Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2500401 A1 2004/04/15

(21) 2 500 401

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2003/09/25
(87) Date publication PCT/PCT Publication Date: 2004/04/15
(85) Entrée phase nationale/National Entry: 2005/03/29
(86) N° demande PCT/PCT Application No.: RU 2003/000421
(87) N° publication PCT/PCT Publication No.: 2004/031212
(30) Priorité/Priority: 2002/10/03 (2002126396) RU

(51) Cl.Int.^{7}/Int.Cl.^{7} C07K 1/00, A61P 31/12, A61K 39/12, A61K 39/02, C07K 14/195, C07K 14/005

(71) Demandeur/Applicant:
STATE RESEARCH CENTER OF VIROLOGY AND BIOTECHNOLOGY "VECTOR", RU

(72) Inventeurs/Inventors:
MAKSYUTOV, AMIR ZAKIEVICH, RU;
RYZHIKOV, ALEXANDER BORISOVICH, RU;
KOLOBOV, ALEXANDER ALEXANDROVICH, RU;
MAKSYUTOV, ZAKI AMIROVICH, RU

(74) Agent: OGILVY RENAULT LLP/S.E.N.C.R.L.,S.R.L.

(54) Titre : PEPTIDES ANTIGENIQUES
(54) Title: ANTIGENIC PEPTIDES

(57) Abrégé/Abstract:
A method for designing antigenic peptide libraries accounts for naturally occurring and potential variability in a group of protein sequences from a variable pathogen. The peptide libraries can elicit an immune response against a range of pathogen variants.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 April 2004 (15.04.2004)

PCT

(10) International Publication Number
WO 2004/031212 A1

(51) International Patent Classification[7]: C07K 1/00, 14/005, 14/195, A61K 39/12, 39/02, A61P 31/12

(21) International Application Number:
PCT/RU2003/000421

(22) International Filing Date:
25 September 2003 (25.09.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2002126396 3 October 2002 (03.10.2002) RU

(71) Applicant *(for all designated States except US)*: **STATE RESEARCH CENTER OF VIROLOGY AND BIOTECHNOLOGY "VECTOR"** [RU/RU]; pgt Koltsovo, Novosibirsky r-n, Novosibirskaya obl., 630559 (RU).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MAKSYUTOV, Amir Zakievich** [RU/RU]; pos. Koltsovo, 12-27, Novosibirskaya obl., 630559 (RU). **RYZHIKOV, Alexander Borisovich** [RU/RU]; pos. Koltsovo, 24-5, Novosibirskaya obl., 630559 (RU). **KOLOBOV, Alexander Alexandrovich** [RU/RU]; Primorskoe shosse, 265-102, Sestroretsk, 188640 (RU). **MAKSYUTOV, Zaki Amirovich** [RU/RU]; pos. Koltsovo, 12-27, Novosibirskaya obl., 630559 (RU).

(74) Agent: **LAW FIRM "GORODISSKY & PARTNERS" LTD.**; Galina EGOROVA , ul. B. Spasskaya, 25, stroenie 3, Moscow, 129010 (RU).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/031212 A1

(54) Title: ANTIGENIC PEPTIDES

(57) Abstract: A method for designing antigenic peptide libraries accounts for naturally occurring and potential variability in a group of protein sequences from a variable pathogen. The peptide libraries can elicit an immune response against a range of pathogen variants.

ANTIGENIC PEPTIDES

CLAIM OF PRIORITY

This application claims priority to Russian Patent Application Serial No. 2002126396, filed on September 27, 2002, the entire contents of which are hereby incorporated by reference.

FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

The U.S. Government may have certain rights in this invention pursuant to Department of Health and Human Services Grant Nos. ISTC #2450 and BTEP #34.

TECHNICAL FIELD

This invention relates to antigenic peptides.

BACKGROUND

A number of pathogens important to human health are variable pathogens, such as HIV-1, HIV-2, hepatitis B and C, influenza, dengue types 1-4 viruses, malaria, and tuberculosis. These and other variable pathogens present a challenge to vaccine development, due to the extremely high genetic variability of the pathogen. In particular, the number of people infected with HIV throughout the world is growing rapidly, especially in countries with poor health care resources. The development of a highly efficient vaccine could help restrict the propagation of this epidemic. One of the main obstacles to developing such a vaccine is the extremely high genetic diversity of HIV. The genetic diversity is observed in various regions of the world in different subtypes, within a single subtype and even, in some cases, within an individual. This variability results from the high rate of mutations in viral proteins occurring during virus replication. See, for example, Coffin, J.M. *Science* 1995, 267:483-489; Delwart et al., *J. Virol.* 1994 68:6672-6683; Markham R.B. et al. *Proc. Natl. Acad. Sci. U.S.A.* 1998, 95:12568-12573; and McCutchan, F.E. et al. *AIDS* 1996 10 Suppl 3:S13-S20, each of which is incorporated by reference in its entirety.

It has been shown in animal models that simultaneous exposure to multiple antigenic determinants often induces immune responses to each of them (Thomson, S.A. et al. *Proc. Natl. Acad. Sci. U.S.A.* 1995, 92:5845-5849; Thomson, S.A. et al. *J. Virol.* 1998, 72: 2246-

2252; and Woodberry, T. et al. *J. Virol.* 1999, 73:5320-5325, each of which is incorporated by reference in its entirety). This indicates that multiple antigenic determinants are suitable as components of a single vaccine construct.

SUMMARY

The protective component of an immune response to a variable pathogen is often directed against a variable region of a protein. If, on the other hand, the protective response was directed against a conserved region of a protein, pathogen variability would not present a problem to immunological control of infections and vaccine prophylaxis. However, an immune response against conserved regions of variable pathogens often proves to be nonprotective. A peptide antigen based on a variable region can induce a protective response, but the variability of these regions is so great that it is impractical to cover the variability by addition of individual peptide sequences. A potential vaccine against a variable pathogen should include a variety of antigens, to produce an immune response to the variants of the pathogen to which an individual may become exposed.

Vaccine prophylaxis or therapy against variable infectious agents requires constructing a vaccine capable of inducing wide-range humoral and cellular immune responses to antigen variable regions. A challenge for vaccine prophylaxis and therapy against variable infectious agents is the development of a vaccine capable of inducing an immune response against variable or hypervariable regions of their protein components.

A library of chimeric peptide antigens containing one or more potentially protective epitopes and vaccine constructs containing such antigens can produce a protective immune response. In particular, chimeric peptide libraries are able to induce a protective immune response including both humoral and cellular responses to HIV-1, HIV-2, hepatitis C and B viruses, influenza virus, dengue types 1-4 viruses, malaria, tuberculosis, or other variable pathogens.

Chimeric peptide libraries can be designed to represent naturally occurring or potential variants of protein epitopes. Peptide libraries developed in this manner can produce a broad immune cross-response and can be applied for constructing vaccines, pharmaceuticals, and diagnostic kits. The design methods can be applied to a broad range of

antigenic determinants of an infectious agent. Different methods can be applied to one infectious agent or one of its proteins or to a single epitope of the protein.

Immunogenic sets of chimeric peptides, referred to as variable chimeric peptide libraries (VCPLs), represent the naturally occurring and potential variability of antigenically active regions. Variable chimeric peptide libraries are sets of homologous peptides variable at one or more positions. They are designed to mimic the genetic diversity of variable protective determinants of an infectious agent. Such VCPLs can induce production of a wide range of antibodies and cytotoxic T-lymphocytes (CTLs) with a joint specificity that covers the diversity of antigenic variants of the variable infectious agent.

In one aspect, a method of manufacturing a family of antigenic peptides includes locating a plurality of variable positions in a region of a pathogen protein, choosing a peptide sequence of the pathogen protein including the plurality of variable positions, selecting one or more substitute amino acid residues for one of the variable positions based on antigenic similarity to amino acid residues naturally occurring at the variable position of the pathogen protein, and preparing a family of antigenic peptides based on the peptide sequence and including the substitute amino acid residues.

In another aspect, a method of designing a family of peptide sequences includes locating a plurality of variable positions in a region of a pathogen protein, choosing a peptide sequence of the pathogen protein including the plurality of variable positions, and selecting one or more substitute amino acid residues for one of the variable positions of the peptide based on antigenic similarity to amino acid residues naturally occurring at the variable position of the pathogen protein, thereby forming a family of peptide sequences.

Selecting can include determining the antigenic similarity using an antigenic similarity matrix. A frequency can be assigned to each substitute amino acid residue in the family of antigenic peptides. Preparing can include weighting the substitute amino acid residues in the family of antigenic peptides based on the assigned frequency. Assigning can include considering the frequency with which the variations naturally occur. The pathogen protein can include a hypervariable region. The pathogen protein can be associated with a virus. The virus can be HIV, hepatitis B virus or hepatitis C virus, or an influenza. The pathogen protein can be HIV gp120. The region can be the V1, V2, V3, V4, or V5 regions.

The pathogen protein can be associated with a malaria pathogen, a dengue pathogen, or a tuberculosis pathogen.

The family of antigenic peptides can include members, such that the members taken together have antigenic similarity to each naturally occurring sequence of the region of the pathogen protein. The family of antigenic peptides can include members, such that the members taken together have antigenic similarity to a non-naturally occurring sequence of the region of the pathogen protein. The method can include identifying peptide sequences of the family, the identified peptide sequences being representative of the sequence diversity of the entire family. Fewer than 500 sequences can be identified as being representative of the sequence diversity of the entire family. Identifying can include calculating a distance between peptide sequences of the family. Calculating a distance can include using an antigenic similarity matrix.

The method can include determining an antigenic similarity between a peptide of the family and a region of a human protein. The method can include removing a peptide from the family of antigenic peptides before preparing the family if the determined antigenic similarity between the peptide of the family and the region of a human protein exceeds a predetermined threshold.

Preparing the family of antigenic peptides can include chemical synthesis of the family of peptides. The chemical synthesis can includes combinatorial synthesis, whereby the peptides are formed as a mixture of different sequences. The separate peptides can be mixed. The chemical synthesis can include parallel synthesis, whereby each peptide is formed separately from other peptides. Preparing the family of antigenic peptides can include expression of the family of peptides by a host organism.

In another aspect, a composition includes a family of antigenic peptides having amino acid sequences having antigenic similarity to amino acid sequences of a variable region of a pathogen protein, wherein each antigenic peptide in the family has at least one amino acid position that varies relative to other antigenic peptides in the family.

One amino acid residue can occur more frequently than another in the position that varies. The family can include greater than 150 or greater than 1,000 mutually unique antigenic peptides. The family can include fewer than 100,000 or fewer than 50,000 mutually unique antigenic peptides. The family can include between 1,000 and 50,000

mutually unique antigenic peptides. The family of antigenic peptides can include sequences having antigenic similarity to sequences from a subtype of HIV. The subtype can be subtype A, subtype B, subtype C, subtype D, subtype F, subtype G, a recombinant subtype, a subtype of HIV group N, a subtype of HIV group O, or combinations thereof. At least two members of the family of antigenic peptides can be mixed together. The family of antigenic peptides can be separated according to sequence. The family can include a multiple antigenic peptide.

In another aspect, a method of eliciting an immune response in a subject includes administering to the subject a composition including a family of antigenic peptides having amino acid sequences having antigenic similarity to amino acid sequences of a variable region of a pathogen protein of a pathogen.

Antigenic similarity can be determined using an antigenic similarity matrix. The composition can be administered to a subject prior to infection by the pathogen. The family of antigenic peptides can have amino acid sequences having antigenic similarity to amino acid sequences of a subtype of the pathogen protein. The family of antigenic peptides can have amino acid sequences having antigenic similarity to amino acid sequences from more than one subtype of the pathogen protein. The composition is administered to a subject infected by the pathogen. The subject can be infected by a subtype of the pathogen. The family of antigenic peptides can have amino acid sequences having antigenic similarity to amino acid sequences from the subtype by which the subject is infected.

In another aspect, a method of diagnosing infection includes contacting a sample with a family of peptides having amino acid sequences having antigenic similarity to amino acid sequences of a variable region of a pathogen protein, wherein each peptide in the family has at least one amino acid position that varies relative to other peptides in the family.

The family of peptides can be antigenically similar to a pathogen subtype, or to more than one pathogen subtype. The family of peptides can be immobilized on a substrate before contacting. The method can include determining if the sample includes antibodies that bind specifically to the family of peptides.

In general, antigenic peptide libraries reflect the existing and potential diversity of a variable region of a pathogen protein. The peptide libraries can be used in prophylactic, therapeutic, and diagnostic applications. The full antigenic diversity of a pathogen epitope can be represented by the library. A library designed to reflect the diversity of all variants of

a pathogen can be a broad-spectrum vaccine or diagnostic. A library designed to reflect the diversity one or more subtypes of pathogen can be a subtype-specific vaccine or diagnostic. Peptide sequences that are highly similar to sequences found in human proteins and are thus likely to induce an autoimmune reaction can be removed from the library. A library can be designed for any variable pathogen protein of known sequence. Examples of variable pathogens include HIV-1, HIV-2, hepatitis B and C, influenza, dengue types 1-4 viruses, malaria, and tuberculosis.

The details of one or more embodiments are set forth in the description below. Other features and advantages will be apparent from the description and from the claims.

DETAILED DESCRIPTION

Much effort has been devoted to development of a vaccine against HIV in particular, but without success to date. To be efficient against the broad range of HIV-1 isolates, potential vaccines should contain immunogens of main isolates representing the entire range of HIV-1 isolates. It is necessary to identify these immunogens and improve their immunogenic properties (McMichael, A.J., and Hanke, T, *Nature Medicine* 2003, 9:874-880; Baltimore, D. and Heilman, C. *Sci. Am.* 1998, 279: 98-103; and van der Groen, G. et al. *AIDS Res. Hum. Retroviruses* 1998, 14 Suppl. 3: S211-S221, each of which is incorporated by reference in its entirety).

Therefore, the resulting vaccine construct should contain key protective B- and T-cell epitopes specific to the broad range of HIV-1 isolates, combined with modern delivery means and adjuvants. Activation of both humoral and cellular branches of the immune system is required (Berzofsky, J.A., and Berkower, I.J., *AIDS* 1995, 9 Suppl. A: S143-S157, which is incorporated by reference in its entirety). The local immune response on mucous membranes and circulating neutralizing antibodies are considered to play key roles in preventing HIV infection, whereas virus-specific CTLs are likely to eliminate infected cells. See, for example, Lehner, T. et al. *Science* 1992, 258:1365-1369; and Goulder, P.J. and Walker, B.D. *Nature Med.* 1999, 5:1233-1235, each of which is incorporated by reference in its entirety.

The amino acid sequence of the envelope glycoprotein (also called env or gp120) of HIV is highly variable between independent isolates and between sequential isolates from a single infected individual. The amino acid variability in gp120 is concentrated in specific

variable regions, numbered V1-V5. The most variable regions often contain neutralizing epitopes so that the virus partially evades the host's immune response and establishes a persistent infection. The principal neutralizing determinant of gp120 is located in the third hypervariable region (V3).

A number of candidate peptide vaccines against HIV-1 have been developed with the recognition of the V3 region as the principal neutralizing determinant in mind. For example, a method of rational design of a polyvalent subunit vaccine against HIV included determination of neutralizing epitopes of the V2 and V3 domains of HIV gp120 isolates from various areas of the world. See, for example, U.S. Patent No. 6,042,836, which is incorporated by reference in its entirety. Another approach to synthetic immunogenic HIV-1 peptides included designing a tandem synthetic peptide containing the sequences of T- and B-cell epitopes and the sequence of the V3 loop from various HIV isolates. See, for example, U.S. Patent No. 5,817,754, which is incorporated by reference in its entirety. However, the high genetic variability of the V3 domain results in a variety of sequences, which cannot be simultaneously introduced into a vaccine construct.

For the same reasons as in the case of vaccines, it is extremely difficult to diagnose antibodies against hypervariable regions of variable infectious agents. These antibodies are the earliest and dominant in the immune response. An example is antibodies against the V3 domain of HIV-1 gp120. See, for example, Baltimore, D. and Heilman, C. *Sci. Am.* 1998, 279: 98-103, Berzofsky, J.A., and Berkower, I.J., *AIDS* 1995, 9 Suppl. A: S143-S157, and Carlos, M.P. et al., *AIDS Res. Hum. Retroviruses* 2000, 16:153-161, each of which is incorporated by reference in its entirety.

The diagnostics of such antibodies against the V3 hypervariable domain of gp120 depends on the genotypes of the antigens involved and the genotype of the infectious agent. It can require using multiple sequences corresponding to a broad range of HIV-1 variants. Otherwise, there is a high risk of missing a rare or modified variant of the virus.

A peptide-based vaccine can include a library of peptides. The peptides can be similar to a variable region of a pathogen protein. In particular, the variable region can include a known antigenic determinant. The peptide sequences in the library are chosen to reflect the naturally occurring and potential variability of the variable region.

Chimeric peptide libraries that mimic the genetic diversity of variable protein regions of a pathogen can induce a protective immune response to a broad spectrum of pathogen variants. To design the chimeric peptide library, a theoretical analysis of protein sequence information is carried out. The theoretical analysis results in a library design that can mimic all known and potential variants of the variable protein region. The library can be a subtype-specific library, mimicking only those variants belonging to a particular subtype of the pathogen, or a broad spectrum library, mimicking variants belonging to more than one subtype. The peptides of the library can bind antibodies that are specific to its various antigenic variants.

