(12) STANDARD PATENT (11) Application No. **AU 2009316629 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Antiviral vaccines with improved cellular immunogenicity

(51) International Patent Classification(s)
C07K 17/00 (2006.01)

(21) Application No: **2009316629** (22) Date of Filing: **2009.11.18**

(87) WIPO No: **WO10/059732**

(30) Priority Data

(31) Number	(32) Date	(33) Country
61/152,184	**2009.02.12**	**US**
61/115,703	**2008.11.18**	**US**
61/248,188	**2009.10.02**	**US**

(43) Publication Date: **2010.05.27**
(44) Accepted Journal Date: **2016.02.04**

(71) Applicant(s)
Beth Israel Deaconess Medical Center;Los Alamos National Security, LLC

(72) Inventor(s)
Barouch, Dan H.;Korber, Bette T.;Fischer, William M.

(74) Agent / Attorney
Pizzeys, PO Box 291, WODEN, ACT, 2606

(56) Related Art
FISCHER, W., et al., "Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants," Nature Medicine, 2007, Vol. 13, No. 1, pages 100-106

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 May 2010 (27.05.2010)

PCT

(10) International Publication Number
WO 2010/059732 A1

(51) International Patent Classification:
C07K 17/00 (2006.01)

(21) International Application Number:
PCT/US2009/064999

(22) International Filing Date:
18 November 2009 (18.11.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/115,703	18 November 2008 (18.11.2008)	US
61/152,184	12 February 2009 (12.02.2009)	US
61/248,188	2 October 2009 (02.10.2009)	US

(71) Applicants *(for all designated States except US)*: **BETH ISRAEL DEACONESS MEDICAL CENTER** [US/US]; 330 Brookline Avenue, Boston, MA 02215 (US). **LOS ALAMOS NATIONAL LABORATORY** [US/US]; P.O. Box 1663, Los Alamos, NM 87545 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BAROUCH, Dan, H.** [US/US]; P.O. Box 1663, Brookline, MA (US). **KORBER, Bette, T.** [US/US]; 330 Brookline Avenue, Los Alamos, NM (US). **FISCHER, William, M.** [US/US]; 330 Brookline Avenue, Los Alamos, NM (US).

(74) Agents: **BIEKER-BRADY, Kristina** et al.; Clark & Elbing LLP, 101 Federal Street, Boston, MA 02110 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

WO 2010/059732 A1

(54) **Title**: ANTIVIRAL VACCINES WITH IMPROVED CELLULAR IMMUNOGENICITY

(57) **Abstract**: The invention provides compositions, methods, and kits for the treatment or prevention of viral infections. The polyvalent (*e.g.*, 2-valent) vaccines described herein incorporate computationally-optimized viral polypeptides that can increase the diversity or breadth and depth of cellular immune response in vaccinated subjects.

ANTIVIRAL VACCINES WITH IMPROVED CELLULAR IMMUNOGENICITY

STATEMENT OF FEDERALLY FUNDED RESEARCH

This research has been sponsored in part by NIH grant numbers U19-AI066305 and U19-AI078526. The government has certain rights to the invention.

FIELD OF THE INVENTION

The invention provides compositions, methods, and kits for the treatment or prevention of viral infections. The polyvalent (e.g., 2-valent) vaccines described herein incorporate computationally-optimized viral polypeptides that can increase the diversity or breadth and depth of cellular immune response in vaccinated subjects.

BACKGROUND OF THE INVENTION

Vaccines that elicit cellular immune responses against viruses must reflect global viral diversity in order to effectively treat or prevent viral infection. For example, the initiation of intense and diverse HIV-1-specific T cell responses is likely crucial for an effective HIV-1 vaccine. Cytotoxic T lymphocyte (CTL) responses are correlated with slow disease progression in humans, and the importance of CTL responses in non-human primate vaccination models is well established. While the highly variable Envelope (Env) is the primary target for neutralizing antibodies against HIV, and vaccine antigens will also need to be tailored to elicit these antibody responses, T cell vaccine components can target more conserved proteins to trigger responses that are more likely to cross-react. But even the most conserved HIV-1 proteins are diverse enough that variation will be an issue. Artificial central-sequence vaccine approaches, such as consensus and ancestral HIV-1 sequences, essentially "split the differences" between strains, can stimulate responses with enhanced cross-reactivity compared to natural strain vaccines. Consensus antigens represent synthetic antigen sequences that are the single best "average" of all circulating strains. While these antigens can elicit directed cellular immune responses, the breadth and intensity of these responses are not substantially improved over previous vaccine strategies.

The development of next-generation vaccines to treat or prevent viral infection must elicit an increased breadth of cellular immunity in order to allow for successful vaccination outcomes. The need for such vaccines is particularly urgent for the treatment or prevention of HIV-1.

SUMMARY OF THE INVENTION

In a first aspect, the invention features a vaccine for treating or reducing the risk of a viral infection in a mammal, such as a human, that includes at least two distinct optimized viral polypeptides (e.g., 2, 3, 4, 5, or more distinct optimized viral polypeptides), wherein the optimized viral polypeptides correspond to the same viral gene product. In one embodiment, the viral infection is caused by a retrovirus, reovirus, picornavirus, togavirus, orthomyxovirus, paramyxovirus, calicivirus, arenavirus, flavivirus, filovirus, bunyavirus, coronavirus, astrovirus, adenovirus, papillomavirus, parvovirus, herpesvirus, hepadnavirus, poxvirus, or polyomavirus. In other embodiments, the retrovirus is human immunodeficiency virus type 1 (HIV-1), and the viral gene products include Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, or Vpu. In a further embodiment, the vaccine includes no more than two optimized viral polypeptides corresponding to one of the Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, or Vpu viral gene products. In another embodiment, the vaccine does not include optimized viral polypeptides corresponding to Gag and Nef. In yet another embodiment, the vaccine includes at least two distinct optimized viral polypeptides (e.g., 2, 3, 4, 5, or more distinct optimized viral polypeptides) for a first viral gene product selected from Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, and Vpu and one or more distinct optimized viral polypeptides (e.g., 2, 3, 4, 5, or more distinct optimized viral polypeptides) for a second viral gene product different from the first viral gene product selected from Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, and Vpu.

In a second aspect, the invention features a vaccine for treating or reducing the risk of human immunodeficiency virus type 1 (HIV-1) infection in a mammal, such as a human, that includes an optimized viral polypeptide that has at least seven contiguous amino acids (e.g., at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 50, 100, 150, 175, 200, 250, 300, 350, 400, 450, 500 or more contiguous amino acids in length) having at least 85% amino acid sequence identity to any one of the sequences set forth in SEQ ID NOS:1-29. In one embodiment, the

optimized viral polypeptide has at least seven contiguous amino acids (e.g., at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 50, 100, 150, 175, 200, 250, 300, 350, 400, 450, 500 or more contiguous amino acids in length) having amino acid sequence identity to any one of the sequences set forth in SEQ ID NOS:1-29. In another embodiment, the optimized viral polypeptide has the amino acid sequence of any one of the sequences set forth in SEQ NOS:1-29. In a further embodiment, the vaccine includes at least two optimized viral polypeptides selected from any one or more of groups a)-k): a) SEQ ID NOS:1 and 2; b) SEQ ID NOS:3, 4, and 5; c) SEQ ID NOS:6 and 7; d) SEQ ID NOS:8-12; e) SEQ ID NOS:13, 14, and 15; f) SEQ ID NOS:16, 17, and 18; g) SEQ ID NOS:19 and 20; h) SEQ ID NOS:21, 22, and 23; i) SEQ ID NOS:24 and 25; j) SEQ ID NOS:26 and 27; k) and SEQ ID NOS:21-22. In another embodiment, the vaccine can include a pair of optimized viral polypeptides selected from any one of groups a)-k) above and one or more different optimized viral polypeptides from the same or a different group a)-k). In other embodiments, the vaccine can include at least three or four or more optimized viral polypeptides from one or more of groups a)-k).

In a third aspect, the invention features a vaccine for treating or reducing the risk of a viral infection in a mammal, such as a human, that includes at least two pairs of distinct optimized viral polypeptides, wherein each pair of optimized viral polypeptides corresponds to the same viral gene product, and wherein no more than two optimized viral polypeptides incorporated in the vaccine correspond to the same viral gene product. In one embodiment, the vaccine includes at least three pairs of distinct optimized viral polypeptides. In another embodiment, the vaccine includes at least four pairs of distinct optimized viral polypeptides. In one embodiment, the viral infection is caused by a retrovirus, reovirus, picornavirus, togavirus, orthomyxovirus, paramyxovirus, calicivirus, arenavirus, flavivirus, filovirus, bunyavirus, coronavirus, astrovirus, adenovirus, papillomavirus, parvovirus, herpesvirus, hepadnavirus, poxvirus, or polyomavirus. In other embodiments, the retrovirus is human immunodeficiency virus type 1 (HIV-1), and the viral gene products include Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, or Vpu. In a further embodiment, the vaccine includes no more than two optimized viral polypeptides corresponding to one of the Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, or Vpu viral gene products. In another embodiment, the vaccine does not include optimized viral polypeptides corresponding to Gag and Nef. In a further embodiment, the vaccine

includes at least three pairs of distinct optimized viral polypeptides corresponding to any three of the Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, or Vpu viral gene products. In another embodiment, the vaccine includes at least four pairs of distinct optimized viral polypeptides corresponding to any four of the Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, or Vpu viral gene products.

In one embodiment of any of the first three aspects of the invention, the vaccine elicits a cellular immune response against a viral gene product. In another embodiment, the vaccine elicits a cellular immune response against HIV-1. In a further embodiment, the nucleotide sequence of at least one distinct optimized viral polypeptide is encoded by a nucleic acid or vector. In one embodiment, the vector is a recombinant adenovirus, such as adenovirus serotype 26 (Ad26), adenovirus serotype 34 (Ad34), adenovirus serotype 35 (Ad35), adenovirus serotype 48 (Ad48), or adenovirus serotype 5 HVR48 (Ad5HVR48). In a further embodiment, the vaccine is in combination with a pharmaceutically acceptable carrier, excipient, or diluent.

In a fourth aspect, the invention features a nucleic acid that includes the nucleotide sequence of an optimized viral polypeptide that has at least seven contiguous amino acids (e.g., at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 50, 100, 150, 175, 200, 250, 300, 350, 400, 450, 500 or more contiguous amino acids in length) having at least 85% amino acid sequence identity to any one of the amino acid sequences set forth in SEQ ID NOS:1-29. In one embodiment, the optimized viral polypeptide has at least seven contiguous amino acids (e.g., at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 50, 100, 150, 175, 200, 250, 300, 350, 400, 450, 500 or more contiguous amino acids in length) having sequence identity to any one of the amino acid sequences set forth in SEQ ID NOS:1-29. In another embodiment, the optimized viral polypeptide has any one of the amino acid sequences set forth in SEQ ID NOS:1-29. In a further embodiment, the nucleic acid includes a vector. In one embodiment, the vector is a recombinant adenovirus, such as adenovirus serotype 26 (Ad26), adenovirus serotype 34 (Ad34), adenovirus serotype 35 (Ad35), adenovirus serotype 48 (Ad48), or adenovirus serotype 5 HVR48 (Ad5HVR48).

In a fifth aspect, the invention features an optimized viral polypeptide that has at least seven contiguous amino acids (e.g., at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 50, 100, 150, 175, 200, 250, 300,

350, 400, 450, 500 or more contiguous amino acids in length) having at least 85% amino acid sequence identity to any one of the amino acid sequences set forth in SEQ ID NOS:1-29. In one embodiment, the optimized viral polypeptide has at least seven contiguous amino acids (e.g., at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 50, 100, 150, 175, 200, 250, 300, 350, 400, 450, 500 or more contiguous amino acids in length) having sequence identity to any one of the amino acid sequences set forth in SEQ ID NOS:1-29. In another embodiment, the optimized viral polypeptide has any one of the amino acid sequences set forth in SEQ ID NOS:1-29.

In a sixth aspect, the invention features a method for treating or reducing the risk of a viral infection in a mammal, such as a human, by administering a vaccine or nucleic acid of the invention. In one embodiment, the viral infection is caused by a retrovirus, reovirus, picornavirus, togavirus, orthomyxovirus, paramyxovirus, calicivirus, arenavirus, flavivirus, filovirus, bunyavirus, coronavirus, astrovirus, adenovirus, papillomavirus, parvovirus, herpesvirus, hepadnavirus, poxvirus, or polyomavirus. In further embodiments, the retrovirus is human immunodeficiency virus type 1 (HIV-1), and the viral gene products include Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, or Vpu. In one embodiment, the vaccine or nucleic acid elicits a cellular immune response against a viral gene product.

In a seventh aspect, the invention features a method of manufacturing a vaccine for treating or reducing the risk of a viral infection in a mammal, such as a human, by synthesizing a vaccine of the invention.

In an eighth aspect, the invention features a method of manufacturing a vaccine for treating or reducing the risk of a viral infection in a mammal, such as a human, by contacting a nucleic acid of the invention with a cell and isolating a optimized viral polypeptide.

In one embodiment of the seventh or eighth aspects of the invention, the optimized viral polypeptide elicits a cellular immune response when administered to a mammal. The cellular immune response can be against a viral gene product. In another embodiment, the viral infection is caused by a retrovirus, reovirus, picornavirus, togavirus, orthomyxovirus, paramyxovirus, calicivirus, arenavirus, flavivirus, filovirus, bunyavirus, coronavirus, astrovirus, adenovirus, papillomavirus, parvovirus, herpesvirus, hepadnavirus, poxvirus, or polyomavirus. In further

embodiments, the retrovirus is human immunodeficiency virus type 1 (HIV-1), and the viral gene products include Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, or Vpu.

In a ninth aspect, the invention features a kit that includes a vaccine of the invention, a pharmaceutically acceptable carrier, excipient, or diluent, and instructions for the use thereof. In one embodiment, the kit also includes an adjuvant.

In a final aspect, the invention features a kit that includes a nucleic acid of the invention, a pharmaceutically acceptable carrier, excipient, or diluent, and instructions for the use thereof. In one embodiment, the kit also includes an adjuvant.

In an embodiment of all aspects of the invention, the optimized viral polypeptide is encoded by a nucleic acid sequence that is optimized for expression in humans (e.g., any one of SEQ ID NOS:5, 10, 11, 12, 15, 18, and 23).

