US 20050181459A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2005/0181459 A1**

Baker et al. (43) **Pub. Date: Aug. 18, 2005**

(54) **METHOD FOR MAPPING AND ELIMINATING T CELL EPITOPES**

(76) Inventors: **Matthew Baker**, Cambridge (GB); **Francis J. Carr**, Balmedie (GB); **Graham Carter**, By Newmachar (GB)

Correspondence Address:
OLSON & HIERL, LTD.
20 NORTH WACKER DRIVE
36TH FLOOR
CHICAGO, IL 60606 (US)

(21) Appl. No.: **11/009,460**

(22) Filed: **Dec. 10, 2004**

Related U.S. Application Data

(63) Continuation-in-part of application No. PCT/EP03/06110, filed on Jun. 11, 2003.

(30) **Foreign Application Priority Data**

Jun. 11, 2002 (EP) 02012919.3

Publication Classification

(51) **Int. Cl.**[7] **G01N 33/53**; G01N 33/567; G06F 19/00; G01N 33/48; G01N 33/50; C12N 5/08

(52) **U.S. Cl.** **435/7.2**; 435/372; 702/19

(57) **ABSTRACT**

The invention provides methods for the identification of immunogenic regions within the amino acid residue sequence of a polypeptide, such as a therapeutic protein or a fragment thereof. The method comprises the steps of: (i) culturing, in vitro, an aliquot of peripheral blood monocyte cells (PBMC) isolated from a donor in the presence of a peptide for a period of up to about 7 days, the amino acid residue sequence of the peptide being identical to at least a portion of the amino acid residue sequence of the polypeptide of interest, the peptide being selected from a library of peptides, the amino acid residue sequences of the individual peptides of the library collectively encompassing the entire amino acid residue sequence of the polypeptide of interest; culturing the T cell aliquot from step (i) for an additional period of up to about 3 days in the presence of a T cell proliferation-stimulating cytokine to expand the number of T cells therein; (iii) culturing the T cell aliquot from step (ii) for a period of about 4 days in the presence of autologous irradiated PBMC from the same donor and in the presence of an additional amount of the peptide sufficient to re-prime the T cells within the PBMC with the peptide; (iv) determining the level of T cell proliferation of the re-primed T cells relative to an established baseline control level of proliferation; and (v) repeating steps (i) through (iv) with each peptide of the library of peptides to thereby identify at least one immunogenic region within the amino acid residue sequence of the polypeptide of interest.

FIGURE 1

INFβ IMMUNOGENIC REGIONS

REGION 1 (R1)

Residue# 49 72
QFQKEDAALTIYEMLQNIFAIFRQ (SEQ ID NO: 116)

Stimulating Peptides:

#17 QFQKEDAAL**TIYEML** (SEQ ID NO: 17)
#18 KEDAAL**TIYEML**QNI (SEQ ID NO: 18)
#19 AAL**TIYEML**QNIFAI (SEQ ID NO: 19)
#20 **TIYEML**QNIFAIFRQ (SEQ ID NO: 20)

REGION 2 (R2)

Residue # 124 145
RYYGRILHYLKAKEYSHCAWT (SEQ ID NO: 117)

Stimulating Peptides:

#42 RYYGRI**LHYLKA** (SEQ ID NO: 42)
#43 GRI**LHYLKA**KEY (SEQ ID NO: 43)
#44 **LHYLKA**KEYSHC (SEQ ID NO: 44)
#45 LKAKEYSHCAWT (SEQ ID NO: 45)

FIGURE 2

IFNα IMMUNOGENIC REGIONS

REGION 1 (R1)

```
Residue#      24                                  57
               ISLFSCLKDRHDFGFPQEEFGNQFQKAETIPVLH (SEQ ID NO: 118)

Stimulating
Peptides:
        #9      SLFSCLKDRHDFGFP (SEQ ID NO: 60)
        #10         SCLKDRHDFGFPQEE (SEQ ID NO: 61)
        #11            KDRHDFGFPQEEFGN (SEQ ID NO: 62)
        #12               HDFGFPQEEFGNQFQ (SEQ ID NO: 63))
        #13                  GFPQEEFGNQFQKAE (SEQ ID NO: 64)
        #14                     QEEFGNQFQKAETIP (SEQ ID NO: 65)
        #15                        FGNQFQKAETIPVLH (SEQ ID NO: 66)
```

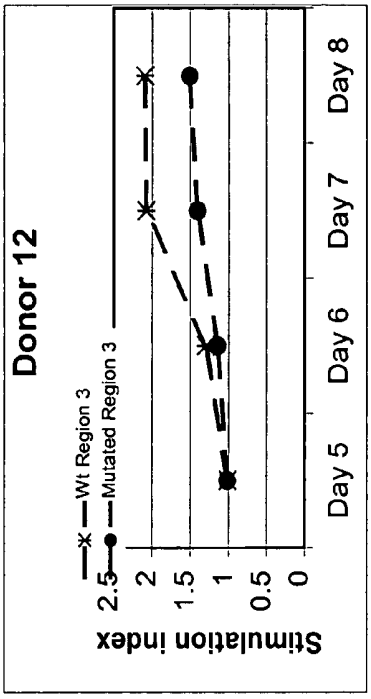
REGION 2 (R2)

```
Residue #            64            78
                      FNLFSTKDSSAAWDE (SEQ ID NO: 119)

Stimulating
Peptides:
          #21      QQIFNLFSTKDSSAA (SEQ ID NO: 72)
          #22          FNLFSTKDSSAAWDE (SEQ ID NO: 73)
          #23             FSTKDSSAAWDETLL (SEQ ID NO: 74)
          #24                 KDSSAAWDETLLDKF (SEQ ID NO: 75)
```

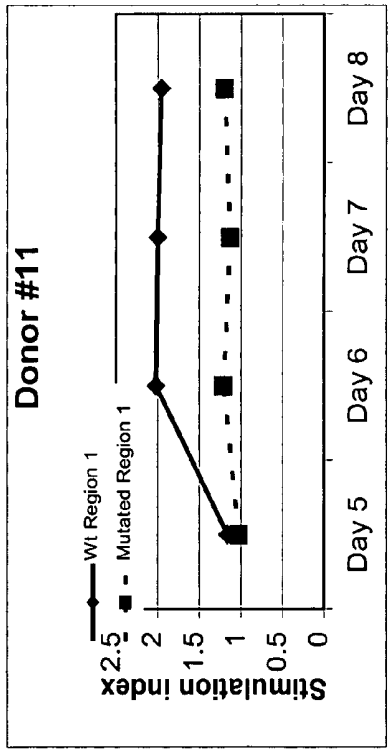
REGION 3 (R3)

```
Residue #        110                 129
                  LMKEDSILAVRKYFQRITLY (SEQ ID NO: 120)

Stimulating
Peptides:
          #36 TETPLMKEDSILAVR (SEQ ID NO: 87)
          #37    PLMKEDSILAVRKYF (SEQ ID NO: 88)
          #38       KEDSILAVRKYFQRI (SEQ ID NO: 89)
          #39          SILAVRKYFQRITLY (SEQ ID NO: 90)
          #40             AVRKYFQRITLYLKE (SEQ ID NO: 91)
```

FIGURE 3



IFN α PEPTIDE	SEQ ID NO:	PEPTIDE SEQUENCE
Region 1	121	QMRRISLFSCLKDRHDFGFP
Mutated Region 1	122	QMRRQSLFSCLKDRHDFGFP
Region 2	123	EMIQQIFNLFSTKDSSAAWDETLLD KFY
Mutated Region 2	124	EMTQQIANLFSTKDSSAAHDETLLD KFY
Region 3	125	TPLMKEDSILAVRKYFQRITLYLKE KKYSPCAW
Mutated Region 3	126	TPLMKEDSRLAVRKYFQRITNYLKE KKYSPCAW



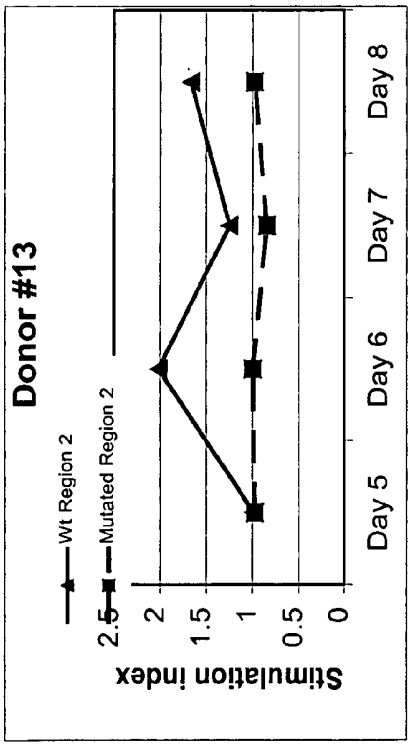
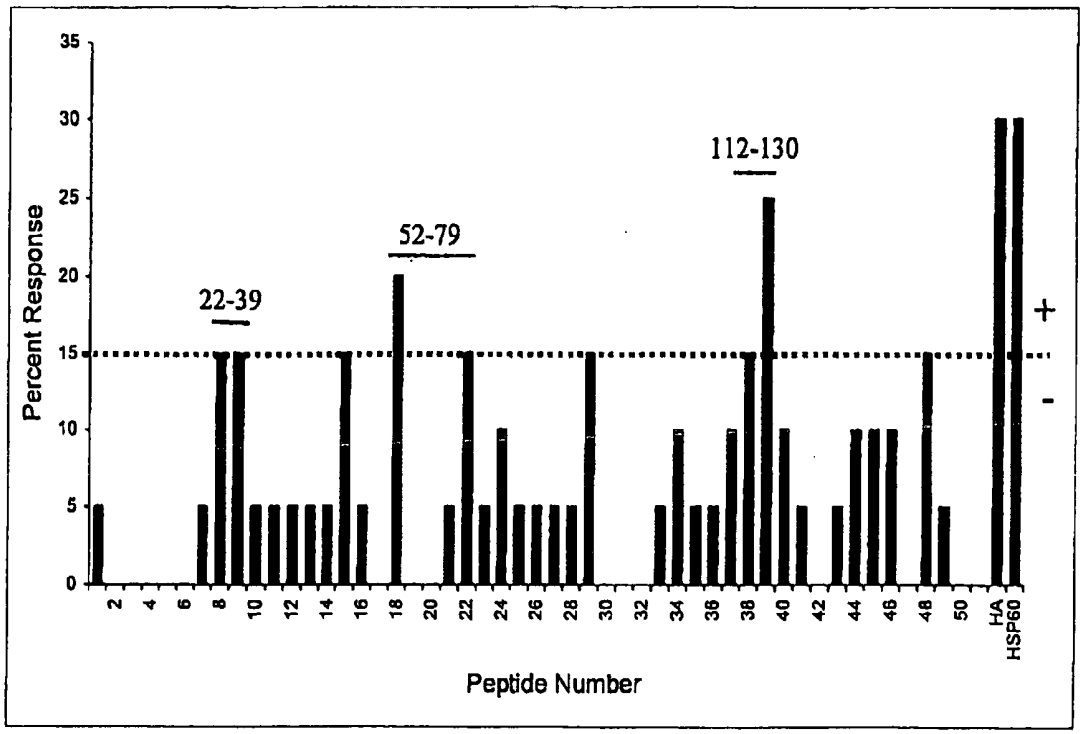


FIGURE 4

A



B

Region 1

```
                                Q
CDLPQTHSLG SRRTLMLLAQ MRRISLFSCL KDRHDFGFPQ EEFGNQFQKA
1
```

Region 2

```
          T    A                H
ETIPVLHEMI QQIFNLFSTK DSSAAWDETL LDKFYTELYQ QLNDLEACVI
51
```

Region 3

```
                R                N
QGVGVTETPL MKEDSILAVR KYFQRITLYL KEKKYSPCAW EVVRAEIMRS
101

FSLSTNLQES LRSKE
151
```

METHOD FOR MAPPING AND ELIMINATING T CELL EPITOPES

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] This application is a continuation-in-part of International Application Serial No. PCT/EP03/06110, filed on Jun. 11, 2003, designating the United States, which is incorporated herein by reference.

FIELD OF THE INVENTION

[0002] The present invention relates to the field of immunology. The invention provides methods for the identification of determinants and epitopes on protein molecules that can evoke an immune response. In particular the invention is concerned with the identification of epitopes for T cells in therapeutic proteins. Additionally, the invention relates to a combined approach of using epitope mapping in concert with identification of MHC class II ligands comprising epitopes from said epitope mapping method and modification of the therapeutic protein to reduce the number epitopes present in the protein sequence.

BACKGROUND OF THE INVENTION

[0003] There are many instances in which the efficacy of a therapeutic protein is limited by an undesirable immune reaction to the therapeutic protein. Several mouse monoclonal antibodies have shown promise as therapeutic agents in a number of human disease settings, but in some cases have failed due to the induction of significant degrees of a human anti-murine antibody (HAMA) response in the patient [Schroff, R. W. et al. (1985) *Cancer Res.* 45: 879-885; Shawler, D. L. et al. (1985) *J. Immunol.* 135: 1530-1535]. For monoclonal antibodies, a number of techniques have been developed in attempt to reduce the HAMA response [WO 89/09622; EP 0239400; EP 0438310; WO 91/06667]. These recombinant DNA approaches have generally reduced the mouse genetic information in the final antibody construct whilst increasing the human genetic information in the final construct. Nonetheless, the resultant "humanized" antibodies have, in several cases, still elicited an immune response in patients [Issacs J. D. (1990) *Sem. Immunol.* 2: 449, 456; Rebello, P. R. et al. (1999) *Transplantation* 68: 1417-1420].

[0004] Antibodies are not the only class of polypeptide molecule administered as a therapeutic agent against which an immune response can be mounted. Even proteins of human origin and with the same amino acid sequences as occur within humans can still induce an immune response in humans. Notable examples amongst others include the therapeutic use of granulocyte-macrophage colony stimulating factor [Wadhwa, M. et al. (1999) *Clin. Cancer Res.* 5: 1353-1361] and interferon alpha 2 [Russo, D. et al. (1996) *Bri. J. Haem.* 94: 300-305; Stein, R. et al. (1988) *New Engl. J. Med.* 318: 1409-1413]. In such situations where these human proteins are immunogenic, there is a presumed breakage of immunological tolerance that would otherwise have been operating in these subjects to these proteins.

[0005] Undesirable immune responses to a human protein are also observed in human patients who are administered the protein as a replacement therapy, for example, in a genetic disease where there is a constitutional lack of the protein, such as hemophilia A, Christmas disease, Gauchers disease, and numerous other conditions. In such cases, the therapeutic replacement protein may function immunologically as a foreign molecule from the outset, and where the individuals are able to mount an immune response to the therapeutic, the efficacy of the therapy is likely to be significantly compromised.

[0006] Irrespective of whether the therapeutic protein is seen by the host immune system as a foreign molecule, or whether an existing tolerance to the molecule is overcome, the mechanism of immune reactivity to the protein is basically the same. The presence, within the protein, of peptide segments that can stimulate the activity of T cells via presentation on MHC class II molecules (i.e., so-called "T cell epitopes") are critical to the induction of an immune response. Such T cell epitopes are commonly defined as any amino acid residue sequence with the ability to bind to MHC Class II molecules. Implicitly, a "T cell epitope" refers to an epitope that can be recognized by a T cell receptor (TCR) when bound to a MHC molecule, and which can, at least in principle, cause the activation of the T cell by engaging a TCR to promote a T cell response.

[0007] MHC Class II molecules are a group of highly polymorphic proteins that play a central role in helper T cell selection and activation. The human leukocyte antigen group DR (HLA-DR) are the predominant isotype of this group of proteins however, isotypes HLA-DQ and HLA-DP perform similar functions. The present invention is applicable to the detection of T cell epitopes presented within the context of DR, DP or DQ MHC Class II. In the human population, individuals bear two to four DR alleles, two DQ and two DP alleles. The structure of a number of DR molecules has been solved and these appear as an open-ended peptide binding groove with a number of hydrophobic pockets which engage hydrophobic residues (pocket residues) of the peptide [Brown et al. *Nature* (1993) 364: 33; Stem et al. (1994) *Nature* 368: 215]. Polymorphism identifying the different allotypes of MHC class II molecule contributes to a wide diversity of different binding surfaces for peptides within the peptide binding groove and at the population level ensures maximal flexibility with regard to the ability to recognize foreign proteins and mount an immune response to pathogenic organisms.

[0008] An immune response to a therapeutic protein proceeds via the MHC class II peptide presentation pathway in which exogenous proteins are engulfed and processed for presentation in association with MHC class II molecules of the DR, DQ or DP type. MHC Class II molecules are expressed by professional antigen presenting cells (APCs), such as macrophages and dendritic cells amongst others. Engagement of a MHC class II peptide complex by a cognate T cell receptor on the surface of the T cell, together with cross-binding of certain other co-receptors, such as the CD4 molecule, can induce an activated state within the T cell. Activation leads to the release of cytokines further activating other lymphocytes, such as B cells to produce antibodies, or activating killer T cells (i.e., cytotoxic T lymphocytes, CTLs) as a full cellular immune response.

[0009] T cell epitope identification is the first step toward elimination of epitopes; however, there are few clear cases in the art where epitope identification and epitope removal are integrated into a single scheme. WO98/52976 and

WO00/34317 teach computational threading approaches to identifying polypeptide sequences with the potential to bind a sub-set of human MHC class II DR allotypes. In these teachings, predicted T cell epitopes are removed by the use of judicious amino acid substitution within the protein of interest. However with this scheme and other computationally based procedures for epitope identification [Godkin, A. J. et al. (1998) *J. Immunol.* 161: 850-858; Stumiolo, T. et al. (1999) *Nat. Biotechnol.* 17: 555-561], peptides predicted to be able to bind MHC class II molecules may not function as T cell epitopes in all situations, particularly in vivo, due to the processing pathways or other phenomena. In addition, the computational approaches to T cell epitope prediction have in general not been capable of predicting epitopes with DP or DQ restriction.

