tanto la rigurosidad de la búsqueda conformacional es una función del intervalo usado en la rotación por etapas del enlace y del límite predeterminado para el solapamiento total. Este último valor puede ser pequeño si se sabe si un bolsillo en particular es rígido, sin embargo, la rigurosidad puede relajarse si se sabe que las posiciones de las cadenas laterales del bolsillo son relativamente flexibles. Por tanto, pueden hacerse concesiones para imitar variaciones en la flexibilidad de los bolsillos del surco de unión. Esta búsqueda conformacional se repite a continuación para cada aminoácido en cada posición de cada estructura cuando se acopla con cada una de las moléculas MHC de clase II para crear la base de datos exhaustiva de conformaciones de cadenas laterales.

Se usa una expresión matemática adecuada para estimar la energía de unión entre los modelos de moléculas MHC de clase II junto con las conformaciones de ligando peptídico que tienen que obtenerse empíricamente analizando la gran base de datos de conformaciones de estructura/cadena lateral descrita anteriormente. Por tanto, se analiza la presencia de posibles epítomos de células T sometiendo a cada posible péptido de longitud variable entre 9 y 20 aminoácidos (aunque la longitud se mantiene constante para cada análisis) a los siguientes cálculos: se selecciona una molécula MHC de clase II junto con una estructura peptídica permitida para esa molécula y se insertan las cadenas laterales correspondientes a la secuencia peptídica deseada. Se recogen datos de identidad atómica y distancia interatómica relacionada con una cadena lateral en particular en una posición particular en la estructura para cada conformación permitida de ese aminoácido (obtenida de la base de datos descrita anteriormente). Esto se repite para cada cadena lateral a lo largo de la estructura y se obtienen puntuaciones del péptido usando una función de puntuación. Se retiene la mejor puntuación para esa estructura y se repite el proceso para cada estructura permitida según el modelo seleccionado. Se comparan las puntuaciones de todas las estructuras permitidas y la puntuación más alta se considera como el valor del péptido para el péptido deseado en ese modelo de MHC de clase II. A continuación se repite este proceso para cada modelo con cada péptido posible derivado de la proteína que se está analizando, y se representan los valores de los péptidos frente a los modelos.

En el contexto de la presente invención, cada ligando que se presenta para el cálculo de la afinidad de unión es un segmento aminoacídico seleccionado entre un péptido o una proteína como se ha descrito anteriormente. Por tanto, el ligando es una extensión de aminoácidos seleccionada de aproximadamente 9 a 20 aminoácidos de longitud derivada de un péptido, polipéptido o proteína de secuencia conocida. Las expresiones “aminoácidos” y “restos” se consideran