Definitions

By "optimized viral polypeptide" or "computationally-optimized viral polypeptide" is meant an immunogenic polypeptide that is not a naturally-occurring viral peptide, polypeptide, or protein. Optimized viral polypeptide sequences are initially generated by modifying the amino acid sequence of one or more naturally-occurring viral gene products (e.g., peptides, polypeptides, and proteins) to increase the breadth, intensity, depth, or longevity of the antiviral immune response (e.g., cellular or humoral immune responses) generated upon immunization (e.g., when incorporated into a vaccine of the invention) of a mammal (e.g., a human). Thus, the optimized viral polypeptide may correspond to a "parent" viral gene sequence; alternatively, the optimized viral polypeptide may not correspond to a specific "parent" viral gene sequence but may correspond to analogous sequences from various strains or quasispecies of a virus. Modifications to the viral gene sequence that can be included in an optimized viral polypeptide include amino acid additions, substitutions, and deletions. In one embodiment of the invention, the optimized viral polypeptide is the composite or merged amino acid sequence of two or more naturally-occurring viral gene products (e.g., natural or clinical viral isolates) in which each potential epitope (e.g., each contiguous or overlapping amino acid sequence of 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more amino acids in length) is analyzed and modified to improve the immunogenicity of the resulting optimized viral polypeptide. Optimized viral

polypeptides that correspond to different viral gene products can also be fused to facilitate incorporation in a vaccine of the invention. Methods of generating an optimized viral polypeptides are described in, e.g., Fisher et al. "Polyvalent Vaccine for Optimal Coverage of Potential T-Cell Epitopes in Global HIV-1 Variants," *Nat. Med.* 13(1):100-106 (2007) and International Patent Application Publication WO 2007/024941, herein incorporated by reference. Once the optimized viral polypeptide sequence is generated, the corresponding polypeptide can be produced or administered by standard techniques (e.g., recombinant viral vectors, such as the adenoviral vectors disclosed in International Patent Application Publications WO 2006/040330 and WO 2007/104792, herein incorporated by reference).

By "pharmaceutically acceptable carrier" is meant a carrier which is physiologically acceptable to the treated mammal while retaining the therapeutic properties of the compound with which it is administered. One exemplary pharmaceutically acceptable carrier is physiological saline. Other physiologically acceptable carriers and their formulations are known to one skilled in the art and described, e.g., in *Remington's Pharmaceutical Sciences* (18th edition, ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA), incorporated herein by reference.

By "vector" is meant a DNA construct that contains a promoter operably linked to a downstream gene or coding region (e.g., a cDNA or genomic DNA fragment, which encodes a polypeptide or polypeptide fragment). Introduction of the vector into a recipient cell (e.g., a prokaryotic or eukaryotic cell, e.g., a bacterium, yeast, insect cell, or mammalian cell, depending upon the promoter within the expression vector) or organism (including, e.g., a human) allows the cell to express mRNA encoded by the vector, which is then translated into the encoded optimized viral polypeptide of the invention. Vectors for *in vitro* transcription/translation are also well known in the art and are described further herein. A vector may be a genetically engineered plasmid, virus, or artificial chromosome derived from, e.g., a bacteriophage, adenovirus, retrovirus, poxvirus, or herpesvirus.

By "viral gene product" is meant any naturally-occurring viral peptide, polypeptide, or protein, or fragment thereof. In one embodiment of the invention, the viral gene product is derived from the human immunodeficiency virus type 1 (HIV-1). HIV-1 viral gene products include the Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, and Vpu polypeptides.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a chart that illustrates the expanded breadth of computationally-optimized HIV-1 Gag, Pol, and Env viral polypeptides against global potential T-cell epitopes (PTE) peptides in Rhesus macaques. Animals immunized with the optimized viral polypeptides (blue) reacted with the greatest number of recall peptide pools.

Figure 2 is a chart that shows that computationally-modified HIV-1 Gag, Pol, and Env viral polypeptides expand the breadth of epitope-specific cellular immune response.

Figure 3 illustrates the breadth of cellular immune responses detected in Rhesus macaques following immunization with HIV-1 viral gene products Gag, Pol, and Env derived from the computationally-modified viral polypeptides of the invention, as well as animals immunized with consensus HIV-1 antigens or HIV-1 clade C isolate antigens. Animals immunized with the optimized viral polypeptides (blue) reacted with the greatest number of recall peptide pools. Since the animals are outbred, the pools differ from animal to animal. Gag, Pol, and Env each elicit many cellular immune responses and can have shared patterns of reactivity.

Figures 4A-4C are graphs showing the potential epitopes shared between the different vaccines tested (2 valent mosaic (Mos2), M consensus (Mcon), and optimized clade C (OptC)) by viral polypeptide (Pol (Fig. 4A), Gag (Fig. 4B), and Env (Fig. 4C)). Figures 4A-C show the relative coverage of the current HIV database full length genome set and the PTE peptides by the different vaccine candidates.

Figure 5 is a graph showing that the number of PTE peptide responses (where each response is considered an independent event regardless of overlap) to the 2 valent mosaic (Mos2) vaccine is greater than the number of responses to the M group consensus (Mcon) vaccine and the natural viral strain vaccine (optimized clade C (C Natural (optimal)), which has been selected to give optimal coverage of the M group collection (OptC) vaccine antigens. Figure 5 shows the number of PTE peptide responses per animal by protein, CD8+ T cell, and CD4+ T cell. Statistically, Mos2 > Mcon ~ OptC (Mcon shows a trend for more response than OptC). The Wilcoxon p-value for Mos2 compared to Mcon: p-value = 0.001058.

Figure 6 is a chart showing the number of PTE peptides that trigger T cell responses. A median number of 16 (range; 12-29) PTE peptides of the 2 valent mosaic (Mos2) vaccine trigger a response in CD8+ T cells, while only a median

number of 6 (range: 0-7) Mcon peptides and only a median number of 3 peptides (range: 0-3) of OptC peptides trigger a response in CD8+ T cells. A median number of 4 (range; 2-6) PTE peptides of the 2 valent mosaic (Mos2) vaccine trigger a response in CD4+ T cells, while only a median number of 1 (range: 0-2) Mcon peptides and only a median number of 0.5 peptides (range: 0-2) of OptC peptides trigger a response in CD4+ T cells. Thus, the trend for responses is Mos2 > Mcon > OptC.

Figure 7 is a schematic summarizing the mapping of all CD8+ T cell Gag PTE peptides that are recognized by T cells from each of the animals studied (see Example 3 below). The animal number, peptide pool and peptide number label the boundaries of each reactive peptide. The symbol signifies the group: *, Mos2; ¥, ConM; ÷, OptC. Gag is included here as an example. There tends to be clustering of CD8 responses even though the animals are outbred. Mosaics have potential advantages over the monovalent vaccines. Mosaics have a better chance of stimulating a response that reacts with more common variants. Mosaics also stimulate multiple responses to the different forms that are present in the cocktail. Thus, mosaics have the potential to block common escape routes. In our study, the mosaic vaccine tended to stimulate T cell responses that recognized more overlapping peptides. There are many hotspots of localization of reactive peptides. PTE peptides are designed to maximize the potential epitope (or 9-mer for a 9 amino acid contiguous stretch) coverage of the HIV-1 M group in the peptide reagents used to assess vaccines. Inevitably, there is a lot of overlap in PTE peptides, but because of the algorithm, overlap is usually an overlap with some variation. Figure 7 discloses SEQ ID NO: 42.

Figure 8 is a schematic summarizing the mapping of CD4+ T cell Gag PTE peptides that are recognized by T cells from each of the animals studied. Figure 8 discloses SEQ ID NO: 43.

Figure 9 is a chart illustrating typical patterns of PTE responses to the ConM vaccine or to the optimal natural vaccine, aligning peptides that elicit a response with the relevant region of the vaccine. Good matches with solid stretches of identity between vaccine and target PTE peptide are necessary to achieve a reaction to these vaccines. Figure 9 discloses SEQ ID NOS 44-57, respectively, in order of appearance.

Figure 10 is a chart illustrating that mosaic vaccines generated many responses that recognized multiple variant overlapping peptides with no apparent antigenic competition and with broad local responses. In particular, four variable PTE peptides were recognized. Moreover, in the region of overlap both mosaic forms were recognized, as well a combination of the two. Finally, a new form (S) was recognized. Figure 10 discloses SEQ ID NOS 58-63, respectively, in order of appearance.

Figure 11 is a chart illustrating a typical pattern of CD8+ PTE peptide responses in mosaic vaccinated animal (361-07). 22 PTE peptides were tested and 8 CD8 responsive regions were identified; 5 regions included variable peptides that match amino acids in one or the other of the mosaics. 5 CD4 responsive regions were identified. Thus, T cell responses to Mosaics see more variable peptides in a given region. This seemed to be true of CD8 T cell responses in particular. This could be the result of triggering multiple T cell clones that recognize variants of epitopes, and these may block fit escape routes. Not only are there more responses, they are deeper and cover more variants. Figure 11 discloses CD8 responses as SEQ ID NOS 64-101, respectively, in order of appearance. CD4 responses disclosed as SEQ ID NOS 102-117, respectively, in order of appearance.

Figure 12 is a graph showing the number of overlapping variable PTE peptides that span regions targeted by vaccine elicited T cells.

Figure 13 is graph showing that the 2 mosaic antigen vaccine yields more T cell responses, relative to the Mcon and OptC vaccines, to regions that contain one or more overlapping PTE peptides. Figure 13 is similar to Figure 5, monkeys shown in the same order from right to left, but with the scale changed to reflect number of responses to regions that contain one or more overlapping PTE peptides rather than single peptides.

Figure 14 is a chart showing the number of T cell responses in animals following administration of 2 valent mosaic (Mos2), Mcon, and OptC vaccines. The 2 valent mosaic (Mos2) vaccine triggers a median number of 8 responses in CD8+ T cells, while only a median number of 3 (range: 0-6) and 1.5 peptides (range: 0-5) CD8+ T cell responses are triggered by Mcon and OptC vaccines, respectively. The 2 valent mosaic (Mos2) vaccine triggers a median number of 3 (range; 2-5) responses in CD4+ T cells, while only a median number of 1 (range: 0-2) and 0.5 (range: 0-2)

CD4+ T cell response are triggered by Mcon and OptC vaccines, respectively. Thus, the trend for responses is Mos2 > Mcon > OptC.

Figure 15 is a graph showing that the mosaic vaccines can elicit more responses that cross-react with C clade natural proteins than can a C clade natural vaccine: GAG pooled peptides representing 5 proteins. Animals vaccinated with the M group consensus or the optimal coverage C clade natural protein had 0-2 responses to the peptides derived from these proteins, while the Mosaic vaccinated animals could respond to 1-5 peptide pools. The Mosaic vaccine elicits more responses to each of the proteins tested than either M con or the optimal C. T cell responses elicited by mosaic vaccines also recognized more pooled peptide sets spanning *actual* Gag proteins. 10-12 Subpools = 10x15mer peptides (except 96ZM Gag, which is 5x20mer peptides).

Figure 16 is a graph showing that the mosaic design is robust to changes in viral polypeptides over time (e.g., Gag M).

Figure 17 is a graph showing that coverage using 9-mer optimization is robust over near (e.g., 8-12 mers) optimization lengths (Gag is shown).

Figure 18 is a graph showing that an increase in the number of variants increases coverage, but has only diminishing returns (Gag is shown).

Figures 19A-19B are graphs showing the breadth and magnitude of epitope-specific T lymphocyte responses to PTE peptides. Figure 19A is a graph showing the numbers of epitope-specific CD4+ (top) and CD8+ (bottom) T lymphocyte responses to individual PTE

peptides following a single immunization of rAd26 vectors expressing mosaic (blue), M consensus (green), clade B + clade C (purple), or optimal natural clade C (red) HIV-1 Gag, Pol, and Env antigens. Individual monkeys are depicted on the x-axis. The different shades of each color reflect responses to the different antigens (Gag, Pol, Env). Figure 19B is a graph showing the numbers of CD4+ (top) and CD8+ (bottom) T lymphocyte response regions.

Figures 20A-20C show a schematic showing CD8+ T lymphocyte responses to PTE peptides at week 4 following immunization mapped on HIV-1 Gag (Figure 20A) (SEQ ID NO: 118), Pol (Figure 20B) (SEQ ID NO: 119), and Env (Figure 20C) (SEQ ID NO: 120) protein sequences. Colors denote monkeys that received the mosaic (blue), M consensus (green), clade B + clade C (purple), or optimal natural clade C (red) HIV-1 Gag, Pol, and Env antigens. For each epitope, the monkey number, antigen (G, Gag; P, Pol; E, Env), subpool number, and individual PTE peptide number are indicated.

Figures 21A-21C show a schematic showing CD4+ T lymphocyte responses to PTE peptides at week 4 following immunization mapped on HIV-1 Gag (Figure 21A) (SEQ ID NO: 121), Pol (Figure 21B) (SEQ ID NO: 122), and Env (Figure 21C) (SEQ ID NO: 123) protein sequences. Colors denote monkeys that received the mosaic (blue), M consensus (green), clade B + clade C (purple), or optimal natural clade C (red) HIV-1 Gag, Pol, and Env antigens. For each epitope, the monkey number, antigen (G, Gag; P, Pol; E, Env), subpool number, and individual PTE peptide number are indicated.

Figure 22 is a schematic showing the alignment of vaccine sequences with reactive PTE peptides in all monkeys at week 4 following immunization with rAd26 vectors expressing mosaic, M consensus, clade B + clade C, or optimal natural clade C HIV-1 Gag, Pol, and Env antigens. For each monkey, vaccine sequences are shown on the top, and reactive PTE peptides are shown beneath the vaccine sequences denoted by the antigen (G, Gag; P, Pol; E, Env) and PTE peptide numbers. The minimal overlap region is shown in bold. Sequence polymorphisms between the two mosaic or the two clade B + clade C antigens are shown in blue. Differences between the vaccine sequences and the reactive PTE peptides are shown in red. Figure 22 discloses SEQ ID NOS 124-640, respectively, in order of appearance.

Minimal regions within the peptides that are likely to contain the immune response epitope, based on overlap between reactive peptides when it occurs, are in bold in the vaccines. If there is no overlapping peptide, we assume the epitope can be anywhere in the peptide, so the whole region is bold. We cannot differentiate between different T cell responses targeting epitopes with different boundaries within a peptide, or more promiscuous clonal T cell responses that can tolerate variation when variants are present; either scenario could be advantageous in a vaccine immune response. The number of targeted regions
corresponds to the minimum number of T cell responses required to account for the data.

Amino acids where the vaccine and the peptides don't match are written in red; if they fall within the region likely to carry the epitope, they are bold red. Amino acid differences outside of the overlapping regions when multiple peptides overlap are marked in red, but not bold.

The vaccines are always at the top. The letter for each protein (Gag is G, Pol is P, Envelope is E) and the peptide number are used to label for each reactive PTE peptide. The protein and HXB2 numbers follow each peptide.

For the mosaic and clade B+C vaccines, there are 2 antigens each and both are included in the alignment; amino acid differences in the vaccines are noted in blue, and if

the reactive peptide carries the variant amino acid in the second mosaic, it is also in blue. In each of the positions where the two vaccine antigens differ, the reactive peptides are also marked in bold to indicate the positions where including two variants may have impacted the vaccine immune response and allowed greater breadth and depth.

For example, the first vaccine summarized is the clade B+C vaccine, and animal 287-95 is the first animal for which responses are listed. There were 3 CD8 responses to PTE peptides, 1 to CD4. Two of the CD8 peptides show substantial overlap, E26 and E282, so both may be targets for the same CTL response; thus we also note there are only 2 CD8 responsive regions, and 1 CD4 responsive region. For each responsive region, we write out the number of overlapping peptides per region (e.g., CD8: 1 2 CD4: 1) to assess depth of responses; the two is red to indicate that the region of overlap is variable in the reactive peptides. If the vaccine differs, like the D/E in the second reactive region, it is marked in blue. Only the region of overlap is bold. The H in E282 was not found in either vaccine so it is marked with red; it is within the region of overlap so it is bold. Each reactive peptide has its protein and corresponding HXB2 numbering noted on the right.