[**0010**] In vitro methods for measuring the ability of synthetic peptides to bind MHC class II molecules, for example using B cell lines of defined MHC allotype as a source of MHC class II binding surface, may be applied to MHC class II ligand identification [see Marshall K. W. et al. (1994) *J. Immunol.* 152:4946-4956; O'Sullivan et al. (1990) *J. Immunol.* 145: 1799-1808; Robadey C. et al. (1997) *J. Immunol* 159: 3238-3246]. However, such techniques are not adapted for the screening multiple potential epitopes to a wide diversity of MHC allotypes, nor can they confirm the ability of a binding peptide to function as a T cell epitope.

[**0011**] Recently, techniques exploiting soluble complexes of recombinant MHC molecules in combination with synthetic peptides have come into use [Kern, F. et al. (1998) *Nature Medicine* 4:975-978; Kwok, W. W. et al. (2001) *TRENDS in Immunol.* 22:583-588]. These reagents and procedures are used to identify the presence of T cell clones from peripheral blood samples from human or experimental animal subjects that are able to bind particular MHC-peptide complexes and are not adapted for the screening multiple potential epitopes to a wide diversity of MHC allotypes.

[**0012**] Biological assays of T cell activation remain the best practical option to providing a reading of the ability of a test peptide/protein sequence to evoke an immune response. Examples of this kind of approach include the work of Petra et al. using T cell proliferation assays to the bacterial protein staphylokinase, followed by epitope mapping using synthetic peptides to stimulate T cell lines [Petra, A. M. et al. (2002) *J. Immunol.* 168: 155-161]. Similarly, T cell proliferation assays using synthetic peptides of the tetanus toxin protein have resulted in definition of immunodominant epitope regions of the toxin [Reece J. C. et al. (1993) *J. Immunol.* 151: 6175-6184]. WO99/53038 discloses an approach whereby T cell epitopes in a test protein may be determined using isolated sub-sets of human immune cells, promoting their differentiation in vitro and culture of the cells in the presence of synthetic peptides of interest and measurement of any induced proliferation in the cultured T cells. The same technique is also described by Stickler et al. [Stickler, M. M. et al. (2000) *J. Immunotherapy* 23:654-660], where in both instances the method is applied to the detection of T cell epitopes within bacterial subtilisin. Such a technique requires careful application of cell isolation techniques and cell culture with multiple cytokine supplements to obtain the desired immune cell sub-sets (dendritic cells, CD4+ and or CD8+ T cells) and is not conducive to rapid through-put screening using multiple donor samples.

[**0013**] In a variation of these approaches, Hiemstra et al. [Hiemstra, H. S. (1997) *Proc. Natl. Acad. Sci USA* 94: 10313-10318] have described a procedure for identifying a peptide epitope capable of stimulating a known T cell. Such a process is valuable in the detection of autoreactive T cell clones for which the (auto)antigen is unknown.

[**0014**] The above examples and other biological assays involving technical variations on the theme of measuring an in vitro T cell activation event, usually by the measurement of an induced proliferation response, abound. However, none of the procedures provide a unified method for the detection of biologically relevant epitopes in proteins of human origin, nor are the known techniques readily applicable to the detection of epitopes of significance to a wide population of MHC allotypes. The present invention provides such a method and provides a basis for the identification and removal of T cell epitopes from a given in principal therapeutically valuable, but originally immunogenic peptide, polypeptide, or protein.

SUMMARY OF THE INVENTION

[**0015**] The present invention relates to the use of a panel of synthetic peptides in a naive T cell assay to map the immunogenic region(s) of a therapeutic protein, preferably, a human protein. The synthetic peptides have amino acid residue sequences corresponding to portions of the sequence of the therapeutic protein, A recall assay of the invention, which involves re-challenging T cells that have already been primed by exposure to the peptide, can then be used to refine the immunogenic map of the therapeutic protein. The results of the assays identify peptide sequences within the therapeutic protein displaying high levels of immunogenicity in vitro. Variants of the highly immunogenic peptides can then be prepared and tested to identify alternative sequences that display reduced immunogenicity in vitro relative to the native therapeutic protein. The sequences identified as having reduced immunogenicity can then be genetically engineered into the therapeutic protein to form a protein variant having reduced immunogenicity, while retaining the therapeutic efficacy of the original therapeutic protein. Alternatively, whole-protein variants of the therapeutic protein can be used in the T cell assays in place of small peptide fragments, if desired, to identify variants displaying minimal immunogenicity in vitro, while retaining therapeutic efficacy.

[**0016**] The T cell assays are preferably biological assays of T cell stimulation, which afford a stimulation index for the test peptide or protein variant as the output of the assay. A stimulation index of less than about 2, preferably less than about 1.8, in a naive T cell assay is an indication of acceptably low immunogenicity to warrant further evaluation.

[**0017**] The invention also relates to the development of T cell lines from individuals to whom a therapeutic protein previously has been administered, and to the use of those T cell lines to map the immunogenic region(s) of the therapeutic protein. In addition, B cell lines can be developed from the same group of individuals in parallel to the T cell lines, so that both the T cell and B cell lines can be utilized to map the immunogenic region(s) of the therapeutic protein. The B cell lines and T cell lines can also provide a source of autologous antigen presenting cells (APC), or optionally can act as binding partners in synthetic peptide binding assays.

[0018] A T cell epitope map of a polypeptide of interest, such as a therapeutic protein, can be prepared by identifying immunogenic regions within the amino acid residue sequence of the polypeptide of interest. A method of the present invention for identifying a T cell epitope within the amino acid sequence of a polypeptide of interest comprises the steps of:

[0019] (i) culturing, in vitro, an aliquot of peripheral blood monocyte cells (PBMC) isolated from a healthy donor in the presence of a peptide for a period of up to about 7 days to form a peptide-primed T cell aliquot, the amino acid residue sequence of the peptide being identical to at least a portion of the amino acid residue sequence of the polypeptide of interest, the peptide being selected from a library of peptides, the amino acid residue sequences of the individual peptides of the library collectively encompassing the entire amino acid residue sequence of the polypeptide of interest;

[0020] (ii) culturing the peptide-primed T cell aliquot from step (i) for an additional period of up to about 3 days in the presence of a T cell proliferation-stimulating cytokine, such as IL-2, to expand the number of T cells therein, forming a T cell-expanded aliquot;

[0021] (iii) culturing the T cell-expanded aliquot from step (ii) for a period of about 4 days in the presence of autologous irradiated PBMC from the same donor and in the presence of an additional amount of the peptide sufficient to re-prime the T cells within the PBMC with the peptide;

[0022] (iv) determining the level of T cell proliferation of the re-primed T cells relative to an established baseline control level of proliferation (e.g., comparison to the level of T cell proliferation in PBMC from the same donor that have not been cultured in the presence of the peptide and calculating a T cell stimulation index for the polypeptide); and

[0023] (v) repeating steps (i) through (iv) with each peptide of the library of peptides to thereby identify at least one immunogenic region within the amino acid residue sequence of the polypeptide of interest.

[0024] Preferably, the method is performed with PBMC from a number of different donor individuals and the T cells are individually primed with a number of different polypeptides whose amino acid residue sequences together overlap at least a putative immunogenic portion of the therapeutic protein.

[0025] Preferably, steps (i) through (v) are repeated for each peptide of the library with PBMC isolated from a plurality of healthy donor individuals, the immunological diversity of the plurality of healthy donor individuals representing more than 90% of MHC class II allotypes.

[0026] Alternatively, the immunogenic region can be identified by a method comprising the steps of:

[0027] (i) culturing, in vitro, an aliquot of peripheral blood monocyte cells (PBMC) isolated from a donor in the presence of a peptide for a period of up to about 7 days to form a peptide-primed T cell aliquot, the amino acid residue sequence of the peptide being identical to at least a portion of the amino acid residue sequence of the polypeptide of interest, the peptide being selected from a library of peptides, the amino acid residue sequences of the individual peptides of the library collectively encompassing the entire amino acid residue sequence of the polypeptide of interest, the donor having an established immune response to the polypeptide of interest;

[0028] (ii) culturing the peptide-primed T cell aliquot from step (i) for an additional period of up to about 3 days in the presence of a T cell proliferation-stimulating cytokine, such as IL-2, to expand the number of T cells therein, forming a T cell-expanded aliquot;

[0029] (iii) culturing the T cell-expanded aliquot from step (ii) for a period of about 4 days in the presence of autologous irradiated PBMC from the same donor and in the presence of an additional amount of the peptide sufficient to re-prime the T cells within the PBMC with the peptide;

[0030] (iv) determining the level of T cell proliferation of the re-primed T cells relative to an established baseline control level of proliferation; and

[0031] (v) repeating steps (i) through (iv) with each peptide of the library of peptides to thereby identify at least one immunogenic region within the amino acid residue sequence of the polypeptide of interest.

[0032] Another aspect of the present invention is a method of preparing a T cell epitope map of a therapeutic protein, as described above, but using an aliquot of a polyclonal or monoclonal cell line derived from a PBMC sample from a donor in place of the aliquot of PBMC. The polyclonal or monoclonal cell line can be derived from PBMC of a healthy donor or a donor having an established immune response to the therapeutic protein of interest.

[0033] Yet another aspect of the invention utilizes a computational method for simulating the binding of a peptide to one or more MHC allotypes to provide a set of putative T cell epitope sequences. Peptides having the putative epitope sequences are then utilized as the polypeptide in one or more of the methods of the present invention described herein. The computational method can also be used to simulate the binding of a peptide with one or more MHC allotypes to identify sequence analogues of the T cell epitopes that no longer bind to a given MHC class II allotype or bind with lowered affinity to a lesser number of MHC allotypes. Peptides having the so-identified sequences can then be synthesized and tested in the methods of the invention, in place of the polypeptide, to quantify level of T cell activation afforded by the peptide. Peptide sequences that are verified to have a lower level of T cell activation can then be incorporated into a variant of the therapeutic protein to provide a new protein having lower immunogenicity than the therapeutic protein, but retaining the therapeutic efficacy thereof. A preferred computational approach is described in WO 02/069232, which is incorporated herein by reference.

[0034] In one preferred embodiment, the T cell activation assays utilize PBMC cells derived from around 20 or more unrelated donors whose immunological diversity encompasses at least about 90% of MHC Class II molecules. Preferably, the immunogenic region is identified by a stimu-

lation index score greater than 1.8, more preferably about 2 or greater, observed for a single polypeptide in PBMC from two or more independent donor samples, preferably at least about 15% of donor samples. Preferably, the T cell epitope has an amino acid residue sequence within a sequence locale of the therapeutic protein indicated to bind with an MHC allotype using a computational method as described herein.

[0035] Identification of protein sequences having a reduced ability to promote an immune response can be achieved using immunologically primed cells from the methods of the invention in combination with a screening process. Multiple variant peptides or a variant protein antigens are evaluated in parallel to pools of reference peptides or in reference to a whole protein antigen containing only wild-type sequences. Peptides or protein variants with a lesser stimulation index compared to the reference pool or wild-type protein are selected for further analysis.

[0036] A method for preparing a variant of a therapeutic protein having substantially the same biological activity and reduced immunogenicity compared to the therapeutic protein comprises the steps of:

[0037] (i) preparing at least one variant of the therapeutic protein, the amino acid residue sequence of the variant differing from the amino acid residue sequence of the therapeutic protein by an amino acid residue within an immunogenic region of the therapeutic protein, the immunogenic region being identified by the method of claim **1**;

[0038] (ii) comparing the biological activity and immunogenicity of the at least one variant to the biological activity and immunogenicity of the therapeutic protein; and

[0039] (iii) selecting a variant having substantially the same biological activity and reduced immunogenicity compared to the therapeutic protein.

[0040] Preferably the amino acid residue sequence of the at least one variant is selected by the steps of:

[0041] (a) calculating a MHC Class II molecule binding score for the immunogenic region using a computational method that sums assigned values for each hydrophobic amino acid residue side chain present in the immunogenic region;

[0042] (b) calculating a binding score for at least one amino acid residue sequence that differs from the amino acid residue sequence of the immunogenic region by an amino acid residue using the same computational method as in step (a); and

[0043] (c) selecting the amino acid residue sequence for the at least one variant having a binding score in step (b) that is lower than the binding score of the immunogenic region of the therapeutic protein.

[0044] Protein sequences or protein preparations from a protein manufacturing process that have an increased ability to promote an immune response can be identifyed using immunologically primed cells of the methods of the invention. In particular, the method of the invention are used in combination with a screening process whereby one or more peptides or whole protein antigens are tested in parallel to reference peptide pools or whole protein antigen affording a known in vitro immune response. Peptides or protein preparations that evoke a different stimulation index profile relative to the reference preparations are selected for further analysis or may be eliminated from the production process.

BRIEF DESCRIPTION OF THE DRAWINGS

[0045] **FIG. 1** shows the immunogenic regions within IFNβ and details the peptide sequences from these regions able to stimulate naive human T cells.

[0046] **FIG. 2** shows the initial determination of immunogenic regions within IFNα and details the peptide sequences from these regions able to stimulate naive human T cells.

[0047] **FIG. 3** provides exemplary data from time course T cell activation assays. Charts plot stimulation index (SI) against time (days) for synthetic peptides derived from the IFNα R1, R2 and R3 epitope regions and analogue peptide sequences containing amino acid substitutions tested in parallel.

[0048] **FIG. 4** illustrates refined immunogenic regions of IFNα2b. Panel A shows a graph of percentage of donor samples that exhibited a positive T cell stimulation response (SI>1.8) for 51 peptides spanning the entire amino acid residue sequence of IFNα2b. Panel B shows the sequence of IFNα2b (SEQ ID NO: 127) with immunogenic regions surrounded by boxes.

DETAILED DESCRIPTION OF THE INVENTION

[0049] According to the first embodiment of the invention there is provided a method whereby a protein antigen may be screened for the presence of determinants within its sequence that are capable of evoking a T cell driven immune response should that protein be introduced into a human subject. The method thereby provides a predictive tool for the identification of T cell epitopes in proteins with therapeutic potential in man where the protein is to be provided for the therapy of an acquired disease state and where that protein may be a human protein.

[0050] It is particularly desired to provide an epitope map of a protein of interest where the map has relevance to a wide spectrum of possible MHC allotypes. It is desired that the map is sufficiently representative to allow the design or selection of a modified protein for which the ability of the protein to evoke a T cell driven immune response is eliminated or at least ameliorated for the majority of patients to whom the protein is likely to be administered. Accordingly, the methods of the present invention preferably utilize PBMC derived T cells from naive donors collected from a pool of donors of sufficient immunological diversity to provide a sample of at least greater than 90% of the MHC class II repertoire (HLA-DR) extant in the human population and preferably greater than 95% of that repertoire. Equivalence to greater than 99% representation is particularly preferred, although it is recognised that there are practical limitations to achieving this goal. Accordingly, where a naive T cell response is to be detected to a given synthetic peptide, the peptide will be contacted with PBMC preparations derived from multiple donors in isolation, the numbers of donors or herein more preferably described as the "donor pool", is for practical purposes not likely to be less than 20 unrelated individuals (pre-selected according to their MHC class II haplotypes).

[0051] The term "naive donor" in the context of the present invention means that the T cells obtained from the individual have not previously been exposed to the protein or peptide antigen of interest, and where the protein antigen is a human protein, the individual has not been in receipt of any therapeutic or exogenous sources of the protein. Thus, according to the first embodiment of the present invention, there is provided a method for T cell epitope mapping exploiting immunologically naive T cells. The T cells are provided from a peripheral blood sample from a plurality of different healthy donors for whom the protein of interest may be an endogenous protein, but whose immune systems have not been exposed to the protein from any exogenous source e.g., administered therapeutically. The assay is conducted using PBMC cultured in vitro using procedures common in the art and involves contacting the PBMC with synthetic peptide species representative of the protein of interest, and following a suitable period of incubation, measurement of peptide induced T cell activation such as cellular proliferation. Measurement is by any suitable means and may for example be conducted using ^{3}H-thymidine incorporation whereby the accumulation of ^{3}H into cellular material is readily measured instrumentally. The degree of cellular proliferation for each combination of PBMC sample and synthetic peptide is examined relative to the level of proliferation observed PBMC that have not been exposed to the peptide. Reference may also be made to the proliferative response seen following treatment with a peptide or peptides for which there is an expected proliferative effect. In this regard it is considered particularly advantageous to use peptide with known broad MHC restriction and especially peptide epitopes with MHC restriction to the DP or DQ isotypes.

[0052] To facilitate assembly of an epitope map for a given protein of interest, a set of synthetic peptides representative of the sequence of the protein are produced. A typical analysis under the methods of the present invention involves the use of peptides containing about 15 amino acid residues, although it will be recognised that a peptide containing not less than 9 amino acid residues is in principle a suitable peptide. Peptides significantly exceeding 15 amino acid residues may also be used but it will equally be recognised that possible secondary structural effects or complexities of intracellular processing may obscure the ability of the peptide to induce a proliferative response. In order to scan the entire length of given protein, a particularly convenient scheme is to produce synthetic peptides each of 15 amino acid residues in length and each overlapping the next peptide in the series by 12 amino acid residues, i.e. each successive peptide in the series incrementally adds a further 3 amino acids to the analysis. In this way any given adjacent pair of peptides will map 18 amino acids of contiguous sequence in the protein of interest. Thus for a protein of interest comprising n amino acid residues, the number of 15-mer synthetic peptides required for a complete scan of the said protein will be $1+(n-12)/3$. Other protocols for selecting peptides to survey the sequence of a therapeutic protein may be utilized, if desired, such as using 9-mer or 13-mer peptides, or varying the number of overlapping residues between the samples, and the like.