a partir de ahora expresiones equivalentes. El ligando, en forma de los aminoácidos consecutivos del péptido que se va a examinar insertados en una estructura de la biblioteca estructural, se coloca en la hendidura de unión de una molécula MHC de clase II de la biblioteca de modelos de moléculas MHC de clase II mediante las coordenadas de los átomos C α de la estructura peptídica y se selecciona una conformación permitida para cada cadena lateral entre la base de datos de conformaciones permitidas. Las identidades atómicas y distancias interatómicas pertinentes también se recuperan de esta base de datos y se usan para calcular el valor de unión del péptido. Los ligandos con una elevada afinidad de unión por el bolsillo de unión de MHC de clase II se marcan como candidatos para mutagénesis dirigida al sitio. Se hacen sustituciones de aminoácidos en el ligando marcado (y por tanto en la proteína de interés) que después se prueban de nuevo usando la función de puntuación para determinar cambios que reducen la afinidad de unión por debajo de un valor umbral predeterminado. A continuación, estos cambios pueden incorporarse a la proteína de interés para eliminar epítomos de células T. La unión entre el ligando peptídico y el surco de unión de moléculas MHC de clase II implica interacciones no covalentes que incluyen, aunque sin limitación: puentes de hidrógeno, interacciones electrostáticas, interacciones hidrófobas (lipófilas) y fuerzas de Van der Waal. Éstas se incluyen en la función de puntuación del péptido como se describe con detalle a continuación. Debe entenderse que un enlace de hidrógeno es un enlace no covalente que puede formarse entre grupos polares o cargados y consta de un átomo de hidrógeno compartido por otros dos átomos. El hidrógeno del donante de hidrógeno tiene una carga positiva donde el aceptor de hidrógeno tiene una carga negativa parcial. Para los propósitos de interacciones péptido/proteína, los donantes del enlace de hidrógeno pueden ser nitrógenos con hidrógeno unido o hidrógenos unidos a oxígeno o nitrógeno. Los átomos aceptores de enlaces de hidrógeno pueden ser oxígenos no unidos a hidrógeno, nitrógenos sin hidrógenos unidos y una o dos conexiones, o azufres con solamente una conexión. Ciertos átomos, como oxígenos unidos a hidrógenos o nitrógenos imina (por ejemplo, C=NH) pueden ser tanto aceptores como donantes de hidrógeno. Las energías del puente de hidrógeno varían de 3 a 7 Kcal/mol y son mucho más fuertes que las fuerzas de Van der Waal, pero más débiles que los enlaces covalentes. Los puentes de hidrógeno también son altamente direccionales y son más fuertes cuando el átomo donante, el átomo de hidrógeno y el átomo aceptor son colineales. Los enlaces electrostáticos se forman entre pares iónicos opuestamente cargados y la fuerza de la interacción es inversamente proporcional al cuadrado de la distancia entre los átomos de acuerdo con la ley de Coulomb. La distancia óptima entre pares iónicos es aproximadamente 2,8 Å. En interacciones proteína/péptido, pueden formarse enlaces electrostáticos entre arginina, histidina o lisina y aspartato o glutamato. La fuerza del enlace dependerá del pKa del grupo ionizante y de la constante dieléctrica del medio aunque son aproximadamente similares en fuerza a los puentes de hidrógeno. Las interacciones lipófilas suponen contactos hidrófobo-hidrófobo favorables que tienen lugar entre la proteína y el ligando peptídico. Habitualmente, éstas se dan entre cadenas laterales de aminoácidos hidrófobos del péptido dispuestos dentro de los bolsillos del surco de unión de modo que no estén expuestos al solvente. La exposición de los restos hidrófobos al solvente es altamente desfavorable ya que las moléculas de solvente adyacentes se ven obligadas a formar puentes de hidrógeno entre sí dando lugar a estructuras de clatrato de tipo jaula. La disminución resultante en la entropía es altamente desfavorable. Los átomos lipófilos pueden ser azufres que no son polares ni aceptores de hidrógeno y átomos de carbono que no son polares. Los enlaces de Van der Waal son fuerzas no específicas que aparecen entre átomos que están alejados de 3 a 4 Å. Son más débiles y menos específicos que los puentes de hidrógeno y los enlaces electrostáticos. La distribución de la carga electrónica alrededor de un átomo cambia con el tiempo y, en un instante, dicha distribución de la carga pasa a no ser simétrica. Esta asimetría transitoria en la carga electrónica induce una asimetría similar en los átomos adyacentes. Las fuerzas de atracción resultantes entre los átomos alcanza un máximo en la distancia de contacto de Van der Waal pero disminuye muy rápidamente a aproximadamente de 1 Å a aproximadamente 2 Å. En cambio, cuando los átomos estando separados por una distancia menor a la distancia de contacto, las fuerzas de repulsión cada vez más fuertes llegan a ser dominantes ya que se solapan las nubes electrónicas externas de los átomos. Aunque las fuerzas de atracción son relativamente débiles en comparación con los enlaces electrostáticos y los puentes de hidrógeno (aproximadamente 0,6 Kcal/mol), las fuerzas de repulsión en particular pueden ser muy importantes para determinar si un ligando peptídico puede unirse satisfactoriamente a una proteína.