Figures 23A-23C are graphs showing the magnitude of all Gag-, Pol-, and Env-specific CD8+ (Figure 23A and 23B) and CD4+ (Figure 23C) T lymphocyte responses arranged from lowest to highest.

Figures 24A-C show the depth of epitope-specific T lymphocyte responses to PTE peptides. Figure 24A is a schematic showing an example of mapped T lymphocyte responses in monkey 366 that received the optimal natural clade C antigens. Figure 24B is a schematic showing an example of mapped T lymphocyte responses in monkey 361 that received the 2-valent mosaic antigens. In Figures 24A and 24B, vaccine sequences are shown on the top (OptC; Mos1, Mos2), and reactive PTE peptides are shown beneath the vaccine sequences denoted by the antigen (G, Gag; P, Pol; E, Env) and the PTE peptide numbers. The minimal overlap region is shown in bold. Sequence polymorphisms between the two mosaic antigens are shown in blue. Differences between the vaccine sequences and the reactive PTE peptides are shown in red. Complete alignments of all positive peptides organized by response regions are shown in Fig. 22. Figure 24C is a graph showing the depth of CD4+ (top) and CD8+ (bottom) T lymphocyte responses following immunization with rAd26 vectors expressing mosaic, M consensus, clade B + clade C, or optimal natural clade C antigens. Individual monkeys are depicted on the x-axis. One response variant (light shade) or >1 response variants (dark shade) are shown for each epitopic region. Figure 24A discloses SEQ ID NOS 641-

650, respectively, in order of appearance. Figure 24B discloses SEQ ID NOS 651-685, respectively, in order of appearance.

Figure 25 is a graph showing the breadth of epitope-specific T lymphocyte responses to HIV-1 Gag peptides from clades A, B, and C. Breadth of cellular immune responses was assessed utilizing subpools of overlapping peptides from the following strains of HIV-1 Gag: clade C DU422, clade C ZM651, consensus C, consensus A, and consensus B. Numbers of positive subpools are shown following a single immunization of rAd26 vectors expressing mosaic (blue), M consensus (green), clade B + clade C (purple), or optimal natural clade C (red) HIV-1 Gag, Pol, and Env antigens. Individual monkeys are depicted on the x-axis.

Figure 26A-D are graphs showing the cellular and humoral immune responses following the boost immunization. Shown are the magnitude (Fig. 26A) and breadth (Fig. 26B) of individual T lymphocyte responses at week 4 post-prime (left side of each panel) and at week 44 post-boost (right side of each panel) for each monkey. Monkeys were primed at week 0 with rAd26 vectors and were boosted at week 40 with rAd5HVR48 vectors expressing mosaic, M consensus, or optimal natural clade C HIV-1 Gag, Pol, and Env antigens. Individual monkeys are depicted on the x axis. In Fig. 26A, red denotes CD8+ T lymphocyte responses, blue denotes CD4+ T lymphocyte responses, lines depict responses observed at both timepoints, and dots depict responses observed at only one timepoint. In Fig. 26B, different shades of each color reflect responses to the different antigens (Gag, Pol, Env). Figure 26C is a graph showing the Env-specific ELISA endpoint titers at weeks 0, 4, and 44. Figure 26D is a graph showing the neutralizing antibody (NAb) titers to the tier 1 clade A (DJ263.8), clade B (SF162.LS), and clade C (MW965.26) viruses at week 44. NAb titers to murine leukemia virus as a negative control were <20 for all samples.

Figure 27 is a graph showing the theoretical coverage of PTE peptides by the various vaccine antigens. Percentage of 9 amino acid PTE peptides that are covered by the mosaic (blue), M consensus (green), clade B + clade C (purple), or optimal natural clade C (red) HIV-1 Gag, Pol, and Env antigens are shown.

DETAILED DESCRIPTION OF THE INVENTION

The invention features optimized viral polypeptides that are computationally derived from naturally-occurring viral gene products. The optimized viral

polypeptides of the invention allow for an increased breadth and depth of virus-specific immunity (e.g., cellular immunity, such as T cell-based immune responses) following immunization of a subject (e.g., a human) with one or more optimized viral polypeptides of the invention or vaccines (e.g., a vector) that incorporate one or more optimized viral polypeptides of the invention. The invention provides vaccines that can be administered to a subject (e.g., a human) infected with or at risk of becoming infected with a viral infection. The vaccines of the invention incorporate at least two distinct optimized viral polypeptides for each corresponding viral gene product represented. The incorporation of at least two distinct optimized viral polypeptides allows for increased coverage and representation of immunogenic epitopes in the vaccine, which the inventors have found results in an increase in the total number of virus-specific immune responses following vaccination of a subject. The present invention also provides methods of administering and manufacturing vaccines, vectors, and optimized viral polypeptides that to a subject (e.g., a human). The compositions, methods, and kits described herein can substantially increase the diversity, breadth, and/or depth of the virus-specific cellular immune responses by providing at least two distinct optimized viral polypeptides.

Optimized Viral Polypeptides of the Invention

The present invention provides for polyvalent (e.g., 2-valent) vaccines that incorporate computationally-optimized viral polypeptides that correspond to and are derived from viral gene products that naturally circulate. Polyvalent mosaic proteins are assembled from natural sequences by *in silico* recombination and optimized to provide maximal coverage of potential T cell epitopes (PTEs) for a given valency. Mosaic antigens are full-length proteins that are designed to preserve natural antigen expression and processing.

The inventors have discovered that immunization with two distinct optimized viral polypeptides corresponding to and derived from a single viral gene product (i.e., a 2-valent vaccine) elicits a substantially higher number of cellular immune responses (e.g., T cell responses) than conventional monovalent or polyvalent vaccines that incorporate naturally-occurring polypeptides derived from the same viral gene product (e.g., sequences based on clinical isolates), or a consensus sequence of such naturally-occurring polypeptides derived from the same viral gene product.

Accordingly, a vaccine that incorporates computationally-optimized viral polypeptides, the sequences of which provide maximum coverage of non-rare short stretches of circulating viral sequences, can increase the breadth and depth of the immune response.

A genetic algorithm is used to create sets of optimized viral polypeptides as "mosaic" blends of fragments of an arbitrary set of naturally-occurring viral gene product sequences provided as inputs. This genetic algorithm strategy uses unaligned protein sequences from a general viral population as an input data set, and thus has the virtue of being "alignment independent." It creates artificial optimized viral polypeptides that resemble viral proteins found in nature, but are not naturally-occurring. The genetic algorithm can be adjusted to optimize viral polypeptides of different lengths, depending on the intended target or desired immune response. As most T cell epitopes are nine amino acids in length, the genetic algorithm utilized to design the optimized viral polypeptides of the invention was based on optimizing each consecutive 9-mer amino acid sequence of a given viral gene product (e.g., HIV-1 Gag). In accordance with this approach, 9-mers (for example) that do not exist in nature or that are very rare can be excluded - this is an improvement relative to consensus sequence-based vaccine strategies since the latter can contain some 9-mers (for example) that occur rarely or not at all in nature. The definition of fitness used for the genetic algorithm is that the most "fit" polyvalent cocktail is the combination of input viral sequences that gives the best coverage (highest fraction of perfect matches) of all of the 9 mers in the population and is subject to the constraint that no 9 mer is absent or rare in the population. The genetic algorithm used to generate the optimized viral polypeptides of the invention is further described in International Patent Application Publication WO 2007/024941, herein incorporated by reference.

In one embodiment, the invention provides polyvalent (e.g., 2-valent) HIV-1 vaccines that incorporate single optimized HIV-1 polypeptides (e.g., the polypeptides set forth in SEQ ID NOS:1-29). In another embodiment, the invention features a polyvalent vaccine that incorporates two or more optimized HIV-1 polypeptides. In each case, the optimized HIV-1 polypeptides are based on all HIV-1 variants in global circulation, known as the HIV-1 Main (M) group. The inventors have generated a set of optimized HIV-1 polypeptides (SEQ ID NOS:1-29) that augment the breadth and depth of cellular immunity based on group M mosaic genes that utilize only two

variants per gene (e.g., two polypeptide sequences each for Gag, Pol, Env, Nef, Tat, Rev, Vif, Vpr, and Vpu). We have obtained the novel and surprising result in Rhesus macaques that the use of these optimized HIV-1 polypeptides in a polyvalent (e.g., 2-valent) HIV-1 group M vaccine elicits a significantly greater breadth and depth of HIV-1-specific cellular immune responses when compared with two other leading vaccine antigen strategies (M consensus antigens and optimal natural clade C antigens).

The invention provides for the fusion of optimized viral polypeptides that correspond to different viral gene products. The genetic algorithm described above can be used to generate fused polypeptides for use in a vaccine of the invention. For example, the optimized HIV-1 polypeptide fusions of Gag/Nef (SEQ ID NOS:19-20), Gag/Pol (SEQ ID NOS:21-27), and Gag/Pol/Nef (SEQ ID NOS:28-29) can be incorporated into a vector of the invention for administration to a subject (e.g., a human) infected with or at risk of being infected with HIV-1. The vaccines of the invention (whether in polypeptide or nucleic acid form) can also include one or more of the non-"mosaic" polypeptides (or sequences encoding them, respectively), such as, e.g., the optimal clade C sequences (SEQ ID NOS: 30-36) or the consensus sequences (SEQ ID NOS: 37-39).

The optimized viral polypeptides disclosed in this invention can be prepared conventionally by chemical synthesis techniques, such as described by Merrifield, *J. Amer. Chem. Soc.* 85:2149 (1963) (see also, e.g., Stemmer et al., 164 *Gene* 49 (1995)). For example, the vaccines can be readily prepared using solid phase peptide synthesis (SPPS). Automated solid phase synthesis can be performed using any one of a number of well known, commercially available automated synthesizers, such as the Applied Biosystems ABI 433A peptide synthesizer. Alternatively, the optimized viral polypeptides of the invention can be recombinantly produced by transfecting or transducing a cell or organism with a nucleic acid or vector (e.g., a viral vector, such an adenovirus) that allows for the intracellular expression of the optimized viral polypeptide. Nucleic acids and vectors that encode the nucleotide sequence of optimized viral polypeptides of the invention can be synthesized by well-known recombinant DNA techniques, including those described herein.

Vaccines of the Invention

The invention also features vaccines that can be administered to a patient infected with or at risk of becoming infected with a virus (e.g., HIV-1). A vaccine of the invention contains at least one of the optimized viral polypeptides of the invention, as discussed herein. The vaccine of the invention can be a nucleic acid encoding the nucleotide sequence of two or more optimized viral polypeptides of the invention (e.g., the immunogenic component of a recombinant (e.g., subunit) or whole-organism (e.g., whole-virus) viral vector). Nucleic acids include vectors (e.g., viral vectors, such as adenoviruses) that incorporate the nucleotide sequence of two or more optimized viral polypeptides of the invention. The optimized viral polypeptides of the invention, as well as vaccines, nucleic acids, and vectors that incorporate optimized viral polypeptides, can be recombinantly expressed in a cell or organism, or can be directly administered to subject (e.g., a human) infected with, or at risk of becoming infected with, a virus.

Vectors of the Invention

The invention also features vectors encoding the nucleotide sequences (e.g., DNA or RNA) of one or more optimized viral polypeptides of the invention. The vector can be a carrier (e.g., a liposome), a plasmid, a cosmid, a yeast artificial chromosome, or a virus that includes a nucleotide sequence encoding one or more optimized viral polypeptides of the invention. The vector can include additional nucleic acid sequences from several sources.

Vectors encoding one or more optimized viral polypeptides of the invention can be constructed using any recombinant molecular biology technique known in the art. The vector, upon transfection or transduction of a target cell or organism, can be extrachromosomal or it can be integrated into the host cell chromosome. The nucleic acid component of a vector can be in single or multiple copy number per target cell, and can be linear, circular, or concatamerized.

Vectors of the invention can also include internal ribosome entry site (IRES) sequences to allow for the expression of multiple peptide or polypeptide chains from a single nucleic acid transcript. For example, a vector of the invention can encode one or more optimized viral polypeptides of the invention as well as another polypeptides (e.g., a detectable label, such as green fluorescent protein (GFP)).

Vectors of the invention further include gene expression elements that facilitate the expression of optimized viral polypeptides of the invention. Gene expression elements useful for the expression of an vector encoding an optimized viral polypeptide of the invention include, but are not limited to (a) regulatory sequences, such as viral transcription promoters and their enhancer elements, such as the SV40 early promoter, Rous sarcoma virus LTR, and Moloney murine leukemia virus LTR; (b) splice regions and polyadenylation sites such as those derived from the SV40 late region; and (c) polyadenylation sites such as in SV40. Also included are plasmid origins of replication, antibiotic resistance or selection genes, multiple cloning sites (e.g., restriction enzyme cleavage loci), and other viral gene sequences (e.g., sequences encoding viral structural, functional, or regulatory elements, such as the HIV long terminal repeat (LTR)).

Vectors of the invention can also include optimized viral polypeptides of the invention that have been optimized for expression in humans, such as, e.g., any one of SEQ ID NOS:11, 14-18, and 23.

Vectors of the invention can also be engineered to include a multiple cloning site (MCS) having the following enzyme cleavage sites: XbaI-EcoRI-Kozak-Start...Stop-BamHI-NheI; and the following sequence: TCTAGA GAATTC GCCACC [ATG gene TAA TGA] GGATCC GCTAGC. Vectors having this MCS can be used with optimized viral polypeptides having no internal XbaI, EcoRI, BamHI, NheI sites and no stretches of 6 or more C's or G's.

In Vivo Administration

The invention features methods for the *in vivo* administration of one or more vaccines of the invention (e.g., a vector encoding two or more optimized viral polypeptides of the invention) to a subject (e.g., a human) to facilitate the expression of two or more optimized viral polypeptides of the invention. Upon administering the vaccine to the subject, one or more optimized viral polypeptides of the invention will be expressed that can elicit protective or therapeutic immune responses (e.g., cellular or humoral immune responses) directed against the viral immunogens.

Several types of vectors can be employed to deliver a nucleotide sequence encoding one or more optimized viral polypeptides of the invention directly to a subject (e.g., a human). Vectors of the invention include viruses, naked DNA,

oligonucleotides, cationic lipids (e.g., liposomes), cationic polymers (e.g., polysomes), virosomes, and dendrimers. The present invention provides for the *ex vivo* transfection or transduction of cells (e.g., blood cells) followed by administration of these cells back into the donor subject to allow for the expression of optimized viral polypeptides of the invention that have immunogenic properties. Cells that can be isolated and transfected or transduced *ex vivo* according to the methods of invention include, but are not limited to, blood cells, skin cells, fibroblasts, endothelial cells, skeletal muscle cells, hepatocytes, prostate epithelial cells, and vascular endothelial cells. Stem cells are also appropriate cells for transduction or transfection with a vector of the invention. Totipotent, pluripotent, multipotent, or unipotent stem cells, including bone marrow progenitor cells and hematopoietic stem cells (HSC), can be isolated and transfected or transduced with an vector encoding one or more optimized viral polypeptides of the invention, and administered to a subject according to the methods of the invention.