[0053] Using the methods outlined above and exemplified in detail within the EXAMPLES herein, the inventors have discovered regions of protein sequence capable of evoking a proliferative response in naive PBMC from different individual healthy donors. The protein sequences in question are sequence strings derived from whole human proteins for which there could be an expectation of immune tolerance but which none the less there is a demonstrable ability to evoke a surrogate immune response in vitro. This ability by extension may also apply in vivo should either of the proteins in question be administered, for example, as therapeutic entities. Specifically these proteins are interferon $\alpha 2$ and interferon β. Both of these proteins are used therapeutically and significantly for both of these molecules, immunogenic responses to these molecules in patients have been recorded [Russo, D. et al. (1996) ibid; Stein, R. et al. (1988) ibid; Myhr, K. M. et al. (2000) *Neurology* 55:1569-1572; Bertolotto, A. et al. (2000) *Immunopharmacology* 48: 95-100]. The present invention therefore provides a generalised scheme for the elucidation of epitope regions within normal human proteins and demonstrates the ability of peptide sequences from these proteins to evoke an in vitro proliferative response in naive PBMC derived from healthy donors.

[0054] A particularly effective method for defining a T cell map using naive T cell assays of the first embodiment is provided in the EXAMPLES 1 and 2 in which initial determinations of immunogenic regions of the molecules interferon beta (IFNβ) and interferon alpha 2 (IFN$\alpha 2$) are disclosed. A particularly preferred method for the identification of T cell epitopes in proteins which are weakly immunogenic in vivo is described in EXAMPLE 3.

[0055] In a second embodiment where the invention provides for the elucidation of a T cell epitope map, such a map may be used to guide the design of a modified version of a therapeutic protein whereby the epitope regions on the protein are suitably modified such that they are no longer able to evoke a proliferative response according to the methods of the invention and the modified protein is less immunogenic to man than the original therapeutic protein.

[0056] According to this second embodiment, suitable modifications to the protein may include amino acid substitution of particular residues or combinations of residues. For the elimination of T cell epitopes, amino acid substitutions are preferably made at appropriate points within the peptide sequence predicted to achieve substantial reduction or elimination of the activity of the T cell epitope. In practice an appropriate point will preferably equate to an amino acid residue binding within one of the pockets provided within the MHC class II binding groove. It is most preferred to alter binding within the first pocket of the cleft at the so-called "P1" or "P1 anchor" position of the peptide. The quality of binding interaction between the P 1 anchor residue of the peptide and the first pocket of the MHC class II binding groove is recognised as being a major determinant of overall binding affinity for the whole peptide. An appropriate substitution at this position of the peptide will be for a residue less readily accommodated within the pocket, for example, substitution to a more hydrophilic residue. Amino acid residues in the peptide at positions equating to binding within other pocket regions within the MHC binding cleft are also considered and fall under the scope of the present.

[0057] It is understood that single amino acid substitutions within a given potential T cell epitope are the most preferred route by which the epitope may be eliminated. Combinations of substitution within a single epitope can be utilized,

if desired, and can be particularly appropriate where individually defined epitopes overlap with each other. Moreover, amino acid substitutions, either singly within a given epitope or in combination within a single epitope, can be made at positions not equating to the "pocket residues" with respect to the MHC class II binding groove, i.e., at any point within the peptide sequence. Substitutions can be made with reference to an homologous structure or structural method produced using in silico techniques known in the art and can be based on known structural features of the molecule. For example, a change can be made to restore tertiary structure or biological activity of the variant molecule to more closely resemble that of the original protein of interest. Such compensatory changes can also include deletion or addition of particular amino acid residues from the polypeptide.

[0058] A particularly effective means of removing epitopes from protein molecules is the concerted use of the naive T cell activation assay methods as outlined herein, together with an in silico computational method, such as the method described in WO 02/069232, which is incorporated fully herein by reference. The software simulates the process of antigen presentation at the level of the peptide MHC class II binding interaction to provide a binding score for any given peptide sequence. The binding score is determined for many of the predominant MHC class II allotypes extant in the population. As this scheme is able to test any peptide sequence, the consequences of amino acid substitutions additions or deletions with respect to the ability of a peptide to interact with a MHC class II binding groove can be predicted. Consequently, compositions can be designed, which contain reduced numbers of peptide segments that are able to interact with the MHC class II, and thereby do not function as immunogenic T cell epitopes. A biological assay using any one given donor sample can assess binding to a maximum of 4 DR allotypes. In contrast, in silico methods can evaluate the same peptide sequence using >40 allotypes simultaneously. In practice, this approach is able to direct the design of new sequence variants which are compromised in the their ability to interact with multiple MHC allotypes.

[0059] By way of an example of the utility of the combined approach to epitope identification and removal, the results of a program involving the modification of human interferon alpha (IFNα) are provided herein. The entire human IFNα sequence was rendered into a set of 51 different 15-mer peptides (listed within Table 2 of EXAMPLE 2). The T cell assay was able to preliminarily define three immunogenic regions (termed R1, R2 and R3) within the molecule, and the software system described in WO 02/069232 was able to identify predicted MHC class II ligands within each of the preliminarily determined epitopes R1-R3. Moreover, the system was further able to identify amino acid substitutions within the epitope regions, which resulted in significant loss of binding affinity between the peptide sequence and essentially all of the MHC class II allotypes represented in the system. A panel of synthetic peptides was constructed encompassing the wild-type epitope regions and variant sequences thereof in which MHC class II binding was eliminated by amino acid substitution. The peptides were used in naive T cell activation assays and the stimulation index determined for each peptide and donor PBMC sample combination. In all instances where a donor sample was found to be responsive to a wild-type peptide, the variant peptide was found not to activate T cells (**FIG. 3**). The epitope regions of IFNα were further refined in EXAMPLE 6. The refined epitope regions R1, R2, and R3 are shown in **FIG. 4**.

[0060] A preferred embodiment of the present invention is to use a modified T cell activation assay in which measurement of a T cell response is performed at different times after adding a test protein or peptide. This novel format for the assay is especially useful for detecting T cell responses in whole proteins or weakly immunogenic polypeptides. The assay format counteracts the complexity of components within the T cell assay mixture comprising a mixture of leukocytes and different molecules including cytokines. For any test protein or peptide, the kinetics of a T cell response in the assay is dependant on a number of factors including the status of T cells within the T cell assay mixture (for example, naive versus memory T cells), the concentration of cytokines at various timepoints, and the rate of generating significant T cell proliferation due to factors such as the concentration of specific peptide-MHC class II complexes. For any given protein or peptide, the peak of T cell proliferation in the assay system may peak before or after day 7 after addition of protein or peptide to the assay mixture such that, by day 7 (the standard assay timepoint), T cell proliferation is not significant. By testing for T cell proliferation over a timecourse, for example on each of days 4, 5, 6, 7, 8 and 9, then T cell responses can be detected which would not necessarily be detected at day 7. An example of a T cell assay timecourse is shown in EXAMPLE 3. For whole proteins, the T cell assay timecourse provides for a sensitive analysis of T cell immunogenicity and thus provides for a sensitive immunogenicity screen for proteins. In addition, as demonstrated in EXAMPLE 3, this assay may also be used to test for the effects of amino acid substitutions on immunogenicity.

[0061] The combined approach of using an in silico tool for the identification of MHC class II ligands and design of sequence analogues lacking MHC class II ligands, in concert with epitope mapping and re-testing using biologically based assays of T cell activation is a particularly effective method and most preferred embodiment of the invention. The general method according to this most preferred embodiment comprises the following steps:

[0062] i) use of naive T cell activation assays and synthetic peptides collectively encompassing the protein sequence of interest to identify epitope regions capable of activating T cells;

[0063] ii) use of a computational scheme simulating the binding of the peptide ligand with one or more MHC allotypes to analyse the epitope regions identified in step (i) and thereby identify MHC class II ligands within the epitope region;

[0064] iii) use of a computational scheme simulating the binding of the peptide ligand with one or more MHC allotypes to identify sequence analogues of the MHC ligands encompassed within the epitope region(s) which no longer bind MHC class II or bind with lowered affinity to a lesser number of MHC allotypes;

[0065] iv) use of naive T cell activation assays and synthetic peptides encompassing entirely or in collection encompassing the epitope regions identified within the protein of interest and testing the sequence ana-

logues in naive T cell activation assay in parallel with the wild-type (parental) sequences;

[0066] It is understood that the software outlined in WO 02/069232 can also be used to define with a high degree of certainty the dataset of all peptides comprising the universe of permissible MHC class ligands for the any human protein such as IFNα. For reasons such as the requirement for proteolytic processing and other physiologic steps leading to the presentation of immunogenic peptides in vivo, it would be clear that a relatively minor sub-set of the entire repertoire of peptides will have ultimate biological relevance. In such situations the inventors have established that ex vivo human T cell activation assays may be used to identify the biologically relevant peptides. Accordingly, synthetic peptides are tested for their ability to evoke a proliferative response in human T cell cultured in vitro. Where this type of approach is conducted using naive human T cells taken from healthy donors, the inventors have established that in the operation of such an assay, a stimulation index equal to or greater than 2 is a useful measure of induced proliferation. The stimulation index (SI) is conventionally derived by division of the proliferation score (e.g. counts per minute of radioactivity if using for example ^{3}H-thymidine incorporation) measured to the test peptide by the score measured in cells not contacted with a test peptide. Peptides which evoke no response give SI=1 although in practice SI values in the range 0.8-1.2 are unremarkable. A number of technical procedures can be incorporated into the operation of such assays in order to ensure confidence in the recorded scores. Typically all determinations are made at least in triplicate and the mean score is computed. In cases where the computed SI is greater than or equal to 2, individual scores of the triplicate repetitions of the assay can be examined for evidence of outlying data. Similarly, control peptides for which there is expectation that the majority of PBMC donor samples will be responsive can be included in each assay plate. The influenza hemagglutinin peptide 307-319, sequence PKYVKQNTLKLA (SEQ ID NO: 114); and the Chlamydia HSP 60 peptide sequence KVVDQIK-KISKPVQH (SEQ ID NO: 115) are particularly suitable control peptides although many other examples may be exploited. Assays should preferably also use a potent whole protein antigen such as hemocyanin from Keyhole Limpet to which all PBMC samples would be expected to exhibit an SI significantly greater than 2.

[0067] According to the methods of the present invention there may be a practical need to test multiple versions of essentially the same peptide sequence in order to establish that the modification, be it a single amino acid substitution or some other change or combination of changes, results in the loss of ability or at least a reduced ability for the peptide(s) to induce a T cell activation effect. This requirement may be met using a number of different practical approaches, one of which could involve the screening of large numbers of variant peptides from the outset and conducting a selection method to identify those in which there is a reduced or absent ability to induce proliferation relative to their parental (e.g. wild-type) peptide sequence. Such an approach could be conducted entirely using naive PBMC samples and run concurrently (i.e. in parallel with) the mapping exercise. It is understood that this approach need not be limited to the screening of synthetic peptide species but may be exploited to the screening of whole protein molecules that for example may comprise a multiplicity of variants produced as a "library" of variants from which a desired member is to be selected. Such a library may be produced for example by recombinant means well known in the art or may comprise species produced using synthetic means for example using the principles of combinatorial chemistry. In any event, the desired property to be selected from the library member in this context would be the inability to induce a proliferative response in a PBMC preparation.

[0068] Alternatively variant peptides may be screened using naive PBMC from entirely different donor pool of samples, i.e. epitope mapping is repeated but using modified peptides where there is an expectation for little or no proliferative induction.

[0069] A further and particularly favoured scheme would involve the testing of modified peptides for their ability to induce a proliferative effect in an immunological recall assay format. This may be achieved for example using PBMC from a known responding donor identified during the initial naive PBMC assay phase and stimulating a sample of those cells using a either synthetic peptides (e.g. in a pool) or whole protein followed by a suitable period of culture in the presence of cytokines such as IL-2. Following this incubation, the culture may be re-stimulated using the synthetic (modified) peptide or modified whole protein of interest and the proliferative effect measured using any suitable means. The inventors have classified this assay format as a "recall" assay, so called as the T cell population responsible for the proliferative response is invoked during a re-stimulation phase.

[0070] The recall type assay is particularly useful in identifying T cell epitopes in protein or peptide antigens that show weak immunogenicity in vivo and can provide corroborating evidence for the existence of a T cell epitope in a given amino acid sequence where the epitope was originally identified by other means, for example by using computational techniques or biological assays. In the operation of such a recall assay, PBMC are isolated from healthy donors or patients with established immune responses to a given therapy. It is necessary to freeze aliquots of autologous PBMC so that they can be used as antigen presenting cells (APC) during subsequent procedures. The assay commences with an antigen priming step. A typical and preferred protocol requires that 2-4×10^{6} PBMC be added to each well of 24 well plate. Either whole protein or peptide antigen, or a peptide pool, is added to the cells at typical concentrations of 1-10 μg/ml and 1-10 μM, respectively (total concentration of peptides in peptide pool would be 1 μM). The final culture volume is 2 ml. The cells are incubated for 7 days where on day 7 10 U/ml IL-2 is added and the cells are incubated for a further 3 days whereupon the cells are ready for the antigen re-challenge phase.

[0071] The antigen re-challenge requires autologous PBMC as APC. The APC are incubated with whole protein or synthetic peptide antigen (for example at a concentration of 1-10 ug/ml) for 1 hour at 37° C. The proliferative capability of the APC is destroyed most preferably using gamma radiation, for example 4000 rads in a round bottom 96 well plate (1×10^{5} PBMC/well). 1-10×10^{4} primed T cells are added to each well containing the APC's. It is important to set up untreated control reactions comprising antigen primed T cells cultured with gamma irradiated APC in the

absence of re-challenge antigen. The cells are incubated for 4 days before pulsing proliferation assessment for example 3H-thymidine incorporation assay. It is understood that such a protocol can equally be conducted using enriched or purified populations of cells.

[0072] In a third embodiment, there is provided a method whereby a protein antigen may be screened for the presence of determinants within its sequence capable of evoking a T cell immune response in individuals for whom the protein of interest is to be administered for therapeutic effect against a genetic (constitutional) disease and where, in effect, the protein antigen due to the nature of the genetic deficit in the individuals will constitute a foreign protein. In this sense, the protein is most likely to represent a potent antigen in vivo and the inventors have established that it is now readily possible to establish polyclonal or mononclonal T cell lines in vitro from the PBMC of such individuals and these lines may be used as effective reagents in the mapping of T cell epitopes within proteins. This is achieved in essentially the same way as the recall assay of the foregoing, with the exception that the T cells are subjected to several rounds of antigen stimulation in vitro followed immediately by expansion in the presence of IL-2. For establishing polyclonal T cell lines 2-3 rounds of antigen stimulation are generally sufficient to generate a large number of antigen specific cells. These are used to screen large numbers of synthetic peptides (for example in the form of peptide pools), and they may be cryogenically stored to be used at a later date. After the initial round of antigen stimulation comprising co-incubation of the antigen and PBMC for 7 days subsequent re-challenges with antigen are performed in the presence of most preferably autologous irradiated PBMC as antigen presenting cells. These rounds of antigen selection are performed for 3-4 days and are interspersed by expansion phases comprising stimulation with IL-2 which may be added every 3 days for a total period of around 9 days. The final re-challenge is performed using T cells that have been "rested", that is T cells which have not been IL-2 stimulated for around 4 days. These cells are stimulated with antigen (e.g. synthetic peptide or whole protein) using most preferably autologous antigen presenting cells as previously for around 4 days and the subsequent proliferative response (if any) is measured thereafter.

[0073] Accordingly, the method of the third embodiment comprises the production of T cell lines or oligoclonal cultures derived from PBMC samples taken from an individual afflicted with the disease of interest, stimulating in vitro said lines or cultures with preparations of synthetic peptides or whole proteins and measuring in vitro the proliferative effect if any of individual synthetic peptides or proteins, producing modified variants of individual synthetic peptides or whole proteins and re-testing said modified peptides or proteins for a continued ability to promote a significant proliferative response in the T cell lines or cultures.

[0074] It is particularly useful to establish T cell lines of oligoclonal cultures from individuals who carry the genetic defect, and in whom therapeutic replacement therapy has been initiated, and in whom the replacement therapy has resulted in the induction of an immune response to the therapeutic protein. A prominent example of this kind of subject is provided by individuals undergoing treatment for hemophillia A, but in whom there is a significant titre of inhibitory antibodies measurable to the therapeutic Factor VIII. Under the scheme of the present invention it would be particularly desired to exploit PBMC samples from this class of so called "inhibitor patients" inasmuch as the epitope map of the Factor VIII protein defined by the T cell repertoire of a significant number of these individuals represents the most prevalent peptide epitopes that are capable of presentation in the in vivo context. In this sense, PBMC from patients in whom there is a previously demonstrated immune response constitute the products of an in vivo priming step and are particularly valuable under the scheme of the present. EXAMPLE 4 herein provides detailed description of an epitope mapping programme conducted on human FVIII exploiting both naive human T cells from healthy donors and PBMCs derived from hemophilia A patients.

[0075] Given that the use of PBMC cell lines from individuals previously in receipt of the immunologically foreign protein is in principle a recall assay, it further provides the practical benefit of there being the capacity for a much larger magnitude of proliferative response to any given stimulating peptide or protein. This reduces the technical challenge of conducting a proliferation measurement and in such a situation may give the opportunity for definition of a possible hierarchy of immunodominant epitopes where multiple epitopes are uncovered to a target protein. This is certain to be the case with particularly large proteins such as Factor VIII although as demonstrated herein, small human protein molecules (e.g. less than 200 amino acid residues) may be expected to harbour multiple or complex (i.e. overlapping) T cell epitopes.

[0076] In a fourth embodiment of the present invention there is provided a scheme whereby the assay format of the foregoing is applied to the screening of production batches of therapeutic biological proteins. The objective of such a screening process is to confirm the consistency of the immunogenic profile of the test biologic and for example may be particularly valuable in situations where the production process for the biologic has been altered by some parameter and although the measured physical properties of the protein may be within accepted ranges, there is a consideration that the potential immunogenic properties of the protein may have been altered. Thus, in order to anticipate the generation of an immunogenic response to any new preparation of the molecule of interest the methods set-out herein are particularly effective in providing such a screening procedure.