In general, the design of an antigenic chimeric peptide library begins by inspecting sequences of a protein from a variable pathogen. A variable region of the protein is identified. The variable region can be a known antigenic determinant. A peptide fragment within the protein sequence including at least one variable position is selected to be the basis of the chimeric peptide library. The fragment includes one or more positions having naturally sequence variations. The peptide fragment can include non-varying positions. The fragment can be 10-50 residues in length, such as 15-45 residues or 20-40 residues in length. For each variable position in the peptide, the residues that occur in that position and the frequency with which each occurs are identified. The choice of residues at each variable position in the library is determined by considerations of antigenic similarity to the residues naturally occurring at that position. The frequency of each residue at a variable position is assigned according to the frequency that it or its corresponding antigenically similar residue occurs naturally. Selection of residues and frequencies is described in more detail below. The residue selection and frequency assignment can be automated, i.e. carried out by a computer program.

The number of sequences in the library (the size of the library) is the product of the number of possible residues at each position in the peptide. The size of the library can include, for example, more than 200,000,000 sequences, fewer than 200,000,000 sequences, fewer than 500,000 sequences, fewer than 200,000 sequences, fewer than 100,000 sequences, fewer than 50,000 sequences, more than 5,000 sequences, more than 10,000 sequences, or more than 20,000 sequences. The size of the library can be reduced by eliminating the residue that occurs with the lowest frequency until a desired size is reached.

Immunogenic sets of chimeric peptides, (variable chimeric peptide libraries or VCPLs) when taken together, represent the naturally occurring and potential variability of antigenically active regions. Variable chimeric peptide libraries are sets of homologous peptides variable at one or more positions. They are designed to mimic the genetic diversity of main variable protective determinants of an infectious agent. Such VCPLs can induce a wide range of antibodies and cytotoxic T-lymphocytes (CTLs), having a joint specificity that covers the diversity of antigenic variants of the infectious agent. Thus, VCPLs induce humoral and cellular immune responses against the diversity of currently existing variants of an infectious agent and against potential variants which may arise in the future.

Libraries of variable chimeric peptides can be constructed so that the conserved positions of these peptides represent amino acids common for all sequences, and the heterogeneous diversity of the variable positions is mimicked by one or several amino acids. The mimicking amino acids can be chimeric, in other words, they mimic the general heterogeneity of variable positions but need not occur at corresponding positions among the known and potential diversity of natural sequences.

The method of designing VCPLs rests upon comprehensive investigation of the whole body of structural evidence on the organization of immunodominant protein epitopes and on the induction of a protective immune response to the target pathogen. The library design begins with collection and analysis of amino acid sequences from the target pathogen, and estimation of the genetic variability and epidemiological significance of pathogen subtypes. The amino acid sequences are searched for conserved, variable, and hypervariable regions that contribute to the development of a protective immune response against the pathogen.

The antigenic properties of the peptide library can be controlled by selecting the pathogen protein sequences that the library is designed from. For example, a library mimicking a particular subtype of pathogen can be designed by selecting only sequences derived from that subtype. In general, the sequences within any one subtype or sub-subtype are more similar to each other than to sequences from other subtypes throughout their genomes. Group M is the main group of viruses in the HIV-1 global pandemic, and it contains multiple subtypes. Group O is the "outlier" group, and group N is a very distinctive form of the virus that is Non-M, Non-O. The HIV-1 subtypes from the M group of HIV-1,

for example, are phylogenetically associated groups or clades of HIV-1 sequences, and are labeled A1, A2, B, C, D, F1, F2, G, H, J and K. In addition, recombinant HIV-1 sequences are possible, where the viral genome includes regions of sequence from more than one subtype. See, for example, the HIV Sequence Database, at hiv-web.lanl.gov. These subtypes represent different lineages of HIV, and have some geographical associations. A library mimicking a particular subtype or recombinant form can be useful, for example, in cases where particular subtypes are found predominantly in particular geographic locations.

The number of potential evolutionary variants of the sequence of an antigenically active region can be enormous. For example, the potential antigenic variability of the V3 region of HIV-1 gp120 is as high as 10^{25}–10^{30} variants. Once the complete variability at each position in the peptide is determined from the naturally occurring sequences, the variability of the desired VCPL can be reduced by recognizing the most significant amino acids that determine the genotypic diversity of the antigenic variants. The rates of occurrence of various amino acids at variable positions of an immunogenic epitope are determined with regard to their relative occurrence in the alignment of the sequences of the region at issue.

The size of the VCPL can be decreased by recognizing the most antigenically significant amino acids at each variable position, and combining them in groups of variable composition with one or more amino acids representing the groups of variable composition. The amino acids representing groups of variable composition are determined on the basis of antigenic similarity.

Antigenic similarity is a quantitative measure of how readily one amino acid residue can substitute for another in a protein-protein interaction, such as, for example, an antibody-antigen interaction. Two residues are very antigenically similar if substituting one for the other in a particular epitope has a small effect on the antigenic behavior of that epitope. Conversely, if the substitution has a large effect, then the two residues have a low antigenic similarity. In general, residues that have similar physical and chemical characteristics will have a high antigenic similarity.

Antigenic similarity can be expressed in a matrix, called an antigenic similarity matrix (ASM). The construction of a matrix is described in, for example, Maksyutov, A.Z., Eroshkin, A.M., and Kulichkov, V.A, *Mol. Biol.* 1987, 21:30-47, which is incorporated by reference in its entirety. The matrix can account for the affinity of two residues, that is, how

frequently two residues are in contact in a globular protein, as determined by examination of 3D structures. See, for example, Warme, P.K., and Morgan, R.S., *J. Mol. Biol.* 1978, 118: 289-304, which is incorporated by reference in its entirety. The matrix is also adjusted for the frequency with which an amino acid occurs in the hypervariable regions of immunoglobulins. See, for example, Bourgarit, J., *J. Ann. Immunol.* 1980, 131D: 267-287, which is incorporated by reference in its entirety.

The values of an initial matrix can be transformed by the formula:

$$ASM_{ij} = 100 - (R_{ij} + R_{ji})/2$$

for each pair of amino acids *i* and *j*, and R_{ij} and R_{ji} are elements of the initial matrix, thus producing a symmetric matrix. To differentiate the effects of coinciding amino acids and increase the weight of the residues that occur frequently in antigenic epitopes, the diagonal elements can be adjusted according to the following formula:

$$ASM_{ii} = ASM_{ii} + 5(h_i + s_i + c)$$

where h_i is the hydrophilicity of the amino acid (Hopp, T.P., and Woods, K.R., *Proc. Natl. Acad. Sci. USA* 1981, 78:3824-3828, which is incorporated by reference in its entirety), s_i is the relative frequency of the amino acid in antigenic determinants of proteins (Maksyutov, A.Z. and Zagrebelnaya, E.S., *Comp. Applic. Biosci.* 1993, 9:291-297, which is incorporated by reference in its entirety), and c is 6.6, the minimum value of $h_i + s_i$, introduced to insure a positive sum ($h_i + s_i + c$). Finally, to satisfy the local alignment program SIM, the quantity 75 is subtracted from each element in the matrix. One such matrix is presented in Table 1. The ASM of Table 1 is aimed at the search for locally similar regions exhibiting close structural features in respect to interaction between an antibody or another binding ligand and the binding site (see, for example, Maksyutov A. Z. et al., *Molecular Biology* 2002, 36: 346-358, which is incorporated by reference in its entirety).

Qualitatively, a matrix value for a pair of amino acids indicates that they are good substitutes for one another, and a low number indicates that they are poor substitutes. For example, valine and leucine have a high antigenic similarity according to the matrix (matrix value of 9), but valine and aspartic acid do not (matrix value of -59), as would be expected from chemical considerations.

Table 1

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	59																			
C	-15	50																		
D	-41	-50	81																	
E	-49	-53	4	79																
F	-7	-12	-49	-51	47															
G	5	-15	-25	-35	-25	49														
H	-30	-34	-42	-49	-34	-35	61													
I	0	-22	-62	-65	-4	-35	-48	37												
K	-38	-49	-48	-62	-46	-35	-12	-55	88											
L	3	-12	-55	-60	4	-30	-41	11	-47	48										
M	-7	-15	-57	-55	-5	-30	-33	-3	-46	4	37									
N	-36	-37	-20	-27	-44	-15	-23	-57	-23	-46	-43	55								
P	-11	-13	-28	-36	-17	-15	-8	-29	-24	-19	-19	-19	59							
Q	-24	-24	-15	-29	-29	-10	-11	-43	-31	-34	-30	3	-10	71						
R	-34	-42	-44	-54	-39	-25	-13	-51	7	-46	-45	-23	-24	-26	69					
S	-23	-31	-15	-35	-34	-5	-3	-45	-19	-38	-40	-10	-4	-5	-15	61				
T	-27	-30	-20	-26	-33	-10	-18	-46	-31	-41	-42	-15	-13	-11	-18	6	62			
V	-1	-16	-59	-62	-4	-30	-44	13	-55	9	-1	-51	-26	-37	-50	-42	-39	47		
W	-22	-21	-38	-39	-18	-40	-31	-27	-44	-19	-22	-23	-9	-19	-38	-23	-17	-19	24	
Y	-25	-30	-26	-20	-20	-25	-25	-37	-31	-27	-23	-10	-13	-12	-32	-19	-18	-32	-8	53

At each position in the peptide, the matrix value for each pair of residues occurring naturally is looked up. A pair having a positive matrix value will be represented in the library by the residue that occurs more frequently in the natural sequences, and its frequency in the library will be the sum of frequencies for each residue that it represents. The remaining residues are combined in subgroups, and for each subgroup, the matrix is searched for a residue that has a positive matrix value with each residue in the subgroup. If such a residue is found, it is assigned a frequency equal to the sum of frequencies of residues in the subgroup. This residue can be one that does not naturally occur in that position. If a naturally occurring residue does not have a positive matrix value with any other naturally occurring residue at that position, and cannot be represented by another residue not naturally occurring at that position, it will represent itself in the library.

As an example of how the matrix can be used to choose residues for a particular position in a peptide library, consider a variable position in where the following six different residues occur naturally: A (5%), D (20%), E (35%), M (22%), V (10%), and Y (8%). The matrix value for each pair of naturally occurring residues is then looked up:

AD	-41
AE	-49
AM	-7
AV	-1
AY	-25
DE	4

DM -7
DV -59
DY -26
EM -55
EV -62
EY -20
MV -1
MY -10
VY -32

Only one pair, DE, has a positive matrix value. That pair of residues will be represented by the more frequent naturally occurring residue, E, and will have a frequency equal to the sum of the frequency of the residues it represents (20+35= 55%). The remaining residues, A, M, V and Y, are divided into subgroups and the matrix is searched for other residues that are antigenically similar to these. The subgroup of residues A, M, and V each have a positive matrix value with L:

AL 3
ML 4
VL 9

Thus, in the library, A, M, and V will be represented by L, which is antigenically similar to all three. The frequency of L in the library is the sum of the frequencies of A, M and V: 5+10+22 = 37%. Finally, since Y is not antigenically similar to another residue, Y represents itself in the library with a frequency of 8%. Table 2 summarizes how the residues from natural sequences are represented in the library for this example.

Table 2

Naturally occurring		Library	
D	20 %	E	55 %
E	35 %		
A	5 %	L	37 %
M	22 %		
V	10 %		
Y	8 %	Y	8 %

Theoretical analysis of the potential cross-reactivity of antibodies raised against peptide libraries shows that in order to induce an immune cross-response sufficiently broad to cover the whole set of antigenic variants of a variable infectious agent would require a library containing tens of thousands of homologous peptides. A smaller library having hundreds of homologous peptides could not induce a sufficiently broad immune response to existing and potential variants of a variable infectious agent. However, increasing the size of the library also increases the danger of that the library will elicit an autoimmune response. The analysis was carried out by methods for estimation of the cross-reactivity of antibodies against homologous antigens. See, for example, Maksyutov A.Z., *Voprosy Virusologii* 1989, 3: 283-288; Maksyutov A.Z., Eroshkin A.M., and Kulichkov V.A., *Molekularnaya Biologiya* 1987, 21: 39-47; Maksyutov, A. Z., and Zagrebelnaya, E. S., *Computer Applications in the Biosciences* 1993, 9: 291-297; and Maksyutov, A. Z., and Zagrebelnyi S.N., *Molekularnaya Biologiya* 1993, 27: 980-991, each of which is incorporated by reference in its entirety.

Typically during the natural infection of an individual and selection of antigenic variants of an infectious agent there is too little time for formation of a sufficient amount of immune effectors capable of cross-reaction with normal human proteins to induce an autoimmune response. Induction of autoimmune processes usually takes a long time, and often requires the constant presence of the sensitizing antigen. Hence, the probability of autosensitization in an individual as a result of local similarity between variants of the hypervariable regions of the variable infectious agent and a human protein may be insignificant. On the other hand, the risk of induction of autoimmune conditions is increased by repeated use of a peptide vaccine for prophylaxis or therapy of an infectious disease. Therefore, during the design of polyepitope vaccine constructs, it should be taken into account that the long persistence of "time-conservative" variants of variable regions of the variable infectious agent can bring about autoimmune conditions, particularly with a vaccine construct containing adjuvants for increasing the immune response. Thus, the design of vaccine constructs on the basis of VCPLs requires comprehensive theoretical consideration of the local similarity between VCPL peptides and the whole set of human proteins.

It can be advantageous to screen the sequences in a library for their potential to generate an autoimmune response. If a peptide in the library is highly similar to a peptide present in a human protein, an immune response could be generated against the human

protein, with deleterious health effects for the person. A measure of the likelihood that a peptide sequence might generate an autoimmune response can be calculated by determining the antigenic similarity of the peptide sequence in the library to sequences found in human proteins.

At the final stage of the method of designing a VCPL, the local similarity between the VCPL peptides and the whole set of human proteins can be analyzed. Candidate vaccines on the based on VCPLs can be screened for potentially autoimmunogenic determinants resulting from local similarity of VCPL peptides to human proteins. These potentially autoimmunogenic determinants can be removed from the library.

The epitopes to be modeled in a VCPL can be chosen so as not to be autoimmunogenic. The method of identifying potentially immunopathogenic regions in proteins of an infectious agent on the basis of recognition of regions displaying close local similarity to human proteins provides a key to rational design of safe polyepitope vaccines against pathogenic microorganisms, because it allows elimination of epitopes with an immunopathogenic potential. In particular, antigenic determinants displaying similarity to human proteins that occur often and/or support important physiological functions of the body should be eliminated.

To search for local similarity between a library sequence and human proteins, the SIM program can be used (Huang, X. and Miller, W., Adv. Appl. Math., 1991, 12, 337–357, which is incorporated by reference in its entirety). The sequences of human proteins are available in public databases, for instance SWISS-PROT (rl. 41, FEB-2003) and SP-TrEMBL (rl. 23, MAR-2003). The comparison can be carried out using the version of SIM available on the European Molecular Biology Laboratory (EMBL) server. The following parameters of the SIM program can be used: gap-open penalty 100, gap-extension penalty 20.

The local similarity search may produce a large number of hits, many of which have a small degree of similarity between a peptide sequence from the library and a region of a human protein. It can be useful to filter the results of the similarity search to identify most relevant peptide-protein similarities. The filtering can be performed as follows. The twenty commonly occurring amino acids are divided into intersecting groups with physicochemical similarities: group 1 - Lys, Arg; group 2 - Asp, Glu; group 3 - Asn, Gln; group 4 - Ser, Thr;

group 5 - Ala, Ile, Leu, Val; group 6 - Ala, Gly; group 7 - Leu, Phe; group 8 - Leu, Met, Val; and group 9 - Phe, Trp, Tyr.

For each individual hit of local similarity between a library peptide and a human protein, the following values are calculated:

index of similarity (Score determined by the SIM program); the percentage of matching amino acids (Match %); the number of matching amino acids (M); the number of mismatching amino acids (MM); the number of gaps (Gap); the number of conservative substitutions (C_{cons}) (that is, the number of substitutions within groups 1, 2, 3 & 4, 5-9); the number of substitutions of charged amino acids (His, Lys, Arg, Asp, Glu) by one of opposite charge (C_{+-}); the number of non-conservative substitutions or deletions of charged amino acids ($C_{\pm}$); the number of non-conservative substitutions or deletions of polar amino acids (Asn, Gln, Ser, Thr) (C_{polar}); and the number of non-conservative substitutions or deletions of aromatic amino acids (Phe, Trp, Tyr) (C_{arom}).

Then an empirical algorithm tests if the following conditions are met:

1)Index of similarity not less than 320 (Score $\geq$ 320).

2)Number of matching amino acids in a given region not less than 6 ($M \geq 6$).

3)Percent of matches not less than 50% (Match % $\geq$ 50%).

4)Number of matching amino acids and conservative substitutions in a given region not less than 8: $M + C_{cons} \geq 8$. However, if within a homologous region there is a subregion which comprises a string of 8 amino acids in a row, which are either conservative or matching then this region is taken as a whole regardless of the percent of match (Match %).

5)$(MM - C_{cons} + Gap)/(M + C_{cons}) \leq \{0.37|_{M=6\div\infty}\}$

6)$1.5*C_{+-} + C_{\pm} + 0.5*(C_{polar} + C_{arom}) \leq 0.1 + 0.5*(M + C_{cons} - 6)$

If all the above criteria are met, then the nature and/or function of the human protein can be considered. If the library sequence is similar to a sequence that occurs frequently in human proteins, or is similar to a sequence in a protein having an important physiological function, that sequence can be removed from the library.