The method of transfection or transduction used to express an optimized viral vector of the invention has a strong influence on the strength and longevity of protein expression in the transfected or transduced cell, and subsequently, in the subject receiving the cell. The present invention provides vectors that are temporal (e.g., adenoviral vectors) or long-lived (e.g., retroviral vectors) in nature. Regulatory sequences (e.g., promoters and enhancers) are known in the art that can be used to regulate protein expression. The type of cell being transfected or transduced also has a strong bearing on the strength and longevity of protein expression. For example, cell types with high rates of turnover can be expected to have shorter periods of protein expression.

***Ex Vivo* Transfection and Transduction**

The invention also features methods for the *ex vivo* transfection and transduction of cells (e.g., blood cells, such as lymphocytes), followed by administration of these cells to a subject (e.g., a human). In one embodiment, the cells are autologous to the treated subject. Cells can be transfected or transduced *ex vivo* with one or more vectors encoding the nucleotide sequence of one or more optimized viral polypeptides of the invention to allow for the temporal or permanent expression of the optimized viral polypeptides in the treated subject. Upon

administering these modified cells to the subject, one or more optimized viral vectors of the invention will be expressed that can elicit protective or therapeutic immune responses (e.g., cellular or humoral immune responses) directed against the viral immunogens.

Several types of vectors can be employed to deliver a nucleotide sequence encoding one or more optimized viral polypeptides of the invention to a cell (e.g., a blood cell, such as a lymphocyte). Vectors of the invention include viruses, naked DNA, oligonucleotides, cationic lipids (e.g., liposomes), cationic polymers (e.g., polysomes), virosomes, and dendrimers. The present invention provides for the *ex vivo* transfection or transduction of cells (e.g., blood cells) followed by administration of these cells back into the donor subject to allow for the expression of optimized viral polypeptides of the invention that have immunogenic properties. Cells that can be isolated and transfected or transduced *ex vivo* according to the methods of invention include, but are not limited to, blood cells, skin cells, fibroblasts, endothelial cells, skeletal muscle cells, hepatocytes, prostate epithelial cells, and vascular endothelial cells. Stem cells are also appropriate cells for transduction or transfection with a vector of the invention. Totipotent, pluripotent, multipotent, or unipotent stem cells, including bone marrow progenitor cells and hematopoietic stem cells (HSC), can be isolated and transfected or transduced with an vector encoding one or more optimized viral polypeptides of the invention, and administered to a subject according to the methods of the invention.

The method of transfection or transduction used to express an optimized viral vector of the invention has a strong influence on the strength and longevity of protein expression in the transfected or transduced cell, and subsequently, in the subject receiving the cell. The present invention provides vectors that are temporal (e.g., adenoviral vectors) or long-lived (e.g., retroviral vectors) in nature. Regulatory sequences (e.g., promoters and enhancers) are known in the art that can be used to regulate protein expression. The type of cell being transfected or transduced also has a strong bearing on the strength and longevity of protein expression. For example, cell types with high rates of turnover can be expected to have shorter periods of protein expression.

Viral Vectors

Viral vectors encoding the nucleotide sequence of one or more optimized viral polypeptides of the invention can be used as a vaccine of the invention. For example, the nucleotide sequence of one or more optimized viral polypeptides of the invention can be inserted recombinantly into that of a natural or modified (e.g., attenuated) viral genome suitable for the transduction of a subject (e.g., *in vivo* administration) or cells isolated from a subject (e.g., for *ex vivo* transduction followed by administration of the cells back to the subject). Additional modifications can be made to the virus to improve infectivity or tropism (e.g., pseudotyping), reduce or eliminate replicative competency, or reduce immunogencity of the viral components (e.g., all components not related to the immunogenic vaccine agent). A vector of the invention can be expressed by the transduced cell and secreted into the extracellular space or remain with the expressing cell (e.g., as an intracellular molecule or displayed on the cell surface). Chimeric or pseudotyped viral vectors can also be used to transduce a cell to allow for expression of one or more optimized viral polypeptides of the invention. Exemplary vectors are described below.

Adenoviruses

Recombinant adenoviruses offer several significant advantages for use as vectors for the expression of one or more optimized viral polypeptides of the invention. The viruses can be prepared to high titer, can infect non-replicating cells, and can confer high-efficiency transduction of target cells *ex vivo* following contact with a target cell population. Furthermore, adenoviruses do not integrate their DNA into the host genome. Thus, their use as expression vectors has a reduced risk of inducing spontaneous proliferative disorders. In animal models, adenoviral vectors have generally been found to mediate high-level expression for approximately one week. The duration of transgene expression (expression of a nucleic acid encoding an optimized viral polypeptide of the invention) can be prolonged by using cell or tissue-specific promoters. Other improvements in the molecular engineering of the adenoviral vector itself have produced more sustained transgene expression and less inflammation. This is seen with so-called "second generation" vectors harboring specific mutations in additional early adenoviral genes and "gutless" vectors in which

virtually all the viral genes are deleted utilizing a Cre-Lox strategy (Engelhardt et al., *Proc. Natl. Acad. Sci. USA* 91:6196 (1994) and Kochanek et al., *Proc. Natl. Acad. Sci. USA* 93:5731 (1996), each herein incorporated by reference).

The rare serotype and chimeric adenoviral vectors disclosed in International Patent Application Publications WO 2006/040330 and WO 2007/104792, each incorporated by reference herein, are particularly useful as vectors of the invention. For example, recombinant adenoviruses rAd26, rAd34, rAd35, rAd48, and rAd5HVR48 can encode one or more optimized viral polypeptides of the invention. One or more recombinant viral vectors encoding optimized viral polypeptides of the invention can be administered to a subject to treat or prevent a viral infection.

Adeno-Associated Viruses (AAV)

Adeno-associated viruses (rAAV), derived from non-pathogenic parvoviruses, can also be used to express optimized viral polypeptides of the invention as these vectors evoke almost no anti-vector cellular immune response, and produce transgene expression lasting months in most experimental systems.

Retroviruses

Retroviruses are useful for the expression of optimized viral polypeptides of the invention. Unlike adenoviruses, the retroviral genome is based in RNA. When a retrovirus infects a cell, it will introduce its RNA together with several enzymes into the cell. The viral RNA molecules from the retrovirus will produce a double-stranded DNA copy, called a provirus, through a process called reverse transcription. Following transport into the cell nucleus, the proviral DNA is integrated in a host cell chromosome, permanently altering the genome of the transduced cell and any progeny cells that may derive from this cell. The ability to permanently introduce a gene into a cell or organism is the defining characteristic of retroviruses used for gene therapy. Retroviruses include lentiviruses, a family of viruses including human immunodeficiency virus (HIV) that includes several accessory proteins to facilitate viral infection and proviral integration. Current, "third-generation," lentiviral vectors feature total replication incompetence, broad tropism, and increased gene transfer

capacity for mammalian cells (see, e.g., Mangeat and and Trono, *Human Gene Therapy* 16(8):913 (2005) and Wiznerowicz and Trono, *Trends Biotechnol.* 23(1):42 (2005), each herein incorporated by reference).

Other Viral Vectors

Besides adenoviral and retroviral vectors, other viral vectors and techniques are known in the art that can be used to express optimized viral polypeptides of the invention in a cell (e.g., a blood cell, such as a lymphocyte) or subject (e.g., a human). Theses viruses include Poxviruses (e.g., vaccinia virus and modified vaccinia virus Ankara or (MVA); see, e.g., U.S. Patent Nos. 4,603,112 and 5,762,938, each incorporated by reference herein), Herpesviruses, Togaviruses (e.g., Venezuelan Equine Encephalitis virus; see, e.g., U.S. Patent No. 5,643,576, incorporated by reference herein), Picornaviruses (e.g., poliovirus; see, e.g., U.S. Patent No. 5,639,649, incorporated by reference herein), Baculoviruses, and others described by Wattanapitayakul and Bauer (*Biomed. Pharmacother.* 54:487 (2000), incorporated by reference herein).

Other Expression Vectors: *Naked DNA and Oligonucleotides*

Naked DNA or oligonucleotides encoding one or more optimized viral polypeptides of the invention can also be used to express these polypeptides in a cell (e.g., a blood cell, such as a lymphocyte) or subject (e.g., a human). See, e.g., Cohen, *Science* 259:1691-1692 (1993); Fynan et al., *Proc. Natl. Acad. Sci. USA*, 90:11478 (1993); and Wolff et al., *BioTechniques* 11:474485 (1991), each herein incorporated by reference. This is the simplest method of non-viral transfection. Efficient methods for delivery of naked DNA exist such as electroporation and the use of a "gene gun," which shoots DNA-coated gold particles into a cell using high pressure gas and carrier particles (e.g., gold).

Lipoplexes and Polyplexes

To improve the delivery of an nucleic acid encoding an optimized viral polypeptide of the invention into a cell or subject, lipoplexes (e.g., liposomes) and polyplexes can be used to protect the vector DNA from undesirable degradation during the transfection process. Plasmid DNA can be covered with lipids in an

organized structure like a micelle or a liposome. When the organized structure is complexed with DNA it is called a lipoplex. There are three types of lipids, anionic (negatively-charged), neutral, or cationic (positively-charged). Lipoplexes that utilize cationic lipids have proven utility for gene transfer. Cationic lipids, due to their positive charge, naturally complex with the negatively- charged DNA. Also as a result of their charge they interact with the cell membrane, endocytosis of the lipoplex occurs, and the DNA is released into the cytoplasm. The cationic lipids also protect against degradation of the DNA by the cell.

Complexes of polymers with DNA are called polyplexes. Most polyplexes consist of cationic polymers and their production is regulated by ionic interactions. One large difference between the methods of action of polyplexes and lipoplexes is that polyplexes cannot release their DNA load into the cytoplasm, so to this end, co-transfection with endosome-lytic agents (to lyse the endosome that is made during endocytosis) such as inactivated adenovirus must occur. However, this is not always the case; polymers such as polyethylenimine have their own method of endosome disruption as does chitosan and trimethylchitosan.

Exemplary cationic lipids and polymers that can be used in combination with an nucleic acid encoding an optimized viral polypeptide of the invention to form lipoplexes, or polyplexes include, but are not limited to, polyethylenimine, lipofectin, lipofectamine, polylysine, chitosan, trimethylchitosan, and alginate.

Hybrid methods

Several hybrid methods of gene transfer combine two or more techniques. Virosomes, for example, combine lipoplexes (e.g., liposomes) with an inactivated virus. This approach has been shown to result in more efficient gene transfer in respiratory epithelial cells than either viral or liposomal methods alone. Other methods involve mixing other viral vectors with cationic lipids or hybridising viruses. Each of these methods can be used to facilitate transfer of an nucleic acid encoding optimized viral polypeptides of the invention into a cell (e.g., a blood cell, such as a lymphocyte) or subject (e.g., a human).

Dendrimers

Dendrimers may be also be used to transfer an nucleic acid encoding an optimized viral polypeptide of the invention into a cell (e.g., a blood cell, such as a lymphocyte) or subject (e.g., a human). A dendrimer is a highly branched macromolecule with a spherical shape. The surface of the particle may be functionalized in many ways, and many of the properties of the resulting construct are determined by its surface. In particular it is possible to construct a cationic dendrimer (i.e. one with a positive surface charge). When in the presence of genetic material such as DNA or RNA, charge complimentarity leads to a temporary association of the nucleic acid with the cationic dendrimer. On reaching its destination the dendrimer-nucleic acid complex is then taken into the cell via endocytosis.

***In Vivo* Administration**

The invention also features *in vivo* methods for immunizing a subject (e.g., a human) with a vaccine of the invention. In one embodiment, one or more vaccines of the invention can be directly administered to a subject to elicit a protective or therapeutic immune response (e.g., a cellular or humoral immune response) against a virus (e.g., HIV-1). Alternatively, a vector encoding one or more optimized viral polypeptides of the invention, as described above, can be directly administered to a subject to prevent or treat a viral infection. A vector (e.g., a viral vector) that efficiently transfects or transduces one or more cells *in vivo* can elicit a broad, durable, and potent immune response in the treated subject. Upon transfer of the nucleic acid component of the expression vector into a host cell (e.g., a blood cell, such as a lymphocyte), the host cell produces and displays or secretes the vaccine of the invention, which then serves to activate components of the immune system such as antigen-presenting cells (APCs), T cells, and B cells, resulting in the establishment of immunity.

Pharmaceutical Compositions

The invention features the vaccines, vectors, and optimized viral polypeptides of the invention in combination with one or more pharmaceutically acceptable excipients, diluents, buffers, or other acceptable carriers. The formulation of a vaccine, vector, or optimized viral polypeptides will employ or allow expression of an effective amount of the optimized viral polypeptide immunogen. That is, there will

be included an amount of antigen which will cause the treated subject (e.g., a human) to produce a specific and sufficient immunological response so as to impart protection to the subject from subsequent exposure to a virus (e.g., HIV-1) or to treat an existing viral infection. For example, a formulation of a vaccine of the invention can allow for the expression of an amount of antigen which will cause the subject to produce a broad and specific cellular immune response. A subject treated with a vaccine, vector, or optimized viral polypeptide of the invention can also produce anti-viral antibodies (e.g., neutralizing antibodies) which can confer a protective or therapeutic benefit to the subject. A vaccine, vector, or optimized viral polypeptide of the invention can be directly administered to a subject, either alone or in combination with any pharmaceutically acceptable carrier, salt or adjuvant known in the art.

Pharmaceutically acceptable salts may include non-toxic acid addition salts or metal complexes that are commonly used in the pharmaceutical industry. Examples of acid addition salts include organic acids such as acetic, lactic, pamoic, maleic, citric, malic, ascorbic, succinic, benzoic, palmitic, suberic, salicylic, tartaric, methanesulfonic, toluenesulfonic, or trifluoroacetic acids or the like; polymeric acids such as tannic acid, carboxymethyl cellulose, or the like; and inorganic acids such as hydrochloric acid, hydrobromic acid, sulfuric acid phosphoric acid, or the like. Metal complexes include zinc, iron, and the like. One exemplary pharmaceutically acceptable carrier is physiological saline. Other physiologically acceptable carriers and their formulations are known to one skilled in the art and described, for example, in Remington's Pharmaceutical Sciences, (18^{th} edition), ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA.

Pharmaceutical formulations of a prophylactically or therapeutically effective amount of a vaccine, vector, or optimized viral polypeptide of the invention can be administered orally, parenterally (e.g., intramuscular, intraperitoneal, intravenous, or subcutaneous injection, inhalation, intradermally, optical drops, or implant), nasally, vaginally, rectally, sublingually, or topically, in admixture with a pharmaceutically acceptable carrier adapted for the route of administration. The concentration of a vaccine, vector, or optimized viral polypeptide of the invention in the formulation can vary from about 0.1-100 wt. %.