[0077] Under the fourth embodiment therefore, T cell lines (polyclonal or mono-clonal) derived as part of the epitope mapping process for the protein of interest, or optionally and in addition, a panel of naive PBMC preparations for which there has been established a population of known responsive preparations, may be used to test the subject protein for immunogenicity in vitro, and the responses scored to the test protein are compared to a reference or "gold-standard" preparation of the protein. In this regard where T cell lines are employed, it is particularly preferred to use lines derived from subjects in whom there has been a demonstrated previous immune response to the reference protein. Such lines provide a high stimulation index score on antigen challenge in vitro and are likely to be representative of the most biologically relevant and immunodominant epitopes within the protein. These lines under the fourth embodiment provide indicators for epitope loss/alteration. By contrast,

under the fourth embodiment, panels of naive PBMC containing a known set of responding allotypes to the target protein provide indication of de novo epitope generation appearing in the test product protein and are equally valuable in predicting an unwanted clinical immunogenic response.

[0078] The term "T cell epitope" as used herein and in the appended claims means an amino acid sequence which is able to bind MHC class II, able to stimulate T cells and/or also to bind (without necessarily measurably activating) T cells in complex with MHC class II.

[0079] The term "peptide" as used herein and in the appended claims, is a compound that includes two or more amino acids linked by peptide bonds. The amino acids are linked together by a peptide bond (defined herein below). There are 20 different naturally occurring amino acids involved in the biological production of peptides, and any number of them may be linked in any order to form a peptide chain or ring. The naturally occurring amino acids employed in the biological production of peptides all have the L-configuration. Synthetic peptides can be prepared employing conventional synthetic methods, utilizing L-amino acids, D-amino acids, or various combinations of amino acids of the two different configurations. Some peptides contain only a few amino acid units. Short peptides, e.g., having less than ten amino acid units, are sometimes referred to as "oligopeptides". Other peptides contain a large number of amino acid residues, e.g. up to 100 or more, and are referred to as "polypeptides". By convention, a "polypeptide" may be considered as any peptide chain containing three or more amino acids, whereas a "oligopeptide" is usually considered as a particular type of "short" polypeptide. Thus, as used herein, it is understood that any reference to a "polypeptide" also includes an oligopeptide. Further, any reference to a "peptide" includes polypeptides, oligopeptides, and proteins. Each different arrangement of amino acids forms different polypeptides or proteins. The number of polypeptides—and hence the number of different proteins—that can be formed is practically unlimited.

[0080] The following Examples are provided as illustrations of various embodiments of the present invention and are not to be construed as limiting the scope of the invention.

EXAMPLE 1

[0081] The interaction between MHC, peptide and T cell receptor (TCR) provides the structural basis for the antigen specificity of T cell recognition. T cell proliferation assays test the binding of peptides to MHC and the recognition of MHC/peptide complexes by the TCR. In vitro T cell proliferation assays of the present example, involve the stimulation of peripheral blood mononuclear cells (PBMCs), containing antigen presenting cells (APCs) and T cells. Stimulation is conducted in vitro using synthetic peptide antigens, and in some experiments whole protein antigen. Stimulated T cell proliferation preferably is measured using ^{3}H-thymidine (^{3}H-Thy) and the presence of incorporated ^{3}H-Thy assessed using scintillation counting of washed fixed cells.

[0082] Buffy coats from human blood stored for less than 12 hours were obtained from the National Blood Service (Addenbrooks Hospital, Cambridge, UK). Ficoll-paque was obtained from Amersham Pharmacia Biotech (Amersham, UK). Serum free AIM V media for the culture of primary human lymphocytes and containing L-glutamine, 50 g/ml streptomycin, 10 μg/ml gentomycin and 0.1% human serum albumin was from Gibco-BRL (Paisley, UK). Synthetic peptides were obtained from Pepscan (The Netherlands) and Babraham Technix (Cambridge, UK).

[0083] Erythrocytes and leukocytes were separated from plasma and platelets by gentle centrifugation of buffy coats. The top phase (containing plasma and platelets) was removed and discarded. Erythrocytes and leukocytes were diluted 1:1 in phosphate buffered saline (PBS) before layering onto 15 ml ficoll-paque (Amersham Pharmacia, Amersham UK). Centrifugation was done according to the manufacturers recommended conditions and PBMCs were harvested from the serum+PBS/ficoll plaque interface. PBMCs were mixed with PBS (1:1) and collected by centrifugation. The supernatant was removed and discarded and the PBMC pellet resuspended in 50 ml PBS. Cells were again pelleted by centrifugation and the PBS supernatant discarded. Cells were resuspended using 50 ml AIM V media and at this point counted and viability assessed using trypan blue dye exclusion. Cells were again collected by centrifugation and the supernatant discarded. Cells were resuspended for cryogenic storage at a density of 3×10^{7} per ml. The storage medium was 90% (v/v) heat inactivated AB human serum (Sigma, Poole, UK) and 10% (v/v) DMSO (Sigma, Poole, UK). Cells were transferred to a regulated freezing container (Sigma) and placed at −70° C. overnight before transferring to liquid N_2 for long term storage. When required for use, cells were thawed rapidly in a water bath at 37° C. before transferring to 10 ml pre-warmed AIM V medium.

[0084] PBMC were stimulated with protein and peptide antigens in a 96 well flat bottom plate at a density of 2×10^{5} PBMC per well. PBMC were incubated for 7 days at 37° C. before pulsing with ^{3}H-Thy (Amersham-Phamacia, Amersham, UK). For the present study, synthetic peptides (15mers) that overlapped by increments of 12 amino acids were generated that spanned the entire sequence of IFNβ. Peptide identification numbers (ID#) and sequences are given in Table 1.

TABLE 1

IFNβ peptides

Peptide ID Number	IFNβ-1a; 15mer sequence	SEQ ID NO:
1	MSYNLLGFLQRSSNF	1
2	NLLGFLQRSSNFQCQ	2
3	GFLQRSSNFQCQKLL	3
4	QRSSNFQCQKLLWQL	4
5	SNFQCQKLLWQLNGR	5
6	QCQKLLWQLNGRLEY	6
7	KLLWQLNGRLEYCLK	7
8	WQLNGRLEYCLKDRM	8
9	NGRLEYCLKDRMNFD	9

TABLE 1-continued

IFNβ peptides

Peptide ID Number	IFNβ-1a; 15mer sequence	SEQ ID NO:
10	LEYCLKDRMNFDIPE	10
11	CLKDRMNFDIPEEIK	11
12	DRMNFDIPEEIKQLQ	12
13	NFDIPEEIKQLQQFQ	13
14	IPEEIKQLQQFQKED	14
15	EIKQLQQFQKEDAAL	15
16	QLQQFQKEDAALTIY	16
17	QFQKEDAALTIYEML	17
18	KEDAALTIYEMLQNI	18
19	AALTIYEMLQNIFAI	19
20	TIYEMLQNIFAIFRQ	20
21	EMLQNIFAIFRQDSS	21
22	QNIFAIFRQDSSSTG	22
23	FAIFRQDSSSTGWNE	23
24	FRQDSSSTGWNETIV	24
25	DSSSTGWNETIVENL	25
26	STGWNETIVENLLAN	26
27	WNETIVENLLANVYH	27
28	TIVENLLANVYHQIN	28
29	ENLLANVYHQINHLK	29
30	LANVYHQINHLKTVL	30
31	VYHQINHLKTVLEEK	31
32	QINHLKTVLEEKLEK	32
33	HLKTVLEEKLEKEDF	33
34	TVLEEKLEKEDFTRG	34
35	EEKLEKEDFTRGKLM	35
36	LEKEDFTRGKLMSSL	36
37	EDFTRGKLMSSLHLK	37
38	TRGKLMSSLHLKRYY	38
39	KLMSSLHLKRYYGRI	39
40	SSLHLKRYYGRILHY	40
41	HLKRYYGRILHYLKA	41
42	RYYGRILHYLKAKEY	42
43	GRILHYLKAKEYSHC	43
44	LHYLKAKEYSHCAWT	44

TABLE 1-continued

IFNβ peptides

Peptide ID Number	IFNβ-1a; 15mer sequence	SEQ ID NO:
45	LKAKEYSHCAWTIVR	45
46	KEYSHCAWTIVRVEI	46
47	SHCAWTIVRVEILRN	47
48	AWTIVRVEILRNFYF	48
49	IVRVEILRNFYFINR	49
50	VEILRNFYFINRLTG	50
51	LRNFYFINRLTGYLR	51

[0085] Each peptide was screened individually against PBMC's isolated from 20 naive donors. Two control peptides that have previously been shown to be immunogenic and a potent non-recall antigen KLH were used in each donor assay. The control antigens used in this study were Flu haemagglutinin 307-319 (sequence: PKYVKQNTLKLAT; SEQ ID NO: 114); Chlamydia HSP 60 peptide (sequence: KVVDQIKKISKPVQH; SEQ ID NO: 115) and Keyhole Limpet hemocyanin.

[0086] Peptides were dissolved in DMSO to a final concentration of 10 mM, these stock solutions were then diluted 1/500 in AIM V media (final concentration 20 μM). Peptides were added to a flat bottom 96 well plate to give a final concentration of 2 and 20 μM in a 100 μl. The viability of thawed PBMC's was assessed by trypan blue dye exclusion, cells were then resuspended at a density of 2×10^6 cells/ml, and 100 μl (2×10^5 PBMC/well) was transferred to each well containing peptides. Triplicate well cultures were assayed at each peptide concentration. Plates were incubated for 7 days in a humidified atmosphere of 5% CO_2 at 37° C. Cells were pulsed for 18-21 hours with 1 μCi ^{3}H-Thy/well before harvesting onto filter mats. CPM values were determined using a Wallac microplate beta top plate counter (Perkin Elmer). Results were expressed as stimulation indices, where the stimulation index (SI) is derived by division of the proliferation score (e.g. counts per minute of radioactivity) measured to the test peptide by the score measured in cells not contacted with a test peptide.

[0087] Mapping T cell epitopes in the IFNβ sequence using the T cell proliferation assay resulted in the identification of two immunogenic regions R1 and R2 resulting, in each case, by responses to four overlapping peptides (**FIG. 1**).

EXAMPLE 2

[0088] An epitope map for the human protein interferon α2 (IFNα) was derived using the method of EXAMPLE 1. In all respects the method was as per EXAMPLE 1 except that synthetic peptides were as given in Table 2 (below) and incubation with the PBMC preparations was at a concentration of 10 μM.

[0089] Mapping T cell epitopes in the IFNα sequence resulted in the initial, preliminary identification of three

immunogenic regions R1, R2, R3. This was determined by T cell proliferation to seven, four and five overlapping peptides respectively as shown in **FIG. 2**. Region 3 is considered to contain a potential immunodominant T cell epitope as proliferation is scored in two thirds of donors that responded to IFNα peptides.

TABLE 2

IFNα peptides

Peptide ID Number	IFNα2b; 15mer sequence	SEQ ID NO:
1	CDLPQTHSLGSRRTL	52
2	PQTHSLGSRRTLMLL	53
3	HSLGSRRTLMLLAQM	54
4	GSRRTLMLLAQMRRI	55
5	RTLMLLAQMRRISLF	56
6	MLLAQMRRISLFSCL	57
7	AQMRRISLFSCLKDR	58
8	RRISLFSCLKDRHDF	59
9	SLFSCLKDRHDFGFP	60
10	SCLKDRHDFGFPQEE	61
11	KDRHDFGFPQEEFGN	62
12	HDFGFPQEEFGNQFQ	63
13	GFPQEEFGNQFQKAE	64
14	QEEFGNQFQKAETIP	65
15	FGNQFQKAETIPVLH	66
16	QFQKAETIPVLHEMI	67
17	KAETIPVLHEMIQQI	68
18	TIPVLHEMIQQIFNL	69
19	VLHEMIQQIFNLFST	70
20	EMIQQIFNLFSTKDS	71
21	QQIFNLFSTKDSSAA	72
22	FNLFSTKDSSAAWDE	73
23	FSTKDSSAAWDETLL	74
24	KDSSAAWDETLLDKF	75
25	SAAWDETLLDKFYTE	76
26	WDETLLDKFYTELYQ	77
27	TLLDKFYTELYQQLN	78
28	DKFYTELYQQLNDLE	79
29	YTELYQQLNDLEACV	80
30	LYQQLNDLEACVIQG	81
31	QLNDLEACVIQGVGV	82

TABLE 2-continued

IFNα peptides

Peptide ID Number	IFNα2b; 15mer sequence	SEQ ID NO:
32	DLEACVIQGVGVTET	83
33	ACVIQGVGVTETPLM	84
34	IQGVGVTETPLMKED	85
35	VGVTETPLMKEDSIL	86
36	TETPLMKEDSILAVR	87
37	PLMKEDSILAVRKYF	88
38	KEDSILAVRKYFQRI	89
39	SILAVRKYFQRITLY	90
40	AVRKYFQRITLYLKE	91
41	KYFQRITLYLKEKKY	92
42	QRITLYLKEKKYSPC	93
43	TLYLKEKKYSPCAWE	94
44	LKEKKYSPCAWEVVR	95
45	KKYSPCAWEVVRAEI	96
46	SPCAWEVVRAEIMRS	97
47	AWEVVRAEIMRSFSL	98
48	VVRAEIMRSFSLSTN	99
49	AEIMRSFSLSTNLQE	100
50	MRSFSLSTNLQESLR	101
51	FSLSTNLQESLRSKE	102

EXAMPLE 3

[0090] Protocol for Conducting a Time Course T Cell Activation Assay

[0091] A general protocol for conducting a time course T cell activation assay comprises the following steps:

[0092] 1. Thaw 1 vial of PBMC per donor

[0093] 2. Resuspend cells at 2-4×10^6 cells/ml (in AIM V).

[0094] 3. Transfer 1 ml to 3 wells of a 24 well plate (giving a final concentration of 2-4×10^6 PBMC/well), since it is usual to test the antigen at two different concentrations and compare against a non-antigen treated control (e.g. 10-50 μg/ml protein or 1-5 μM peptide).

[0095] 4. Make stock solutions of antigens typically 100 μg/ml for proteins and 2-10 μM for peptides. Add 1 ml of antigen to each well to give a final concentration 10-50 μg/ml protein or 1-5 μM peptide.

[0096] 5. Incubate for 5 days.

[0097] 6. Gently resuspend the cells in the 2 ml cultures by pipetting and from each condition remove 100 μl cells and place into a well of 96 well plate (round bottom), repeat this three time of reach culture condition (total of 300 ul removed from each culture condition per time point).

[0098] 7. To each well of cells in the 96 well plate add 1 μCi/well 3H[Thy] in 100 ul AIM V.

[0099] 8. Incubate overnight and harvest.

[0100] 9. Repeat stage 6-8 for days 6, 7, and 8 (day 9 can be included if necessary).

[0101] 10. Make SI determinations and plot the SI versus time for each antigen.

[0102] **FIG. 3** shows typical results for the timecourse assay for immunogenicity of long peptides spanning the immunogenic regions of interferon α2 (compare EXAMPLE 2). This novel timecourse method is especially useful for analysis of whole proteins as a screen for T cell immunogenicity (SI's>1.8) and to analyse the effects on immunogenicity of amino acid modifications within the protein.

EXAMPLE 4

[0103] Method for Establishment of T Cell Lines and Clones.

[0104] Peripheral blood mononuclear cells (PBMC) were isolated from blood obtained from hemophiliac patients, and cryogenically stored under liquid nitrogen. Blood samples were provided with fully informed consent and working under local ethical approval of the Addenbrooke's Health Care Trust.

[0105] T cell lines were established by stimulating antigen specific T cells in bulk cultures using FVIII followed by several cycles of IL-2 induced expansion. Initially PBMC were incubated (at 37° C. in a humidified atmosphere of 5% CO_2) at 2×10^6 in 2 ml AIM V media containing 4 ug/ml FVIII (Refacto™) in 24 well plates. After 7 days incubation 100 U/ml IL-2 was added and cultures were incubated for further 3 days.

[0106] T-blasts were collected and counted upon completion of the 10 day antigen/IL-2 stimulation. In order to retain antigen specificity T blasts were subjected to a second round of antigen stimulation using γ-irradiated autologous PBMC as antigen presenting cells. This was achieved by incubating 1×10^6 autologous PBMC/well in a 24 well plate with 4 μg/ml FVIII for 1 hour in 0.75 ml AIM V (containing 5% heat inactivated human AB serum) before being subjected to 4000 rads γ-irradiation. Autologous T blasts were added in 0.25 ml AIM V at 4×10^5 cells/ml to the γ-irradiated antigen presenting cells (pre-loaded with FVIII) and incubated for 3 days. T blasts were expanded by stimulating cells with 100 U/ml IL-2 for 3 days; cultures were then supplied with fresh IL-2 (final concentration of 100 U/ml) at 3 day intervals for a total of 9 days. To ensure that all expanded T blasts were antigen specific a third round of antigen stimulation was performed, where T blasts were collected and resuspended at 4×10^5 cells/ml in AIM V media. As described before antigen presenting cells were generated by incubating 1×10^6 γ-irradiated autologous PBMC in a 24 well plate with 4 μg/ml FVIII for 1 hour in 0.75 ml AIM V (containing 5% heat inactivated human AB serum). Autologous T blasts in 0.25 ml AIM V at 4×10^5 cells/ml were added to the γ-irradiated antigen presenting cells and incubated for 3 days. A final expansion in 10 U/ml IL-2 was performed 3 days before T blasts were collected and used to screen peptide pools.