In certain circumstances, it can be desirable for the library to have a small number of sequences, such as fewer than 1,000 sequences, fewer than 500 sequences, or fewer than 250 sequences. The library is representative of the full antigenic diversity found in the naturally

occurring sequences. Such a representative chimeric peptide library (RCPL) can be created by selecting sequences from a large library (e.g., a VCPL) such that the selected sequences are representative of the full antigenic diversity found in the naturally occurring sequences and non-natural sequences generated from the large library. The non-natural sequences can reflect potential sequences, for example, by including in a single sequence variants at different positions that have not been found together naturally.

Because an RCPL has a smaller number of peptides than a VCPL, it is inherently less likely to induce an autoimmune response. The sequences chosen for the RCPL can be biased towards certain residues at certain positions, analogous to the residue frequencies in a VCPL. An RCPL can also be biased toward particular discrete sequences. In other words, if an RCPL includes 100 distinct peptide sequences, those peptides can be combined in any arbitrary ratio. Despite the much smaller number of sequences, the antigenic diversity of an RCPL can be as great as that of VCPL from which the RCPL is created. In this way, a mixture of a relatively small number of chemical species (i.e. peptide sequences) can produce an equally broad immune response as a mixture of a very large number of chemical species.

RCPLs are sets of homologous peptides evenly covering the diverse variants of hypervariable regions of infectious agents. RCPLs are characterized by smaller numbers of peptides in the libraries (compared to VCPLs), which evenly cover the set of antigenic variants of hypervariable regions of variable infectious agents, including potential variants. RCPLs are designed to mimic the antigenic diversity of main variable protective determinants of an infectious agent. RCPLs allow production of a wide range of antibodies and cytotoxic T-lymphocytes (CTLs), whose combined specificity covers the set of antigenic variants of the variable infectious agent. Thus, RCPLs can induce a humoral and cellular response against the diversity of current and potential variants of variable infectious agents.

To select the specified number, N, of peptides in the RCPL, a large number (e.g. >1,000, >10,000, >100,000, >1,000,000, or >10,000,000) of peptides sequences are chosen at random from the whole diversity of variants of the mimicked hypervariable antigenically active region (e.g., a VCPL), accounting for the rates of occurrence of amino acids at particular positions of the mimicked region. The set of randomly chosen peptide sequences is arranged in the order of the relative rates of their occurrence in the whole diversity of the variants of the mimicked hypervariable antigenic region. The first 37–100% of the

sequences from the set of randomly chosen peptides are used. The percentage used is determined by empirical considerations.

The peptide sequence having amino acids with the greatest rates of occurrence at each position is taken as the first sequence of the RCPL. Alternatively, any peptide of the set of variants of the mimicked hypervariable antigenic region may be used as the first, or reference, sequence.

Each subsequent peptide sequence added to the RCPL is chosen from the set of randomly chosen peptides such that the distance between the new sequence and all sequences previously added to the library is maximal or approaches the maximum. The distance between two peptides, R_{ij}, can be determined with the use of the antigenic similarity matrix (see, e.g., Table 1). R_{ij} is calculated according to:

$$R_{ij} = \sum_{k=1}^{l} F(i_k, j_k)$$

where i_k is the k^{th} residue of peptide i, j_k is the k^{th} residue of peptide j, and l is the number of amino acid residues in each peptide of the library. The function F can be described by an antigenic similarity matrix:

$$F(i_k, j_k) = \mathrm{ASM}(i_k, j_k)$$

where the value of $\mathrm{ASM}(i_k, j_k)$ is the matrix value for the pair of residues represented by i_k and j_k.

For each peptide n from the randomly chosen peptides, the sum of distances $\boldsymbol{S_n}$ between the first N peptides selected (i.e., those already in the RCPL) and peptide n is calculated according to the formula:

$$S_n = \sum_{i=1}^{N} R_{in}$$

where R_{in} is calculated as desibed above. The peptide n that has the greatest value of $\boldsymbol{S_n}$ is added to the library. The selection of sequences from the large set of randomly chosen sequences continues until the RCPL has the desired number of sequences.

The function F can alternatively be used simply to count the number of substitutions between two peptides. In this case, the value of $F(i_k, j_k)$ is 1 when i_k equals j_k (i.e. both peptides have the same residue at position k), and $F(i_k, j_k)$ is 0 when i_k and j_k are different.

When selected in this way, the sequences of the RCPL should evenly cover the whole set of variants of the mimicked hypervariable antigenically active region.

Each peptide included in the RCPL can be screened for autoimmunogenic properties as described above. If a peptide sequence is determined to have the potential to induce an autoimmune response, it can be excluded from the RCPL. Another candidate, virtually identical in the required properties, is instead chosen as the next peptide in the RCPL. The choice algorithm is designed so that the each new peptide added to the RCPL is chosen from a number of sequences of equal value (i.e. of equal distance from the previous peptide sequence in the library).

A sufficient number of peptides chosen as described here can evenly cover the whole diversity of the antigenic variants of hypervariable epitopes of variable infectious agents. The greater is the number of peptides making up an RCPL, the "denser" the immune response that it can induce. Because the algorithm of choosing N peptides is incremental, the first N/2 peptides also evenly cover the whole diversity of the variants, although less densely.

Libraries of representative chimeric peptides can be constructed as candidate vaccines. Taken together, the peptide sequences in the library cover the whole diversity of variants of hypervariable antigenically active regions of a variable infectious agent. The sequences are designed to mimic the genetic diversity of main variable protective determinants of infectious agents. Also, the mimicking peptides can be chimeric, because, although they mimic the heterogeneous set of variants, the sequences in the RCPL actually occur in few, if any, of the known and potential variants of epitope sequences.

In general, peptide libraries can be synthesized manually or by using an automated synthesizer as described, e.g., in Lebl M. and Krchnak V. (1997) Solid-Phase Peptide Synthesis, *Methods in Enzymology*, 289, Academic Press, Minneapolis, Minnesota, which is incorporated by reference in its entirety. The synthesis can be a combinatorial synthesis, in which mixtures of amino acid are coupled to the growing peptide, and a mixture of peptide sequences is formed. In a combinatorial synthesis, amino acids are added to a coupling reaction in proportions dictated by the desired ratio of residues at each position in the peptide, adjusted for the relative coupling rate for each amino acid. Alternatively, the library can be produced by parallel synthesis, in which each sequence is synthesized separately, and

the library can be a group of individually pure peptide sequences. Such a library can be converted to a mixture by mixing the pure peptides.

Automated combinatorial synthesis can be preferable when a large library is to be produced. Split synthesis can be preferable when a small library of individual sequences (e.g. an RCPL) is to be produced.

The peptide libraries can be part of a multiple antigenic peptide (MAP). A multiple antigenic peptide includes a branched core from which several peptides extend. The branched core can be based on branching repeats of lysine. The MAP can include for example 4 or 8 peptide chains. The MAP can include a hydrophobic group, such as an alkyl chain derived from palmitic acid. The hydrophobic group can facilitate assembly of the MAP with liposomes or virus-like particles.

The peptides can include the 20 common naturally occurring L-amino acids. The peptides can optionally include corresponding D-amino acids, or mixtures of D- and L-amino acids. Non-natural amino acids can likewise be included in the peptides.

The peptide library can be produced by recombinant methods. A DNA construct encoding the peptides of the library can be introduced into a host organism capable of protein expression from the construct. Suitable host organisms can include bacteria such as *E. coli*, *Lactobacillus*, and *Bifida*, and plants, such as tobacco.

The peptide libraries can be used in diagnostic applications, for example to determine the presence of antibodies, or to determine CTL function. For example, the libraries may be used an antigens for an enzyme-linked immunosorbent assay (ELISA) to determine the presence of antibodies in a subject sample. In a solid-phase assay such as an ELISA, the peptides of the library are immobilized on a substrate (for example, in a polystyrene well). The substrate is then exposed to a test sample (e.g., serum from a subject) under conditions that allow specific peptide-antibody complex formation. Any unbound antibodies are then washed away. Antibodies associated with the substrate by virtue of a specific peptide-antibody interaction are then detected. Detection can include exposing the antibodies to a secondary antibody specific for immunoglobulins from the subject. The secondary antibody can be radiolabelled, or coupled to an enzyme such as horseradish peroxidase or alkaline phosphatase, that catalyzes the formation of a colored product.

The diagnostic peptide library can be a broad spectrum diagnostic reagent, or a subtype specific reagent. If the library was designed to be antigenically similar to all known subtypes of a pathogen, then it can be used to detect that a subject has been infected by the pathogen, regardless of what subtype the subject is infected with. If instead the library was designed to be antigenically similar to only one subtype, then it can be used to detect if a subject was infected by the same subtype. A group of subtype specific libraries can be used together to determine which of several subtypes a subject was infected by.

Vaccines based on antigenic peptides are known. See, for example, U.S. Patent Nos. 4,601,903, 4,599,230, and 4,599,231, each of which is incorporated by reference in its entirety. Vaccines may be prepared as injectables, as liquid solutions or emulsions. The peptides may be mixed with pharmaceutically acceptable excipients that are compatible with the peptides. Excipients may include water, saline, dextrose, glycerol, ethanol, and combinations thereof. The vaccine may further contain auxiliary substances such as wetting or emulsifying agents, pH buffering agents, or adjuvants to enhance the effectiveness of the vaccines. Methods of achieving adjuvant effect for the vaccine include the use of agents such as aluminum hydroxide or phosphate (alum), commonly used as 0.05 to 0.1 percent solution in phosphate buffered saline. Vaccines may be administered parenterally, by injection subcutaneously or intramuscularly. Alternatively, other modes of administration including suppositories and oral formulations may be desirable. For suppositories, binders and carriers may include, for example, polyalkylene glycols or triglycerides. Oral formulations may include normally employed incipients such as, for example, pharmaceutical grades of saccharine, cellulose and magnesium carbonate. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as is therapeutically effective, protective and immunogenic. The quantity to be administered depends on the subject to be treated, including, for example, the capacity of the individual's immune system to synthesize antibodies, and to produce a cell-mediated immune response. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner. However, suitable dosage ranges are readily determinable by one skilled in the art and may be of the order of micrograms of the peptides.

Suitable regimes for initial administration and booster doses are also variable, but may include an initial administration followed by subsequent administrations. The dosage of the vaccine may also depend on the route of administration and will vary according to the size of the host.

Nucleic acid molecules encoding the peptides of the library may also be used directly for immunization by administration of the nucleic acid molecules directly, for example by injection, or by first constructing a live vector, such as Salmonella, BCG, adenovirus, poxvirus, vaccinia or poliovirus, and administering the vector.

The vaccine can be a prophylactic vaccine that induces a protective immune response in a subject, before the subject is exposed to the pathogen. In general, a prophylactic vaccine includes a broad-spectrum peptide library. In some cases, a prophylactic vaccine can include a subtype-specific peptide library, for instance, when the subject to be vaccinated is likely to be exposed to only one subtype of the pathogen. The vaccine can be a therapeutic vaccine, designed to induce or enhance an immune response in a subject that is infected by the pathogen. If the subject is known to be infected with a particular subtype of pathogen, the therapeutic vaccine can be one designed to induce an immune response against that subtype.

Examples

Peptide libraries were designed based on each of the variable regions V1-V5 of HIV gp120. The sequences used to generate the libraries were from HIV subtype A, B, C, D, F, G, or sequences from all these subtypes were used together. A peptide library generated using sequences from only one subtype can be used as a subtype-specific vaccine, whereas a library generated from multiple subtypes can be used as a broad spectrum vaccine. Each library was initially designed to include fewer than 200,000,000 sequences. From the initial library, smaller libraries were designed to include fewer than 500,000 sequences and fewer than 50,000 sequences. Smaller libraries are designed by removing infrequently occurring residues from the larger library. In Tables 3-37 below, the identities and relative frequencies of residues are given. For example, in Table 3, residue X^{11} can be S, N or G in the largest library; S or N in the next smaller library, and S in the smallest library. The numbers represent the relative frequencies of each residue at that position. In the largest library,

position X^{11} is occupied by S, N or G in an 80:13:7 ratio; in the next smaller library, X^{11} is occupied by S or N in an 80:13 ratio; and in the smallest library, every peptide has S at X^{11}.

Chimeric Peptide Libraries – V1 region of HIV gp120

Peptide libraries based on the V1 region of HIV gp120 from subtypes A, B, C, D, F, and G were designed. The libraries had the structure:

C-T-X^1-X^2-X^3-X^4-X^5-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-X^{16}-X^{17}-E-X^{18}-K-N

where X^1-X^{18} are defined according to Table 3.

Table 3

HIV gp120 V1 region, subtypes ABCDFG																		
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}
50,000	D 68	L 81	K 43	N 87	T 33	T 81	N 67	T 69	N 70	S 61	S 80	S 77	G 49	M 59	E 88	K 58	G 79	I 73
	N 32		T 22		N 26		T 33		T 30	N 20		N 15	S 33	T 15		E 16	E 21	M 27
			N 18		A 24								N 18	I 14				
					D 17													
500,000				T 13				N 13			N 13				K 12			
200,000,000		W 12	G 8			N 10		V 11		A 12	G 7	E 8		E 12		G 10		
		Y 7	E 8			V 10		A 7		I 7						N 9		
																T 8		

Peptide libraries based on the V1 region of HIV gp120 from subtype A were designed. The libraries had the structure:

C-T-X^1-X^2-X^3-X^4-X^5-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-X^{16}-E-X^{17}-K-N

where X^1-X^{17} are defined according to Table 4.

Table 4

HIV gp120 V1 region, subtype A																	
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}
50,000	N 87	V 67	N 70	V 76	T 65	S 53	T 62	S 42	N 45	T 34	T 57	E 32	G 37	D 29	K 62	E 69	I 82
	D 13	A 33	T 30	S 24	N 24	N 32	N 20	N 36	A 40	I 29		T 26	E 34	M 21	E 19	G 31	
												N 22		N 20			
500,000									S 15	N 19			S 16		Q 19		M 18
										D 18							
200,000,000					V 11	V 15	V 10	V 14			I 13	V 13	N 13	G 14			
							A 8	A 8			R 13	Y 8		T 9			
											D 9			I 7			
											N 7						

Peptide libraries based on the V1 region of HIV gp120 from subtype B were designed. The libraries had the structure:

C-T-X^1-X^2-X^3-N-X^4-X^5-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-E-X^{16}-X^{17}-E-X^{18}-K-N

where X^1-X^{18} are defined according to Table 5.

Table 5

HIV gp120 V1 region, subtype B																		
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}
50,000	D 85	L 79	K 55	A 30	T 91	N 74	T 90	N 66	S 74	S 83	S 88	G 49	E 29	I 34	M 67	K 69	G 92	I 75
			N 14	T 30		T 26		T 34	N 19			S 36	M 20	T 22	I 14	G 12		M 25
			T 12	N 21								N 15	G 17	M 18				
				D 19									T 17	K 18				
500,000	N 15	W 14									N 12				T 12			
200,000,000		Y 7	E 11		N 9		N 10		A 7	N 8			I 9	E 8	E 7	E 11	E 8	
			G 9							G 8			K 8			T 8		

Peptide libraries based on the V1 region of HIV gp120 from subtype C were designed. The libraries had the structure:

C-X^1-X^2-X^3-X^4-X^5-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-X^{16}-X^{17}-E-X^{18}-K-N

where X^1-X^{18} are defined according to Table 6.

Table 6

HIV gp120 V1 region, subtype C																		
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}
50,000	T 73	N 86	A 48	T 76	R 28	N 63	G 48	N 68	V 43	T 84	Y 52	N 67	N 59	T 80	M 58	E 30	G 57	I 61
			V 41	N 24	S 24	T 30	N 22	T 32	T 33		V 21		D 21		I 24	K 27	E 34	M 39
					N 20		S 22		A 24							N 26		
500,000	R 19				V 18						S 18			N 20	T 18			
200,000,000	V 8	D 14	T 11		A 10	D 7	V 9			N 16	N 9	T 18	S 11			T 17	K 9	
												D 15	G 9					

Peptide libraries based on the V1 region of HIV gp120 from subtype D were designed. The libraries had the structure:

C-X^1-X^2-X^3-X^4-X^5-X^6-X^7-T-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-X^{16}-E-X^{17}-K-N

where X^1-X^{17} are defined according to Table 7.

Table 7

HIV gp120 V1 region, subtype D																	
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}
50,000	T 87	D 85	W 38	K 45	N 47	N 73	T 35	T 39	N 44	N 55	N 47	T 45	T 75	M 32	E 64	G 71	M 91
			A 20	W 18	G 28	E 28	A 29	N 24	G 25	S 29	T 24	V 37	D 25	I 21	G 36		
					S 25				K 17		K 18			E 19			
500,000							N 18	D 19				A 18				I 17	
							Y 18	G 18									
200,000,000	I 13	N 15	V 17	T 16					T 14	K 16	A 11			N 10		T 12	V 9
			E 15	N 11										K 9			
			Y 11	L 11										T 9			

Peptide libraries based on the V1 region of HIV gp120 from subtype F were designed. The libraries had the structure:

C-X^1-X^2-X^3-X^4-X^5-X^6-T-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-T-L-K-X^{14}-X^{15}-X^{16}-X^{17}-X^{18}-X^{19}-X^{20}-N

where X^1-X^{20} are defined according to Table 8.