En una realización, la función de puntuación de Böhm (enfoque SCORE1) se usa para estimar la constante de unión. (Böhm, H.J., *J. Comput Aided Mol. Des.*, 8(3): 243-256 (1994) que se incorpora por la presente en su totalidad). En otra realización, la función de puntuación (enfoque SCORE2) se usa para estimar las afinidades de unión como indicador de un ligando que contiene un epítipo de células T (Böhm, H.J., *J. Comput Aided Mol. Des.*, 12(4): 309-323 (1998) que se incorpora por la presente en su totalidad). Sin embargo, las funciones de puntuación de Böhm que se describen en las referencias anteriores se usan para estimar la afinidad de unión de un ligando a una proteína cuando ya se sabe que el ligando se une satisfactoriamente a la proteína y se ha resuelto la estructura del complejo proteína/ligando, estando presente la estructura resuelta en el Banco de datos de proteínas ("BDP"). Por lo tanto, la función de puntuación se ha desarrollado con el provecho de los datos de unión positiva conocidos. Para permitir la discriminación entre enlazadores positivos y negativos, debe añadirse un término de repulsión a la ecuación. Además, se consigue una estimación más satisfactoria de la energía de unión calculando las interacciones lipófilas por pares en lugar de usando el término de energía basado en el área de las funciones de Böhm anteriores. Por lo tanto, en una realización preferida, la energía de unión se estima usando una función de puntuación de Böhm modificada. En la función de puntuación de Böhm modificada, la energía de unión entre la proteína y el ligando ($\Delta G_{\text{unión}}$) se estima considerando los siguientes parámetros: la reducción de la energía de unión debida a la pérdida global de entropía de traslación y rotación del ligando (ΔG_0); las contribuciones de los puentes de hidrógeno ideales (ΔG_{hb}) donde al menos un miembro del par es neutro; las contribuciones de interacciones iónicas no perturbadas ($\Delta G_{\text{iónica}}$); las interacciones lipófilas entre átomos del ligando lipófilo y átomos del aceptor lipófilo (ΔG_{lipo}); la pérdida de energía de unión debida a la congelación de los grados de libertad internos del ligando, es decir, se reduce la libertad de rotación alrededor de cada enlace C-C (ΔG_{rot}); la energía de la interacción entre la proteína y el ligando (E_{vdw}). La consideración de estos términos da lugar a ecuación 1:

ES 2 309 168 T3

$$(\Delta G_{\text{unión}}) = (\Delta G_0) + (\Delta G_{\text{hb}} \times N_{\text{hb}}) + (\Delta G_{\text{iónica}} \times N_{\text{iónica}}) + (\Delta G_{\text{lipo}} \times N_{\text{lipo}}) + (\Delta G_{\text{rot}} + N_{\text{rot}}) + (E_{\text{vdw}}).$$

Donde N es el número de interacciones aceptables para un término específico de la ecuación y, en una realización, ΔG_0 , ΔG_{hb} , $\Delta G_{\text{iónica}}$, ΔG_{lipo} y ΔG_{rot} son constantes que tienen los valores: 5,4, -4,7, -4,7, -0,17, y 1,4, respectivamente.

El término N_{hb} se calcula de acuerdo con la ecuación 2:

$$N_{\text{hb}} = \sum_{\text{enlaces-h}} f(\Delta R, \Delta \alpha) \times f(N_{\text{adyac}}) \times f_{\text{pcs}}$$

$f(\Delta R, \Delta \alpha)$ es una función de penalización que representa grandes desviaciones de los puentes de hidrógeno en condiciones ideales y se calcula de acuerdo con la ecuación 3:

$$f(\Delta R, \Delta \alpha) = f_1(\Delta R) \times f_2(\Delta \alpha)$$

Donde: $f_1(\Delta R) = 1$ si $\Delta R \leq \text{TOL}$

o $= 1 - (\Delta R - \text{TOL})/0,4$ si $\Delta R \leq 0,4 + \text{TOL}$

o $= 0$ si $\Delta R > 0,4 + \text{TOL}$

Y: $f_2(\Delta \alpha) = 1$ si $\Delta \alpha < 30^\circ$

o $= 1 - (\Delta \alpha - 30)/50$ si $\Delta \alpha \leq 80^\circ$

o $= 0$ si $\Delta \alpha > 80^\circ$

TOL es la desviación tolerada de la longitud del enlace de hidrógeno = 0,25 Å

ΔR es la desviación de la longitud del enlace de hidrógeno H-O/N desde el valor ideal = 1,9 Å

$\Delta \alpha$ es la desviación del ángulo del puente de hidrógeno LN/O-H desde su valor ideal de 180°

$f(N_{\text{adyac}})$ distingue entre partes cóncavas y convexas de una superficie proteica y por lo tanto asigna mayor peso a las interacciones polares encontradas en bolsillos que a las que aparecen en la superficie proteica. Esta función se calcula de acuerdo con la siguiente ecuación 4:

$$f(N_{\text{adyac}}) = (N_{\text{adyac}}/N_{\text{adyac},0})^\alpha \quad \text{donde } \alpha=0,5$$

N_{adyac} es el número de átomos de la proteína que no son hidrógeno que están más cerca de 5 Å. de cualquier átomo dado de la proteína.