[0107] Cloning from Bulk Cultures

[0108] After the third stimulation with FVIII antigen T blasts were collected and resuspended by serial dilution to a density of 4×10^2-1×10^4 cells/ml (2× final culture density). Autologous PBMC were thawed and resuspended to 2×10^6 cells/ml (2× final culture density) in a polyproplene tube. PBMC were then exposed to 4000 rads of γ-radiation and were used as antigen presenting cells to select antigen reactive T cell clones by limiting dilution. The γ-irradiated antigen presenting cells (1×10^6 final density) were mixed with the T blasts (2×10^2-5×10^3 final density), 1-10 μg/ml FVIII antigen and 100 U/ml IL-2. T cell clones were established in Terasaki plates by adding 20 μl of the APC, T blast, FVIII and IL-2 mixture to each well. Limiting dilution cloning was performed using 2-50 T blasts/well of a Terasaki plate.

[0109] Selection and Maintenance of T Cell Clones

[0110] T blasts were incubated with FVIII antigen, IL-2 and γ-irradiated autologous antigen presenting cells for approximately 14 days. After identifying wells that contained cells showing unequivocal growth, T blasts were transferred to a single well of a round bottom 96 well plate containing 1×10^5 γ-irradiated allogenic PBMC, 100 U/ml IL-2 and 1 μg/ml phytohaemaglutinin (PHA) in a final volume 200 μl AIM V (with 1% heat inactivated human AB serum). T cell clones were split when cells became confluent, and ultimately transferred to a single well of 24 well plate containing 1×10^6 γ-irradiated allogenic PBMC (feeder cells), 100 U/ml IL-2 and 1 μg/ml phytohaemaglutinin (PHA) in a final volume of 2 ml AIM V (with 1% heat inactivated human AB serum). Routine maintenance of T cell clones involved stimulation with fresh PHA and allogenic feeder cells every 2-3 weeks (depending on cell growth) and twice weekly stimulation with 100 U/ml IL-2. Only T cell clones that proved to be FVIII specific were expanded and used to screen FVIII peptides.

[0111] EBV Transformation of Autologous B Cells.

[0112] B cells from PBMC preparations were immortalized to generate B lymphoblastoid cell lines (BLCL) by adding 3 ml of filtered (0.45μ) B95.8 supernatant to 4×10^6 PBMC and incubating at 37° C. for 1 hour. PBMC were pelleted and resuspended in 2 ml RPMI containing 5% heat-inactive fetal calf serum (FCS) and 1 μg/ml cyclosporin A. After 7 days incubation 1 ml of culture media was replaced with fresh RPMI containing 5% FCS and 2 μg/ml cyclosporin A (to give a final concentration of 1 μg/ml cyclosporin A). This feeding regime was repeated on days 14 and 21 after which cells were split when necessary using RPMI containing 5% FCS and expanded into tissue culture flasks.

[0113] Screening FVIII Peptides Using T Cell Lines/Clones

[0114] Peptides of 15 residues in length and overlapping with the previous peptide by increments of 12 amino acids were synthesized (Pepscan, Netherlands). Peptides were

initially solubilized at 10 mM in 100% dimethylsulphoxide (DMSO) for storage. Peptide pools were generated to simultaneously screen a large number of peptides against FVIII specific T cell lines. Pools were organized such that each pool contained overlapping peptides of subsequent pools by using this approach T cell epitopes that overlap two peptides will result in inducing proliferation two separate pools. Each pool typically consisted of 8 peptides with each peptide being tested at either 1 or 5 μM.

[0115] Autologous PBMC (for T cell lines) or EBV transformed BLCL (for T cell clones) were used as antigen presenting cells by re-suspending 1×10^5 PBMC or BLCL in 50 μl AIM V media which was then added to each well of a round bottom 96 well plate. Peptide pools were added in triplicate wells for each pool at both concentrations (1 or 5 μM). Antigen presenting cells and peptide pools were incubated for 1 hour at 37° C. before exposure to 4000 rads γ-irradiation. BLCL were pre-treated with 1 μg/ml Mitomycin C for 1 hour at 37° C. followed by washing 4 times in AIM V when used as antigen presenting cells (instead of γ-irradiated autologous PBMC) for T cell clones. Antigen specific T cell lines or T cell clones were then added at 5×10^4 cells per well and the cultures were incubated for 3 days. On the third day each well was pulsed with 1 μCi [^{3}H]-Thymidine for a minimum of 8 hours. After harvesting the plates onto filtermats the cpm/well was determined using a Wallac Microplate Beta counter.

[0116] Naive T Cell Epitope Map Using PBMC from Healthy Donors

[0117] Blood from 40 healthy HLA-DR typed donors was used to isolate PBMC which were used to screen individual FVIII peptides at two concentrations (1 and 5 μM). Since there were insufficient numbers of PBMC from each donor to screen all FVIII peptides, donors were split into two groups where the first 20 donors were used to screen peptides spanning the first half of the molecule and the second set of donors used to screen the remaining peptides. Donors were selected according to MHC class II allotypes expressed in order to cover a large number of allotypes present in the world population. MHC allotypes were detected using the tissue types for all PBMC samples were assayed using a commercially available reagent system (Dynal, Wirral, UK). Assays were conducted in accordance with the suppliers recommended protocols and standard ancillary reagents and agarose electrophoresis systems. PBMC contain physiological numbers of naive T cells and antigen presenting cells. These cells were used at a density of 2×10^5 cells/well (96 flat bottom plate) to screen peptides at 1 and 5 μM in triplicate 200 μl cultures. Cells were incubated with peptides at 37° C. for 6 days before pulsing each well with 1 μCi [^{3}H]-Thymidine for a minimum of 8 hours. Cultures were harvested onto filtermats and the cpm/well was determined using a Wallac Microplate Beta counter.

[0118] TABLE 3 shows an epitope map for human B-domain deleted FVIII generated using T cell lines from haemophiliacs and naive T cell preparations from healthy individuals. Where T blasts and naive PBMC derived T cells were used to identify peptide pools containing T cell epitopes, those pools were then decoded to identify the individual peptide containing the T cell epitope.

TABLE 3

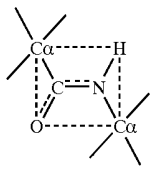
Residue #*	Peptide Sequence	SEQ ID NO:
196	ILLFAVFDEGKSWSH	103
406	SYKSQYLNNGPQRIG	104
415	GPQRIGRKYKKVRFM	105
511	YKWTVTVRDGPTKSD	106
610	ASNIMHSINGYVFDS	107
634	VAYWYILSIGAQTDF	108
817	MSSSPHVLRNRAQSG	109
1009	CNIQMEDPTFKENYR	110
1117	STLFLVYSNKCQTPL	111
1204	ISQFIIMYSLDGKKW	112
1251	IARYIRLHPTHYSIRSTLRM	113

*Sequence numbering according to B domain deleted sequence

EXAMPLE 5

[0119] Computational Scheme

[0120] There are a number of factors that play important roles in determining the total structure of a protein or polypeptide. First, the peptide bond, i.e., that bond which joins the amino acids in the chain together, is a covalent bond. This bond is planar in structure, essentially a substituted amide. An "amide" is any of a group of organic compounds containing the grouping —CONH—. The planar peptide bond linking Cα of adjacent amino acids may be represented as depicted below:



[0121] Because the O═C and the C—N atoms lie in a relatively rigid plane, free rotation does not occur about these axes. Hence, a plane schematically depicted by the interrupted line is sometimes referred to as an "amide" or "peptide plane" plane wherein lie the oxygen (O), carbon (C), nitrogen (N), and hydrogen (H) atoms of the peptide backbone. At opposite corners of this amide plane are located the Cα atoms. Since there is substantially no rotation about the O═C and C—N atoms in the peptide or amide plane, a polypeptide chain thus comprises a series of planar peptide linkages joining the Cα atoms.

[0122] A second factor that plays an important role in defining the total structure or conformation of a polypeptide or protein is the angle of rotation of each amide plane about the common Cα linkage. The terms "angle of rotation" and "torsion angle" are hereinafter regarded as equivalent terms. Assuming that the O, C, N, and H atoms remain in the amide

plane (which is usually a valid assumption, although there may be some slight deviations from planarity of these atoms for some conformations), these angles of rotation define the N and R polypeptide's backbone conformation, i.e., the structure as it exists between adjacent residues. These two angles are known as ϕ and ψ. A set of the angles ϕ_1, ψ_1, where the subscript i represents a particular residue of a polypeptide chain, thus effectively defines the polypeptide secondary structure. The conventions used in defining the Φ, ψ angles, i.e., the reference points at which the amide planes form a zero degree angle, and the definition of which angle is ϕ, and which angle is ψ, for a given polypeptide, are defined in the literature. See, e.g, Ramachandran et al. *Adv. Prot. Chem.* 23:283-437 (1968), at pages 285-94, which pages are incorporated herein by reference.

[0123] The present method can be applied to any protein, and is based in part upon the discovery that in humans the primary Pocket 1 anchor position of MHC Class II molecule binding grooves has a well designed specificity for particular amino acid side chains. The specificity of this pocket is determined by the identity of the amino acid at position 86 of the beta chain of the MHC Class II molecule. This site is located at the bottom of Pocket 1 and determines the size of the side chain that can be accommodated by this pocket. Marshall, K. W., *J. Immunol.,* 152:4946-4956 (1994). If this residue is a glycine, then all hydrophobic aliphatic and aromatic amino acids (hydrophobic aliphatics being: valine, leucine, isoleucine, methionine and aromatics being: phenylalanine, tyrosine and tryptophan) can be accommodated in the pocket, a preference being for the aromatic side chains. If this pocket residue is a valine, then the side chain of this amino acid protrudes into the pocket and restricts the size of peptide side chains that can be accommodated such that only hydrophobic aliphatic side chains can be accommodated. Therefore, in an amino acid residue sequence, wherever an amino acid with a hydrophobic aliphatic or aromatic side chain is found, there is the potential for a MHC Class II restricted T cell epitope to be present. If the side-chain is hydrophobic aliphatic, however, it is approximately twice as likely to be associated with a T cell epitope than an aromatic side chain (assuming an approximately even distribution of Pocket 1 types throughout the global population).

[0124] A computational method embodying the present invention profiles the likelihood of peptide regions to contain T cell epitopes as follows:

[0125] (1) The primary sequence of a peptide segment of predetermined length is scanned, and all hydrophobic aliphatic and aromatic side chains present are identified; (2) The hydrophobic aliphatic side chains are assigned a value greater than that for the aromatic side chains; preferably about twice the value assigned to the aromatic side chains, e.g., a value of 2 for a hydrophobic aliphatic side chain and a value of 1 for an aromatic side chain; (3) The values determined to be present are summed for each overlapping amino acid residue segment (window) of predetermined uniform length within the peptide, and the total value for a particular segment (window) is assigned to a single amino acid residue at an intermediate position of the segment (window), preferably to a residue at about the midpoint of the sampled segment (window). This procedure is repeated for each sampled overlapping amino acid residue segment (window). Thus, each amino acid residue of the peptide is assigned a value that relates to the likelihood of a T cell epitope being present in that particular segment (window); (4) The values calculated and assigned as described in Step 3, above, can be plotted against the amino acid coordinates of the entire amino acid residue sequence being assessed; (5) All portions of the sequence which have a score of a predetermined value, e.g., a value of 1, are deemed likely to contain a T cell epitope and can be modified, if desired.

[0126] This particular aspect of the present invention provides a general method by which the regions of peptides likely to contain T cell epitopes can be described. Modifications to the peptide in these regions have the potential to modify the MHC Class II binding characteristics.

[0127] According to another aspect of the present invention, T cell epitopes can be predicted with greater accuracy by the use of a more sophisticated computational method which takes into account the interactions of peptides with models of MHC Class II alleles.

[0128] The computational prediction of T cell epitopes present within a peptide according to this particular aspect contemplates the construction of models of at least 42 MHC Class II alleles based upon the structures of all known MHC Class II molecules and a method for the use of these models in the computational identification of T cell epitopes, the construction of libraries of peptide backbones for each model in order to allow for the known variability in relative peptide backbone alpha carbon (Cα) positions, the construction of libraries of amino-acid side chain conformations for each backbone dock with each model for each of the 20 amino-acid alternatives at positions critical for the interaction between peptide and MHC Class II molecule, and the use of these libraries of backbones and side-chain conformations in conjunction with a scoring function to select the optimum backbone and side-chain conformation for a particular peptide docked with a particular MHC Class II molecule and the derivation of a binding score from this interaction.

[0129] Models of MHC Class II molecules can be derived via homology modeling from a number of similar structures found in the Brookhaven Protein Data Bank ("PDB"). These may be made by the use of semi-automatic homology modeling software (Modeller, Sali A. & Blundell T L., 1993. *J. Mol Biol* 234:779-815) which incorporates a simulated annealing function, in conjunction with the CHARM forcefield for energy minimization (available from Molecular Simulations Inc., San Diego, Calif.). Alternative modeling methods can be utilized as well.

[0130] The present method differs significantly from other computational methods which use libraries of experimentally derived binding data of each amino-acid alternative at each position in the binding groove for a small set of MHC Class II molecules (Marshall, K. W., et al., *Biomed. Pept. Proteins Nucleic Acids,* 1(3): 157-162) (1995) or yet other computational methods which use similar experimental binding data in order to define the binding characteristics of particular types of binding pockets within the groove, again using a relatively small subset of MHC Class II molecules, and then 'mixing and matching' pocket types from this pocket library to artificially create further 'virtual' MHC

Class II molecules (Sturniolo T., et al., *Nat. Biotech,* 17(6): 555-561 (1999). Both prior methods suffer the major disadvantage that, due to the complexity of the assays and the need to synthesize large numbers of peptide variants, only a small number of MHC Class II molecules can be experimentally scanned. Therefore the first prior method can only make predictions for a small number of MHC Class II molecules. The second prior method also makes the assumption that a pocket lined with similar amino-acids in one molecule will have the same binding characteristics when in the context of a different Class II allele and suffers further disadvantages in that only those MHC Class II molecules can be 'virtually' created which contain pockets contained within the pocket library. Using the modeling approach described herein, the structure of any number and type of MHC Class II molecules can be deduced, therefore alleles can be specifically selected to be representative of the global population. In addition, the number of MHC Class II molecules scanned can be increased by making further models further than having to generate additional data via complex experimentation.

[**0131**] The use of a backbone library allows for variation in the positions of the Cα atoms of the various peptides being scanned when docked with particular MHC Class II molecules. This is again in contrast to the alternative prior computational methods described above which rely on the use of simplified peptide backbones for scanning amino-acid binding in particular pockets. These simplified backbones are not likely to be representative of backbone conformations found in 'real' peptides leading to inaccuracies in prediction of peptide binding. The present backbone library is created by superposing the backbones of all peptides bound to MHC Class II molecules found within the Protein Data Bank and noting the root mean square (RMS) deviation between the Cα atoms of each of the eleven amino-acids located within the binding groove. While this library can be derived from a small number of suitable available mouse and human structures (currently 13), in order to allow for the possibility of even greater variability, the RMS FIGURE for each C"-α position is increased by 50%. The average Cα position of each amino-acid is then determined and a sphere drawn around this point whose radius equals the RMS deviation at that position plus 50%. This sphere represents all allowed Cα positions.

[**0132**] Working from the Cα with the least RMS deviation (that of the amino-acid in Pocket 1 as mentioned above, equivalent to Position 2 of the 11 residues in the binding groove), the sphere is three-dimensionally gridded, and each vertex within the grid is then used as a possible location for a Cα of that amino-acid. The subsequent amide plane, corresponding to the peptide bond to the subsequent amino-acid is grafted onto each of these Cαs and the ϕ and ψ angles are rotated step-wise at set intervals in order to position the subsequent Cα. If the subsequent Cα falls within the 'sphere of allowed positions' for this Cα than the orientation of the dipeptide is accepted, whereas if it falls outside the sphere then the dipeptide is rejected. This process is then repeated for each of the subsequent Cα positions, such that the peptide grows from the Pocket 1 Cα'seed', until all nine subsequent Cαs have been positioned from all possible permutations of the preceding Cαs. The process is then repeated once more for the single Cα preceding pocket 1 to create a library of backbone Cα positions located within the binding groove.

[**0133**] The number of backbones generated is dependent upon several factors: The size of the 'spheres of allowed positions'; the fineness of the gridding of the 'primary sphere' at the Pocket 1 position; the fineness of the step-wise rotation of the ϕ and ψ angles used to position subsequent Cαs. Using this process, a large library of backbones can be created. The larger the backbone library, the more likely it will be that the optimum fit will be found for a particular peptide within the binding groove of an MHC Class II molecule. Inasmuch as all backbones will not be suitable for docking with all the models of MHC Class II molecules due to clashes with amino-acids of the binding domains, for each allele a subset of the library is created comprising backbones which can be accommodated by that allele. The use of the backbone library, in conjunction with the models of MHC Class II molecules creates an exhaustive database consisting of allowed side chain conformations for each amino-acid in each position of the binding groove for each MHC Class II molecule docked with each allowed backbone. This data set is generated using a simple steric overlap function where a MHC Class II molecule is docked with a backbone and an amino-acid side chain is grafted onto the backbone at the desired position. Each of the rotatable bonds of the side chain is rotated step-wise at set intervals and the resultant positions of the atoms dependent upon that bond noted. The interaction of the atom with atoms of side-chains of the binding groove is noted and positions are either accepted or rejected according to the following criteria: The sum total of the overlap of all atoms so far positioned must not exceed a pre-determined value. Thus the stringency of the conformational search is a function of the interval used in the step-wise rotation of the bond and the pre-determined limit for the total overlap. This latter value can be small if it is known that a particular pocket is rigid, however the stringency can be relaxed if the positions of pocket side-chains are known to be relatively flexible. Thus allowances can be made to imitate variations in flexibility within pockets of the binding groove. This conformational search is then repeated for every amino-acid at every position of each backbone when docked with each of the MHC Class II molecules to create the exhaustive database of side-chain conformations.