Formulations for parenteral administration of compositions containing a vaccine, vector, or optimized viral polypeptide of the invention include sterile

aqueous or non-aqueous solutions, suspensions, or emulsions. Examples of suitable vehicles include propylene glycol, polyethylene glycol, vegetable oils, gelatin, hydrogenated naphalenes, and injectable organic esters, such as ethyl oleate. Such formulations may also contain adjuvants, such as preserving, wetting, emulsifying, and dispersing agents. Biocompatible, biodegradable lactide polymer, lactide/glycolide copolymer, or polyoxyethylene-polyoxypropylene copolymers may be used to control the release of the compounds. Other potentially useful parenteral delivery systems for compositions containing a vaccine, vector, or optimized viral polypeptide of the invention include ethylene-vinyl acetate copolymer particles, osmotic pumps, implantable infusion systems, and liposomes.

Liquid formulations can be sterilized by, for example, filtration through a bacteria-retaining filter, by incorporating sterilizing agents into the compositions, or by irradiating or heating the compositions. Alternatively, they can also be manufactured in the form of sterile, solid compositions, which can be dissolved in sterile water or some other sterile injectable medium immediately before use.

Compositions containing vaccine, vector, or optimized viral polypeptide of the invention for rectal or vaginal administration are preferably suppositories which may contain, in addition to active substances, excipients such as coca butter or a suppository wax. Compositions for nasal or sublingual administration are also prepared with standard excipients known in the art. Formulations for inhalation may contain excipients, for example, lactose, or may be aqueous solutions containing, for example, polyoxyethylene-9-lauryl ether, glycocholate and deoxycholate, or may be oily solutions for administration in the form of nasal drops or spray, or as a gel.

The amount of active ingredient in the compositions of the invention can be varied. One skilled in the art will appreciate that the exact individual dosages may be adjusted somewhat depending upon a variety of factors, including the peptide being administered, the time of administration, the route of administration, the nature of the formulation, the rate of excretion, the nature of the subject's conditions, and the age, weight, health, and gender of the patient. In addition, the severity of the condition treated by the vaccine, vector, or optimized viral polypeptide will also have an impact on the dosage level. Generally, dosage levels of between 0.1 μg/kg to 100 mg/kg of body weight are administered daily as a single dose or divided into multiple doses. Preferably, the general dosage range is between 250 μg/kg to 5.0 mg/kg of body

weight per day. Wide variations in the needed dosage are to be expected in view of the differing efficiencies of the various routes of administration. For instance, oral administration generally would be expected to require higher dosage levels than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, which are well known in the art. In general, the precise prophylactically or therapeutically effective dosage can be determined by the attending clinician in consideration of the above-identified factors.

The amount of a vaccine, vector, or optimized viral polypeptide of the invention present in each dose given to a patient is selected with regard to consideration of the patient's age, weight, sex, general physical condition and the like. The amount of a vaccine, vector, or optimized viral polypeptide required to induce an immune response (e.g., a cellular immune response) or produce an exogenous effect in the patient without significant adverse side effects varies depending upon the pharmaceutical composition employed and the optional presence of an adjuvant. Initial doses can be optionally followed by repeated boosts, where desirable. The method can involve chronically administering the vaccine, vector, or optimized viral polypeptide of the invention. For therapeutic use or prophylactic use, repeated dosages of the immunizing vaccine, vector, or optimized viral polypeptide can be desirable, such as a yearly booster or a booster at other intervals. The dosage administered will, of course, vary depending upon known factors such as the pharmacodynamic characteristics of the particular vaccine, vector, or optimized viral polypeptide, and its mode and route of administration; age, health, and weight of the recipient; nature and extent of symptoms, kind of concurrent treatment, frequency of treatment, and the effect desired. A vaccine, vector, or optimized viral polypeptide of the invention can be administered in chronic treatments for subjects at risk of acute infection due to needle sticks or maternal infection. A dosage frequency for such "acute" infections may range from daily dosages to once or twice a week i.v. or i.m., for a duration of about 6 weeks. The vaccine, vector, or optimized viral polypeptide can also be employed in chronic treatments for infected patients, or patients with advanced infection with a virus (e.g., HIV-1). In infected patients, the frequency of

chronic administration can range from daily dosages to once or twice a week i.v. or i.m., and may depend upon the half-life of immunogen present in the vaccine, vector, or optimized viral polypeptide of the invention.

Adjuvants

A vaccine of the invention used to vaccinate a mammal (e.g., a human) in need thereof against a virus can be administered concurrent with or in series with one or more pharmaceutically acceptable adjuvants to increase the immunogenicity of the vaccine. Adjuvants approved for human use include aluminum salts (alum). These adjuvants have been useful for some vaccines including hepatitis B, diphtheria, polio, rabies, and influenza. Other useful adjuvants include Complete Freund's Adjuvant (CFA), Incomplete Freund's Adjuvant (IFA), muramyl dipeptide (MDP), synthetic analogues of MDP, N-acetylmuramyl-L-alanyl-D-isoglutamyl-L-alanine-2-[1,2-dipalmitoyl-s-gly- cero-3-(hydroxyphosphoryloxy)]ethylamide (MTP-PE) and compositions containing a metabolizable oil and an emulsifying agent, wherein the oil and emulsifying agent are present in the form of an oil-in-water emulsion having oil droplets substantially all of which are less than one micron in diameter.

Kits

The invention provides kits that include a pharmaceutical composition containing a vaccine, vector, or optimized viral polypeptide of the invention, and a pharmaceutically-acceptable carrier, in a therapeutically effective amount for preventing or treating a viral infection. The kits include instructions to allow a clinician (e.g., a physician or nurse) to administer the composition contained therein.

Preferably, the kits include multiple packages of the single-dose pharmaceutical composition(s) containing an effective amount of a vaccine, vector, or optimized viral polypeptide of the invention. Optionally, instruments or devices necessary for administering the pharmaceutical composition(s) may be included in the kits. For instance, a kit of this invention may provide one or more pre-filled syringes containing an effective amount of a vaccine, vector, or optimized viral polypeptide of the invention. Furthermore, the kits may also include additional components such as

instructions or administration schedules for a patient infected with or at risk of being infected with a virus to use the pharmaceutical composition(s) containing a vaccine, vector, or optimized viral polypeptide of the invention.

It will be apparent to those skilled in the art that various modifications and variations can be made in the compositions, methods, and kits of the present invention without departing from the spirit or scope of the invention. Thus, it is intended that the present invention cover the modifications and variations of this invention provided they come within the scope of the appended claims and their equivalents.

Examples

The present invention is illustrated by the following examples, which are in no way intended to be limiting of the invention.

Example 1

The mosaic antigen Gag, Pol, Nef, and Env sequences (SEQ ID NOS:1-8) were constructed using the genetic algorithm discussed above. These sequences were then modified to make them practical for vaccine development by eliminating cleavage/fusion activity in Env (SEQ ID NOS:9-11), eliminating catalytic activity in Pol (SEQ ID NOS:12-14), eliminating myristylation sites in Nef (SEQ ID NOS:16-18), and constructing fusion constructs including GagNef, GagPol, or GagPolNef (SEQ ID NOS:19-29). The comparator optimal natural clade C genes are also depicted (SEQ ID NOS:30-36).

Example 2

Twenty rhesus monkeys were immunized with $3x10^{10}$ vp rAd26 vectors expressing Gag, Pol, and Env genes from M consensus (Group 1), 2-valent M mosaic (Group 2), or optimal natural clade C (Group 3) sequences. The M consensus sequences represent synthetic sequences that represent the single best "average" of circulating viruses worldwide. The 2-valent M mosaic sequences are described above. The optimal natural clade C sequences are naturally occurring sequences from actual clade C HIV-1 viruses that are the most "consensus-like" in character. Cellular immune breadth was assessed by evaluating the number of responding peptides from

the global potential T cell epitope (PTE) peptide set. The PTE peptides represent >85% of global HIV-1 sequences and are freely available from the NIH.

The results show that the novel 2-valent M mosaic sequences dramatically outperformed these other two leading antigen concepts. As shown in Table 1, the 2-valent M mosaic antigens elicited significantly increased breadth of Gag-specific, Env-specific, Pol-specific, and total T lymphocyte responses as compared with M consensus antigens and optimal natural clade C antigens. (Mean represents the average # epitopes in each group of monkeys; SEM represents the standard error of the mean).

Table 1: Mosaic HIV-1 Gag/Pol/Env Antigens Expand Breadth Against Global PTE Peptides in Rhesus Monkeys

readth	Group I: M Consensus		Group II: 2-valent M Mosaic		Group III: Natural Cla	
	Mean	SEM	Mean	SEM	Mean	SEM
Gag	2.0	0.4	7.7	0.9	2.2	0.5
Env	2.0	0.4	4.0	0.6	1.6	0.5
Pol	2.7	0.5	8.1	1.4	2.4	0.5
Total	6.7	0.7	19.9	1.9	6.1	1.1

Example 3

Macaque monkeys were immunized IM with $3x10^{10}$ vp rAd26 vectors expressing Gag, Pol, and Env genes from M consensus (Group 1; n=7), 2-valent M mosaic (Group 2; n=7), or optimal natural clade C (Group 3; n=6) sequences described in Example 2. Cellular immune breadth was assessed by evaluating the number of responding peptides from the global potential T cell epitope (PTE) peptide set.

As a readout, we assessed the CD4/CD8 IFNγ Elispot responses to pooled PTE peptides (magnitude). The epitopes were comprehensively mapped using 15 mer PTE peptides to assess the number of positives (positives were defined as 55 spot forming cells (SFC) per 10^6 PBMC and 4x background). Pooled sets of overlapping peptides spanning 5 Gag proteins were also tested to compare responses to a set of complete proteins.

The results show that the 2-valent M mosaic sequences dramatically outperformed the other two leading antigen concepts (Mcon and OptC).

Example 4

We used modeling to validate our observation that T cell responses increase as a result of the mosaic vaccine. We fit Poisson regression models that predicted the number of reactive peptides as a function of vaccine, polypeptide, and T cell type and then did a stepwise elimination of interactions. We observed that, although the mosaic vaccine produced a highly significant enhancement in the number of positive PTE responses, it did so more-or-less uniformly across all polyproteins and T-cell types. Thus, one may predict the number of peptides having a positive effect in an animal by combining contributions that depend, separately, on the type of T-cell, the polypeptide, and the vaccine the animal received.

These models also included random effects to account for animal-to-animal variation. This is a precaution designed to make for more credible *p*-values, by properly apportioning the predictive power of the model.

We observed the following effects:

a) There are many more CD8 responses than CD4 responses, by a factor of 4.37, $p < 2 \times 10^{-16}$;

b) There are fewer responses in gp160 than in gag or pol, by a factor of 0.54, $p = 0.000830$, and no significant difference between gag and pol (even when normalized by sequence length as pol is twice as long as Gag and so has more opportunity to react); and

c) The mosaic vaccine generates significantly more positive responses than Mcon (by a factor of 3.6, $p = 6.26 \times 10^{-11}$) while OptC generates fewer, though the Mcon-OptC difference is not significant.

Example 5

If one considers just the minimal number of responses elicited by a vaccine and detected by PTE peptides, so that all peptide that overlap by ≥ 8 amino acids regardless of variation are counted just 1 time, the mosaic vaccines still generate a greater number of responses to distinct regions.

For CD8, counting each overlapping peptide set just once:

Statistical summary:

Mos2 > Mcon ~ OptC (Mcon shows a trend for more response than OptC)

Wilcoxon p-value for Mos2 compared to Mcon: p-value = 0.0009992

Wilcoxon p-value for Mcon compared to Optimal C: p-value = 0.2351

Summary of the groups:

Vaccine	Min.	1stQu.	Median	Mean	3rdQu.	Max.
Mos2.cd8	7	7.5	8	9.4	11	14
Mcon.cd8	0	3	3	3.3	4	6
OptC.cd8	0	1	1.5	2	4.25	5

For CD4, counting each overlapping peptide set just once (there is very little overlap in CD4, so this is almost the same as the first count).

Statistical summary:

Mos2 >> Mcon ~ OptC (Mcon shows a trend for more response than OptC)

Wilcoxon p-value for Mos2 compared to Mcon: p-value = 0.00198

Wilcoxon p-value for Mcon compared to Optimal C: 0.099

Summary of the groups:

Vaccine	Min.	1stQu.	Median	Mean	3rdQu.	Max.
Mos2.cd4	2	2.5	3	3.4	4.5	5
Mcon.cd4	0	1	1	1.3	2	2
OptC.cd4	0	0	0.5	0.67	1	2

Example 6: Poisson regression counting each overlapping peptide set just once

Using overlapping PTE peptides, we determined the following, which are in broad agreement with the results discussed in Example 4 above, where each positive PTE response counted separately:

a) There are many more CD8 responses than CD4 responses, by a factor of about 2.8, $p \approx 1 \times 10^{-7}$;

b) The mosaic vaccine generates significantly more positive responses than Mcon (by a factor of 2.84, $p \approx 4.3 \times 10^{-7}$), while OptC generates fewer, though the Mcon-OptC difference is not significant; and

c) There are more responses to Pol than to Gag and more to Gag than gp160, but only the Pol-gp160 difference, a factor of about 2, was significant, $p < 0.001$.

Example 7:

The following table is a tally of the total responses to Gag, Pol, and Env responses to the three vaccines in the 7 animals vaccinated with 2 Mosaic (Mos2) or Mcon, and the 6 animals vaccinated with the Optimal Natural C clade (OptC):

		CD8			CD4	
	Env	Gag	Pol	Env	Gag	Pol
2Mos	13	20	33	3	10	11
ConM	8	7	8	2	3	4
OptC	4	5	5	1	2	1

The OptC vaccine yielded an average response across all monkeys that was slightly less than the CD8+ T cell response per protein. The Mcon vaccine exhibited ~ 1 response per protein. Only with Mos2 do we observe a difference in the proteins, where Env typically has fewer responses than either Gag or Pol.

Each of the proteins in the Mos2 vaccine elicited many responses and contributed to the overall response. The relative length of the consensus proteins after the modifications to inactivate pol and the deletion of the cleavage and fusion domain in Env was: 671 amino acids of Env, 851 of Pol, 498 of Gag (1.35 : 1.7 : 1).

Summary

Breadth: The 2 mosaic vaccines elicit T cell responses that are capable of recognizing many more epitope-regions than the M consensus or a single optimal natural strain.

Depth: The diversity of the PTE peptides recognized suggests both forms in the 2 mosaics are eliciting different T cell responses to the variant peptides, increasing the cross-reactive potential.

Example 8:

Mosaic HIV-1 vaccines of the invention expand the breadth and depth of cellular immune responses in Rhesus monkeys. We constructed mosaic HIV-1 Gag, Pol, and Env antigens that optimized PTE coverage of HIV-1 M group sequences, which include all major HIV-1 clades and recombinant lineages in the Los Alamos HIV-1 sequence

database. A 2-valent mosaic strategy was utilized to balance the competing issues of theoretical coverage and practical utility. 2-valent mosaic HIV-1 Gag, Pol, and Env antigens substantially expanded the breadth and magnitude (depth) of epitope-specific CD8+ and CD4+ T lymphocyte responses in rhesus monkeys, relative to the immune response observed using consensus and natural sequence HIV-1 antigens in rhesus monkeys.