$N_{\text{adyac},0}$ es una constante = 25

f_{pcs} es una función que considera el área de superficie de contacto polar por enlace de hidrógeno y, por lo tanto, distingue entre puentes de hidrógeno fuertes y débiles y su valor se determina de acuerdo con los siguientes criterios:

$f_{\text{pcs}} = \beta$ cuando $A_{\text{polar}}/N_{\text{HB}} < 10 \text{ Å}^2$ o

$f_{\text{pcs}} = 1$ cuando $A_{\text{polar}}/N_{\text{HB}} > 10 \text{ Å}^2$

A_{polar} es el tamaño de la superficie de contacto proteína polar-ligando

N_{HB} es el número de puentes de hidrógeno

β es una constante cuyo valor = 1,2

Para la aplicación de la función de puntuación de Böhm modificada, se calculan las contribuciones de las interacciones iónicas, $\Delta G_{\text{iónica}}$, de un modo similar al descrito anteriormente para los puentes de hidrógeno ya que se supone la misma dependencia geométrica.

El término N_{lipo} se calcula de acuerdo con la siguiente ecuación 5:

$$N_{\text{lipo}} = \sum_{\text{IL}} (r_{\text{IL}})$$

$f(r_{IL})$ se calcular para todos los átomos del ligando lipófilo, 1, y todos los átomos de la proteína lipófila, L, de acuerdo con los siguientes criterios:

$$f(r_{IL}) = 1 \text{ cuando } r_{IL} \leq R1 \quad f(r_{IL}) = (r_{IL} - R1)/(R2 - R1) \text{ cuando } R2 < r_{IL} < R1$$

$$f(r_{IL}) = 0 \text{ cuando } r_{IL} \geq R2$$

Donde:

$$R1 = r_1^{vdw} + r_L^{vdw} + 0,5$$

$$R2 = R1 + 3,0$$

r_1^{vdw} es el radio de Van der Waal del átomo 1 y

r_L^{vdw} es el radio de Van der Waal del átomo L

El término N_{rot} es el número de enlaces rotatorios de la cadena lateral del aminoácido y se acepta que es la cantidad de enlaces sp^3-sp^3 y sp^3-sp^2 acíclicos. Las rotaciones de $-CH_3$ o $-NH_3$ terminales no se tienen en cuenta.

El término final, E_{vdw} , se calcula de acuerdo con la siguiente ecuación 6:

$$E_{vdw} = \varepsilon_1 \varepsilon_2 ((r_1^{vdw} + r_2^{vdw})^{12}/r^{12} - (r_1^{vdw} + r_2^{vdw})^6/r^6),$$

donde:

ε_1 y ε_2 son constantes dependientes de la identidad atómica

$r_1^{vdw} + r_2^{vdw}$ son los radios atómicos de Van der Waal

r es la distancia entre un par de átomos.

Con respecto a la Ecuación 6, en una realización, las constantes ε_1 y ε_2 tiene los valores atómicos: C: 0,245, N: 0,283, O: 0,316 y S: 0,316, respectivamente (es decir, para átomos de carbono, nitrógeno, oxígeno y azufre, respectivamente). Con respecto a las ecuaciones 5 y 6, los radios de Van der Waal tienen los valores atómicos C: 1,85, N: 1,75, O: 1,60 y S: 2,00 Å.

Debe entenderse que todos los valores predeterminados y constantes dados en las ecuaciones anteriores se determinan dentro de las limitaciones del conocimiento actual de las interacciones proteína ligando en particular con respecto al tipo de cálculos que se está llevando a cabo en este documento. Por lo tanto, es posible que, conforme se mejore esta función de puntuación adicionalmente, puedan cambiar estos valores y constantes por lo que puede usarse cualquier valor numérico adecuado que ofrezca los resultados deseados en términos de estimación de la energía de unión de una proteína a un ligando y por tanto están dentro del alcance de la presente invención.

Como se ha descrito anteriormente, la función de puntuación se aplica a datos extraídos de la base de datos de conformaciones de cadenas laterales, identidades atómicas y distancias interatómicas. Para los propósitos de la presente descripción, el número de moléculas MHC de clase II incluidas en esta base de datos es de 42 modelos más cuatro estructuras resueltas. Debe ser evidente a partir de las anteriores descripciones que la naturaleza modular de la construcción del método informático de la presente invención significa que los nuevos modelos pueden simplemente añadirse y analizarse con la biblioteca de estructuras peptídicas y la función de búsqueda conformacional de cadena lateral para crear conjuntos de datos adicionales que puedan procesarse mediante la función de puntuación de péptido como se ha descrito anteriormente. Esto permite que el repertorio de moléculas MHC de clase II analizadas aumente fácilmente, o que se replacen estructuras y datos asociados si se dispone de datos que permitan crear modelos más precisos de los alelos existentes.