[**0134**] A suitable mathematical expression is used to estimate the energy of binding between models of MHC Class II molecules in conjunction with peptide ligand conformations which have to be empirically derived by scanning the large database of backbone/side-chain conformations described above. Thus a protein is scanned for potential T cell epitopes by subjecting each possible peptide of length varying between 9 and 20 amino-acids (although the length is kept constant for each scan) to the following computations: An MHC Class II molecule is selected together with a peptide backbone allowed for that molecule and the side-chains corresponding to the desired peptide sequence are grafted on. Atom identity and interatomic distance data relating to a particular side-chain at a particular position on the backbone are collected for each allowed conformation of that amino-acid (obtained from the database described above). This is repeated for each side-chain along the backbone and peptide scores derived using a scoring function. The best score for that backbone is retained and the process repeated for each allowed backbone for the selected model. The scores from all allowed backbones are compared and the highest score is deemed to be the peptide score for the desired peptide in that MHC Class II model. This process

is then repeated for each model with every possible peptide derived from the protein being scanned, and the scores for peptides versus models are displayed.

[0135] In the context of the present invention, each ligand presented for the binding affinity calculation is an amino-acid segment selected from a peptide or protein as discussed above. Thus, the ligand is a selected stretch of amino acids about 9 to 20 amino acids in length derived from a peptide, polypeptide or protein of known sequence. The terms "amino acids" and "residues" are hereinafter regarded as equivalent terms. The ligand, in the form of the consecutive amino acids of the peptide to be examined grafted onto a backbone from the backbone library, is positioned in the binding cleft of an MHC Class II molecule from the MHC Class II molecule model library via the coordinates of the C"-α atoms of the peptide backbone and an allowed conformation for each side-chain is selected from the database of allowed conformations. The relevant atom identities and interatomic distances are also retrieved from this database and used to calculate the peptide binding score. Ligands with a high binding affinity for the MHC Class II binding pocket are flagged as candidates for site-directed mutagenesis. Amino-acid substitutions are made in the flagged ligand (and hence in the protein of interest) which is then retested using the scoring function in order to determine changes which reduce the binding affinity below a predetermined threshold value. These changes can then be incorporated into the protein of interest to remove T cell epitopes.

[0136] Binding between the peptide ligand and the binding groove of MHC Class II molecules involves non-covalent interactions including, but not limited to: hydrogen bonds, electrostatic interactions, hydrophobic (lipophilic) interactions and Van der Walls interactions. These are included in the peptide scoring function as described in detail below. It should be understood that a hydrogen bond is a non-covalent bond which can be formed between polar or charged groups and consists of a hydrogen atom shared by two other atoms. The hydrogen of the hydrogen donor has a positive charge where the hydrogen acceptor has a partial negative charge. For the purposes of peptide/protein interactions, hydrogen bond donors may be either nitrogens with hydrogen attached or hydrogens attached to oxygen or nitrogen. Hydrogen bond acceptor atoms may be oxygens not attached to hydrogen, nitrogens with no hydrogens attached and one or two connections, or sulphurs with only one connection. Certain atoms, such as oxygens attached to hydrogens or imine nitrogens (e.g. C═NH) may be both hydrogen acceptors or donors. Hydrogen bond energies range from 3 to 7 Kcal/mol and are much stronger than Van der Waal's bonds, but weaker than covalent bonds. Hydrogen bonds are also highly directional and are at their strongest when the donor atom, hydrogen atom and acceptor atom are co-linear. Electrostatic bonds are formed between oppositely charged ion pairs, and the strength of the interaction is inversely proportional to the square of the distance between the atoms according to Coulomb's law. The optimal distance between ion pairs is about 2.8 Å. In protein/peptide interactions, electrostatic bonds may be formed between arginine, histidine or lysine and aspartate or glutamate. The strength of the bond will depend upon the pKa of the ionizing group and the dielectric constant of the medium although they are approximately similar in strength to hydrogen bonds.

[0137] Lipophilic interactions are favorable hydrophobic-hydrophobic contacts that occur between the protein and peptide ligand. Usually, these will occur between hydrophobic amino acid side chains of the peptide buried within the pockets of the binding groove such that they are not exposed to solvent. Exposure of the hydrophobic residues to solvent is highly unfavorable since the surrounding solvent molecules are forced to hydrogen bond with each other forming cage-like clathrate structures. The resultant decrease in entropy is highly unfavorable. Lipophilic atoms may be sulphurs which are neither polar nor hydrogen acceptors and carbon atoms which are not polar.

[0138] Van der Waal's bonds are non-specific forces found between atoms which are 3-4 Å apart. They are weaker and less specific than hydrogen and electrostatic bonds. The distribution of electronic charge around an atom changes with time and, at any instant, the charge distribution is not symmetric. This transient asymmetry in electronic charge induces a similar asymmetry in neighboring atoms. The resultant attractive forces between atoms reaches a maximum at the Van der Waal's contact distance but diminishes very rapidly at about 1 Å to about 2 Å. Conversely, as atoms become separated by less than the contact distance, increasingly strong repulsive forces become dominant as the outer electron clouds of the atoms overlap. Although the attractive forces are relatively weak compared to electrostatic and hydrogen bonds (about 0.6 Kcal/mol), the repulsive forces in particular may be very important in determining whether a peptide ligand may bind successfully to a protein.

[0139] In one embodiment, the Böhm scoring function (SCORE1 approach) is used to estimate the binding constant. (Böhm, H. J., *J. Comput Aided Mol. Des.,* 8(3):243-256 (1994) which is hereby incorporated in its entirety). In another embodiment, the scoring function (SCORE2 approach) is used to estimate the binding affinities as an indicator of a ligand containing a T cell epitope (Böhm, H. J., *J. Comput Aided Mol. Des.,* 12(4):309-323 (1998) which is hereby incorporated in its entirety). However, the Böhm scoring functions as described in the above references are used to estimate the binding affinity of a ligand to a protein where it is already known that the ligand successfully binds to the protein and the protein/ligand complex has had its structure solved, the solved structure being present in the Protein Data Bank ("PDB"). Therefore, the scoring function has been developed with the benefit of known positive binding data. In order to allow for discrimination between positive and negative binders, a repulsion term must be added to the equation. In addition, a more satisfactory estimate of binding energy is achieved by computing the lipophilic interactions in a pairwise manner rather than using the area based energy term of the above Böhm functions. Therefore, in a preferred embodiment, the binding energy is estimated using a modified Böhm scoring function. In the modified Böhm scoring function, the binding energy between protein and ligand (ΔG_{bind}) is estimated considering the following parameters: The reduction of binding energy due to the overall loss of translational and rotational entropy of the ligand (ΔG_0); contributions from ideal hydrogen bonds (ΔG_{hb}) where at least one partner is neutral; contributions from unperturbed ionic interactions (ΔG_{ionic}); lipophilic interactions between lipophilic ligand atoms and lipophilic acceptor atoms (ΔG_{lipo}); the loss of binding energy due to the freezing of internal degrees of freedom in the ligand, i.e., the freedom of rotation about each C—C

bond is reduced (ΔG_{rot}); the energy of the interaction between the protein and ligand (E_{vdW}). Consideration of these terms gives equation 1:

$$(\Delta G_{bind})=(\Delta G_0)+(\Delta G_{hb}\times N_{hb})+(\Delta G_{ionic}\times N_{ionic})+(\Delta G_{lipo}\times N_{lipo})+(\Delta G_{rot}+N_{rot})+(E_{vdW}).$$

[0140] Where N is the number of qualifying interactions for a specific term and, in one embodiment, ΔG_0, ΔG_{hb}, ΔG_{ionic}, ΔG_{lipo} and ΔG_{rot} are constants which are given the values: 5.4, −4.7, −4.7, −0.17, and 1.4, respectively.

[0141] The term N_{hb} is calculated according to equation 2:

$$N_{hb}=\Sigma_{h\text{-}bonds}f(\Delta R,\Delta\alpha)\times f(N_{neighb})\times f_{pcs}$$

[0142] $f(\Delta R, \Delta\alpha)$ is a penalty function which accounts for large deviations of hydrogen bonds from ideality and is calculated according to equation 3:

$$f(\Delta R,\Delta-\square)=f1(\Delta R)\times f2(\Delta\alpha)$$

[0143] Where:

[0144] f1(ΔR)=1 if ΔR<=TOL

[0145] or =1−(ΔR−TOL)/0.4 if ΔR<=0.4+TOL

[0146] or =0 if ΔR>0.4+TOL

[0147] And:

[0148] f2(Δα)=1 if Δα<30°

[0149] or =1−(Δα−30)/50 if Δα<=80°

[0150] or =0 if Δα>80°

[0151] TOL is the tolerated deviation in hydrogen bond length=0.25 Å

[0152] ΔR is the deviation of the H—O/N hydrogen bond length from the ideal value=1.9 Å

[0153] Δα is the deviation of the hydrogen bond angle $\angle_{N/O\text{—}H..O/N}$ from its idealized value of 180° $f(N_{neighb})$ distinguishes between concave and convex parts of a protein surface and therefore assigns greater weight to polar interactions found in pockets rather than those found at the protein surface. This function is calculated according to equation 4 below:

$$f(N_{neighb})=(N_{neighb}/N_{neighb,0})^{\alpha} \text{ where } \alpha=0.5$$

[0154] N_{neighb} is the number of non-hydrogen protein atoms that are closer than 5 Å to any given protein atom.

[0155] $N_{neighb,0}$ is a constant=25

[0156] f_{pcs} is a function which allows for the polar contact surface area per hydrogen bond and therefore distinguishes between strong and weak hydrogen bonds and its value is determined according to the following criteria:

[0157] $f_{pcs}=\beta$ when $A_{polar}/N_{HB}<10$ Å^2;

[0158] or $f_{pcs}=1$ when $\Delta_{polar}/N_{HB}>10$ Å^2;

[0159] A_{polar} is the size of the polar protein-ligand contact surface;

[0160] N_{HB} is the number of hydrogen bonds; and

[0161] β is a constant whose value=1.2.

[0162] For the implementation of the modified Böhm scoring function, the contributions from ionic interactions, ΔG_{ionic}, are computed in a similar fashion to those from hydrogen bonds described above since the same geometry dependency is assumed.

[0163] The term N_{lipo} is calculated according to equation 5 below:

$$N_{lipo}=\Sigma_{lL}f(r_{lL});$$

[0164] $f(r_{lL})$ is calculated for all lipophilic ligand atoms, l, and all lipophilic protein atoms, L, according to the following criteria:

[0165] $f(r_{lL})=1$ when r_{lL}<=R1$f(r_{lL})=(r_{lL}-R1)/(R2-R1)$ when R2<r_{lL}>R1;

[0166] $f(r_{lL})=0$ when r_{lL}>=R2;

[0167] Where: $R1=r_l^{vdw}+r_L^{vdw}+0.5$;

[0168] and R2=R1+3.0; _o**9** and r_l^{vdw} is the Van der Waal's radius of atom l;

[0169] and r_L^{vdw} is the Van der Waal's radius of atom L.

[0170] The term N_{rot} is the number of rotatable bonds of the amino acid side chain and is taken to be the number of acyclic sp^3-sp^3 and sp^3-sp^2 bonds. Rotations of terminal —CH_3 or —NH_3 are not taken into account.

[0171] The final term, E_{vdW}, is calculated according to equation 6 below:

$$E_{vdW}=\epsilon_1\epsilon_2((r_1^{vdw}+r_2^{vdw})^{12}/r^{12}-(r_1^{vdw}+r_2^{vdw})^6/r^6),$$

where:

[0172] ϵ_1 and ϵ_2 are constants dependant upon atom identity;

[0173] $r_1^{vdw}+r_2^{vdw}$ are the Van der Waal's atomic radii;

[0174] r is the distance between a pair of atoms.

[0175] With regard to Equation 6, in one embodiment, the constants ϵ_1 and ϵ_2 are given the atom values: C, 0.245, N, 0.283, O: 0.316, S: 0.316, respectively (i.e. for atoms of Carbon, Nitrogen, Oxygen and Sulphur, respectively). With regards to equations 5 and 6, the Van der Waal's radii are given the atom values C, 1.85, N, 1.75, O: 1.60, S: 2.00 Å.

[0176] It should be understood that all predetermined values and constants given in the equations above are determined within the constraints of current understandings of protein ligand interactions with particular regard to the type of computation being undertaken herein. Therefore, it is possible that, as this scoring function is refined further, these values and constants may change hence any suitable numerical value which gives the desired results in terms of estimating the binding energy of a protein to a ligand may be used and hence fall within the scope of the present invention.

[0177] As described above, the scoring function is applied to data extracted from the database of side-chain conformations, atom identities, and interatomic distances. For the purposes of the present description, the number of MHC Class II molecules included in this database is 42 models plus four solved structures. It should be apparent from the above descriptions that the modular nature of the construction of the computational method of the present invention means that new models can simply be added and scanned with the peptide backbone library and side-chain conformational search function to create additional data sets which can be processed by the peptide scoring function as

described above. This allows for the repertoire of scanned MHC Class II molecules to easily be increased, or structures and associated data to be replaced if data are available to create more accurate models of the existing alleles.

[0178] The present prediction method can be calibrated against a data set comprising a large number of peptides whose affinity for various MHC Class II molecules has previously been experimentally determined. By comparison of calculated versus experimental data, a cut of value can be determined above which it is known that all experimentally determined T cell epitopes are correctly predicted.

[0179] It should be understood that, although the above scoring function is relatively simple compared to some sophisticated methodologies that are available, the calculations are performed extremely rapidly. It should also be understood that the objective is not to calculate the true binding energy per se for each peptide docked in the binding groove of a selected MHC Class II protein. The underlying objective is to obtain comparative binding energy data as an aid to predicting the location of T cell epitopes based on the primary structure (i.e. amino acid sequence) of a selected protein. A relatively high binding energy or a binding energy above a selected threshold value would suggest the presence of a T cell epitope in the ligand. The ligand may then be subjected to at least one round of amino-acid substitution and the binding energy recalculated. Due to the rapid nature of the calculations, these manipulations of the peptide sequence can be performed interactively within the program's user interface on cost-effectively available computer hardware. Major investment in computer hardware is thus not required.

[0180] It would be apparent to one skilled in the art that other available software could be used for the same purposes. In particular, more sophisticated software which is capable of docking ligands into protein binding-sites may be used in conjunction with energy minimization. Examples of docking software are: DOCK (Kuntz et al., *J. Mol. Biol.,* 161:269-288 (1982)), LUDI (Böhm, H. J., *J. Comput Aided Mol. Des.,* 8:623-632 (1994)) and FLEXX (Rarey M., et al., *ISMB,* 3:300-308 (1995)). Examples of molecular modeling and manipulation software include: AMBER (Tripos) and CHARM (Molecular Simulations Inc.). The use of these computational methods would severely limit the throughput of the method of this invention due to the lengths of processing time required to make the necessary calculations. However, it is feasible that such methods could be used as a 'secondary screen' to obtain more accurate calculations of binding energy for peptides which are found to be 'positive binders' via the method of the present invention.

[0181] The limitation of processing time for sophisticated molecular mechanic or molecular dynamic calculations is one which is defined both by the design of the software which makes these calculations and the current technology limitations of computer hardware. It may be anticipated that, in the future, with the writing of more efficient code and the continuing increases in speed of computer processors, it may become feasible to make such calculations within a more manageable time-frame. Further information on energy functions applied to macromolecules and consideration of the various interactions that take place within a folded protein structure can be found in: Brooks, B. R., et al., *J. Comput. Chem.,* 4:187-217 (1983) and further information concerning general protein-ligand interactions can be found in: Dauber-Osguthorpe et al., *Proteins* 4(1):31-47(1988), which are incorporated herein by reference in their entirety. Useful background information can also be found, for example, in Fasman, G. D., ed., *Prediction of Protein Structure and the Principles of Protein Conformation*, Plenum Press, New York, ISBN: 0-306 4313-9.

EXAMPLE 6

[0182] Refinement of Immunogenic Regions for IFN α

[0183] The 51 peptides shown in Table 2 were screened in a T cell activation assay of the invention utilizing PBMC from 20 healthy donor individuals and 20 individuals suffering from a chronic HCV infection who had previously been treated with IFN α2b according to National Institute of Clinical Excellence, (UK) guidelines (blood samples provided by Dr. G. Alexander, Addenbrooke's Hospital, Cambridge, UK).

[0184] Bulk cultures of $2\text{-}4\times10^6$ PBMC per well of a 24-well plate were incubated for 6-9 days with the individual peptides of Table 2. Positive control peptides having SEQ ID NO: 114 (influenza hemataggglutinin peptide 307-319) AND SEQ ID NO: 115 (Chlamydia HSP 60 peptide), were assessed, as well. Proliferation was assessed at various time points by gently resuspending the bulk cultures and removing samples of PBMC that were then incubated in triplicate in wells of U-bottomed, 96-well plates with about 1 μCi/well of tritiated thymidine for 18 hours. The SI for each peptide was determined as described in EXAMPLES 1 and 2. The percentage of donors whose PBMC responded to a peptide to afford an SI of greater than about 1.95 are plotted against the peptides in **FIG. 4**, Panel A. The immunogenic regions R1, R2, and R3 for IFN α were refined by analyzing the amino acid sequences of the peptides to which greater than 15% of the donor PBMC afforded an SI of >1.95 and comparing those sequences to the full sequence of IFN α (SEQ ID NO: 127). The positions (relative to full length IFN α, SEQ ID NO: 127) of the residues spanned by the significantly responding peptides are indicated in Panel A by underlined ranges of numbers over the responsive peptides. These positions define the refined immunogenic regions of the IFN α. The refined regions for R1, R2, and R3 of IFN α are shown in boxes in Panel B of **FIG. 4**. Amino acid residue substitutions in the immunogenic regions that lower immunogenicity are shown as bold letters above underlined amino acid residues of SEQ ID NO: 127). Refined immunogenic regions R1, R2, and R3 substantially overlay with the corresponding preliminary immunogenic regions set forth in EXAMPLE 2.