Table 8

HIV gp120 V1 region, subtype F																				
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}	X^{20}
50,000	T 60	N 47	A 37	T 53	N 61	A 84	N 73	D 40	T 60	I 54	T 37	N 53	T 56	E 90	E 89	P 70	G 80	A 65	I 86	Q 67
	R 28	D 32	T 25	I 47	T 27		S 27	N 27	N 27	T 46	N 32	G 35	D 31					E 35		K 33
			N 25								A 31									
500,000						T 16		G 20								Q 18	E 20			
200,000,000	A 12	A 11	I 13		V 12			S 13	A 13			D 12	P 13	I 10	S 11	S 12			M 14	
		S 11																		

Peptide libraries based on the V1 region of HIV gp120 from subtype G were designed. The libraries had the structure:

C-T-X^1-X^2-X^3-X^4-X^5-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-X^{16}-X^{17}-N

where X^1-X^{17} are defined according to Table 9.

Table 9

HIV gp120 V1 region, subtype G																	
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}
50,000	N 77	V 88	T 65	N 67	N 71	S 34	T 70	K 35	N 72	V 62	T 56	E 51	K 44	E 53	E 52	I 73	K 78
	D 23		N 35		T 29	C 26	N 16	T 25		S 22	E 17	S 19	E 24	N 31	R 29	M 27	
						N 24		N 16						A 17	N 19		
500,000						G 16				C 16	C 15	G 16	G 17				
												N 14	N 15				
200,000,000		A 12		Y 12			A 14	E 14	E 14		V 11						T 12
				K 10				G 10	T 14								E 10
				T 10													

Chimeric Peptide Libraries – V2 region of HIV gp120

Peptide libraries based on the V2 region of HIV gp120 from subtypes A, B, C, D, F, and G were designed. The libraries had the structure:

S-X^1-N-X^2-T-T-X^3-X^4-R-X^5-K-X^6-X^7-X^8-X^9-X^{10}-X^{11}-L-F-Y-X^{12}-L-D-V-V-X^{13}-I-X^{14}-X^{15}-X^{16}-X^{17}-X^{18}-X^{19}-Y-X^{20}-L-X^{21}-X^{22}-C

where X^1-X^{22} are defined according to Table 10.

Table 10

HIV gp120 V2 region, subtypes A, B, C, D, F, G																					
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}	X^{20}	X^{21}	X^{22}
50,000	F 91	I 84	E 37	I 91	D 95	V 45	Q 69	K 89	E 62	Y 89	A 79	K 93	P 73	D 63	N 78	D 54	S 51	T 91	S 76	R 95	I 95	S 51
		M 16	S 32			K 40			V 33		S 21		Q 27	N 22	D 22	N 32	N 49		N 10			N 49
			N 18			M 15										S 15						
500,000				M 9			K 31	Q 11		H 11								N 9				
200,000,000	Y 9		V 7		N 5				Q 5			S 7		G 8					E 8	T 5	T 5	
			G 6											K 7					R 6			

Peptide libraries based on the V2 region of HIV gp120 from subtype A were designed. The libraries had the structure:

S-X^1-X^2-X^3-T-T-X^4-L-R-D-K-K-X^5-X^6-V-X^7-X^8-L-F-Y-R-L-D-V-V-X^9-I-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-X^{16}-X^{17}-X^{18}-Y-R-L-I-X^{19}-C

where X^1-X^{19} are defined according to Table 11.

Table 11

HIV gp120 V2 region, subtype A																			
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}
50,000	F 63	N 91	M 58	E 92	Q 62	K 90	Y 79	S 83	Q 59	N 61	E 68	N 62	N 71	S 86	S 82	N 69	S 59	E 60	N 96
	Y 37		I 42		K 38		H 15	A 17	P 35	D 20		S 15	T 20		N 18	S 18	N 32	Q 18	
										K 13		G 10				D 13		S 16	
500,000						Q 10					Q 11		D 9						
											A 11								
											K 9								
200,000,000		S 9		V 8			S 6		K 6	T 6		K 8		N 9			E 4	K 6	S 4
												D 4		A 5			K 4		

Peptide libraries based on the V2 region of HIV gp120 from subtype B were designed. The libraries had the structure:

S-F-X^1-I-T-T-X^2-X^3-X^4-X^5-K-X^6-X^7-X^8-X^9-X^{10}-A-X^{11}-F-X^{12}-X^{13}-L-D-V-V-X^{14}-I-X^{15}-X^{16}-X^{17}-X^{18}-T-X^{19}-Y-X^{20}-L-X^{21}-X^{22}-C

where X^1-X^{22} are defined according to Table 12.

Table 12

HIV gp120 V2 region, subtype B																					
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}	X^{20}	X^{21}	X^{22}
50,000	N 94	S 53	I 85	R 94	D 85	V 68	Q 80	K 95	E 93	Y 95	L 95	Y 95	K 91	P 88	D 79	N 76	D 70	N 61	S 87	R 92	I 86	S 81
		N 28	M 15		N 8	M 24	K 20						S 9	Q 13	G 11	D 18	N 22	T 39		T 8	T 7	N 19
		G 11				R 8																
		D 9																				
500,000									Q 7								S 7		R 7		R 7	
200,000,000	K 6			S 6	G 7			E 5		H 5	T 5	N 5			K 5	K 7			N 6			
															N 5							

Peptide libraries based on the V2 region of HIV gp120 from subtype C were designed. The libraries had the structure:

S-F-N-X^1-T-T-E-L-R-D-K-X^2-X^3-X^4-X^5-X^6-A-L-F-Y-R-X^7-D-I-V-X^8-L-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-Y-X^{15}-L-I-X^{16}-C

where X^1-X^{16} are defined according to Table 13.

Table 13

HIV gp 120 V2 region, subtype C																
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}
50,000	I 45	K 88	K 49	K 61	V 53	Y 60	L 94	P 87	N 63	N 71	N 56	S 89	S 71	E 76	R 90	N 96
	A 21		Q 42	Q 26	E 35	H 23		Q 8	D 28	S 18	S 32		G 14	T 10	I 10	
	T 17			T 13	A 11				K 9				R 9			
	M 16															
500,000		T 7								D 7		G 7		K 7		
200,000,000		Q 5	E 4			N 6	P 6	S 5		K 4	D 7	N 4	N 5	N 6		S 4
			H 4			S 5					A 5					
						R 5										

Peptide libraries based on the V2 region of HIV gp120 from subtype D were designed. The libraries had the structure:

S-F-N-I-T-T-X^1-V-X^2-D-K-X^3-X^4-X^5-X^6-X^7-A-L-X^8-Y-R-X^9-D-V-V-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-X^{16}-X^{17}-X^{18}-Y-R-L-I-X^{19}-C

where X^1-X^{19} are defined according to Table 14.

Table 14

HIV gp120 V2 region subtype D																			
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}
50,000	V 58	R 93	K 91	K 86	Q 86	V 57	Y 50	F 97	L 96	Q 77	I 85	D 83	N 79	Q 62	N 80	S 83	S 87	S 83	N 96
	E 38		Q 9		K 10	E 34	H 35			P 23	M 15	N 11	T 12	S 22	T 20	Q 13	A 8	N 14	
							Q 10							K 13					
500,000		I 7		E 7								G 6	D 8						
				Q 7															
200,000,000	A 4				E 4	K 5	S 5	Y 3	P 4					D 3		G 2	N 5	D 3	H 4
						A 4										K 2			

Peptide libraries based on the V2 region of HIV gp120 from subtype F were designed. The libraries had the structure:

S-F-N-X^1-T-X^2-E-V-R-D-K-X^3-X^4-K-X^5-X^6-A-L-F-Y-R-L-D-I-X^7-X^8-I-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-Y-R-L-X^{16}-X^{17}-X^{18}

where X^1-X^{18} are defined according to Table 15.

Table 15

HIV gp120 region V2, subtype F																		
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}
50,000	I 50	T 90	Q 53	K 47	V 67	H 54	V 85	P 75	N 56	N 72	S 35	S 50	N 73	S 42	T 47	I 95	N 57	C 95
	M 30		K 47	L 29	E 19	Q 21		Q 25	S 24		N 30	D 28	S 27	R 37	E 38		S 24	
	T 20			Q 24							D 20	N 22						
500,000						Y 15	E 15				G 15				I 15		H 14	
200,000,000		M 10			Q 14	S 11			K 10	K 14				G 10		T 5	G 5	I 5
									G 10	A 7				N 5				
										M 7				D 5				

Peptide libraries based on the V2 region of HIV gp120 from subtype G were designed. The libraries had the structure:

S-F-N-X^1-T-X^2-X^3-X^4-X^5-D-K-X^6-K-X^7-E-Y-A-L-F-Y-R-X^8-D-V-V-X^9-I-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-Y-X^{16}-L-X^{17}-X^{18}-C

where X^1-X^{18} are defined according to Table 16.

Table 16

HIV gp120 V2 region, subtype G																		
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}
50,000	I 94	T 95	E 97	I 97	R 93	K 80	T 51	L 95	P 91	N 60	D 85	N 49	N 54	N 63	N 66	R 87	I 95	N 97
					Q 7	Q 10	K 29			D 28	G 15	S 24	S 27	S 31	S 27	V 10		
						T 7	A 8			T 8		G 13	D 10	G 6	D 7			
							E 7					D 9	R 6					
500,000							Q 5	T 5	Q 5								T 5	
200,000,000	M 3	A 5	A 3	M 3		E 3			S 4	R 4		K 4	I 3			T 3		H 3
	T 3																	

Chimeric Peptide Libraries – V3 region of HIV gp120

Peptide libraries based on the V3 region of HIV gp120 from subtypes A, B, C, D, F, and G were designed. The libraries had the structure:

X^1-C-X^2-R-P-X^3-N-N-T-R-X^4-X^5-X^6-X^7-X^8-G-X^9-G-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-T-G-
X^{15}-I-X^{16}-G-X^{17}-I-R-X^{18}-A-X^{19}-C-X^{20}

where X^1-X^{20} are defined according to Table 17.

Table 17

HIV gp120 V3 region, subtypes A, B, C, D, F, G																				
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}	X^{20}
50,000	N 78	T 89	N 77	K 90	S 78	I 92	H 45	I 93	P 95	R 53	A 72	F 90	Y 92	A 60	D 68	I 96	D 86	Q 85	H 85	N 95
	V 9	I 11	S 10		G 15		R 29			Q 47	T 23	W 10		T 40	Q 14		N 14	K 15	Y 15	
	T 8						P 12								A 11					
							S 8													
500,000			G 8		R 7		N 6	M 7												
200,000,000				Q 5		T 4							F 4							
	E 4		Y 4	T 4		M 4			L 5		V 4		H 3		R 7	T 4				T 5

Peptide libraries based on the V3 region of HIV gp120 from subtype A were designed. The libraries had the structure:

X^1-C-X^2-R-P-X^3-N-N-X^4-R-X^5-X^6-V-X^7-I-G-P-G-X^8-X^9-F-X^{10}-X^{11}-X^{12}-X^{13}-
X^{14}-I-X^{15}-G-X^{16}-I-R-X^{17}-A-X^{18}-C-X^{19}

where X^1-X^{19} are defined according to Table 18. A dash "-" indicates no residue at that position. For example, if X^{12} is represeented by a dash, then no residue occupies X12, and the residue of X^{11} is bound directly to the residue of X^{13}.

Table 18

HIV gp120 V3 region, subtype A																		
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}
50,000	N 76	T 72	N 68	T 97	K 78	S 85	R 66	Q 93	A 56	Y 95	A 95	T 91	G 89	D 79	I 92	D 84	Q 76	H 76	N 87
	T 19	I 28	G 23		T 19	G 11	H 34	R 7	T 41				- 6	A 12	T 8	N 16	K 22	Y 24	T 7
			S 9																
500,000											T 5	R 5	D 5						E 5
200,000,000	D 5			K 3	Q 3	R 2			V 3	F 3		A 2		N 4			E 2		
						N 2				H 2		- 2		R 3					
														V 2					

Peptide libraries based on the V3 region of HIV gp120 from subtype B were designed. The libraries had the structure:

X^1-C-X^2-R-P-X^3-N-N-T-R-K-X^4-I-X^5-X^6-G-X^7-G-X^8-X^9-X^{10}-X^{11}-X^{12}-T-X^{13}-
X^{14}-I-X^{15}-G-X^{16}-I-R-X^{17}-A-X^{18}-C-X^{19}

where X^1-X^{19} are defined according to Table 19.

Table 19

HIV gp120 V3 region, subtype B																			
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}
50,000	N 94	T 91	N 86	S 75	H 58	I 89	P 94	R 85	A 90	F 80	Y 91	T 59	G 94	E 62	I 97	D 88	Q 89	H 91	N 95
		I 9	S 10	G 18	P 18	M 11		Q 9		W 15		A 41		Q 21		N 13	K 11	Y 9	T 5
				R 7	N 11									R 9					
					T 11									G 9					
500,000									V 5		F 4								
									T 5		H 4								
200,000,000	T 4		G 4		Y 3		L 3	G 3		V 4			E 3		T 3				
	H 2						W 3	S 3					R 3						

Peptide libraries based on the V3 region of HIV gp120 from subtype C were designed. The library had the structure:

X^1-C-X^2-R-P-X^3-N-N-T-R-X^4-X^5-X^6-X^7-I-G-P-G-Q-X^8-F-X^9-X^{10}-T-X^{11}-X^{12}-I-X^{13}-G-X^{14}-I-R-X^{15}-A-X^{16}-C-X^{17}

where X^1-X^{17} are defined according to Table 20.

Table 20

HIV gp120 V3 region, subtype C																	
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}
50,000	V 57	T 91	N 77	K 88	S 98	I 76	R 97	T 82	Y 96	A 98	G 85	D 77	I 98	D 87	Q 90	H 82	N 92
	N 26	I 6	G 12	E 9		M 24	G 3	A 16	F 4		N 11	G 20		N 13	K 6	Y 17	T 4
	E 9		S 10														
	T 4																
500,000		A 3		Q 3									T 2		E 2		I 2
200,000,000	M 2		H 1		G 2			V 2		T 2	D 2	N 1			L 1	N 1	D 1
	R 2										K 1	S 1			H 1		H 1
											T 1	I 1					

Peptide libraries based on the V3 region of HIV gp120 from subtype D were designed. The libraries had the structure:

X^1-C-X^2-R-P-X^3-X^4-X^5-X^6-R-X^7-X^8-X^9-X^{10}-I-G-X^{11}-G-X^{12}-X^{13}-X^{14}-X^{15}-X^{16}-T-X^{17}-X^{18}-G-X^{19}-I-X^{20}-X^{21}-A-X^{22}-C-X^{23}

where X^1-X^{23} are defined according to Table 21.

Table 21

HIV gp120 V3 region, subtype D																						
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}	X^{20}	X^{21}	X^{22}	X^{23}
50,000	N 78	T 83	Y 74	N 87	N 89	T 72	Q 71	S 37	T 74	H 61	P 64	Q 70	A 92	L 80	Y 80	T 92	I 82	I 83	D 84	R 92	Q 84	H 67	N 82
	T 22	I 17	N 26				R 29	R 35	I 26	P 19	L 16	R 30		Y 20	F 20		K 18		N 16		K 16	Y 33	
								G 28															
500,000				K 13		I 16					S 11												
						K 12																	
200,000,000					K 11					S 10	Q 8		T 8			A 8		T 10		G 8			T 11
										R 9								K 8					K 8

Peptide libraries based on the V3 region of HIV gp120 from subtype F were designed. The libraries had the structure:

X^1-C-T-R-P-X^2-N-N-X^3-R-K-X^4-I-X^5-L-G-P-G-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-I-X^{14}-G-X^{15}-I-R-X^{16}-A-X^{17}-C-X^{18}

where X^1-X^{18} are defined according to Table 22.

Table 22

HIV gp120 V3 region, subtype F																		
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}
50,000	N 93	N 96	T 97	S 91	H 73	Q 69	A 87	F 97	Y 92	A 73	T 95	G 92	D 87	I 96	D 91	K 92	H 86	N 83
	T 7			G 6	R 10	R 28	V 10			T 27	A 5	D 5	A 8	T 4	N 9	Q 8	Y 14	I 11
					Q 6								N 5					
500,000		S 4			P 4				H 4									
					Y 4				F 4									
200,000,000			I 3	R 3	S 3	H 3	T 3	I 3				S 3						D 3
																		K 3

Peptide libraries based on the V3 region of HIV gp120 from subtype G were designed. The libraries had the structure:

X^1-C-X^2-R-P-X^3-N-N-T-R-K-S-X^4-X^5-X^6-G-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-T-X^{14}-I-X^{15}-G-X^{16}-I-R-X^{17}-A-X^{18}-C-X^{19}

where X^1-X^{19} are defined according to Table 23.

Table 23

HIV gp120 V3 region, subtype G																			
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}
50,000	N 44	T 90	N 88	I 98	T 32	F 66	P 93	G 98	Q 86	A 76	F 93	Y 98	A 95	G 91	I 98	D 76	Q 94	H 83	N 96
	I 19	I 10	S 12		R 25	I 34			R 7	T 17	I 5		T 5	D 5		N 24		Y 15	
	T 15				H 22					V 7									
	M 12				P 15														
	R 9				N 5														
500,000							T 5		H 5					S 4					K 2
200,000,000				K 2			L 2	R 2	G 2		S 2	H 2			T 2		L 2	F 2	T 2
																	K 2		
																	P 2		

Chimeric Peptide Libraries – V4 region of HIV gp120

Peptide libraries based on the V4 region of HIV gp120 from subtypes A, B, C, D, F and G were designed. The libraries had the structure:

C-N-T-T-X^1-L-F-N-S-T-W-X^2-X^3-X^4-X^5-W-X^6-T-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-
X^{14}-X^{15}-X^{16}-X^{17}-I-X^{18}-L-X^{19}-C

where X^1-X^{19} are defined according to Table 24.