We immunized 27 outbred rhesus monkeys with a single injection of recombinant adenovirus serotype 26 (rAd26) vectors expressing the following antigens: (i) 2-valent mosaic (N=7), (ii) M consensus (N=7), (iii) 2-valent combined clade B and clade C (N=7), or (iv) optimal natural clade C (N=6) HIV-1 Gag, Pol, and Env antigens. A total dose of $3x10^{10}$ viral particles of rAd26 vectors expressing these antigens was administered once i.m. to each animal. The optimal clade C antigens were the natural strain sequences selected to provide maximal PTE coverage of clade C sequences in the Los Alamos HIV-1 sequence database (discussed in the Materials and Methods below). We assessed the breadth and magnitude (depth) of vaccine-elicited HIV-1-specific T lymphocyte responses by IFN-γ ELISPOT assays at week 4 following immunization utilizing pools and subpools of peptides that included all PTEs found in at least 15% of HIV-1 M group sequences. All individual peptide responses were resolved, and cell-depleted IFN-γ ELISPOT assays were performed to determine if reactive peptides represented CD8+ or CD4+ T lymphocyte epitopes.

The total number of Gag-, Pol-, and Env-specific cellular immune responses to PTE peptides elicited by the mosaic antigens was 3.8-fold higher than the number of responses induced by the consensus or natural sequence antigens (Fig. 19A; P = 1 x 10-11, comparing the mosaic with the consensus antigens, the next highest group, based on a Poisson regression model). There were 4.4-fold more CD8+ than CD4+ T lymphocyte responses (P < 10-11) and fewer responses to Env than to Gag or Pol (P < 0.0007). The median number of CD8+ T lymphocyte responses was highest for the

mosaic vaccine, followed by the consensus, the combined B+C, and the natural clade C vaccines (medians of 16, 5, 3, and 2 responses per animal in each group, respectively). Although there were fewer CD4+ T lymphocyte responses overall, the same relative pattern emerged with the highest number of CD4+ T lymphocyte responses to the mosaic vaccine, followed by the consensus, the combined B+C, and the natural clade C vaccines (medians of 4, 1, 1, and 0.5 responses per animal in each group, respectively). The numbers of CD8+ and CD4+ T lymphocyte responses elicited by the consensus, the combined B+C, and the natural clade C vaccines were not statistically distinguishable.

PTE peptides include multiple overlapping sequences that reflect naturally occurring HIV-1 sequence polymorphisms, and thus the PTE peptide responses encompass both the recognition of a particular epitope (breadth) and the cross-recognition of variants of that epitope (depth). We performed a conservative analysis of breadth by assessing the number of reactive epitopic regions per monkey in which all reactive PTE peptides that overlapped by 8 or more amino acids were counted as one event. In this conservative analysis, we still observed that the mosaic antigens elicited 3.1-fold increased numbers of Gag, Pol, and Env reactive epitopic regions as compared with the consensus antigens or natural sequence antigens (Fig. 19B; $P = 1.6 \times 10^{-7}$, Poisson regression). Epitopic regions exhibited some clustering across animals, as evidenced by regions of high epitope density (Figs. 20A-20C and Figs. 21A-21C). Complete alignments of all positive peptides organized by response regions are shown in Fig. 22.

These data show that the mosaic antigens substantially increased the breadth of cellular immune responses as compared with M consensus and natural clade C antigens. The 2-valent mosaic antigens also proved superior to the 2-valent combination of clade B and clade C antigens (Fig. 19A and 19B), indicating that the enhanced breadth was due to the mosaic sequence design and did not simply reflect the use of two distinct antigenic sequences per protein. To determine if the increased breadth induced by mosaic antigens compromised the potency of the responses, we assessed the magnitude of all individual CD8+ and CD4+ T lymphocyte responses. The magnitude of these responses proved comparable among all groups (Fig.23; $P = 0.58$ and $P = 0.99$, respectively, two-sided Kolmogorov-Smirnov tests). Thus, mosaic antigens expanded cellular immune breadth without compromising the magnitude of

individual epitope-specific responses, indicating that antigenic competition and immunodominance constraints did not limit the immunogenicity of the mosaic antigens in this study.

We next characterized the depth of the cellular immune responses elicited by the various vaccine regimens. We defined depth as the number of simultaneously elicited variant PTE peptides for a particular epitopic region. Inducing responses to multiple common epitope variants may increase immunologic coverage of infecting virus sequences, block common escape routes in vivo, or force the virus down tertiary escape routes that incur high fitness costs. The consensus and natural sequence antigens elicited responses that were characterized by a high degree of sequence identity between the vaccine sequences and the reactive PTE peptides, as exemplified by the responses in monkey 366 that received the natural clade C antigens (Fig. 24A; see also Fig. 22). In contrast, the mosaic antigens elicited responses that were characterized by multiple reactive PTE peptides in particular epitopic regions. These peptides represented common variants and often reflected the polymorphisms contained in the mosaic vaccine sequences, as exemplified by the responses in monkey 361 (Fig. 24B; see also Fig. 22). A summary of all epitope-specific responses in these animals demonstrates that the mosaic antigens increased the frequency of cellular immune responses to peptides with two or more targeted variants as compared with the consensus or natural sequence antigens (Fig. 24C; $P = 0.001$, Wilcoxon rank-sum test comparing the mosaic with the consensus antigens, the next highest group).

To complement the analysis utilizing PTE peptides, we also assessed the breadth of cellular immune responses in the vaccinated monkeys with traditional overlapping peptides covering 5 different Gag sequences: clade C DU422, clade C ZM651, consensus C, consensus A, and consensus B. Cellular immune breadth was determined by assessing reactivity to subpools of 10 overlapping peptides spanning each Gag sequence. The mosaic antigens elicited greater breadth of T lymphocyte responses as compared with the consensus or natural sequence antigens against all Gag sequences that were tested (Fig. 25; $P = 1 \times 10^{-7}$, binomial regression). Thus, the mosaic antigens augmented cellular immune breadth not only to PTE peptides but also to actual Gag peptides from clades A, B, and C. The mosaic antigens even proved superior to the optimal natural clade C antigens for inducing responses against

clade C Gag peptides. Moreover, the mosaic antigens elicited comparable responses to Gag peptides from multiple clades, whereas the natural clade C antigens exhibited diminished responses to clade A and clade B Gag peptides (Fig. 25).

To assess the durability of these observations, we boosted the monkeys that received the mosaic, consensus, and optimal natural clade C antigens at week 40 with a total dose of $3x10^{10}$ viral particles of the heterologous vector rAd5HVR48 expressing HIV-1 Gag, Pol, and Env antigens that matched the sequences utilized in the initial immunization. Cellular immune breadth was determined by assessing reactivity to subpools of 10 PTE peptides at week 4 (post-prime) and at week 44 (post-boost). The majority of CD8+ and CD4+ T lymphocyte responses that were observed after the priming immunization expanded following the boost (Fig. 26A, red and blue lines), and a number of new responses were also detected (Fig. 26A, red and blue dots). At week 44, the magnitude of individual cellular immune responses proved comparable among groups (Fig. 26A). The number of subpool responses elicited by the mosaic antigens (median 27 responses per animal), however, remained substantially higher than the number of subpool responses induced by the consensus antigens (median 11 responses per animal) or the optimal natural clade C antigens (median 10 responses per animal) following the boost immunization (Fig. 26B). Both before and after the boost, there were more responses per animal elicited by the mosaic vaccine than by the consensus or natural clade C vaccines ($P < 0.001$, Wilcoxon rank-sum tests for all pairwise comparisons).

We also measured Env-specific humoral immune responses following the boost immunization by ELISAs (Fig. 26C) and luciferase-based pseudovirus neutralization assays (Fig. 26D). All groups exhibited comparable ELISA titers to clade C gp140 and comparable neutralizing antibody (NAb) responses to the tier 1 clade C virus MW965.26. The mosaic antigens elicited slightly higher Nab responses to the tier 1 clade B virus SF162.LS as compared with the consensus or natural clade C antigens ($P = 0.02$, Wilcoxon rank-sum test), although we did not detect any NAb responses to tier 2 viruses in any group.

Our data demonstrate that mosaic HIV-1 Gag, Pol, and Env antigens augmented both the breadth and depth of epitope-specific cellular immune responses as compared with consensus or natural sequence antigens in rhesus monkeys, in good agreement with theoretical predictions (Fig. 27). The striking results with mosaic

antigens in this study may have reflected the fact that rAd26 vectors are particularly efficient at eliciting CD8+ T lymphocyte responses as well as the fact that mosaic antigens appear particularly effective at augmenting CD8+ T lymphocyte breadth (Fig. 19A and 19B). We also observed enhanced CD4+ T lymphocyte breadth with mosaic antigens, although there were substantially lower numbers of these responses.

The breadth of Gag-specific cellular immune responses has been shown to be critical for SIV control in rhesus monkeys and for HIV-1 control in humans. Moreover, in the phase 2b STEP study, the rAd5-based HIV-1 vaccine candidate expressing natural clade B Gag, Pol, and Nef antigens elicited only a limited breadth of HIV-1-specific cellular immune responses, and no vaccine benefit was observed. Vaccinees in the STEP study developed a median of only 2-3 epitope-specific T lymphocyte responses, including a median of only 1 epitope-specific response to Gag, and this very narrow breadth of cellular immune responses likely provided insufficient immunologic coverage of the diversity of infecting viruses. Viral escape from CD8+ T lymphocytes has also been reported to occur rapidly during acute HIV-1 infection, and thus vaccine-elicited cellular immune responses against common epitope variants may also prove critical. Taken together, these studies emphasize the need to develop HIV-1 vaccine strategies that augment cellular immune breadth and depth.

Since we evaluated mosaic HIV-1 antigens in the present study, we were unable to assess the protective efficacy of these vaccine regimens against SIV challenges. However, we have previously reported that the breadth of SIV-specific cellular immune responses elicited by rAd vectors correlated with protective efficacy against SIV challenges in rhesus monkeys (Liu et al., *Nature* 457:87, 2009). We have also shown that cellular immune responses against variant epitopes can block SIV mutational evolution in rhesus monkeys *in vivo (*Barouch et al., *Nat. Immunol.* 6:247, 2005), suggesting the biologic relevance of expanding cellular immune depth. Modeling the protective efficacy of mosaic vaccines against SIV challenges in nonhuman primates has intrinsic limitations, since the observed diversity of SIV and HIV-1 M group sequences differs substantially and is influenced by different underlying biology. For example, CD8+ T lymphocyte selection pressure in natural hosts such as sooty mangabees appears substantially less than that in humans.

Thus, the further evaluation of mosaic antigens as candidate HIV-1 vaccines can be benefited by clinical trials.

In summary, we demonstrate that 2-valent mosaic HIV-1 Gag, Pol, and Env antigens substantially expanded cellular immune breadth and depth in rhesus monkeys. These findings have major implications for HIV-1 vaccine development, since global virus diversity and viral escape from cellular immune responses represent critical hurdles in the development of a T cell-based HIV-1 vaccine. A 2-valent cocktail of mosaic antigens is also practical and potentially feasible for clinical development. Increasing the valency of mosaic antigens may further improve coverage. Finally, the mosaic antigen strategy is generalizable and could be utilized for other genetically diverse pathogens in addition to HIV-1.

Materials and Methods

Antigen design and vector production. 2-valent mosaic Gag, Pol, and Env antigens were constructed to provide optimal coverage of HIV-1 M group sequences in the Los Alamos HIV-1 sequence database essentially as described (*1, 2*). Optimal natural clade C antigens were selected to be the sequences that provide optimal PTE coverage of clade C sequences in the Los Alamos HIV-1 sequence database (C.IN.-.70177 Gag, C.ZA.04.04ZASK208B1 Pol, C.SN.90.90SE_364 Env). Clade B antigens were selected to be near-consensus or consensus sequences (B.CAM-1 Gag, B.IIIB Pol, B.Con Env) and were used to complement the optimal clade C antigens for the 2-valent clade B + C vaccine approach. Pol antigens contained RT and IN without PR and included point mutations to eliminate catalytic activity as described (Priddy et al., *Clinical infectious diseases* 46:1769, 2008). Env gp140 antigens contained point mutations to eliminate cleavage and fusion activity. Vaccine sequences are depicted in Fig. 27. Recombinant, replication-incompetent adenovirus serotype 26 (rAd26) and hexon-chimeric rAd5HVR48 vectors expressing these antigens were grown in PER.55K cells and purified by double CsCl gradient sedimentation essentially as described (Abbink et al., *J. Virol.* 81:4654, 2007, and Roberts et al., *Nature* 441:239, 2006).

Animals and immunizations. 27 outbred rhesus monkeys that did not express the MHC class I allele *Mamu-A*01* were housed at New England Primate Research Center (NEPRC), Southborough, MA. Immunizations involved 3 x 10^{10}

viral particles rAd26 or rAd5HVR48 vectors expressing mosaic, M consensus, clade B + clade C, or optimal natural clade C HIV-1 Gag, Pol, and Env antigens delivered as 1 ml injections i.m. in both quadriceps muscles at weeks 0 and 40. All animal studies were approved by our Institutional Animal Care and Use Committees (IACUC).

IFN-γ ELISPOT assays. HIV-1-specific cellular immune responses in vaccinated monkeys were assessed by interferon-γ (IFN-γ) ELISPOT assays essentially as described (Roberts et al., *Nature* 441:239, 2006, and Liu et al., *Nature* 457:87, 2009). HIV-1 Gag, Pol, and Env potential T cell epitope (PTE) peptides that included all PTEs found in at least 15% of HIV-1 M group sequences as well as HIV-1 Gag peptides from clade C DU422, clade C ZM651, consensus C, consensus A, and consensus B strains were obtained from the NIH AIDS Research and Reference Reagent Program. 96-well multiscreen plates (Millipore) were coated overnight with 100 µl/well of 10 µg/ml anti-human IFN-γ (BD Biosciences) in endotoxin-free Dulbecco's PBS (D-PBS). The plates were then washed three times with D-PBS containing 0.25% Tween-20 (D-PBS/Tween), blocked for 2 h with D-PBS containing 5% FBS at 37°C, washed three times with D-PBS/Tween, rinsed with RPMI 1640 containing 10% FBS to remove the Tween-20, and incubated with 2 µg/ml each peptide and 2 x 10^5 PBMC in triplicate in 100 µl reaction volumes. Following an 18 h incubation at 37oC, the plates were washed nine times with PBS/Tween and once with distilled water. The plates were then incubated with 2 µg/ml biotinylated anti-human IFN-γ (BD Biosciences) for 2 h at room temperature, washed six times with PBS/Tween, and incubated for 2 h with a 1:500 dilution of streptavidin-alkaline phosphatase (Southern Biotechnology Associates). Following five washes with PBS/Tween and one with PBS, the plates were developed with nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl-phosphate chromogen (Pierce), stopped by washing with tap water, air dried, and read using an ELISPOT reader (Cellular Technology Ltd). Spot-forming cells (SFC) per 10^6 PBMC were calculated. Media backgrounds were typically < 15 SFC per 10^6 PBMC. Positive responses were defined as > 55 SFC per 10^6 PBMC and > 4-fold background.