[0185] The peptide sequences included in each of the refined regions were determined as follows: The most immunogenic region (R3) was located in the region spanned by peptides 38 and 39 (spanning residues 112-130 of SEQ ID NO: 127), to which 15% and 25% of donors responded, respectively. Region R2 was initially located at peptide 18, but then expanded to include peptide 22, which was borderline at 15% response, Peptides 19 and 20 were insoluble, so the presence of immunogenic residues could not be ruled out. Accordingly, refined R2 spans amino acid residues 52-79 of SEQ ID NO: 127. No other peptides showed responses in greater than 15% of donors, however, peptides 8 and 9, spanning residues 22-39 of SEQ ID NO: 127, were

border line at 15%, thus defining R1 as the least immunogenic among the three regions R1, R2, and R3. The magnitude of the SI for peptide 39 suggested that the portion of region R3 spanned by this peptide (residues 116-130 of SEQ ID NO: 127) may include an immunodominant T cell epitope.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 127

<210> SEQ ID NO 1
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 1

Met Ser Tyr Asn Leu Leu Gly Phe Leu Gln Arg Ser Ser Asn Phe
1 5 10 15

<210> SEQ ID NO 2
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 2

Asn Leu Leu Gly Phe Leu Gln Arg Ser Ser Asn Phe Gln Cys Gln
1 5 10 15

<210> SEQ ID NO 3
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 3

Gly Phe Leu Gln Arg Ser Ser Asn Phe Gln Cys Gln Lys Leu Leu
1 5 10 15

<210> SEQ ID NO 4
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 4

Gln Arg Ser Ser Asn Phe Gln Cys Gln Lys Leu Leu Trp Gln Leu
1 5 10 15

<210> SEQ ID NO 5
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 5

Ser Asn Phe Gln Cys Gln Lys Leu Leu Trp Gln Leu Asn Gly Arg
1 5 10 15

-continued

<210> SEQ ID NO 6
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 6

Gln Cys Gln Lys Leu Leu Trp Gln Leu Asn Gly Arg Leu Glu Tyr
1 5 10 15

<210> SEQ ID NO 7
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 7

Lys Leu Leu Trp Gln Leu Asn Gly Arg Leu Glu Tyr Cys Leu Lys
1 5 10 15

<210> SEQ ID NO 8
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 8

Trp Gln Leu Asn Gly Arg Leu Glu Tyr Cys Leu Lys Asp Arg Met
1 5 10 15

<210> SEQ ID NO 9
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 9

Asn Gly Arg Leu Glu Tyr Cys Leu Lys Asp Arg Met Asn Phe Asp
1 5 10 15

<210> SEQ ID NO 10
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 10

Leu Glu Tyr Cys Leu Lys Asp Arg Met Asn Phe Asp Ile Pro Glu
1 5 10 15

<210> SEQ ID NO 11
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 11

Cys Leu Lys Asp Arg Met Asn Phe Asp Ile Pro Glu Glu Ile Lys
1 5 10 15

-continued

<210> SEQ ID NO 12
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 12

Asp Arg Met Asn Phe Asp Ile Pro Glu Glu Ile Lys Gln Leu Gln
1 5 10 15

<210> SEQ ID NO 13
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 13

Asn Phe Asp Ile Pro Glu Glu Ile Lys Gln Leu Gln Gln Phe Gln
1 5 10 15

<210> SEQ ID NO 14
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 14

Ile Pro Glu Glu Ile Lys Gln Leu Gln Gln Phe Gln Lys Glu Asp
1 5 10 15

<210> SEQ ID NO 15
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 15

Glu Ile Lys Gln Leu Gln Gln Phe Gln Lys Glu Asp Ala Ala Leu
1 5 10 15

<210> SEQ ID NO 16
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 16

Gln Leu Gln Gln Phe Gln Lys Glu Asp Ala Ala Leu Thr Ile Tyr
1 5 10 15

<210> SEQ ID NO 17
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 17

-continued

Gln Phe Gln Lys Glu Asp Ala Ala Leu Thr Ile Tyr Glu Met Leu
1 5 10 15

<210> SEQ ID NO 18
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 18

Lys Glu Asp Ala Ala Leu Thr Ile Tyr Glu Met Leu Gln Asn Ile
1 5 10 15

<210> SEQ ID NO 19
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 19

Ala Ala Leu Thr Ile Tyr Glu Met Leu Gln Asn Ile Phe Ala Ile
1 5 10 15

<210> SEQ ID NO 20
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 20

Thr Ile Tyr Glu Met Leu Gln Asn Ile Phe Ala Ile Phe Arg Gln
1 5 10 15

<210> SEQ ID NO 21
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 21

Glu Met Leu Gln Asn Ile Phe Ala Ile Phe Arg Gln Asp Ser Ser
1 5 10 15

<210> SEQ ID NO 22
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 22

Gln Asn Ile Phe Ala Ile Phe Arg Gln Asp Ser Ser Ser Thr Gly
1 5 10 15

<210> SEQ ID NO 23
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

-continued

<400> SEQUENCE: 23

Phe Ala Ile Phe Arg Gln Asp Ser Ser Ser Thr Gly Trp Asn Glu
1 5 10 15

<210> SEQ ID NO 24
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 24

Phe Arg Gln Asp Ser Ser Ser Thr Gly Trp Asn Glu Thr Ile Val
1 5 10 15

<210> SEQ ID NO 25
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 25

Asp Ser Ser Ser Thr Gly Trp Asn Glu Thr Ile Val Glu Asn Leu
1 5 10 15

<210> SEQ ID NO 26
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 26

Ser Thr Gly Trp Asn Glu Thr Ile Val Glu Asn Leu Leu Ala Asn
1 5 10 15

<210> SEQ ID NO 27
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 27

Trp Asn Glu Thr Ile Val Glu Asn Leu Leu Ala Asn Val Tyr His
1 5 10 15

<210> SEQ ID NO 28
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 28

Thr Ile Val Glu Asn Leu Leu Ala Asn Val Tyr His Gln Ile Asn
1 5 10 15

<210> SEQ ID NO 29
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 29

Glu Asn Leu Leu Ala Asn Val Tyr His Gln Ile Asn His Leu Lys
1 5 10 15

<210> SEQ ID NO 30
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 30

Leu Ala Asn Val Tyr His Gln Ile Asn His Leu Lys Thr Val Leu
1 5 10 15

<210> SEQ ID NO 31
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 31

Val Tyr His Gln Ile Asn His Leu Lys Thr Val Leu Glu Glu Lys
1 5 10 15

<210> SEQ ID NO 32
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 32

Gln Ile Asn His Leu Lys Thr Val Leu Glu Glu Lys Leu Glu Lys
1 5 10 15

<210> SEQ ID NO 33
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 33

His Leu Lys Thr Val Leu Glu Glu Lys Leu Glu Lys Glu Asp Phe
1 5 10 15

<210> SEQ ID NO 34
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 34

Thr Val Leu Glu Glu Lys Leu Glu Lys Glu Asp Phe Thr Arg Gly
1 5 10 15

<210> SEQ ID NO 35
<211> LENGTH: 15
<212> TYPE: PRT

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 35

Glu Glu Lys Leu Glu Lys Glu Asp Phe Thr Arg Gly Lys Leu Met
1 5 10 15

<210> SEQ ID NO 36
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 36

Leu Glu Lys Glu Asp Phe Thr Arg Gly Lys Leu Met Ser Ser Leu
1 5 10 15

<210> SEQ ID NO 37
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 37

Glu Asp Phe Thr Arg Gly Lys Leu Met Ser Ser Leu His Leu Lys
1 5 10 15

<210> SEQ ID NO 38
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 38

Thr Arg Gly Lys Leu Met Ser Ser Leu His Leu Lys Arg Tyr Tyr
1 5 10 15

<210> SEQ ID NO 39
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 39

Lys Leu Met Ser Ser Leu His Leu Lys Arg Tyr Tyr Gly Arg Ile
1 5 10 15

<210> SEQ ID NO 40
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 40

Ser Ser Leu His Leu Lys Arg Tyr Tyr Gly Arg Ile Leu His Tyr
1 5 10 15

<210> SEQ ID NO 41

-continued

<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 41

His Leu Lys Arg Tyr Tyr Gly Arg Ile Leu His Tyr Leu Lys Ala
1 5 10 15

<210> SEQ ID NO 42
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 42

Arg Tyr Tyr Gly Arg Ile Leu His Tyr Leu Lys Ala Lys Glu Tyr
1 5 10 15

<210> SEQ ID NO 43
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 43

Gly Arg Ile Leu His Tyr Leu Lys Ala Lys Glu Tyr Ser His Cys
1 5 10 15

<210> SEQ ID NO 44
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 44

Leu His Tyr Leu Lys Ala Lys Glu Tyr Ser His Cys Ala Trp Thr
1 5 10 15

<210> SEQ ID NO 45
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 45

Leu Lys Ala Lys Glu Tyr Ser His Cys Ala Trp Thr Ile Val Arg
1 5 10 15

<210> SEQ ID NO 46
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 46

Lys Glu Tyr Ser His Cys Ala Trp Thr Ile Val Arg Val Glu Ile
1 5 10 15

-continued

<210> SEQ ID NO 47
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 47

Ser His Cys Ala Trp Thr Ile Val Arg Val Glu Ile Leu Arg Asn
1 5 10 15

<210> SEQ ID NO 48
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 48

Ala Trp Thr Ile Val Arg Val Glu Ile Leu Arg Asn Phe Tyr Phe
1 5 10 15

<210> SEQ ID NO 49
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 49

Ile Val Arg Val Glu Ile Leu Arg Asn Phe Tyr Phe Ile Asn Arg
1 5 10 15

<210> SEQ ID NO 50
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 50

Val Glu Ile Leu Arg Asn Phe Tyr Phe Ile Asn Arg Leu Thr Gly
1 5 10 15

<210> SEQ ID NO 51
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 51

Leu Arg Asn Phe Tyr Phe Ile Asn Arg Leu Thr Gly Tyr Leu Arg
1 5 10 15

<210> SEQ ID NO 52
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 52

Cys Asp Leu Pro Gln Thr His Ser Leu Gly Ser Arg Arg Thr Leu

-continued

1 5 10 15

<210> SEQ ID NO 53
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 53

Pro Gln Thr His Ser Leu Gly Ser Arg Arg Thr Leu Met Leu Leu
1 5 10 15

<210> SEQ ID NO 54
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 54

His Ser Leu Gly Ser Arg Arg Thr Leu Met Leu Leu Ala Gln Met
1 5 10 15

<210> SEQ ID NO 55
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 55

Gly Ser Arg Arg Thr Leu Met Leu Leu Ala Gln Met Arg Arg Ile
1 5 10 15

<210> SEQ ID NO 56
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 56

Arg Thr Leu Met Leu Leu Ala Gln Met Arg Arg Ile Ser Leu Phe
1 5 10 15

<210> SEQ ID NO 57
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 57

Met Leu Leu Ala Gln Met Arg Arg Ile Ser Leu Phe Ser Cys Leu
1 5 10 15

<210> SEQ ID NO 58
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 58

-continued

Ala Gln Met Arg Arg Ile Ser Leu Phe Ser Cys Leu Lys Asp Arg
1 5 10 15

<210> SEQ ID NO 59
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 59

Arg Arg Ile Ser Leu Phe Ser Cys Leu Lys Asp Arg His Asp Phe
1 5 10 15

<210> SEQ ID NO 60
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 60

Ser Leu Phe Ser Cys Leu Lys Asp Arg His Asp Phe Gly Phe Pro
1 5 10 15

<210> SEQ ID NO 61
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 61

Ser Cys Leu Lys Asp Arg His Asp Phe Gly Phe Pro Gln Glu Glu
1 5 10 15

<210> SEQ ID NO 62
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 62

Lys Asp Arg His Asp Phe Gly Phe Pro Gln Glu Glu Phe Gly Asn
1 5 10 15

<210> SEQ ID NO 63
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 63

His Asp Phe Gly Phe Pro Gln Glu Glu Phe Gly Asn Gln Phe Gln
1 5 10 15

<210> SEQ ID NO 64
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

-continued

<400> SEQUENCE: 64

Gly Phe Pro Gln Glu Glu Phe Gly Asn Gln Phe Gln Lys Ala Glu
1 5 10 15

<210> SEQ ID NO 65
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 65

Gln Glu Glu Phe Gly Asn Gln Phe Gln Lys Ala Glu Thr Ile Pro
1 5 10 15

<210> SEQ ID NO 66
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 66

Phe Gly Asn Gln Phe Gln Lys Ala Glu Thr Ile Pro Val Leu His
1 5 10 15

<210> SEQ ID NO 67
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 67

Gln Phe Gln Lys Ala Glu Thr Ile Pro Val Leu His Glu Met Ile
1 5 10 15

<210> SEQ ID NO 68
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 68

Lys Ala Glu Thr Ile Pro Val Leu His Glu Met Ile Gln Gln Ile
1 5 10 15

<210> SEQ ID NO 69
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 69

Thr Ile Pro Val Leu His Glu Met Ile Gln Gln Ile Phe Asn Leu
1 5 10 15

<210> SEQ ID NO 70
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 70

Val Leu His Glu Met Ile Gln Gln Ile Phe Asn Leu Phe Ser Thr
1 5 10 15

<210> SEQ ID NO 71
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 71

Glu Met Ile Gln Gln Ile Phe Asn Leu Phe Ser Thr Lys Asp Ser
1 5 10 15

<210> SEQ ID NO 72
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 72

Gln Gln Ile Phe Asn Leu Phe Ser Thr Lys Asp Ser Ser Ala Ala
1 5 10 15

<210> SEQ ID NO 73
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 73

Phe Asn Leu Phe Ser Thr Lys Asp Ser Ser Ala Ala Trp Asp Glu
1 5 10 15

<210> SEQ ID NO 74
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 74

Phe Ser Thr Lys Asp Ser Ser Ala Ala Trp Asp Glu Thr Leu Leu
1 5 10 15

<210> SEQ ID NO 75
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 75

Lys Asp Ser Ser Ala Ala Trp Asp Glu Thr Leu Leu Asp Lys Phe
1 5 10 15

<210> SEQ ID NO 76
<211> LENGTH: 15

-continued

<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 76

Ser Ala Ala Trp Asp Glu Thr Leu Leu Asp Lys Phe Tyr Thr Glu
1 5 10 15

<210> SEQ ID NO 77
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 77

Trp Asp Glu Thr Leu Leu Asp Lys Phe Tyr Thr Glu Leu Tyr Gln
1 5 10 15

<210> SEQ ID NO 78
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 78

Thr Leu Leu Asp Lys Phe Tyr Thr Glu Leu Tyr Gln Gln Leu Asn
1 5 10 15

<210> SEQ ID NO 79
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 79

Asp Lys Phe Tyr Thr Glu Leu Tyr Gln Gln Leu Asn Asp Leu Glu
1 5 10 15

<210> SEQ ID NO 80
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 80

Tyr Thr Glu Leu Tyr Gln Gln Leu Asn Asp Leu Glu Ala Cys Val
1 5 10 15

<210> SEQ ID NO 81
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 81

Leu Tyr Gln Gln Leu Asn Asp Leu Glu Ala Cys Val Ile Gln Gly
1 5 10 15

-continued

<210> SEQ ID NO 82
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 82

Gln Leu Asn Asp Leu Glu Ala Cys Val Ile Gln Gly Val Gly Val
1 5 10 15

<210> SEQ ID NO 83
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 83

Asp Leu Glu Ala Cys Val Ile Gln Gly Val Gly Val Thr Glu Thr
1 5 10 15

<210> SEQ ID NO 84
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 84

Ala Cys Val Ile Gln Gly Val Gly Val Thr Glu Thr Pro Leu Met
1 5 10 15

<210> SEQ ID NO 85
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 85

Ile Gln Gly Val Gly Val Thr Glu Thr Pro Leu Met Lys Glu Asp
1 5 10 15

<210> SEQ ID NO 86
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 86

Val Gly Val Thr Glu Thr Pro Leu Met Lys Glu Asp Ser Ile Leu
1 5 10 15

<210> SEQ ID NO 87
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 87

Thr Glu Thr Pro Leu Met Lys Glu Asp Ser Ile Leu Ala Val Arg
1 5 10 15

-continued

<210> SEQ ID NO 88
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 88

Pro Leu Met Lys Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe
1 5 10 15

<210> SEQ ID NO 89
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 89

Lys Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile
1 5 10 15

<210> SEQ ID NO 90
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 90

Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr Leu Tyr
1 5 10 15

<210> SEQ ID NO 91
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 91

Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr Leu Tyr Leu Lys Glu
1 5 10 15

<210> SEQ ID NO 92
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 92

Lys Tyr Phe Gln Arg Ile Thr Leu Tyr Leu Lys Glu Lys Lys Tyr
1 5 10 15

<210> SEQ ID NO 93
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 93

-continued

Gln Arg Ile Thr Leu Tyr Leu Lys Glu Lys Lys Tyr Ser Pro Cys
1 5 10 15

<210> SEQ ID NO 94
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 94

Thr Leu Tyr Leu Lys Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu
1 5 10 15

<210> SEQ ID NO 95
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 95

Leu Lys Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val Arg
1 5 10 15

<210> SEQ ID NO 96
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 96

Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val Arg Ala Glu Ile
1 5 10 15