Table 24

HIV gp120 V4 region, subtypes ABCDFG																			
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}
50,000	Q 67	N 52	N 78	S 61	T 90	N 37	E 58	G 70	S 79	N 86	N 59	T 90	E 67	G 78	N 70	D 63	T 65	T 85	P 84
	K 33	F 18		N 22		S 33		E 18	L 21		S 24			S 22	S 30	N 13	N 25	I 15	Q 16
		T 16		G 17							D 18								
500,000		V 14					T 15	S 12		D 14									
							N 14												
							K 13												
200,000,000			S 11		A 10	D 11						N 10	K 12			G 12	I 9		
			G 10			V 9							T 11			S 12			
						G 8							N 10						

Peptide libraries based on the V4 region of HIV gp120 from subtype A were designed. The libraries had the structure:

C-X^1-T-S-X^2-L-F-N-X^3-T-W-X^4-X^5-X^6-X^7-X^8-X^9-X^{10}-X^{11}-T-X^{12}-X^{13}-X^{14}-X^{15}-
X^{16}-I-X^{17}-L-X^{18}-C

where X^1-X^{18} are defined according to Table 25.

Table 25

HIV gp120 V4 region, subtype A																		
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}
50,000	N 87	G 56	S 90	T 43	N 62	I 26	Q 48	E 58	S 85	N 91	S 85	E 60	S 80	N 85	D 72	T 66	T 91	P 66
		D 22		N 32	S 20	T 25	T 41	S 17	L 15			K 17	L 20		G 28	N 28		Q 34
				E 25		M 23		G 14				G 16						
						G 14												
500,000	D 13		G 10		A 10	N 12	E 10	V 10										
200,000,000		S 8			D 8					D 9	D 9	N 7		S 9		K 6	I 9	
		K 7									N 8			D 6				
		N 7									I 7							

Peptide libraries based on the V4 region of HIV gp120 from subtype B were designed. The libraries had the structure:

C-N-T-T-X^1-L-F-N-S-T-W-X^2-X^3-X^4-X^5-X^6-X^7-T-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-G-X^{15}-X^{16}-X^{17}-I-X^{18}-L-X^{19}-C

where X^1-X^{19} are defined according to Table 26.

Table 26

HIV gp120 region V4, subtype B																			
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}	X^{18}	X^{19}
50,000	Q 64	N 55	N 89	S 62	T 90	W 90	N 43	E 61	G 74	S 79	N 85	N 59	T 89	E 69	N 72	D 65	T 63	T 83	P 86
	K 27	T 16		N 20			S 34	T 15	E 17	L 21		D 16		K 12	S 28	N 13	N 27	I 17	
		F 16		G 17				K 15				S 16							
500,000		V 13					D 13				D 15					S 11			Q 14
200,000,000	P 9		G 11		A 10	G 10	V 10	N 9	R 10			G 9	N 11	T 10		G 11	I 9		
														N 9					

Peptide libraries based on the V4 region of HIV gp120 from subtype C were designed. The libraries had the structure:

C-N-T-S-X^1-L-F-N-X^2-T-Y-X^3-X^4-X^5-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-X^{15}-I-X^{16}-L-X^{17}-C

where X^1-X^{17} are defined according to Table 27.

Table 27

HIV gp 120 V4 region, subtype C																	
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}	X^{17}
50,000	K 36	S 65	N 47	N 58	G 63	T 91	Y 50	N 78	S 58	T 79	E 45	S 59	N 78	S 70	T 71	T 92	P 77
	G 24	G 35	M 27	S 25	S 18		S 20		G 19	N 21	S 28	D 15	S 22		N 16		Q 23
	N 20		S 18	P 17							N 15						
500,000							F 16							A 16	I 13		
							N 14							E 15			
200,000,000	S 10		R 8		D 10	L 9		I 11	N 13		G 12	N 11				I 8	
	D 9				N 8			T 11	D 10			V 9					
												G 7					

Peptide libraries based on the V4 region of HIV gp120 from subtype D were designed. The libraries had the structure:

C-X^1-T-S-X^2-L-F-N-X^3-X^4-X^5-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-I-X^{15}-C

where X^1-X^{15} are defined according to Table 28

Table 28

HIV gp 120 V4 region, subtype D															
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}
50,000	N 94	K 62	S 83	T 92	W 85	D 40	N 47	N 32	T 78	G 43	D 60	N 74	I 70	T 54	P 67
		G 26				N 26	T 42	G 26		W 25	N 17	T 26	M 30	N 24	S 21
		N 13				S 23		S 23		N 17	S 12			I 13	Q 11
								E 12		T 10					
500,000					L 11	K 9			N 8		G 7			K 9	
									L 8						
200,000,000	D 6		R 7	E 5	Y 4	G 4	I 6	K 7	E 5	D 5	V 4				
			N 6	V 3		I 3	D 2								
			G 5				K 2								

Peptide libraries based on the V4 region of HIV gp120 from subtype F were designed. The libraries had the structure:

X^1-X^2-T-X^3-X^4-L-F-X^5-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-I-X^{14}-L-X^{15}-C

where X^1-X^{15} are defined according to Table 29.

Table 29

HIV gp120 V4 region, subtype F															
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}
50,000	C 95	N 91	S 91	G 53	N 85	D 44	T 92	N 81	D 46	T 84	N 86	G 53	T 81	T 82	P 89
		D 9		E 20	T 9	N 31		T 6	N 28	A 10	T 6	D 32	N 8	I 18	L 8
				K 10		S 17			S 10			S 9	I 6		
				Q 8											
500,000			D 6	I 6	K 6	A 5			G 5			N 6	P 5		
200,000,000	Y 5		A 3	S 3		H 3	A 4	E 5	M 4	N 3	K 4				Q 3
							V 4	I 5	K 3	I 3	D 4				
								A 3	I 3						

Peptide libraries based on the V4 region of HIV gp120 from subtype G were designed. The libraries had the structure:

C-N-T-S-X^1-L-F-X^2-X^3-X^4-X^5-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-I-X^{15}-L-X^{16}-C

where X^1-X^{16} are defined according to Table 30.

Table 30

HIV gp120 V4 region, subtype G																
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}	X^{13}	X^{14}	X^{15}	X^{16}
50,000	G 54	N 86	N 54	S 88	N 63	S 57	N 61	T 39	S 88	N 35	N 61	N 74	T 45	T 78	T 87	P 95
	K 26		S 25		S 16	D 22	S 34	I 23		E 29	S 27	D 15	D 34	N 15		
			E 21			N 20		N 14		K 14	E 12		G 21			
								E 12		T 14						
500,000	E 11				K 12					A 8		S 11				
	Q 10				D 9											
200,000,000		R 8		N 6			D 5	A 7	A 7					I 6	I 7	Q 5
		T 6		E 6				R 4	N 6						M 6	

Chimeric Peptide Libraries – V5 region of HIV gp120

Peptide libraries based on the V5 region of HIV gp120 from subtypes A, B, C, D, F, and G were designed. The libraries had the structure:

X^1-X^2-X^3-X^4-X^5-X^6-E-X^7-F-R-P-X^8

where X^1-X^8 are defined according to Table 31.

Table 31

HIV gp120 V5 region, subtypes ABCDFG								
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8
50,000	N 75	S 42	N 42	N 68	E 35	T 75	I 53	G 99
	D 9	N 25	T 36	S 18	T 23	N 20	T 47	
	G 6	E 16	D 13	E 6	N 18	P 2		
	T 5	G 10	K 5	G 5	G 13	I 2		
	K 3	K 5	G 4	K 3	K 10			
	P 2	I 1						
500,000					M 1	E 1		E 1
					H 1			

Peptide libraries based on the V5 region of HIV gp120 from subtype A were designed. The libraries had the structure:

X^1-X^2-X^3-X^4-X^5-X^6-X^7-X^8-X^9-R-P-X^{10}

where X^1-X^{10} are defined according to Table 32.

Table 32

HIV gp120 V5 region, subtype A										
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}
50,000	G 31	N 60	N 76	S 76	T 57	N 61	E 99	T 85	F 99	G 78
	N 24	T 22	S 17	N 16	E 16	T 27		I 15		I 16
	D 24	G 6	D 3	R 4	N 10	E 7				T 4
	S 8	D 5			A 9	P 3				
	K 6	I 4			I 4					
	V 6	K 3								
500,000				G 2	K 2	K 2	Q 1		S 1	R 2
					M 1					
200,000,000			R 1	E 1	P 1					
			G 1	F 1						
			I 1							
			H 1							

Peptide libraries based on the V5 region of HIV gp120 from subtype B were designed. The libraries had the structure:

X^1-X^2-X^3-X^4-X^5-X^6-E-X^7-F-R-P-G

where X^1-X^7 are defined according to Table 33.

Table 33

HIV gp120 V5 region, subtype B							
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7
50,000	N 78	S 44	N 40	N 69	E 37	T 77	I 57
	D 9	N 25	T 37	S 16	T 21	N 19	T 43
	G 4	E 16	D 15	G 6	N 18	P 2	
	T 4	G 9	K 5	E 6	G 13	I 2	
	K 3	K 6	G 4	K 3	K 10		
	P 2	M 1			M 2		

Peptide libraries based on the V5 region of HIV gp120 from subtype C were designed. The libraries had the structure:

X^1-X^2-X^3-X^4-X^5-E-X^6-X^7-X^8-X^9-X^{10}

where X^1-X^{10} are defined according to Table 34.

Table 34

HIV gp120 V5 region, subtype C										
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}
50,000	E 45	T 44	N 81	D 28	T 85	T 55	F 99	R 98	P 99	G 89
	N 18	N 16	T 13	N 26	N 6	I 42				E 5
	T 12	D 13	D 4	T 85	E 5	K 2				I 2
	G 11	G 13		K 6	G 2					Q 2
	K 7	K 7		G 9	I 2					T 2
	L 5	I 5								
500,000	P 2	P 2	G 2	I 2				Q 1	T 1	
200,000,000						N 1	I 1	S 1		

Peptide libraries based on the V5 region of HIV gp120 from subtype D were designed. The libraries had the structure:

X^1-X^2-X^3-X^4-X^5-X^6-X^7-E-X^8-X^9-R-X^{10}-X^{11}

where X^1-X^{11} are defined according to Table 35.

Table 35

HIV gp120 V5 region, subtype D											
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}
50,000	N 58	N 92	N 44	S 70	N 52	Q 50	N 78	T 94	F 97	P 98	G 99
	T 13	G 3	S 40	N 25	S 32	E 22	S 18	I 6	I 3		
	A 13	D 3	D 6		R 7	H 12					
	K 9		G 6		G 4	S 10					
	E 3				H 3	P 3					
	H 3										
500,000			I 2	G 2	D 2	R 2	H 2			T 2	
			H 2	R 2			R 2				
200,000,000	Y 1	H 1		D 1		G 1					Q 1
		K 1									

Peptide libraries based on the V5 region of HIV gp120 from subtype F were designed. The libraries had the structure:

X^1-X^2-X^3-X^4-X^5-X^6-X^7-F-R-P-X^8

where X^1-X^8 are defined according to Table 36.

Table 36

HIV gp120 V5 region, subtype F								
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8
50,000	N 43	N 65	N 46	T 51	N 60	E 97	T 83	G 80
	L 30	D 14	S 35	K 14	S 29	Q 3	I 17	E 10
	E 16	G 7	G 7	D 14	D 7			I 7
	S 8	K 7	E 6	N 14	I 4			Q 3
	K 4	T 7	K 6	G 7				

Peptide libraries based on the V5 region of HIV gp120 from subtype G were designed. The libraries had the structure:

X^1-X^2-X^3-X^4-X^5-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}

where X^1-X^{12} are defined according to Table 37.

Table 37

HIV gp120 V5 region, subtype G												
Library size <	X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}	X^{12}
50,000	N 40	N 78	N 75	S 53	T 85	N 63	E 98	T 82	F 91	R 93	P 87	G 88
	A 18	T 9	T 10	N 35	G 7	T 25		I 17	I 5		A 7	R 8
	T 16	A 8	D 10	E 4		E 6					T 4	
	D 15			A 4								
	R 9											
500,000			A 3		E 3	G 4			N 2	E 4		
									T 2	S 2		
200,000,000	Y 2	K 2	K 2	I 2	K 2	K 1	N 1	D 1		G 1	D 1	E 1
		D 2		K 1	N 2	F 1	S 1				K 1	W 1
		P 1		H 1	F 1							Y 1
												V 1

Synthesis and Antigenic Properties of a VCPL

A peptide library mimicking the naturally occurring diversity of sequences in the main hypervariable V3 region of the HIV-1 (subtype B) gp120 envelope protein was designed. The identity and frequency of the amino acids in the library were chosen based on sequence alignments and antigenic similarity. HIV-1 sequences are available from the HIV Sequence Database (hiv-web.lanl.gov). The peptide library had the structure:

N-N-N-T-R-K-X^1-I-X^2-X^3-X^4-X^5-G-X^6-X^7-X^8-Y-X^9-T-G-X^{10}-I-I-G-X^{11}-I-R-Q

where X^1- X^{11} were defined according to Table 38.

Table 38

X^1	X^2	X^3	X^4	X^5	X^6	X^7	X^8	X^9	X^{10}	X^{11}
S 75	H 60	I 82	G 97	P 95	R 77	A 93	F 76	T 58	E 39	D 88
G 18	P 18	M 12	A 3	W 5	Q 9	T 7	W 15	A 42	D 28	N 12
R 7	N 11				K 8		L 9		Q 24	
	S 11				S 3				R 9	
					G 3					

The peptides of the library were synthesized on Applied Biosystems 430A and Vega Coupler C250 automatic synthesizers with the use of the Boc/Bzl strategy of peptide chain elongation. A styrene copolymer with 1% divinylbenzene was used as the polymer support, and a phenacylamidomethyl (PAM) group served as the anchor group (Sparrow, J.T. *J. Org. Chem.* 1976, 41:1350-1353, which is incorporated by reference in its entirety). Condensation was done with 1-hydroxybenzotriazole esters prepared directly before use by mixing a 3-fold molar excess relative to the polymers of the Boc product of amino acid, 1-hydroxybenzotriazole and diisopropylcarbodiimide. The completeness of the condensation was monitored (Kaiser, E. et al. *Anal. Biochem.* 1970, 34:395-398, which is incorporated by reference in its entirety). In the event of a positive ninhydrin test after neutralization of the peptidyl polymer, condensation was repeated. If the test was negative, then the temporary Boc protective groups were removed with undiluted trifluoroacetic acid, and neutralization was performed by adding N,N'-diisopropylethylamine directly to the reaction mixture (Schnolzer, M. et al. *Int. J. Pept. Prot. Res.* 1992, 40: 180-183, which is incorporated by reference in its entirety). After removal of the Boc groups the following amino acid residue was attached. Permanent protective groups used for side chains were: 2-chlorobenzyloxycarbonyl groups for lysine, dichlorobenzyl for tyrosine, the corresponding benzyl esters for aspartic acid, glutamic acid, serine and threonine, formyl for tryptophan, tosyl for arginine, and methionine was introduced in the form of the corresponding sulfoxide. Upon completion of assembly of the protective polypeptide chain, the end products were deblocked with a one-time cleavage from the polymer with the aid of anhydrous hydrogen

fluoride in the presence of scavengers (Tam, J.P. et al. *J. Am. Chem. Soc.* 1983, 105: 6442-6444, which is incorporated by reference in its entirety).

During the synthesis of peptide libraries a mixture of protected amino acid residues was added in the condensation stage in variable positions relative to the peptidyl polymer in a certain quantitative relationship, which provided a given incorporation of different amino acid residues to the growing polypeptide chain. Because different amino acids have different rates of amino group acylation, a number of model compounds were synthesized and reaction rate coefficients were derived. The Boc products of the amino acids which were introduced into the mixture, were used in ratios adjusted for the reaction rate coefficients, to provide peptide products having the desired ratios (as shown in Table 38) in the variable positions.

The peptide library was purified by gel permeation chromatography using Sepharose G-50 and 50% acetic acid as the eluent. The synthesized compounds were characterized by analytical single-photon high-performance liquid chromatography (Gilson chromatograph, France; Xterra RP18 column, 125 Å, 3.9 x 150 mm, 5 μm, flow rate 1 mL/min, eluent 0.1% trifluoroacetic acid, 10-40% and 20-40% acetonitrile gradient in 16 min for pentarphin and cyclopentarphin, respectively), and by amino acid analysis (hydrolysis 6N HCl, 24 hours, 110 °C; LKB amino acid analyzer, Alpha Plus, Sweden). The purity of the end products according to analytic HPLC was not less than 95%. The amino acid composition of peptide hydrolysates (6N HCl, 120 °C, 24 hours, LKB 4151 analyzer, Alpha Plus) corresponded to the theoretical.

BALB/c line mice ages 5-6 weeks were immunized 3 times intraperitoneally with the peptide library diluted in a phosphate buffer solution with Complete Freund's Adjuvant (CFA). The mice were administered 0.2 mL peptide solution in a dose of 50 μg/head. The second and third immunization were conducted 2 and 4 weeks after the first immunization, respectively. A control group of animals were administered a peptide solution mimicking the antigenic portion of protein E of the tick encephalitis virus (peptide sequence: KRDQSDRGWGNHAGLFGKGSIVT) according to the same scheme and in the same doses. Ten days after the last immunization a blood sample was taken from the retroorbital vein of the animals for determination of the titer of specific antibodies in the serum.