Epitope mapping. Comprehensive CD8+ and CD4+ T lymphocyte epitope mapping was performed utilizing Gag, Pol, and Env PTE peptides that were obtained from the NIH AIDS Research and Reference Reagent Program. IFN-γ ELISPOT

assays were conducted at week 4 following immunization initially with complete peptide pools as well as with subpools containing 10 PTE peptides. All peptide subpools with positive responses were deconvoluted, and epitopes were confirmed with individual 15 amino acid PTE peptides. Cell-depleted IFN-γ ELISPOT assays were then performed to determine if reactive peptides represented CD8+ or CD4+ T lymphocyte epitopes. Partial epitope mapping utilizing PTE subpools was also performed 4 weeks following the boost immunization at week 44. All borderline responses were retested and only considered positive if confirmed. Partial epitope mapping utilizing subpools containing 10 overlapping Gag peptides was also performed to assess breadth to HIV-1 Gag from various clades.

Humoral immune assays. Env-specific humoral immune responses were evaluated by direct ELISAs utilizing HIV-1 clade C Env gp140 and luciferase-based pseudovirus neutralization assays essentially as described (Montefiori, *Evaluating neutralizing antibodies against HIV, SIV and SHIV in luciferase reporter gene assays. Current Protocols in Immunology,* Coligan, Kruisbeek, Margulies, Shevach, Strober, and Coico, Ed. (John Wiley & Sons, 2004, pp. 1-15).

Statistical analyses. All statistical analyses were done using the package R (Team, *Foundation for Statisical Computing, Vienna, Austria*, 2009). To analyze the breadth of cellular immune responses to mapped PTE peptides (Fig. 19A), we fit Poisson regression models that predicted the number of reactive peptides as a function of vaccine group, antigen (Gag, Pol, Env), and lymphocyte subpopulation (CD4, CD8). Our models included random effects to accommodate animal-to-animal variation and were fit with the lme4 library (Pinheiro, *Springer, New York* (2000)) of the package R. The data fit the models well (dispersion parameter 1.0), and there were no significant interactions among the three explanatory factors. For example, the 3.8-fold enhancement in the number of PTE peptides recognized by monkeys that received the mosaic antigens as compared to those that received the consensus or natural sequence antigens (Fig. 19A) applied equally to PTEs from Gag, Pol, and Env and held for responses by CD8+ as well as CD4+ T lymphocytes. The analysis of the number of reactive epitopic regions (Fig. 19B) also included Poisson regression models with random effects and again fit well (dispersion parameter 0.87) without any significant interactions. Comparisons of the magnitude of CD8+ and CD4+ T lymphocyte responses (Fig. 23) were performed utilizing 2-sided Kolmogorov-

Smirnov tests. Non-parametric tests to compare the breadth and depth of responses per monkey between different vaccines were also performed (Figs. 19A and 24C). We initially employed Kruskal-Wallis tests to determine if there was a difference among the 4 vaccine groups. In each case this was highly significant, and we then assessed all pairwise comparisons between the 4 vaccine groups using Wilcoxon rank-sum tests. In each of these comparisons, the mosaic vaccine elicited significantly more responses per monkey than the other 3 vaccines. To analyze the breadth of responses to HIV-1 Gag from various clades (Fig. 25), we fit the data to binomial regression models. These models used the vaccine group as an explanatory variable and included random effects to account for animal-to-animal and strain-to-strain variation. The data were slightly underdispersed, but the animals that received the mosaic vaccine still elicited a significantly larger number of responses. PTE coverage assessment was performed using tools available at the Los Alamos HIV-1 sequence database.

SEQUENCE APPENDIX

I. 2-VALENT M MOSAIC ENV GP160, GAG, POL, NEF SEQUENCES

MOSAIC ENV1 GP160 (AA SEQUENCE)

SEQ ID NO:1

MRVTGIRKNYQHLWRWGTMLLGILMICSAAGKLWVTVYYGVPVWKEATTTLFCASDA
KAYDTEVHNVWATHACVPTDPNPQEVVLENVTENFNMWKNNMVEQMHEDIISLWDQS
LKPCVKLTPLCVTLNCTDDVRNVTNNATNTNSSWGEPMEKGEIKNCSFNITTSIRNK
VQKQYALFYKLDVVPIDNDSNNTNYRLISCNTSVITQACPKVSFEPIPIHYCAPAGF
AILKCNDKKFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSENFTNN
AKTIMVQLNVSVEINCTRPNNNTRKSIHIGPGRAFYTAGDIIGDIRQAHCNISRANW
NNTLRQIVEKLGKQFGNNKTIVFNHSSGGDPEIVMHSFNCGGEFFYCNSTKLFNSTW
TWNNSTWNNTKRSNDTEEHITLPCRIKQIINMWQEVGKAMYAPPIRGQIRCSSNITG
LLLTRDGGNDTSGTEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQ**R**
EKRAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLLSGIVQQQNNLLRAIEAQQHL
LQLTVWGIKQLQARVLAVERYLKDQQLLGIWGCSGKLICTTTVPWNASWSNKSLDKI
WNNMTWMEWEREINNYTSLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISNWL
WYIKIFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTRLPAPRGPDRPEGIEEE
GGERDRDRSVRLVDGFLVLIWDDLQSLCLFSYHRLRDLLLIVELLGRRGWEALKYWW
NLLQYWSQELKNSAISLLNATAVAVAEGTDRVIEALQRACRAILHIPRRIRQGLERL
LL

MOSAIC ENV2 GP160 (AA SEQUENCE)

SEQ ID NO:2

MRVRGIQRNWPQWWIWGILGFWMIIICRVMGNLWVTVYYGVPVWKEAKTTLFCASDA
KAYEKEVHNVWATHACVPTDPNPQEMVLENVTENFNMWKNDMVDQMHEDIIRLWDQS
LKPCVKLTPLCVTLECRNVRNVSSNGTYNIIHNETYKEMKNCSFNATTVVEDRKQKV
HALFYRLDIVPLDENNSSEKSSENSSEYYRLINCNTSAITQACPKVSFDPIPIHYCA
PAGYAILKCNNKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLLNGSLAEEEIIIRSEN
LTNNAKTIIVHLNETVNITCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQAHCNLS
RDGWNKTLQGVKKKLAEHFPNKTINFTSSSGGDLEITTHSFNCRGEFFYCNTSGLFN
GTYMPNGTNSNSSSNITLPCRIKQIINMWQEVGRAMYAPPIAGNITCRSNITGLLLT
RDGGSNNGVPNDTETFRPGGGDMRNNWRSELYKYKVVEVKPLGVAPTEAKRRVVE**RE**
KRAVGIGAVFLGILGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLRAIEAQQHML
QLTVWGIKQLQTRVLAIERYLQDQQLLGLWGCSGKLICTTAVPWNTSWSNKSQTDIW
DNMTWMQWDKEIGNYTGEIYRLLEESQNQQEKNEKDLLALDSWKNLWNWFDITNWLW
YIKIFIMIVGGLIGLRIILGVLSIVRRVRQGYSPLSFQTLTPNPRGLDRLGRIEEEG
GEQDRDRSIRLVNGFLALAWDDLRSLCLFSYHQLRDFILIVARAVELLGRSSLRGLQ
RGWEALKYLGNLVQYWGLELKKGAISLLDTIAIAVAEGTDRIIELIQSICRAIRNIP
RRIRQGFEASLL

MOSAIC GAG1 (AA SEQUENCE)

SEQ ID NO:3

MGARASVLSGGELDRWEKIRLRPGGKKKYRLKHIVWASRELERFAVNPGLLETSEGC
RQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALEKIEEEQNKSKKK
AQQAAADTGNSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPM
FSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQM
REPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPVSILDI
RQGPKEPFRDYVDRFYKTLRAEQASQDVKNWMTETLLVQNANPDCKTILKALGPAAT
LEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKTVKCFNCGKEG
HIAKNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSNKGRPGNFLQNRPEPT
APPEESFRFGEETTTPSQKQEPIDKEMYPLASLKSLFGNDPSSQ

MOSAIC GAG2 (AA SEQUENCE)

SEQ ID NO:4

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLVWASRELERFALNPGLLETSEGC
KQIIKQLQPALQTGTEELRSLFNTVATLYCVHAEIEVRDTKEALDKIEEEQNKSQQK
TQQAKEADGKVSQNYPIVQNLQGQMVHQPISPRTLNAWVKVIEEKAFSPEVIPMFTA
LSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGPVAPGQMREP
RGSDIAGTTSNLQEQIAWMTSNPPIPVGDIYKRWIILGLNKIVRMYSPTSILDIKQG
PKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNANPDCKTILRALGPGATLEE
MMTACQGVGGPSHKARVLAEAMSQTNSTILMQRSNFKGSKRIVKCFNCGKEGHIARN
CRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPAE
SFRFEETTPAPKQEPKDREPLTSLRSLFGSDPLSQ

MOSAIC POL1 (AA SEQUENCE)

SEQ ID NO:5

FFRENLAFQQGEAREFPSEQTRANSPTSRELQVRGDNPHSEAGAERQGTLNFPQITL
WQRPLVSIKVGGQIREALL**D**TGADDTVLEDINLPGKWKPKMIGGIGGFIKV**R**QYDQI
LIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNFPISPIETVPVKLKPGMDGPR
VKQWPLTEEKIKALTAICEEMEKEGKITKIGPENPYNTPVFAIKKKDSTKWRKLVDF
RELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEGFRKYTAFTIPS
TNNETPGIRYQYNVLPQGWKGSPAIFQCSMTRILEPFRAKNPEIVIYQYMD**D**LYVGS
DLEIGQHRAKIEELREHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQLPE
KDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGAKALTDIVPLTEEAELELAE
NREILKEPVHGVYYDPSKDLIAEIQKQGHDQWTYQIYQEPFKNLKTGKYAKMRTAHT
NDVKQLTEAVQKIAMESIVIWGKTPKFRLPIQKETWETWWTDYWQATWIPEWEFVNT
PPLVKLWYQLEKDPIAGVETFYVDGAANRETKLGKAGYVTDRGRQKIVSLTETTNQK
TELQAIYLALQDSGSEVNIVTDSQYALGIIQAQPDKSESELVNQIIEQLIKKERVYL
SWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKAQEEHEKYHSNWRAMASDFNLPP
VVAKEIVASCDQCQLKGEAMHGQVDCSPGIWQL**D**CTHLEGKIILVAVHVASGYIEAE
VIPAETGQETAYFILKLAGRWPVKVIHTDNGSNFTSAAVKAACWWAGIQQEFGIPYN
PQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERII
DIIATDIQTKELQKQIIKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSDIK
VVPRRKVKIIKDYGKQMAGADCVAGRQDED

MOSAIC POL2 (AA SEQUENCE)

SEQ ID NO:6

FFRENLAFPQGKAREFSSEQTRANSPTRRELQVWGRDNNSLSEAGADRQGTVSFSFP
QITLWQRPLVTIKIGGQLKEALL**D**TGADDTVLEEMNLPGRWKPKMIGGIGGFIKV**R**Q
YDQIPIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPGM
DGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPIFAIKKKDSTKWRK
LVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAF
TIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMD**D**L
YVGSDLEIGQHRTKIEELRQHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPI
VLPEKDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRGTKALTEVVPLTEEAEL
ELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMR
GAHTNDVKQLTEAVQKIATESIVIWGKTPKFKLPIQKETWEAWWTEYWQATWIPEWE
FVNTPPLVKLWYQLEKEPIVGAETFYVDGAANRETKLGKAGYVTDRGRQKVVSLTDT
TNQKTELQAIHLALQDSGLEVNIVTDSQYALGIIQAQPDKSESELVSQIIEQLIKKE
KVYLAWVPAHKGIGGNEQVDKLVSRGIRKVLFLDGIDKAQEEHEKYHSNWRAMASEF
NLPPIVAKEIVASCDKCQLKGEAIHGQVDCSPGIWQL**D**CTHLEGKVILVAVHVASGY
IEAEVIPAETGQETAYFLLKLAGRWPVKTIHTDNGSNFTSATVKAACWWAGIKQEFG
IPYNPQSQGVVESINKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGEYSAG
ERIVDIIASDIQTKELQKQITKIQNFRVYYRDSRDPLWKGPAKLLWKGEGAVVIQDN
SDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED

MOSAIC NEF1 (AA SEQUENCE)

SEQ ID NO:7

MGGKWSKSSVVGWPAIRERMRRAEPAADGVGAVSRDLEKHGAITSSNTAANNADCAW
LEAQEEEEVGFPVRPQVPLRPMTYKGALDLSHFLKEKGGLEGLIYSQKRQDILDLWV
YHTQGYFPDWQNYTPGPGIRYPLTFGWCFKLVPVEPEKIEEANEGENNSLLHPMSQH
GMDDPEKEVLMWKFDSRLAFHHMARELHPEYYKDC

MOSAIC NEF2 (AA SEQUENCE)

SEQ ID NO:8

MGGKWSKSSIVGWPAVRERIRRAEPAAEGVGAASQDLDKYGALTSSNTAATNADCAW
LEAQEDEEVGFPVKPQVPLRPMTYKAAFDLSFFLKEKGGLDGLIYSKKRQEILDLWV
YNTQGFFPDWQNYTPGPGVRYPLTFGWCFKLVPVDPREVEEANKGENNCLLHPMNLH
GMDDPEREVLVWRFDSRLAFHHMAREKHPEYYKNC

II. 2-VALENT M MOSAIC ENV GP140 SEQUENCES (CLEAVAGE/FUSION-DEFECTIVE)

MOSAIC ENV1 GP140 (AA SEQUENCE)

SEQ ID NO:9

MRVTGIRKNYQHLWRWGTMLLGILMICSAAGKLWVTVYYGVPVWKEATTTLFCASDA
KAYDTEVHNVWATHACVPTDPNPQEVVLENVTENFNMWKNNMVEQMHEDIISLWDQS

LKPCVKLTPLCVTLNCTDDVRNVTNNATNTNSSWGEPMEKGEIKNCSFNITTSIRNK
VQKQYALFYKLDVVPIDNDSNNTNYRLISCNTSVITQACPKVSFEPIPIHYCAPAGF
AILKCNDKKFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSENFTNN
AKTIMVQLNVSVEINCTRPNNNTRKSIHIGPGRAFYTAGDIIGDIRQAHCNISRANW
NNTLRQIVEKLGKQFGNNKTIVFNHSSGGDPEIVMHSFNCGGEFFYCNSTKLFNSTW
TWNNSTWNNTKRSNDTEEHITLPCRIKQIINMWQEVGKAMYAPPIRGQIRCSSNITG
LLLTRDGGNDTSGTEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQ**S**
EKSAVGIGAVFLGFLGAAGSTMGAASMTLTVQARLLLSGIVQQQNNLLRAIEAQQHL
LQLTVWGIKQLQARVLAVERYLKDQQLLGIWGCSGKLICTTTVPWNASWSNKSLDKI
WNNMTWMEWEREINNYTSLIYTLIEESQNQQEKNEQELLELDKWASLWNWFDISNWL
W

MOSAIC ENV2 GP140 (AA SEQUENCE)