El presente método de predicción puede calibrarse con respecto a un conjunto de datos que comprende una gran cantidad de péptidos cuya afinidad por diversas moléculas MHC de clase II se ha determinado previamente a nivel experimental. Por comparación de datos calculados frente a datos experimentales, puede determinarse un punto de corte para el valor por encima del cual se sabe que todos los epítomos de células T determinados experimentalmente se han predicho correctamente.

Debe entenderse que, aunque la función de puntuación anterior es relativamente simple en comparación con algunas metodologías sofisticadas que están disponibles, los cálculos se realizan de forma extremadamente rápida. También debe entenderse que el objetivo no es calcular la energía de unión verdadera *per se* para cada péptido acoplado en el surco de unión de una proteína MHC de clase II seleccionada. El objetivo subyacente es obtener datos de energía de unión comparativos como ayuda para predecir la localización de epítomos de células T en base a la estructura primaria

(es decir, la secuencia de aminoácidos) de una proteína seleccionada. Una energía de unión relativamente elevada o una energía de unión por encima de un valor umbral seleccionado sugeriría la presencia de un epítipo de células T en el ligando. A continuación, el ligando puede someterse a al menos una ronda de sustitución de aminoácidos, tras lo cual se recalcula la energía de unión. Debido a la naturaleza rápida de los cálculos, estas manipulaciones de la secuencia peptídica pueden realizarse de forma interactiva dentro de la interfaz del usuario del programa en un hardware informático disponible rentable. Por tanto no se requiere una inversión grande en hardware informático.

Sería evidente para un especialista en la materia que podría usarse otro software disponible con el mismo propósito. En particular, puede usarse un software más sofisticado que sea capaz de acoplar ligandos en sitios de unión de proteínas junto con una reducción al mínimo de la energía. Los ejemplos de software de acoplamiento son: DOCK (Kuntz y col., *J. Mol. Biol.*, 161: 269-288 (1982)), LUDI (Böhm, H.J., *J. Comput Aided Mol. Des.*, 8: 623-632 (1994)) y FLEXX (Rarey M., y col., *ISMB*, 3: 300-308 (1995)). Los ejemplos de software de modelado y manipulación molecular incluyen: AMBER (Tripos) y CHARMM (Molecular Simulations Inc.). El uso de estos métodos informáticos limitaría en gran medida el rendimiento del método de esta invención debido a las extensiones de tiempo de procesamiento requeridas para hacer los cálculos necesarios. Sin embargo, es factible que dichos métodos puedan usarse como un “filtro secundario” para obtener cálculos más precisos de la energía de unión en aquellos péptidos que se descubre son “enlazadores positivos” mediante el método de la presente invención. La limitación del tiempo de procesamiento para cálculos mecánicos moleculares o dinámicos moleculares sofisticados es aquella que está definida tanto por el diseño del software que hace estos cálculos como por las limitaciones tecnológicas actuales del hardware informático. Debe preverse que, en el futuro, con la escritura de un código más eficaz y los aumentos continuos de la velocidad de los procesadores informáticos, pueda llegar a ser factible hacer dichos cálculos dentro de un intervalo de tiempo más manejable. Puede encontrarse información adicional sobre las funciones de energía aplicadas a macromoléculas y la consideración de las diversas interacciones que tienen lugar en una estructura proteica plegada en: Brooks, B.R., y col., *J. Comput. Chem.*, 4:187-217 (1983) y puede encontrarse información adicional referente a interacciones proteína-ligando generales en: Dauber-Osguthorpe y col., *Proteins* 4(1): 31-47 (1988), que se incorporan en este documento por referencia en su totalidad. También pueden encontrarse la información general útil por ejemplo, en Fasman, G.D., ed., *Prediction of Protein Structure and the Principles of Protein Conformation*, Plenum Press, Nueva York, ISBN: 0-306 4313-9.

REIVINDICACIONES

1. Un factor estimulante de colonias de granulocitos (G-CSF) humano que es sustancialmente no inmunogénico o menos inmunogénico que cualquier molécula no modificada con la misma actividad biológica cuando se utiliza *in vivo*, en el que dicha pérdida de inmunogenicidad se consigue realizando una sustitución de aminoácidos seleccionada a partir de la Tabla 2 en la posición 153 de la molécula original no modificada en un posible epítipo de células T dado, que es un ligando de MHC de clase II o secuencia peptídica que muestra la capacidad de estimular o unirse a células T mediante la presentación en el contexto del MHC de clase II.