The following antigens were immobilized in separate wells of microtiter plates for immunoenzyme analysis (Medpolimer, Moscow): the chimeric-peptide library, and each of

15 synthetic peptides representing the amino acid sequences of the V3 region of gp120 from different HIV-1 subtypes. The 15 synthetic peptides represent the consensus sequences of appropriate subtypes. Where several peptides are present from one subtype, they represent different variants.

The antigens were dissolved in 0.05 M carbonate-bicarbonate buffer solution, pH 9.6 to a final concentration of 10 μg/mL, placed in each well of the microtiter plate in 0.1 mL amounts and incubated overnight at a temperature of 37 °C. The wells were washed 3 times with phosphate buffer solution with Tween 20, and non-specific binding was blocked with a 0.5% solution of bovine serum albumin, fraction 5 (Sigma). Mouse blood sera samples were serially diluted two-fold in a Tris-HCl buffer solution, and placed in 0.1 mL amounts into the wells with immobilized antigen and incubated at 37 °C for 30 min. After incubation the wells were washed 5 times with a phosphate buffer solution. The antigen-antibody complexes were determined by adding 0.1 mL quantities of horseradish peroxidase-conjugated goat antibodies against mouse IgG (Sigma) in phosphate buffer solution. Following incubation at 37 °C for 30 min, a substrate mixture containing orthophenylene diamine was added. The enzyme reaction was halted by the addition of a sulfuric acid solution, and the optical density at 492 nm was measured using a Multiscan EX microplate photometer. The results of titer determination of the antigen-specific antibodies in the mouse blood sera are presented in Table 39.

Immunization of mice with the chimeric-peptide library caused formation of serum antibodies specific both for the chimeric-peptide library itself (the immunizing antigens) and for peptides forming antigenic determinants in the V3 region of some known HIV-1 subtypes. In particular, the antibody titers for subtype B peptides were higher than for other subtypes, as would be expected for a peptide library design to mimic the diversity of subtype B and not other subtypes. The chimeric peptide library was immunogenic, and induced formation of antibodies that interact with antigenic determinants of a broad spectrum of HIV-1 variants.

Table 39

Immobilized antigen (peptides corresponding to region V3 gp120 HIV-1, subtype)	Mean titer of specific antibodies
Chimeric-peptide library	1:1600
Peptide № 1, subtype A	1:400
Peptide № 2, subtype B	1:3200
Peptide № 3, subtype B	1:1600
Peptide № 4, subtype B	1:800
Peptide № 5, subtype B	1:1600
Peptide № 6, subtype B	1:1600
Peptide № 7, subtype C	1:400
Peptide № 8, subtype D	1:400
Peptide № 9, subtype E	1:400
Peptide № 10, subtype E	1:200
Peptide № 11, subtype F	1:100
Peptide № 12, subtype G	<1:100
Peptide № 13, subtype H	<1:100
Peptide № 14, subtype I	1:100
Peptide № 15, subtype J	1:100
Peptide of protein E of the tick encephalitis virus (control)	<1:100

The peptide library described above was tested as a diagnostic reagent in solid-phase immunoenzyme analysis (i.e., ELISA) of serum samples from 50 HIV-infected patients and 30 healthy donors for the purpose of determining specific antibodies. Preliminarily sera of HIV-infected patients were analyzed with the aid of a number of commercial test systems. In addition, the peptide library was assessed for its ability to detect antibodies to HIV-1 with the aid of standard sera panels anti-HIV-1 series 010 and series 007.

The chimeric-peptide library was dissolved in 0.05 M carbonate-bicarbonate buffer solution, pH 9.6 to a final concentration of 1.5 micrograms/mL. The solutions were added in 0.1 mL quantities to the wells of the microtiter plates and incubated at room temperature overnight, to immobilize the peptides in the wells. Then the wells were washed 3-5 times with a solution of FSB-T (phosphate-salt buffer solution containing Tween 20). Henceforth the method completely corresponds to the instructions for the use of the immunoenzyme test system for detecting antibodies to HIV-1 produced by DGUEPP "Vektor-BiAlgam" (Kol'tsovo, Novosibirsk Oblast).

Blood sera samples were diluted 20 fold in an FSB-T solution containing 0.2% casein and were placed in wells of the microtiter plate in 0.1 mL amounts and incubated at 37 °C for

30 min, to bind antibodies to the immobilized peptides. Then the wells were washed 5-7 times with FSB-T solution. To the wells were added 0.1 mL amounts of a monoclonal antibody to IgG conjugated to horseradish peroxidase in a dilution of 1:20 in FSB-T containing 0.2% casein. The microtiter plate was incubated at 37 °C for 30 min, then washed 5-7 times with an FSB-T solution. Next, 0.1 mL amounts of a solution of substrate mixture of 0.05% TMB and 0.005% hydrogen peroxide in 0.05 M phosphate-citrate buffer solution, pH 5.0 was added to the wells. The microtiter plate was kept at room temperature (18-22° C) for 15-20 min in the dark. The reaction was stopped by addition of 0.05 mL amounts of 2 M sulfuric acid solution to the wells. Calculation of the results is accomplished by measurement of the optical density of the samples on a Multiscan-type automatic spectrophotometer at a wavelength of 450 nm.

As a control for ruling out non-specific results in setting up the reaction a substrate mixture (BCP + substrate) is used, along with a control conjugate (BCP + conjugate + substrate), and a control serum which does not contain antibodies to HIV according to data from testing in Russian and foreign third generation screening test systems.

The test sera are judged to have antibodies to HIV-1 if the values of the optical density (OD) of the corresponding investigative solution in the wells with the investigative sera exceed the critical level ODcrit, calculated according to the formula:

$$ODcrit = OD(C')_{mean} + 0.2$$

where $OD(C')_{mean}$ is the mean optical density for negative control sera.

The results of the study of the anti-HIV-1 standard panel sera series 010 and 007 with the use of the chimeric-peptide library showed an 81% sensitivity and 93% specificity. In addition, a study was undertaken of the effectiveness of the chimeric-peptide library for detection of specific antibodies to HIV-1 in blood sera samples with the use of 50 sera from HIV-infected patients and 30 sera from healthy donors. The results of the tests showed more than 80% detection of positive samples containing specific antibodies to HIV-1 antigens on the basis of results with commercial third generation test systems.

To test the immunogenicity of the chimeric peptide library in different antigen forms, the library was synthesized as linear peptides, a multiple antigen peptide of 4 linear peptides (MAP4), or multiple antigen peptide of 8 linear peptides (MAP8). Rabbits were immunized subcutaneously with 300 micrograms of the constructions described above mixed with either

complete Freund's adjuvant (first immunization) or incomplete Freund's adjuvant (second and third immunizations) at 12 day intervals. At day 12 after the third immunization, the rabbits were bled and all blood sera were tested by ELISA against the linear, MAP4 and MAP8 antigens. Table 40 illustrates that immunization of rabbits with the HIV-1 V3 library in different antigen forms induces a strong humoral immune response to the highly variable V3 loop sequences from chimeric peptide library.

Table 40

	Antibody titers in ELISA for different antigen forms		
Immunizing dose	linear	MAP4	MAP8
linear	5,120	-	-
MAP4	81,920	163,840	163,840
MAP8	10,240	81,920	81,920

The immunogenicity of the MAP4 chimeric peptide library in different formulations was tested. The library was formulated in Freund's adjuvant, artificial virus like particles (VLP) or in phosphate buffered solution (PBS). In addition, the MAP4 construction was modified with palmitic acid (MAP4-P) and formulated in liposomes or microemulsion. The liposomes included 100 micrograms of MAP4-P and 250 micrograms of dsRNA, and the microemulsion included 100 micrograms of MAP4-P and 100 micrograms of dsRNA.

The VLPs contain a DNA molecule covered with polypeptides carrying the MAP4 peptide library. The target polypeptides are exposed on the surface of the particle and are attached to DNA via spermidine-polygluquine conjugates. The dimensions of the particle will allow the target antigens to conjugate on the surface in copy numbers ranging from one to several hundred. The core of the VLP can contain DNA segments as large as 10,000 bp. See, for example, Lebedev, L. R. et al. (2000) *Molecular genetics, microbiology and virology (Russian)*(3), 36-40; Lebedev, L. R., et al. (2000). *Molecular Biology (Russian)* 34(3), 480-485; and Lebedev, L. R. et al. (2001) *Biotechnology (Russian)*(1), 3-12; and Sizov, A. A. et al. (2001) *Biotechnology (Russian)*(1), 13-18, each of which is incorporated by reference in its entirety.

The liposomes forms were prepared with the MAP4-P library. The palmitic acid modification increases the lipophilicity of the peptides. The liposomal compositions of the

peptide were made by solubilization-injection. The size of the bulk of liposomes was within 200 nm.

The microemulsions were also made with the MAP4-P peptide construct and contained (in 1 mL) 100 micrograms of peptide, 250 micrograms of ridostin, and 800 microgams of the commercial (ICN) cationic amphiphile dimethyldioctadecyl ammonium bromide (DDAB).

Mice (five mice in each group) were immunized intraperitoneally with the constructions described above three times at 12 days intervals. The data in Table 41 indicate the high diversity of individual immune response in conventional mice, but nevertheless strong humoral immune response was detected for all systems after 3rd immunization, except for the peptide library solution in PBS. The antigen formulation of MAP4-P in liposomes induced high antibody levels after the second immunization.

Table 41

Antigen Form	Antigen Delivery System	Total Dose µg	Im. No.	AB titers in mouse serum in ELISA for MAP4 in solid phase					AB titer, Average	AB Titer, Stand. Deviation	AB Titer, I95
				Mouse #							
				1	2	3	4	5			
MAP4	PBS	40	2nd	0	0	0	0	-	0	0	
		60	3rd	0	0	0	0	-	0	0	
MAP4	Freund's adjuvant	40	2nd	0	3200	1600	1600	12800	3840	5135	4501
		60	3rd	51200	12800	102400	25600	204800	79360	78070	68430
MAP4-P	Liposome	40	2nd	0	51200	51200	25600	-	32000	24510	24019
		60	3rd	51200	102400	51200	102400	-	76800	29560	28969
MAP4-P	Microemulsion	40	2nd	0	0	6400	25600	-	8000	12115	11872
		60	3rd	0	0	204800	51200	-	64000	96920	94980
MAP4-P	VLP	100	2nd	12800	12800	6400	-	-	10667	3695	4181
		150	3rd	51200	204800	25600	-	-	93867	96920	109673

To determine the polarization of the immune response of mice immunized with the MAP4 library in different delivery systems described above, on day 10 after the 3rd immunization, splenocytes were distinguished. The splenocytes were stimulated *in vitro* with the library of linear peptides. A number of splenocytes produced IFN-gamma or IL-4 cytokines, markers of Th1 or Th2 immune response, respectively, as measured by ELISPOT.

The number of cells that produced one of two cytokines was noted, after stimulation or without stimulation (Table 42). The results presented in Table 42 show that immunization of mice with a solution of MAP4 peptide library with Freund's Adjuvant does not result in a major difference (significance level 0.05) in the number of stimulated cells from the control measurement (without stimulation), which confirms the absence of induction of the T-helper response with immunization by antigens in a mixture with Freund's Adjuvant. The highest number of cells producing IFN-gamma after stimulation in vitro using the linear peptide library was noted in mice immunized with a microemulsion containing the MAP4 library. The immunization using VLP resulted in the formation of a less pronounced immune response of the Th1 type. Comparison of the number of cells producing IL-4 and IFN-gamma reveals that immunization of mice with the MAP4 library in VLPs results in a balanced T-helper response, i.e. the number of cells producing IFN-gamma and IL-4 does not significantly differ. On the other hand, immunization of mice with the same antigens in a microemulsion results in a significant predominance of the Th1 type immune response after stimulation with the library of linear peptides. Thus, the above example shows the possibility of inducing a strong T-helper response using the MAP4 library of chimeric peptides.

Table 42

Antigen delivery system	Splenocytes stimulated in vitro ?	Number of cells producing IFN-gamma Determined for 500,000 splenocytes in ELISPOT								
		Cell sample ##						Average No. of cells	Standard Deviation	I95 of cell number
		1	2	3	4	5	6			
VLP	Yes	59	99	272	-	55	29	102.80	97.84	85.76
	No	2	5	1	-	3	5	3.20	1.79	1.57
CFA/IFA	Yes	6	3	33	37	12	10	16.83	14.47	11.58
	No	1	0	7	6	1	1	2.67	3.01	2.41
Micro emulsion	Yes	165	145	525	450	600	600	414.17	208.41	166.76
	No	4	0	1	2	3	1	1.83	1.47	1.18
		Number of cells producing IL-4. Determined for 500,000 splenocytes in ELISPOT								
VLP	Yes	21	36	420	490	61	43	178.50	215.70	172.59
	No	11	8	9	13	3	0	7.33	4.93	3.94
CFA/IFA	Yes	47	24	39	54	2	5	28.50	21.81	17.45
	No	27	33	26	31	4	4	20.83	13.29	10.63
Micro emulsion	Yes	56	42	159	135	79	79	91.67	45.76	36.62
	No	11	4	17	16	3	2	8.83	6.74	5.39

VLP – artificial virus-like particles
CFA/IFA – complete Freund's adjuvant/incomplete Freund's adjuvant

An RCPL having 30 peptides, each 28 residues in length, that models the antigenic diversity of HIV-1 subtype B variants was synthesized in the form of MAP4. The antigenic properties of the RCPL was compared to a VCPL which also modeled the antigenic diversity of HIV-1 subtype B variants.

RCPL sequences (a period '.' indicates that the residue unchanged from the first sequence listed):

NNNTRKSIHIGPGQAFYATGDIIGDIRQ
.....T..RL...RT..T..E.....K
S.....GV.......W....Q.......
S.......R....R..FR......N...
.......V......V..T..E.T....K
.....RR.RM...RT.............
G.......T....R.L.T..E...N...
.....RG.P......L..........K
.....T.V.L...RT..T..Q.......
.........RM..........G.T.N...
S.......P.....T....D.......K
......G......R.WHT..Q.......
S......V.L....V..........N...
.....TG......R..FT..E.....K
G....R..R..L.R......E.......
........TR......L.T....T.....
S......VT....RT..R.........K
.....T...L....V.F...E.......
......G.R..........D....N..K
......R..M...RT.....Q...N...
S......VT....R.L.T..G.......
.....R...L....VW.T..E.......
G.......R......L......T....K
.....RGV.....RT..T......N...
S....T.VP....R......Q......K
.........RL.W..T..T.........K
..K..R..........H...........
S.....R..L...R...T..E...N...
..............T...R.Q.T.....
.....TG.P....R.W.T..Q.......

A rabbit was immunized 3 times with the MAP4 VCPL and CFA (first immunization) or IFA (second immunization). The antibody titer of the sera was tested by ELISA using either the immunizing antigen, or the MAP4 RCPL as the immobilized peptide. The titer of rabbit serum in ELISA with MAP4 VCPL on solid phase was 1:204,800, and with MAP4 RCPL on solid phase was 1:102,400.

A number of embodiments have been described. Nevertheless, it will be understood that various modifications may be made. For example, chimeric peptide libraries can be designed to mimic the antigenic diversity of other HIV proteins, or of proteins from other variable pathogens. Accordingly, other embodiments are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method of manufacturing a family of antigenic peptides comprising:

locating a plurality of variable positions in a region of a pathogen protein;

choosing a peptide sequence of the pathogen protein including the plurality of variable positions;

selecting one or more substitute amino acid residues for one of the variable positions based on antigenic similarity to amino acid residues naturally occurring at the variable position of the pathogen protein; and

preparing a family of antigenic peptides based on the peptide sequence and including the substitute amino acid residues.

2. The method of claim 1, wherein selecting includes determining the antigenic similarity using an antigenic similarity matrix.

3. The method of claim 1, further comprising assigning a frequency to each substitute amino acid residue in the family of antigenic peptides.

4. The method of claim 3, wherein preparing includes weighting the substitute amino acid residues in the family of antigenic peptides based on the assigned frequency.

5. The method of claim 3, wherein assigning further includes considering the frequency with which the variations naturally occur.

6. The method of claim 1, wherein the pathogen protein includes a hypervariable region.

7. The method of claim 1, wherein the pathogen protein is associated with a virus.

8. The method of claim 7, wherein the virus is HIV.

9. The method of claim 1, wherein the pathogen protein is HIV gp120.

10. The method of claim 9, wherein the region is selected from the group consisting of the V1, V2, V3, V4, and V5 regions.

11. The method of claim 7, wherein the virus is hepatitis B virus or hepatitis C virus.

12. The method of claim 7, wherein the virus is an influenza virus.

13. The method of claim 7, wherein the virus a dengue virus.

14. The method of claim 1, wherein the pathogen protein is associated with a malaria pathogen.

15. The method of claim 1, wherein the pathogen protein is associated with a tuberculosis pathogen.

16. The method of claim 1, further comprising determining an antigenic similarity between a peptide of the family and a region of a human protein.

17. The method of claim 16, further comprising removing a peptide from the family of antigenic peptides before preparing the family if the determined antigenic similarity between the peptide of the family and the region of a human protein exceeds a predetermined threshold.

18. The method of claim 1, wherein the family of antigenic peptides includes members, such that the members taken together have antigenic similarity to each naturally occurring sequence of the region of the pathogen protein.

19. The method of claim 1, wherein the family of antigenic peptides includes members, such that the members taken together have antigenic similarity to a non-naturally occurring sequence of the region of the pathogen protein.