<210> SEQ ID NO 97
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 97

Ser Pro Cys Ala Trp Glu Val Val Arg Ala Glu Ile Met Arg Ser
1 5 10 15

<210> SEQ ID NO 98
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 98

Ala Trp Glu Val Val Arg Ala Glu Ile Met Arg Ser Phe Ser Leu
1 5 10 15

<210> SEQ ID NO 99
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

-continued

<400> SEQUENCE: 99

Val Val Arg Ala Glu Ile Met Arg Ser Phe Ser Leu Ser Thr Asn
1 5 10 15

<210> SEQ ID NO 100
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 100

Ala Glu Ile Met Arg Ser Phe Ser Leu Ser Thr Asn Leu Gln Glu
1 5 10 15

<210> SEQ ID NO 101
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 101

Met Arg Ser Phe Ser Leu Ser Thr Asn Leu Gln Glu Ser Leu Arg
1 5 10 15

<210> SEQ ID NO 102
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 102

Phe Ser Leu Ser Thr Asn Leu Gln Glu Ser Leu Arg Ser Lys Glu
1 5 10 15

<210> SEQ ID NO 103
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 103

Ile Leu Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp Ser His
1 5 10 15

<210> SEQ ID NO 104
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 104

Ser Tyr Lys Ser Gln Tyr Leu Asn Asn Gly Pro Gln Arg Ile Gly
1 5 10 15

<210> SEQ ID NO 105
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 105

Gly Pro Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met
1 5 10 15

<210> SEQ ID NO 106
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 106

Tyr Lys Trp Thr Val Thr Val Arg Asp Gly Pro Thr Lys Ser Asp
1 5 10 15

<210> SEQ ID NO 107
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 107

Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val Phe Asp Ser
1 5 10 15

<210> SEQ ID NO 108
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 108

Val Ala Tyr Trp Tyr Ile Leu Ser Ile Gly Ala Gln Thr Asp Phe
1 5 10 15

<210> SEQ ID NO 109
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 109

Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln Ser Gly
1 5 10 15

<210> SEQ ID NO 110
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 110

Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys Glu Asn Tyr Arg
1 5 10 15

<210> SEQ ID NO 111
<211> LENGTH: 15
<212> TYPE: PRT

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 111

Ser Thr Leu Phe Leu Val Tyr Ser Asn Lys Cys Gln Thr Pro Leu
1 5 10 15

<210> SEQ ID NO 112
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 112

Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu Asp Gly Lys Lys Trp
1 5 10 15

<210> SEQ ID NO 113
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 113

Ile Ala Arg Tyr Ile Arg Leu His Pro Thr His Tyr Ser Ile Arg Ser
1 5 10 15

Thr Leu Arg Met
20

<210> SEQ ID NO 114
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 114

Pro Lys Tyr Val Lys Gln Asn Thr Leu Lys Leu Ala Thr
1 5 10

<210> SEQ ID NO 115
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell Epitopes

<400> SEQUENCE: 115

Lys Val Val Asp Gln Ile Lys Lys Ile Ser Lys Pro Val Gln His
1 5 10 15

<210> SEQ ID NO 116
<211> LENGTH: 24
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell epitope region

<400> SEQUENCE: 116

Gln Phe Gln Lys Glu Asp Ala Ala Leu Thr Ile Tyr Glu Met Leu Gln
1 5 10 15

-continued

Asn Ile Phe Ala Ile Phe Arg Gln
20

<210> SEQ ID NO 117
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell epitope region

<400> SEQUENCE: 117

Arg Tyr Tyr Gly Arg Ile Leu His Tyr Leu Lys Ala Lys Glu Tyr Ser
1 5 10 15

His Cys Ala Trp Thr
20

<210> SEQ ID NO 118
<211> LENGTH: 34
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell epitope region

<400> SEQUENCE: 118

Ile Ser Leu Phe Ser Cys Leu Lys Asp Arg His Asp Phe Gly Phe Pro
1 5 10 15

Gln Glu Glu Phe Gly Asn Gln Phe Gln Lys Ala Glu Thr Ile Pro Val
20 25 30

Leu His

<210> SEQ ID NO 119
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell epitope region

<400> SEQUENCE: 119

Phe Asn Leu Phe Ser Thr Lys Asp Ser Ser Ala Ala Trp Asp Glu
1 5 10 15

<210> SEQ ID NO 120
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell epitope region

<400> SEQUENCE: 120

Leu Met Lys Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg
1 5 10 15

Ile Thr Leu Tyr
20

<210> SEQ ID NO 121
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell epitope region

<400> SEQUENCE: 121

-continued

Gln Met Arg Arg Ile Ser Leu Phe Ser Cys Leu Lys Asp Arg His Asp
1 5 10 15

Phe Gly Phe Pro
20

<210> SEQ ID NO 122
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell epitope region

<400> SEQUENCE: 122

Gln Met Arg Arg Gln Ser Leu Phe Ser Cys Leu Lys Asp Arg His Asp
1 5 10 15

Phe Gly Phe Pro
20

<210> SEQ ID NO 123
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell epitope region

<400> SEQUENCE: 123

Glu Met Ile Gln Gln Ile Phe Asn Leu Phe Ser Thr Lys Asp Ser Ser
1 5 10 15

Ala Ala Trp Asp Glu Thr Leu Leu Asp Lys Phe Tyr
20 25

<210> SEQ ID NO 124
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell epitope region

<400> SEQUENCE: 124

Glu Met Thr Gln Gln Ile Ala Asn Leu Phe Ser Thr Lys Asp Ser Ser
1 5 10 15

Ala Ala His Asp Glu Thr Leu Leu Asp Lys Phe Tyr
20 25

<210> SEQ ID NO 125
<211> LENGTH: 33
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell epitope region

<400> SEQUENCE: 125

Thr Pro Leu Met Lys Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe
1 5 10 15

Gln Arg Ile Thr Leu Tyr Leu Lys Glu Lys Lys Tyr Ser Pro Cys Ala
20 25 30

Trp

<210> SEQ ID NO 126
<211> LENGTH: 33

-continued

<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Potential T-cell epitope region

<400> SEQUENCE: 126

Thr Pro Leu Met Lys Glu Asp Ser Arg Leu Ala Val Arg Lys Tyr Phe
1 5 10 15

Gln Arg Ile Thr Asn Tyr Leu Lys Glu Lys Lys Tyr Ser Pro Cys Ala
20 25 30

Trp

<210> SEQ ID NO 127
<211> LENGTH: 165
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 127

Cys Asp Leu Pro Gln Thr His Ser Leu Gly Ser Arg Arg Thr Leu Met
1 5 10 15

Leu Leu Ala Gln Met Arg Arg Ile Ser Leu Phe Ser Cys Leu Lys Asp
20 25 30

Arg His Asp Phe Gly Phe Pro Gln Glu Glu Phe Gly Asn Gln Phe Gln
35 40 45

Lys Ala Glu Thr Ile Pro Val Leu His Glu Met Ile Gln Gln Ile Phe
50 55 60

Asn Leu Phe Ser Thr Lys Asp Ser Ser Ala Ala Trp Asp Glu Thr Leu
65 70 75 80

Leu Asp Lys Phe Tyr Thr Glu Leu Tyr Gln Gln Leu Asn Asp Leu Glu
85 90 95

Ala Cys Val Ile Gln Gly Val Gly Val Thr Glu Thr Pro Leu Met Lys
100 105 110

Glu Asp Ser Ile Leu Ala Val Arg Lys Tyr Phe Gln Arg Ile Thr Leu
115 120 125

Tyr Leu Lys Glu Lys Lys Tyr Ser Pro Cys Ala Trp Glu Val Val Arg
130 135 140

Ala Glu Ile Met Arg Ser Phe Ser Leu Ser Thr Asn Leu Gln Glu Ser
145 150 155 160

Leu Arg Ser Lys Glu
165

We claim:

1. A method of identifying a T cell epitope within the amino acid sequence of a polypeptide of interest, the method comprising the steps of:

(i) culturing, in vitro, an aliquot of peripheral blood monocyte cells (PBMC) isolated from a healthy donor in the presence of a peptide for a period of up to about 7 days to form a peptide-primed T cell aliquot, the amino acid residue sequence of the peptide being identical to at least a portion of the amino acid residue sequence of the polypeptide of interest, the peptide being selected from a library of peptides, the amino acid residue sequences of the individual peptides of the library collectively encompassing the entire amino acid residue sequence of the polypeptide of interest;

(ii) culturing the peptide-primed T cell aliquot from step (i) for an additional period of up to about 3 days in the presence of a T cell proliferation-stimulating cytokine to expand the number of T cells therein, forming a T cell-expanded aliquot;

(iii) culturing the T cell-expanded aliquot from step (ii) for a period of about 4 days in the presence of autologous irradiated PBMC from the same donor and in the presence of an additional amount of the peptide sufficient to re-prime the T cells within the PBMC with the peptide;

(iv) determining the level of T cell proliferation of the re-primed T cells relative to an established baseline control level of proliferation; and

(v) repeating steps (i) through (iv) with each peptide of the library of peptides to thereby identify at least one immunogenic region within the amino acid residue sequence of the polypeptide of interest.

2. The method of claim 1 wherein the PBMC have been isolated from a healthy donor whose immune system has not previously been exposed to the polypeptide of interest or any antigenic portion thereof.

3. The method of claim 1 further comprising repeating steps (i) through (v) for each peptide of the library with PBMC isolated from a plurality of healthy donor individuals, the immunological diversity of the plurality of healthy donor individuals representing more than 90% of MHC class II allotypes.

4. The method of claim 1 wherein the polypeptide of interest is a therapeutic protein or a fragment thereof.

5. The method of claim 4 wherein the healthy donor is a human and the therapeutic protein is a human protein.

6. The method of claim 1 wherein each peptide in the library of peptides consists of 9 to 15 amino acid residues.

7. The method of claim 1 wherein each peptide in the library of peptides consists of 15 amino acid residues.

8. The method of claim 1 wherein the cytokine is IL-2.

9. The method of claim 1 wherein the level of T cell proliferation of the re-primed T cells in step (iv) is repeatedly determined over a pre-selected time course protocol.

10. The method of claim 1 wherein the baseline level of T cell proliferation is established by determining a level of T cell proliferation for PBMC from the same donor that have been cultured in the absence of the peptide.

11. The method of claim 10 wherein a stimulation index is calculated for the peptide, the stimulation index being equal to the level of T cell proliferation of the re-primed T cells divided by the level of T cell proliferation of PBMC from the same donor cultured in the absence of the peptide.

12. The method of claim 11 wherein a stimulation index of greater than 1.8 for a given peptide indicates that the portion of the amino acid residue sequence of the polypeptide of interest encompassed by the amino acid residue sequence of the given peptide is a potential immunogenic region of the polypeptide of interest.

13. The method of claim 1 wherein the level of T cell proliferation is determined by culturing the re-primed T cells with tritiated thymidine and measuring the level of tritiated thymidine taken up by the re-primed T cells.

14. The method of claim 1 wherein a monoclonal or polyclonal T cell line isolated from PBMC of the donor are utilized in place of the aliquot of PBMC in step (i).

15. A method of identifying a T cell epitope within the amino acid sequence of a polypeptide of interest, the method comprising the steps of:

(i) culturing, in vitro, an aliquot of peripheral blood monocyte cells (PBMC) isolated from a donor in the presence of a peptide for a period of up to about 7 days to form a peptide-primed T cell aliquot, the amino acid residue sequence of the peptide being identical to at least a portion of the amino acid residue sequence of the polypeptide of interest, the peptide being selected from a library of peptides, the amino acid residue sequences of the individual peptides of the library collectively encompassing the entire amino acid residue sequence of the polypeptide of interest, the donor having an established immune response to the polypeptide of interest;

(ii) culturing the peptide-primed T cell aliquot from step (i) for an additional period of up to about 3 days in the presence of a T cell proliferation-stimulating cytokine to expand the number of T cells therein, forming a T cell-expanded aliquot;

(iii) culturing the T cell-expanded aliquot from step (ii) for a period of about 4 days in the presence of autologous irradiated PBMC from the same donor and in the presence of an additional amount of the peptide sufficient to re-prime the T cells within the PBMC with the peptide;

(iv) determining the level of T cell proliferation of the re-primed T cells relative to an established baseline control level of proliferation; and

(v) repeating steps (i) through (iv) with each peptide of the library of peptides to thereby identify at least one immunogenic region within the amino acid residue sequence of the polypeptide of interest.

16. The method of claim 15 wherein the polypeptide of interest is a therapeutic protein or a fragment thereof.

17. The method of claim 16 wherein the healthy donor is a human and the therapeutic protein is a human protein.

18. The method of claim 15 wherein each peptide in the library of peptides consists of 9 to 15 amino acid residues.

19. The method of claim 15 wherein each peptide in the library of peptides consists of 15 amino acid residues.

20. The method of claim 15 wherein the cytokine is IL-2.

21. The method of claim 15 wherein the level of T cell proliferation of the re-primed T cells in step (iv) is repeatedly determined over a pre-selected time course protocol.

22. The method of claim 15 wherein the baseline level of T cell proliferation is established by determining a level of T cell proliferation for PBMC from the same donor, but which has been cultured in the absence of the peptide.

23. The method of claim 22 wherein a stimulation index is calculated for the peptide, the stimulation index being equal to the level of T cell proliferation of the re-primed T cells divided by the level of T cell proliferation of PBMC from the same donor cultured in the absence of the peptide.

24. The method of claim 23 wherein a stimulation index of greater than 1.8 for a given peptide indicates that the portion of the amino acid residue sequence of the polypeptide of interest encompassed by the amino acid residue sequence of the given peptide is a potential immunogenic region of the polypeptide of interest.

25. The method of claim 15 wherein the level of T cell proliferation is determined by culturing the re-primed T cells with tritiated thymidine and measuring the level of tritiated thymidine taken up by the re-primed T cells.

26. The method of claim 15 wherein a monoclonal or polyclonal T cell line isolated from PBMC of the donor are utilized in place of the aliquot of PBMC in step (i).

27. A method for preparing a variant of a therapeutic protein having substantially the same biological activity and reduced immunogenicity compared to the therapeutic protein, the method comprising the steps of:

(i) preparing at least one variant of the therapeutic protein, the amino acid residue sequence of the variant differing from the amino acid residue sequence of the therapeutic protein by an amino acid residue within an immunogenic region of the therapeutic protein, the immunogenic region being identified by the method of claim 1;

(ii) comparing the biological activity and immunogenicity of the at least one variant to the biological activity and immunogenicity of the therapeutic protein; and

(iii) selecting a variant having substantially the same biological activity and reduced immunogenicity compared to the therapeutic protein.

28. The method of claim 27 wherein the amino acid residue sequence of the at least one variant is selected by the steps of:

(a) calculating a MHC Class II molecule binding score for the immunogenic region using a computational method that sums assigned values for each hydrophobic amino acid residue side chain present in the immunogenic region;

(b) calculating a binding score for at least one amino acid residue sequence that differs from the amino acid residue sequence of the immunogenic region by an amino acid residue using the same computational method as in step (a); and

(c) selecting the amino acid residue sequence for the at least one variant having a binding score in step (b) that is lower than the binding score of the immunogenic region of the therapeutic protein.

29. A method for preparing a variant of a therapeutic protein having substantially the same biological activity and reduced immunogenicity compared to the therapeutic protein, the method comprising the steps of:

(i) preparing at least one variant of the therapeutic protein, the amino acid residue sequence of the variant differing from the amino acid residue sequence of the therapeutic protein by an amino acid residue within an immunogenic region of the therapeutic protein, the immunogenic region being identified by the method of claim 15;

(ii) comparing the biological activity and immunogenicity of the at least one variant to the biological activity and immunogenicity of the therapeutic protein; and

(iii) selecting a variant having substantially the same biological activity and reduced immunogenicity compared to the therapeutic protein.

30. The method of claim 29 wherein the amino acid residue sequence of the at least one variant is selected by the steps of:

(a) calculating a MHC Class II molecule binding score for the immunogenic region using a computational method that sums assigned values for each hydrophobic amino acid residue side chain present in the immunogenic region;

(b) calculating a binding score for at least one amino acid residue sequence that differs from the amino acid residue sequence of the immunogenic region by an amino acid residue using the same computational method as in step (a); and

(c) selecting the amino acid residue sequence for the at least one variant having a binding score in step (b) that is lower than the binding score of the immunogenic region of the therapeutic protein.

* * * * *

专利名称(译)	用于定位和消除T细胞表位的方法		
公开(公告)号	US20050181459A1	公开(公告)日	2005-08-18
申请号	US11/009460	申请日	2004-12-10
[标]申请(专利权)人(译)	BAKER MATTHEW CARR FRANCISĴ CARTER GRAHAM		
申请(专利权)人(译)	BAKER MATTHEW CARR FRANCIS J. CARTER GRAHAM		
当前申请(专利权)人(译)	BAKER MATTHEW CARR FRANCIS J. CARTER GRAHAM		
[标]发明人	BAKER MATTHEW CARR FRANCIS J CARTER GRAHAM		
发明人	BAKER, MATTHEW CARR, FRANCIS J. CARTER, GRAHAM		
IPC分类号	C12N5/08 G01N33/48 G01N33/50 G01N33/53 G01N33/567 G01N33/68 G06F19/00		
CPC分类号	G01N33/6878 G01N33/5047		
优先权	PCT/EP2003/006110 2003-06-11 WO 2002012919 2002-06-11 EP		
外部链接	Espacenet USPTO		

摘要(译)

本发明提供了鉴定多肽的氨基酸残基序列内的免疫原性区域的方法，例如治疗性蛋白质或其片段。该方法包括以下步骤：（i）在肽存在下，在体外培养从供体分离的等分试样的外周血单核细胞（PBMC），持续最多约7天，氨基酸残基序列该肽与目的多肽的至少一部分氨基酸残基序列相同，该肽选自肽文库，该文库的各个肽的氨基酸残基序列共同包含整个氨基酸。目标多肽的残基序列;在T细胞增殖刺激细胞因子的存在下，将来自步骤（i）的T细胞等分试样再培养至多约3天，以扩增其中T细胞的数量;（iii）在来自相同供体的自体辐射的PBMC存在下，在存在足以重新引发T细胞的额外量的肽的情况下，将步骤（ii）的T细胞等分试样培养约4天。在具有肽的PBMC内;（iv）相对于确定的基线对照增殖水平确定再引发的T细胞的T细胞增殖水平;（v）用肽文库的每个肽重复步骤（i）至（iv），从而鉴定目的多肽的氨基酸残基序列内的至少一个免疫原性区域。