2. Una molécula G-CSF modificada según la reivindicación 1, en la que el resto de tipo salvaje valina en la posición 153 está sustituido con alanina.

3. Una molécula G-CSF modificada según la reivindicación 2 ó 3, en la que adicionalmente se realizan una o más de las sustituciones de restos aminoácidos que se indican en la Tabla 3 para la reducción del número de alotipos MHC capaces de unirse a péptidos derivados de dicha molécula.

4. Una molécula G-CSF modificada según cualquiera de las reivindicaciones 1 a 3, en la que adicionalmente se realiza una sustitución posterior de aminoácidos para recuperar la actividad biológica de dicha molécula.

5. Una molécula G-CSF modificada según cualquiera de las reivindicaciones 1 a 5, en la que la alteración de aminoácidos se realiza en referencia a técnicas de modelado *in silico*.

6. Una secuencia de ADN que codifica una molécula factor estimulante de colonias de granulocitos (G-CSF) modificada de cualquiera de las reivindicaciones 1 a 5.

7. Una composición farmacéutica que comprende una molécula modificada que tiene la actividad biológica del factor estimulante de colonias de granulocitos (G-CSF) según se ha definido en cualquiera de las reivindicaciones citadas anteriormente, opcionalmente junto con un vehículo, diluyente o excipiente farmacéuticamente aceptable.

8. Un método para obtener una molécula modificada que tiene la actividad biológica del factor estimulante de colonias de granulocitos (G-CSF) humano como se ha definido en cualquiera de las reivindicaciones 1 a 5,

que comprende las siguientes etapas:

(i) determinar la secuencia de aminoácidos del polipéptido o parte del mismo;

(ii) identificar uno o más posibles epítopos de células T dentro de la secuencia de aminoácidos de la proteína seleccionando una región del péptido que tiene una secuencia de restos aminoácidos conocida; (b) muestreando secuencialmente segmentos de restos aminoácidos solapantes de tamaño uniforme predeterminado y constituidos por al menos tres restos aminoácidos de la región seleccionada; (c) calculando la puntuación de la unión a la molécula MHC de clase II de cada uno de dichos segmentos muestreados sumando los valores asignados a cada cadena lateral de resto aminoácido hidrófobo presente en dicho segmento de restos aminoácidos muestreado y (d) identificando al menos uno de dichos segmentos adecuado para la modificación, en función de la puntuación de unión de la molécula MHC de clase II calculada para ese segmento, para cambiar el valor de unión a MHC de clase II global para el péptido sin reducir sustancialmente la utilidad terapéutica del mismo;

(iii) diseñar nuevas variantes de secuencia con uno o más aminoácidos dentro de los posibles epítopos de células T identificados y modificados de modo que se reduzca sustancialmente o elimine la actividad del epítipo de células T según se determina mediante la unión de los péptidos a moléculas MHC usando técnicas *in vitro* o *in silico* o ensayos biológicos, o por la unión de complejos péptido-MHC a células T; donde la modificación se realiza sustituyendo el resto aminoácido de la posición 153 del G-CSF humano;

(iv) construir dichas variantes de secuencia mediante tecnología de ADN recombinante y probar dichas variantes para identificar una o más variantes con las propiedades deseadas; y

(v) opcionalmente repetir las etapas (ii)-(iv).

9. Un método de la reivindicación 8, en el que adicionalmente se lleva a cabo una o más sustituciones de restos aminoácidos según se indica en la Tabla 3 para reducir el número de alotipos MHC capaces de unirse a péptidos derivados de dicha molécula.

10. Una secuencia de un péptido epitópico de células T dentro de la secuencia de aminoácidos del G-CSF humano identificada según la etapa (ii) de la reivindicación 8 y seleccionada entre el grupo compuesto por las siguientes secuencias solapantes de 13 mer, que incluyen la posición 153 (V) del G-CSF humano:

ES 2 309 168 T3

SAFQRRAGGVLYA

GGVLYASHLQSFL

5 GVLVASHLQSFLE

VLVASHLQSFLEV

10 11. Una secuencia peptídica compuesta de al menos 9 restos aminoacídicos consecutivos de un péptido epitópico de células T de 13 mer según la reivindicación 10.

12. Uso de una secuencia peptídica según la reivindicación 10 u 11 para obtener un factor estimulante de colonias de granulocitos (G-CSF) humano que tiene sustancialmente ninguna o menos inmunogenicidad que cualquier molécula
15 no modificada con la misma actividad biológica cuando se utiliza *in vivo*.

20

25

30

35

40

45

50

55

60

65