20. The method of claim 1, further comprising identifying peptide sequences of the family, the identified peptide sequences being representative of the sequence diversity of the entire family.

21. The method of claim 20, wherein fewer than 500 sequences are identified.

22. The method of claim 20, wherein identifying includes calculating a distance between peptide sequences of the family.

23. The method of claim 22, wherein calculating a distance includes using an antigenic similarity matrix.

24. The method of claim 1, wherein preparing the family of antigenic peptides includes chemical synthesis of the family of peptides.

25. The method of claim 24, wherein the chemical synthesis includes combinatorial synthesis, whereby the peptides are formed as a mixture of different sequences.

26. The method of claim 24, wherein the chemical synthesis includes parallel synthesis, whereby each peptide is formed separately from other peptides.

27. The method of claim 26, further comprising mixing the separate peptides.

28. The method of claim 1, wherein preparing the family of antigenic peptides includes expression of the family of peptides by a host organism.

29. A composition comprising a family of antigenic peptides having amino acid sequences having antigenic similarity to amino acid sequences of a variable region of a pathogen protein, wherein each antigenic peptide in the family has at least one amino acid position that varies relative to other antigenic peptides in the family.

30. The composition of claim 29, wherein one amino acid residue occurs more frequently than another in the position that varies.

31. The composition of claim 29, wherein the family includes greater than 150 mutually unique antigenic peptides.

32. The composition of claim 29, wherein the family includes greater than 1,000 mutually unique antigenic peptides.

33. The composition of claim 29, wherein the family includes fewer than 100,000 mutually unique antigenic peptides.

34. The composition of claim 29, wherein the family includes fewer than 50,000 mutually unique antigenic peptides.

35. The composition of claim 29, wherein the family includes between 1,000 and 50,000 mutually unique antigenic peptides.

36. The composition of claim 29, wherein the pathogen protein is HIV gp120.

37. The composition of claim 36, wherein the variable region is selected from the group consisting of the V1 region, the V2 region, the V3 region, the V4 region and the V5 region.

38. The composition of claim 37, wherein the family of antigenic peptides includes sequences having antigenic similarity to sequences from a subtype of HIV.

39. The composition of claim 38, wherein the subtype is selected from the group consisting of subtype A, subtype B, subtype C, subtype D, subtype F, subtype G, a recombinant subtype, a subtype of HIV group N, a subtype of HIV group O, and combinations thereof.

40. The composition of claim 29, wherein at least two members of the family of antigenic peptides are mixed together.

41. The composition of claim 29, wherein the family of antigenic peptides are separated according to sequence.

42. The composition of claim 29, wherein the family includes a multiple antigenic peptide.

43. A peptide library comprising a family of peptides including the fragment:

-N-N-T-R-X^4-X^5-X^6-X^7-X^8-G-X^9-G-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-T-G-X^{15}-I-X^{16}-G-X^{17}-I-R-

wherein each X^4-X^{17} is a fragment zero, one, two or three amino acid residues in length.

44. The peptide library of claim 43, wherein the family has antigenic similarity to the V3 region of HIV gp120.

45. The peptide library of claim 44, wherein the family has antigenic similarity to the V3 region of HIV gp120 of HIV subtypes A, B, C, D, F, and G.

46. The peptide library of claim 43, wherein the family of peptides have the formula:

X^1-C-X^2-R-P-X^3-N-N-T-R-X^4-X^5-X^6-X^7-X^8-G-X^9-G-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-T-G-X^{15}-I-X^{16}-G-X^{17}-I-R-X^{18}-A-X^{19}-C-X^{20}

wherein each X^1-X^{20} is a fragment zero, one, two or three amino acid residues in length.

47. The peptide library of claim 46, wherein for each peptide of the family, X^1 independently is N, V, T or E.

48. The peptide library of claim 46, wherein for each peptide of the family, X^2 independently is T or I.

49. The peptide library of claim 46, wherein for each peptide of the family, X^3 independently is N, S, G or Y.

50. The peptide library of claim 46, wherein for each peptide of the family, X^4 independently is K, Q or T.

51. The peptide library of claim 46, wherein for each peptide of the family, X^5 independently is S, G or R.

52. The peptide library of claim 46, wherein for each peptide of the family, X^6 independently is I, T or M.

53. The peptide library of claim 46, wherein for each peptide of the family, X^7 independently is H, R, P, S or N.

54. The peptide library of claim 46, wherein for each peptide of the family, X^8 independently is I or M.

55. The peptide library of claim 46, wherein for each peptide of the family, X^9 independently is P or L.

56. The peptide library of claim 46, wherein for each peptide of the family, X^{10} independently is R or Q.

57. The peptide library of claim 46, wherein for each peptide of the family, X^{11} independently is A, T or V.

58. The peptide library of claim 46, wherein for each peptide of the family, X^{12} independently is F or W.

59. The peptide library of claim 46, wherein for each peptide of the family, X^{13} independently is Y, F or H.

60. The peptide library of claim 46, wherein for each peptide of the family, X^{14} independently is A or T.

61. The peptide library of claim 46, wherein for each peptide of the family, X^{15} independently is D, Q, A or R.

62. The peptide library of claim 46, wherein for each peptide of the family, X^{16} independently is I or T.

63. The peptide library of claim 46, wherein for each peptide of the family, X^{17} independently is D or N.

64. The peptide library of claim 46, wherein for each peptide of the family, X^{18} independently is Q or K.

65. The peptide library of claim 46, wherein for each peptide of the family, X^{19} independently is H or Y.

66. The peptide library of claim 46, wherein for each peptide of the family, X^{20} independently is N or T.

67. The peptide library of claim 46, wherein for each peptide of the family, X^1 independently is N, V, T or E; X^2 independently is T or I; X^3 independently is N, S, G or Y; X^4 independently is K, Q or T; X^5 independently is S, G or R; X^6 independently is I, T or M; X^7 independently is H, R, P, S or N; X^8 independently is I or M; X^9 independently is P or L; X^{10} independently is R or Q; X^{11} independently is A, T or V; X^{12} independently is F or W; X^{13} independently is Y, F or H; X^{14} independently is A or T; X^{15} independently is D, Q, A or R; X^{16} independently is I or T; X^{17} independently is D or N; X^{18} independently is Q or K; X^{19} independently is H or Y; and X^{20} independently is N or T.

68. The peptide library of claim 46, wherein at least two members of the family of peptides are mixed together.

69. The peptide library of claim 46, wherein the family of peptides are separated according to sequence.

70. The peptide library of claim 46, wherein the family includes greater than 150 mutually unique peptide sequences.

71. The peptide library of claim 46, wherein the family includes fewer than 100,000 mutually unique peptide sequences.

72. The peptide library of claim 46, wherein the family includes fewer than 500 mutually unique peptide sequences, the sequences being representative of the entire sequence diversity available.

73. A peptide library comprising a family of peptides including the fragment:

-N-N-X^4-R-X^5-X^6-V-X^7-I-G-P-G-X^8-X^9-F-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-I-X^{15}-G-X^{16}-I-R-

wherein each X^4-X^{16} is a fragment zero, one, two or three amino acid residues in length.

74. The peptide library of claim 73, wherein the family has antigenic similarity to the V3 region of HIV gp120.

75. The peptide library of claim 74, wherein the family has antigenic similarity to the V3 region of HIV gp120 of HIV subtype A.

76. The peptide library of claim 73, wherein the family of peptides have the formula:

X^1-C-X^2-R-P-X^3-N-N-X^4-R-X^5-X^6-V-X^7-I-G-P-G-X^8-X^9-F-X^{10}-X^{11}-X^{12}-X^{13}-X^{14}-I-X^{15}-G-X^{16}-I-R-X^{17}-A-X^{18}-C-X^{19}

wherein each X^1-X^{19} is a fragment zero, one, two or three amino acid residues in length.

77. The peptide library of claim 76, wherein for each peptide of the family, X^1 independently is N, T, or D.

78. The peptide library of claim 76, wherein for each peptide of the family, X^2 independently is T or I.

79. The peptide library of claim 76, wherein for each peptide of the family, X^3 independently is N, G, or S.

80. The peptide library of claim 76, wherein for each peptide of the family, X^4 independently is T or K.

81. The peptide library of claim 76, wherein for each peptide of the family, X^5 independently is K, T, or Q.

82. The peptide library of claim 76, wherein for each peptide of the family, X^6 independently is S, G, R or N.

83. The peptide library of claim 76, wherein for each peptide of the family, X^7 independently is R or H.

84. The peptide library of claim 76, wherein for each peptide of the family, X^8 independently is Q or R.

85. The peptide library of claim 76, wherein for each peptide of the family, X^9 independently is A, T, or V.

86. The peptide library of claim 76, wherein for each peptide of the family, X^{10} independently is Y, F, or H.

87. The peptide library of claim 76, wherein for each peptide of the family, X^{11} independently is A or T.

88. The peptide library of claim 76, wherein for each peptide of the family, X^{12} independently is T, R, A, or no residue.

89. The peptide library of claim 76, wherein for each peptide of the family, X^{13} independently is G, D, or no residue.

90. The peptide library of claim 76, wherein for each peptide of the family, X^{14} independently is D, A, N, R, or V.

91. The peptide library of claim 76, wherein for each peptide of the family, X^{15} independently is I or T.

92. The peptide library of claim 76, wherein for each peptide of the family, X^{16} independently is D or N.

93. The peptide library of claim 76, wherein for each peptide of the family, X^{17} independently is Q, K, or E.

94. The peptide library of claim 76, wherein for each peptide of the family, X^{18} independently is H or Y.

95. The peptide library of claim 76, wherein for each peptide of the family, X^{19} independently is N, T, or E.

96. The peptide library of claim 76, wherein for each peptide of the family X^1 independently is N, T, or D; X^2 independently is T or I; X^3 independently is N, G, or S; X^4 independently is T or K; X^5 independently is K, T, or Q; X^6 independently is S, G, R or N; X^7 independently is R or H; X^8 independently is Q or R; X^9 independently is A, T, or V; X^{10} independently is Y, F, or H; X^{11} independently is A or T; X^{12} independently is T, R, A, or no residue; X^{13} independently is G, D, or no residue; X^{14} independently is D, A, N, R, or V; X^{15} independently is I or T; X^{16} independently is D or N; X^{17} independently is Q, K, or E; X^{18} independently is H or Y; and X^{19} independently is N, T, or E.

97. The peptide library of claim 76, wherein at least two members of the family of peptides are mixed together.

98. The peptide library of claim 76, wherein the family of peptides are separated according to sequence.

99. The peptide library of claim 76, wherein the family includes greater than 150 mutually unique peptide sequences.

100. The peptide library of claim 76, wherein the family includes fewer than 100,000 mutually unique peptide sequences.

101. The peptide library of claim 76, wherein the family includes fewer than 500 mutually unique peptide sequences, the sequences being representative of the entire sequence diversity available.

102. A peptide library comprising a family of peptides including the fragment:

-N-N-T-R-K-X^4-I-X^5-X^6-G-X^7-G-X^8-X^9-X^{10}-X^{11}-X^{12}-T-X^{13}-X^{14}-I-X^{15}-G-X^{16}-I-R-

wherein each X^4-X^{16} is a fragment zero, one, two or three amino acid residues in length.

103. The peptide library of claim 102, wherein the family has antigenic similarity to the V3 region of HIV gp120.

104. The peptide library of claim 103, wherein the family has antigenic similarity to the V3 region of HIV gp120 of HIV subtype B.

105. The peptide library of claim 102, wherein the family of peptides have the formula:

X^1-C-X^2-R-P-X^3-N-N-T-R-K-X^4-I-X^5-X^6-G-X^7-G-X^8-X^9-X^{10}-X^{11}-X^{12}-T-X^{13}-X^{14}-I-X^{15}-G-X^{16}-I-R-X^{17}-A-X^{18}-C-X^{19}

wherein each X^1-X^{19} is a fragment zero, one, two or three amino acid residues in length.

106. The peptide library of claim 105, wherein for each peptide of the family, X^1 independently is N, T, or H.

107. The peptide library of claim 105, wherein for each peptide of the family, X^2 independently is T or I.

108. The peptide library of claim 105, wherein for each peptide of the family, X^3 independently is N, S, or G.

109. The peptide library of claim 105, wherein for each peptide of the family, X^4 independently is S, G, or R.

110. The peptide library of claim 105, wherein for each peptide of the family, X^5 independently is H, P, N, T or Y.

111. The peptide library of claim 105, wherein for each peptide of the family, X^6 independently is I or M.

112. The peptide library of claim 105, wherein for each peptide of the family, X^7 independently is P, L, or W.

113. The peptide library of claim 105, wherein for each peptide of the family, X^8 independently is R, Q, G or S.

114. The peptide library of claim 105, wherein for each peptide of the family, X^9 independently is A, V or T.

115. The peptide library of claim 105, wherein for each peptide of the family, X^{10} independently is F, W, or V.

116. The peptide library of claim 105, wherein for each peptide of the family, X^{11} independently is Y, F or H.

117. The peptide library of claim 105, wherein for each peptide of the family, X^{12} independently is T or A.

118. The peptide library of claim 105, wherein for each peptide of the family, X^{13} independently is G, E or R.

119. The peptide library of claim 105, wherein for each peptide of the family, X^{14} independently is E, Q, R or G.

120. The peptide library of claim 105, wherein for each peptide of the family, X^{15} independently is I or T.

121. The peptide library of claim 105, wherein for each peptide of the family, X^{16} independently is D or N.

122. The peptide library of claim 105, wherein for each peptide of the family, X^{17} independently is Q or K.

123. The peptide library of claim 105, wherein for each peptide of the family, X^{18} independently is H or Y.

124. The peptide library of claim 105, wherein for each peptide of the family, X^{19} independently is N or T.

125. The peptide library of claim 105, wherein for each peptide of the family, X^1 independently is N, T or H; X^2 independently is T or I; X^3 independently is N, S, or G; X^4 independently is S, G, or R; X^5 independently is H, P, N, T or Y; X^6 independently is I or M; X^7 independently is P, L, or W; X^8 independently is R, Q, G or S; X^9 independently is A, V or T; X^{10} independently is F, W, or V; X^{11} independently is Y, F or H; X^{12} independently is T or A; X^{13} independently is G, E or R; X^{14} independently is E, Q, R or G; X^{15} independently is I or T; X^{16} independently is D or N; X^{17} independently is Q or K; X^{18} independently is H or Y; and X^{19} independently is N or T.

126. The peptide library of claim 105, wherein at least two members of the family of peptides are mixed together.

127. The peptide library of claim 105, wherein the family of peptides are separated according to sequence.

128. The peptide library of claim 105, wherein the family includes greater than 150 mutually unique peptide sequences.

129. The peptide library of claim 105, wherein the family includes fewer than 100,000 mutually unique peptide sequences.

130. The peptide library of claim 105, wherein the family includes fewer than 500 mutually unique peptide sequences, the sequences being representative of the entire sequence diversity available.

131. A peptide library comprising a family of peptides including the fragment:

-N-N-T-R-X^4-X^5-X^6-X^7-I-G-P-G-Q-X^8-F-X^9-X^{10}-T-X^{11}-X^{12}-I-X^{13}-G-X^{14}-I-R-

wherein each X^4-X^{14} is a fragment zero, one, two or three amino acid residues in length.

132. The peptide library of claim 131, wherein the family has antigenic similarity to the V3 region of HIV gp120.

133. The peptide library of claim 132, wherein the family has antigenic similarity to the V3 region of HIV gp120 of HIV subtype C.

134. The peptide library of claim 131, wherein the family of peptides have the formula:

X^1-C-X^2-R-P-X^3-N-N-T-R-X^4-X^5-X^6-X^7-I-G-P-G-Q-X^8-F-X^9-X^{10}-T-X^{11}-X^{12}-I-X^{13}-G-X^{14}-I-R-X^{15}-A-X^{16}-C-X^{17}

wherein each X^1-X^{17} is a fragment zero, one, two or three amino acid residues in length.

135. The peptide library of claim 134, wherein for each peptide of the family, X^1 independently is V, N, E, T, M or R.

136. The peptide library of claim 134, wherein for each peptide of the family, X^2 independently is T, I, or A.

137. The peptide library of claim 134, wherein for each peptide of the family, X^3 independently is N, G, S, or H.

138. The peptide library of claim 134, wherein for each peptide of the family, X^4 independently is K, E or Q.

139. The peptide library of claim 134, wherein for each peptide of the family, X^5 independently is S or G.

140. The peptide library of claim 134, wherein for each peptide of the family, X^6 independently is I or M.

141. The peptide library of claim 134, wherein for each peptide of the family, X^7 independently is R or G.

142. The peptide library of claim 134, wherein for each peptide of the family, X^8 independently is T, A, or V.

143. The peptide library of claim 134, wherein for each peptide of the family, X^9 independently is Y or F.

144. The peptide library of claim 134, wherein for each peptide of the family, X^{10} independently is A or T.

145. The peptide library of claim 134, wherein for each peptide of the family, X^{11} independently is G, N, D, K or T.

146. The peptide library of claim 134, wherein for each peptide of the family, X^{12} independently is D, G, N, S, or I.

147. The peptide library of claim 134, wherein for each peptide of the family, X^{13} independently is I or T.

148. The peptide library of claim 134, wherein for each peptide of the family, X^{14} independently is D or N.

149. The peptide library of claim 134, wherein for each peptide of the family, X^{15} independently is Q, K, E, L or H.

150. The peptide library of claim 134, wherein for each peptide of the family, X^{16} independently is H, Y, or N.

151. The peptide library of claim 134, wherein for each peptide of the family, X^{17} independently is N, T, I, D or H.

152. The peptide library of claim 134, wherein for each peptide of the family, X^1 independently is V, N, E, T, M or R; X^2 independently is T, I, or A; X^3 independently is N, G, S, or H; X^4 independently is K, E or Q; X^5 independently is S or G; X^6 independently is I or M; X^7 independently is R or G; X^8 independently is T, A, or V; X^9 independently is Y or F; X^{10} independently is A or T; X^{11} independently is G, N, D, K or T; X^{12} independently is D, G, N, S, or I; X^{13} independently is I or T; X^{14} independently is D or N; X^{15} independently is Q, K, E, L or H; X^{16} independently is H, Y, or N; and X^{17} independently is N, T, I, D or H.

153. The peptide library of claim 134, wherein at least two members of the family of peptides are mixed together.

154. The peptide library of claim 134, wherein the family of peptides are separated according to sequence.

155. The peptide library of claim 134, wherein the family includes greater than 150 mutually unique peptide sequences.

156. The peptide library of claim 134, wherein the family includes fewer than 100,000 mutually unique peptide sequences.

157. The peptide library of claim 134, wherein the family includes fewer than 500 mutually unique peptide sequences, the sequences being representative of the entire sequence diversity available.

158. A peptide library comprising a family of peptides including the fragment:
-R-P-X^3-X^4-X^5-X^6-R-X^7-X^8-X^9-X^{10}-I-G-X^{11}-G-X^{12}-X^{13}-X^{14}-X^{15}-X^{16}-T-X^{17}-X^{18}-G-X^{19}-I-
wherein each X^3-X^{19} is a fragment zero, one, two or three amino acid residues in length.

159. The peptide library of claim 158, wherein the family has antigenic similarity to the V3 region of HIV gp120.

160. The peptide library of claim 159, wherein the family has antigenic similarity to the V3 region of HIV gp120 of subtype D.

161. The peptide library of claim 158, wherein the family of peptides have the formula:

X^1-C-X^2-R-P-X^3-X^4-X^5-X^6-R-X^7-X^8-X^9-X^{10}-I-G-X^{11}-G-X^{12}-X^{13}-X^{14}-X^{15}-X^{16}-T-X^{17}-X^{18}-G-X^{19}-I-X^{20}-X^{21}-A-X^{22}-C-X^{23}

wherein each X^1-X^{23} is a fragment zero, one, two or three amino acid residues in length.

162. The peptide library of claim 161, wherein for each peptide of the family, X^1 independently is N or T.

163. The peptide library of claim 161, wherein for each peptide of the family, X^2 independently is T or I.

164. The peptide library of claim 161, wherein for each peptide of the family, X^3 independently is Y or N.

165. The peptide library of claim 161, wherein for each peptide of the family, X^4 independently is N or K.

166. The peptide library of claim 161, wherein for each peptide of the family, X^5 independently is N or K.

167. The peptide library of claim 161, wherein for each peptide of the family, X^6 independently is T, I or K.

168. The peptide library of claim 161, wherein for each peptide of the family, X^7 independently is Q or R.

169. The peptide library of claim 161, wherein for each peptide of the family, X^8 independently is S, R or G.

170. The peptide library of claim 161, wherein for each peptide of the family, X^9 independently is T or I.

171. The peptide library of claim 161, wherein for each peptide of the family, X^{10} independently is H, P, S or R.

172. The peptide library of claim 161, wherein for each peptide of the family, X^{11} independently is P, L, S or Q.

173. The peptide library of claim 161, wherein for each peptide of the family, X^{12} independently is Q or R.

174. The peptide library of claim 161, wherein for each peptide of the family, X^{13} independently is A or T.

175. The peptide library of claim 161, wherein for each peptide of the family, X^{14} independently is L or Y.

176. The peptide library of claim 161, wherein for each peptide of the family, X^{15} independently is Y or F.

177. The peptide library of claim 161, wherein for each peptide of the family, X^{16} independently is T or A.

178. The peptide library of claim 161, wherein for each peptide of the family, X^{17} independently is I or K.

179. The peptide library of claim 161, wherein for each peptide of the family, X^{18} independently is I, T or K.

180. The peptide library of claim 161, wherein for each peptide of the family, X^{19} independently is D or N.

181. The peptide library of claim 161, wherein for each peptide of the family, X^{20} independently is R or G.

182. The peptide library of claim 161, wherein for each peptide of the family, X^{21} independently is Q or K.

183. The peptide library of claim 161, wherein for each peptide of the family, X^{22} independently is H or Y.

184. The peptide library of claim 161, wherein for each peptide of the family, X^{23} independently is N, T or K.

185. The peptide library of claim 161, wherein for each peptide of the family, X^1 independently is N or T; X^2 independently is T or I; X^3 independently is Y or N; X^4 independently is N or K; X^5 independently is N or K; X^6 independently is T, I or K; X^7 independently is Q or R; X^8 independently is S, R or G; X^9 independently is T or I; X^{10} independently is H, P, S or R; X^{11} independently is P, L, S or Q; X^{12} independently is Q or R; X^{13} independently is A or T; X^{14} independently is L or Y; X^{15} independently is Y or F; X^{16} independently is T or A; X^{17} independently is I or K; X^{18} independently is I, T or K; X^{19} independently is D or N; X^{20} independently is R or G; X^{21} independently is Q or K; X^{22} independently is H or Y; and X^{23} independently is N, T or K.

186. The peptide library of claim 161, wherein at least two members of the family of peptides are mixed together.

187. The peptide library of claim 161, wherein the family of peptides are separated according to sequence.

188. The peptide library of claim 161, wherein the family includes greater than 150 mutually unique peptide sequences.

189. The peptide library of claim 161, wherein the family includes fewer than 100,000 mutually unique peptide sequences.

190. The peptide library of claim 161, wherein the family includes fewer than 500 mutually unique peptide sequences, the sequences being representative of the entire sequence diversity available.

191. A peptide library comprising a family of peptides including the fragment:

-N-N-X^3-R-K-X^4-I-X^5-L-G-P-G-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-I-X^{14}-G-X^{15}-I-R-

wherein each X^3-X^{15} is a fragment zero, one, two or three amino acid residues in length.

192. The peptide library of claim 191, wherein the family has antigenic similarity the V3 region of HIV gp120.

193. The peptide library of claim 192, wherein the family has antigenic similarity the V3 region of HIV gp120 of HIV subtype F.

194. The peptide library of claim 191, wherein the family of peptides have the formula:

X^1-C-T-R-P-X^2-N-N-X^3-R-K-X^4-I-X^5-L-G-P-G-X^6-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-I-X^{14}-G-X^{15}-I-R-X^{16}-A-X^{17}-C-X^{18}

wherein each X^1-X^{18} is a fragment zero, one, two or three amino acid residues in length.

195. The peptide library of claim 194, wherein for each peptide of the family, X^1 independently is N or T.

196. The peptide library of claim 194, wherein for each peptide of the family, X^2 independently is N or S.

197. The peptide library of claim 194, wherein for each peptide of the family, X^3 independently is T or I.

198. The peptide library of claim 194, wherein for each peptide of the family, X^4 independently is S, G, or R.

199. The peptide library of claim 194, wherein for each peptide of the family, X^5 independently is H, R, Q, P, Y, or S.

200. The peptide library of claim 194, wherein for each peptide of the family, X^6 independently is Q, R or H.

201. The peptide library of claim 194, wherein for each peptide of the family, X^7 independently is A, V or T.

202. The peptide library of claim 194, wherein for each peptide of the family, X^8 independently is F or I.

203. The peptide library of claim 194, wherein for each peptide of the family, X^9 independently is Y, H or F.

204. The peptide library of claim 194, wherein for each peptide of the family, X^{10} independently is A or T.

205. The peptide library of claim 194, wherein for each peptide of the family, X^{11} independently is T or A.

206. The peptide library of claim 194, wherein for each peptide of the family, X^{12} independently is G, D or S.

207. The peptide library of claim 194, wherein for each peptide of the family, X^{13} independently is D, A, or N.

208. The peptide library of claim 194, wherein for each peptide of the family, X^{14} independently is I or T.

209. The peptide library of claim 194, wherein for each peptide of the family, X^{15} independently is D or N.

210. The peptide library of claim 194, wherein for each peptide of the family, X^{16} independently is K or Q.

211. The peptide library of claim 194, wherein for each peptide of the family, X^{17} independently is H or Y.

212. The peptide library of claim 194, wherein for each peptide of the family, X^{18} independently is N, I, D, or K.

213. The peptide library of claim 194, wherein for each peptide of the family, X^1 independently is N or T; X^2 independently is N or S; X^3 independently is T or I; X^4 independently is S, G, or R; X^5 independently is H, R, Q, P, Y, or S; X^6 independently is Q, R or H; X^7 independently is A, V or T; X^8 independently is F or I; X^9 independently is Y, H or F; X^{10} independently is A or T; X^{11} independently is T or A; X^{12} independently is G, D or

S; X^{13} independently is D, A, or N; X^{14} independently is I or T; X^{15} independently is D or N; X^{16} independently is K or Q; X^{17} independently is H or Y; and X^{18} independently is N, I, D, K or X.

214. The peptide library of claim 194, wherein at least two members of the family of peptides are mixed together.

215. The peptide library of claim 194, wherein the family of peptides are separated according to sequence.

216. The peptide library of claim 194, wherein the family includes greater than 150 mutually unique peptide sequences.

217. The peptide library of claim 194, wherein the family includes fewer than 100,000 mutually unique peptide sequences.

218. The peptide library of claim 194, wherein the family includes fewer than 500 mutually unique peptide sequences, the sequences being representative of the entire sequence diversity available.

219. A peptide library comprising a family of peptides including the fragment:

-N-N-T-R-K-S-X^4-X^5-X^6-G-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-T-X^{14}-I-X^{15}-G-X^{16}-I-R-

wherein each X^4-X^{16} is a fragment zero, one, two or three amino acid residues in length.

220. The peptide library of claim 219, wherein the family has antigenic similarity to the V3 region of HIV gp120.

221. The peptide library of claim 220, wherein the family has antigenic similarity to the V3 region of HIV gp120 of HIV subtype G.

222. The peptide library of claim 219, wherein the family of peptides have the formula:

X^1-C-X^2-R-P-X^3-N-N-T-R-K-S-X^4-X^5-X^6-G-X^7-X^8-X^9-X^{10}-X^{11}-X^{12}-X^{13}-T-X^{14}-I-X^{15}-G-X^{16}-I-R-X^{17}-A-X^{18}-C-X^{19}

wherein each X^1-X^{19} is a fragment zero, one, two or three amino acid residues in length.

223. The peptide library of claim 222, wherein for each peptide of the family, X^1 independently is N, I, T, M or R.

224. The peptide library of claim 222, wherein for each peptide of the family, X^2 independently is T or I.

225. The peptide library of claim 222, wherein for each peptide of the family, X^3 independently is N or S.

226. The peptide library of claim 222, wherein for each peptide of the family, X^4 independently is I or K.

227. The peptide library of claim 222, wherein for each peptide of the family, X^5 independently is T, R, H, P or N.

228. The peptide library of claim 222, wherein for each peptide of the family, X^6 independently is F or I.

229. The peptide library of claim 222, wherein for each peptide of the family, X^7 independently is P, T, or L.

230. The peptide library of claim 222, wherein for each peptide of the family, X^8 independently is G or R.

231. The peptide library of claim 222, wherein for each peptide of the family, X^9 independently is Q, R, H or G.

232. The peptide library of claim 222, wherein for each peptide of the family, X^{10} independently is A, T or V.

233. The peptide library of claim 222, wherein for each peptide of the family, X^{11} independently is F, I or S.

234. The peptide library of claim 222, wherein for each peptide of the family, X^{12} independently is Y or H.

235. The peptide library of claim 222, wherein for each peptide of the family, X^{13} independently is A or T.

236. The peptide library of claim 222, wherein for each peptide of the family, X^{14} independently is G, D or S.

237. The peptide library of claim 222, wherein for each peptide of the family, X^{15} independently is I or T.

238. The peptide library of claim 222, wherein for each peptide of the family, X^{16} independently is D or N.

239. The peptide library of claim 222, wherein for each peptide of the family, X^{17} independently is Q, L, K or P.

240. The peptide library of claim 222, wherein for each peptide of the family, X^{18} independently is H, Y or F.

241. The peptide library of claim 222, wherein for each peptide of the family, X^{19} independently is N, K or T.

242. The peptide library of claim 222, wherein for each peptide of the family, X^1 independently is N, I, T, M or R; X^2 independently is T or I; X^3 independently is N or S; X^4 independently is I or K; X^5 independently is T, R, H, P or N; X^6 independently is F or I; X^7 independently is P, T, or L; X^8 independently is G or R; X^9 independently is Q, R, H or G; X^{10} independently is A, T or V; X^{11} independently is F, I or S; X^{12} independently is Y or H; X^{13} independently is A or T; X^{14} independently is G, D or S; X^{15} independently is I or T; X^{16} independently is D or N; X^{17} independently is Q, L, K or P; X^{18} independently is H, Y or F; X^{19} independently is N, K or T.

243. The peptide library of claim 222, wherein at least two members of the family of peptides are mixed together.

244. The peptide library of claim 222, wherein the family of peptides are separated according to sequence.

245. The peptide library of claim 222, wherein the family includes greater than 150 mutually unique peptide sequences.

246. The peptide library of claim 222, wherein the family includes fewer than 100,000 mutually unique peptide sequences.

247. The peptide library of claim 222, wherein the family includes fewer than 500 mutually unique peptide sequences, the sequences being representative of the entire sequence diversity available.

248. A method of eliciting an immune response in a subject comprising administering to the subject a composition including a family of antigenic peptides having amino acid sequences having antigenic similarity to amino acid sequences of a variable region of a pathogen protein of a pathogen.

249. The method of claim 248, wherein the antigenic similarity is determined using an antigenic similarity matrix.

250. The method of claim 248, wherein the composition is administered to a subject prior to infection by the pathogen.

251. The method of claim 250, wherein the family of antigenic peptides has amino acid sequences having antigenic similarity to amino acid sequences of a subtype of the pathogen protein.

252. The method of claim 250, wherein the family of antigenic peptides has amino acid sequences having antigenic similarity to amino acid sequences from more than one subtype of the pathogen protein.

253. The method of claim 248, wherein the composition is administered to a subject infected by the pathogen.

254. The method of claim 253, wherein the subject is infected by a subtype of the pathogen.

255. The method of claim 254, wherein the family of antigenic peptides has amino acid sequences having antigenic similarity to amino acid sequences from the subtype by which the subject is infected.

256. A method of designing a family of peptide sequences, comprising:

locating a plurality of variable positions in a region of a pathogen protein;

choosing a peptide sequence of the pathogen protein including the variable positions; and

selecting one or more substitute amino acid residues for one of the variable positions of the peptide based on antigenic similarity to amino acid residues naturally occurring at the variable position of the pathogen protein, thereby forming a family of peptide sequences.

257. The method of claim 256, wherein selecting includes determining the antigenic similarity using an antigenic similarity matrix.

258. The method of claim 256, further comprising assigning a frequency to each substitute amino acid residue in the family of antigenic peptide sequences.

259. The method of claim 258, wherein assigning further includes considering the frequency with which the variations naturally occur.

260. The method of claim 256, further comprising determining an antigenic similarity between a peptide of the family and a region of a human protein.

261. The method of claim 260, further comprising removing a peptide sequence from the family of peptide sequences if the calculated antigenic similarity between the peptide of the family and the region of a human protein exceeds a predetermined threshold.

262. The method of claim 256, wherein the family of peptide sequences includes members, such that the members taken together have antigenic similarity to each naturally occurring sequence of the region of the pathogen protein.

263. The method of claim 256, wherein the family of antigenic peptides includes members, such that the members taken together have antigenic similarity to a non-naturally occurring sequence of the region of the pathogen protein.

264. The method of claim 256, further comprising identifying peptide sequences of the family, the identified peptide sequences being representative of the sequence diversity of the entire family.

265. The method of claim 264, wherein fewer than 500 sequences are identified.

266. The method of claim 264, wherein identifying includes calculating a distance between peptide sequences of the family.

267. The method of claim 266, wherein calculating a distance includes using an antigenic similarity matrix.

268. The method of claim 256, further comprising identifying peptide sequences of the family having antigenic similarity to a subtype of the pathogen protein.

269. The method of claim 256, wherein the pathogen protein includes a hypervariable region.

270. The method of claim 256, wherein the pathogen is a virus.

271. The method of claim 270, wherein the virus is HIV.

272. The method of claim 271, wherein the pathogen protein is gp120.

273. The method of claim 272, wherein the region is selected from the group consisting of the V1 region, the V2 region, the V3 region, the V4 region, and the V5 region.

274. A method of diagnosing infection comprising contacting a sample with a family of peptides having amino acid sequences having antigenic similarity to amino acid sequences of a variable region of a pathogen protein, wherein each peptide in the family has at least one amino acid position that varies relative to other peptides in the family.

275. The method of claim 274, wherein antigenic similarity is determined using an antigenic similarity matrix.

276. The method of claim 274, wherein the family of peptides is antigenically similar to a pathogen subtype.

277. The method of claim 274, wherein the family of peptides is antigenically similar to more than one pathogen subtype.

278. The method of claim 274, further comprising immobilizing the family of peptides on a substrate before contacting.

279. The method of claim 274, further comprising determining if the sample includes antibodies that bind specifically to the family of peptides.