SEQ ID NO:10

MRVRGIQRNWPQWWIWGILGFWMIIICRVMGNLWVTVYYGVPVWKEAKTTLFCASDA
KAYEKEVHNVWATHACVPTDPNPQEMVLENVTENFNMWKNDMVDQMHEDIIRLWDQS
LKPCVKLTPLCVTLECRNVRNVSSNGTYNIIHNETYKEMKNCSFNATTVVEDRKQKV
HALFYRLDIVPLDENNSSEKSSENSSEYYRLINCNTSAITQACPKVSFDPIPIHYCA
PAGYAILKCNNKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLLNGSLAEEEIIIRSEN
LTNNAKTIIVHLNETVNITCTRPNNNTRKSIRIGPGQTFYATGDIIGDIRQAHCNLS
RDGWNKTLQGVKKKLAEHFPNKTINFTSSSGGDLEITTHSFNCRGEFFYCNTSGLFN
GTYMPNGTNSNSSSNITLPCRIKQIINMWQEVGRAMYAPPIAGNITCRSNITGLLLT
RDGGSNNGVPNDTETFRPGGGDMRNNWRSELYKYKVVEVKPLGVAPTEAKRRVVE**SE**
KSAVGIGAVFLGILGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLRAIEAQQHML
QLTVWGIKQLQTRVLAIERYLQDQQLLGLWGCSGKLICTTAVPWNTSWSNKSQTDIW
DNMTWMQWDKEIGNYTGEIYRLLEESQNQQEKNEKDLLALDSWKNLWNWFDITNWLW

MOS3 ENV GP140 (AA SEQUENCE)

678 AA

SEQ ID NO:11

MRVKGIRKNYQHLWKWGTMLLGMLMICSAAEQLWVTVYYGVPVWRDAET
TLFCASDAKAYEREVHNIWATHACVPTDPNPQEIVLENVTEEFNMWKNDMV
EQMHTDIISLWDESLKPCVKLAPLCVTLNCTNANLNCTNDNCNRTVDKMREE
IKNCSFNMTTELRDKKQKVYALFYKLDIVPIEKNSSEYRLINCNTSTITQACPK
VTFEPIPIHYCTPAGFAILKCKDKKFNGTGPCKNVSTVQCTHGIKPVISTQLLL
NGSLAEGEIIIRSENITNNAKTIIVQLNESVVINCTRPGNNTRKSVRIGPGQAFY
ATGEIIGDIRQAYCNISRAKWNNTLKQIVTKLKEQFKNKTIVFNQSSGGDPEIT
THSFNCGGEFFYCNTTQLFNSTWNSNSTWNDTTGSVTEGNDTITLPCRIKQIV
NMWQRVGQAMYAPPIEGNITCKSNITGLLLVRDGGNINRTNETFRPGGGNMK
DNWRSELYKYKVVEIKPLGVAPTRAKRRVVE**SEKS**AVGLGAVFLGFLGTAG
STMGAASLTLTVQARQVLSGIVQQQSNLLKAIEAQQHLLKLTVWGIKQLQAR
ILAVERYLRDQQLLGIWGCSGKLICTTNVPWNSSWSNKSQEEIWNNMTWMQ
WDREISNYTDTIYRLLEDSQNQQEKNEQDLLALDKWASLWNWFSITNWLW

III. 2-VALENT M MOSAIC POL SEQUENCES (EXTENSIVELY INACTIVATED, PR-DELETED, 9 A INACTIVATION MUTATIONS TO ELIMINATE CATALYTIC ACTIVITY)

MOSAIC POL1 (AA SEQUENCE)

SEQ ID NO:12

MAPISPIETVPVKLKPGMDGPRVKQWPLTEEKIKALTAICEEMEKEGKITKIGPENP
YNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVL**A**VG
DAYFSVPLDEGFRKYTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQCSMTRILE
PFRAKNPEIVIYQYM**AA**LYVGSDLEIGQHRAKIEELREHLLKWGFTTPDKKHQKEPP
FLWMGYELHPDKWTVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLL
RGAKALTDIVPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGHDQWTYQ
IYQEPFKNLKTGKYAKMRTAHTNDVKQLTEAVQKIAMESIVIWGKTPKFRLPIQKET
WETWWTDYWQATWIPEWEFVNTPPLVKLWYQLEKDPIAGVETFYV**A**GAANRETKLGK
AGYVTDRGRQKIVSLTETTNQKT**A**LQAIYLALQDSGSEVNIVT**A**SQYALGIIQAQPD
KSESELVNQIIEQLIKKERVYLSWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKA
QEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDQCQLKGEAMHGQVDCSPGIWQL**A**C
THLEGKIILVAVHVASGYIEAEVIPAETGQETAYFILKLAGRWPVKVIHT**A**NGSNFT
SAAVKAACWWAGIQQEFGIPYNPQSQGVV**A**SMNKELKKIIGQVRDQAEHLKTAVQMA
VFIHNFKRKGGIGGYSAGERIIDIIATDIQTKELQKQIIKIQNFRVYYRDSRDPIWK
GPAKLLWKGEGAVVIQDNSDIKVVPRRKVKIIKDYGKQMAGADCVAGRQDED

MOSAIC POL2 (AA SEQUENCE)

SEQ ID NO:13

MAPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPENP
YNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVL**A**VG
DAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILE
PFRKQNPDIVIYQYM**AA**LYVGSDLEIGQHRTKIEELRQHLLRWGFTTPDKKHQKEPP
FLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLL
RGTKALTEVVPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQ
IYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKIATESIVIWGKTPKFKLPIQKET
WEAWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETFYV**A**GAANRETKLGK
AGYVTDRGRQKVVSLTDTTNQKT**A**LQAIHLALQDSGLEVNIVT**A**SQYALGIIQAQPD
KSESELVSQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSRGIRKVLFLDGIDKA
QEEHEKYHSNWRAMASEFNLPPIVAKEIVASCDKCQLKGEAIHGQVDCSPGIWQL**A**C
THLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGRWPVKTIHT**A**NGSNFT
SATVKAACWWAGIKQEFGIPYNPQSQGVV**A**SINKELKKIIGQVRDQAEHLKTAVQMA
VFIHNFKRKGGIGEYSAGERIVDIIASDIQTKELQKQITKIQNFRVYYRDSRDPLWK
GPAKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQDED

MOS3 POL V3 (AA SEQUENCE)

851 AA

SEQ ID NO:14

MAPISPIDTVPVTLKPGMDGPKIKQWPLTEEKIKALTEICTEMEKEGKISRIGPENP
YNTPVFAIKKKNSTRWRKLVDFRELNKKTQDFWEVQLGIPHPAGLKKKRSVTVL**A**VG
DAYFSVPLDKDFRKYTAFTIPSVNNETPGVRYQYNVLPQGWKGSPAIFQCSMTKILE
PFRAQNPEIVIYQYV**AA**LYVGSDLEIEQHRTKIEELRAHLLSWGFTTPDKKHQREPP
FLWMGYELHPDRWTVQPIELPEKESWTVNDIQKLVGKLNWASQIYPGIKVKQLCRLL
RGAKALTEVIPLTKEAELELAENREILREPVHGVYYDPSKDLVAEIQKQGQDQWTYQ
IYQEPYKNLKTGKYARKRSAHTNDVRQLTEAVQKIALESIVIWGKIPKFRLPIQRET
WETWWTEYWQATWIPDWEFVNTPPLVKLWYQLEKEPIAGAETFYV**A**GASNRETKIGK
AGYVTDKGRQKVVSLTETTNQKA**A**LQAIQLALQDSGPEVNIVT**A**SQYVLGIIQAQPD
RSESELVNQIIEELIKKEKVYLSWVPAHKGIGGNEQVDKLVSAGIRKILFLDGIDKA
QEEHERYHSNWRTMASDFNLPPIVAKEIVANCDKCQLKGEAMHGQVDCSPGMWQL**A**C
THLEGKIIIVAVHVASGYMEAEVIPAETGQETAYYILKLAGRWPVKVVHT**A**NGSNFT
STTVKAACWWANVTQEFGIPYNPQSQGVI**A**SMNKELKKIIGQVREQAEHLKTAVQMA
VLIHNFKRRGGIGGYSAGERIVDIIATDIQTRELQKQIIKIQNFRVYFRDSRDPVWK
GPAKLLWKGEGAVVIQDNSEIKVVPRRKVKIIRDYGKQMAGDDCVAGRQDEDQ

IV. 2-VALENT M MOSAIC GAG SEQUENCE

MOS3 GAG (AA SEQUENCE)

508 AA

SEQ ID NO:15

MGARASVLSGGKLDAWEKIRLRPGGKKKYKLKHIVWASRELDRFALNPGLLETAEGC
QQIIEQLQPALQTGSEELKSLYNTVAVLYCVHQRIDVKDTKEALDKIEEIQNKSKQK
TQQAAADTGSSSKVSQNYPIVQNAQGQMVHQALSPRTLNAWVKVVEEKGFNPEVIPM
FSALAEGATPQDLNMMLNIVGGHQAAMQILKDTINEEAADWDRLHPVHAGPIPPGQM
REPRGSDIAGTTSTPQEQIGWMTSNPPVPVGEIYKRWIIMGLNKIVRMYSPVSILDI
KQGPKESFRDYVDRFFKVLRAEQATQEVKNWMTETLLIQNANPDCKSILRALGPGAS
LEEMMTACQGVGGPSHKARILAEAMSQANNTNIMMQRGNFKGQKRIKCFNCGKEGHL
ARNCRAPRKRGCWKCGREGHQMKDCNERQANFLGKIWPSSKGRPGNFPQSRPEPTAP
LEPTAPPAEPTAPPAESFGFGEEITPSPKQEQKDREPLTSLKSLFGSDPLLQ

V. 2-VALENT M MOSAIC NEF SEQUENCES (POSITION 2 G TO A TO DELETE MYRISTYLATION SITE

MOS1 NEF

(206 AA)

SEQ ID NO:16

M**A**GKWSKSSVVGWPAIRERMRRAEPAADGVGAVSRDLEKHGAITSSNTAANNADCAW
LEAQEEEEVGFPVRPQVPLRPMTYKGALDLSHFLKEKGGLEGLIYSQKRQDILDLWV
YHTQGYFPDWQNYTPGPGIRYPLTFGWCFKLVPVEPEKIEEANEGENNSLLHPMSQH
GMDDPEKEVLMWKFDSRLAFHHMARELHPEYYKDC

MOS2 NEF

(206 AA) - POSITION 2 G TO A TO DELETE MYRISTYLATION SITE

SEQ ID NO:17

MAGKWSKSSIVGWPAVRERIRRAEPAAEGVGAASQDLDKYGALTSSNTAATNADCAW
LEAQEDEEVGFPVKPQVPLRPMTYKAAFDLSFFLKEKGGLDGLIYSKKRQEILDLWV
YNTQGFFPDWQNYTPGPGVRYPLTFGWCFKLVPVDPREVEEANKGENNCLLHPMNLH
GMDDPEREVLVWRFDSRLAFHHMAREKHPEYYKNC

MOS3 NEF

(208 AA)

SEQ ID NO:18

MAGKWSKRSVVGWPAVRERMRRTEPAAEGVGAVSQDLDKHGALTSSNTAHNNADCAW
LQAQEEEEEVGFPVRPQVPVRPMTYKAAVDLSHFLKEKGGLEGLIHSQKRQEILDLW
VYHTQGFFPDWHNYTPGPGTRFPLTFGWCYKLVPVDPKEVEEANEGENNCLLHPMSQ
HGMEDEDREVLKWKFDSSLARRHMARELHPEFYKDCL

VI. 2-VALENT M MOSAIC GAGNEF FUSION SEQUENCES

MOSAIC GAGNEF1 (AA SEQUENCE)

SEQ ID NO:19

MGARASVLSGGELDRWEKIRLRPGGKKKYRLKHIVWASRELERFAVNPGLLETSEGC
RQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALEKIEEEQNKSKKK
AQQAAADTGNSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPM
FSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQM
REPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPVSILDI
RQGPKEPFRDYVDRFYKTLRAEQASQDVKNWMTETLLVQNANPDCKTILKALGPAAT
LEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKTVKCFNCGKEG
HIAKNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSNKGRPGNFLQNRPEPT
APPEESFRFGEETTTPSQKQEPIDKEMYPLASLKSLFGNDPSSQAGKWSKSSVVGWP
AIRERMRRAEPAADGVGAVSRDLEKHGAITSSNTAANNADCAWLEAQEEEEVGFPVR
PQVPLRPMTYKGALDLSHFLKEKGGLEGLIYSQKRQDILDLWVYHTQGYFPDWQNYT
PGPGIRYPLTFGWCFKLVPVEPEKIEEANEGENNSLLHPMSQHGMDDPEKEVLMWKF
DSRLAFHHMARELHPEYYKDC

MOSAIC GAGNEF2 (AA SEQUENCE)

SEQ ID NO:20

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLVWASRELERFALNPGLLETSEGC
KQIIKQLQPALQTGTEELRSLFNTVATLYCVHAEIEVRDTKEALDKIEEEQNKSQQK
TQQAKEADGKVSQNYPIVQNLQGQMVHQPISPRTLNAWVKVIEEKAFSPEVIPMFTA
LSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGPVAPGQMREP
RGSDIAGTTSNLQEQIAWMTSNPPIPVGDIYKRWIILGLNKIVRMYSPTSILDIKQG

PKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNANPDCKTILRALGPGATLEE
MMTACQGVGGPSHKARVLAEAMSQTNSTILMQRSNFKGSKRIVKCFNCGKEGHIARN
CRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPAE
SFRFEETTPAPKQEPKDREPLTSLRSLFGSDPLSQAGKWSKSSIVGWPAVRERIRRA
EPAAEGVGAASQDLDKYGALTSSNTAATNADCAWLEAQEDEEVGFPVKPQVPLRPMT
YKAAFDLSFFLKEKGGLDGLIYSKKRQEILDLWVYNTQGFFPDWQNYTPGPGVRYPL
TFGWCFKLVPVDPREVEEANKGENNCLLHPMNLHGMDDPEREVLVWRFDSRLAFHHM
AREKHPEYYKNC

VII. 2-VALENT M MOSAIC GAGPOL FUSION SEQUENCES (VERSION 3; POL EXTENSIVELY INACTIVATED, PR-DELETED, 9 A INACTIVATION MUTATIONS TO ELIMINATE CATALYTIC ACTIVITY)

MOSAIC GAGPOL1 V3 (AA SEQUENCE)

SEQ ID NO:21

MGARASVLSGGELDRWEKIRLRPGGKKKYRLKHIVWASRELERFAVNPGLLETSEGC
RQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALEKIEEEQNKSKKK
AQQAAADTGNSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPM
FSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQM
REPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPVSILDI
RQGPKEPFRDYVDRFYKTLRAEQASQDVKNWMTETLLVQNANPDCKTILKALGPAAT
LEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKTVKCFNCGKEG
HIAKNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSNKGRPGNFLQNRPEPT
APPEESFRFGEETTTPSQKQEPIDKEMYPLASLKSLFGNDPSSQ**MA**PISPIETVPVK
LKPGMDGPRVKQWPLTEEKIKALTAICEEMEKEGKITKIGPENPYNTPVFAIKKKDS
TKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVL**A**VGDAYFSVPLDEGFR
KYTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQCSMTRILEPFRAKNPEIVIYQ
YM**AA**LYVGSDLEIGQHRAKIEELREHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKW
TVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGAKALTDIVPLT
EEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGHDQWTYQIYQEPFKNLKTGK
YAKMRTAHTNDVKQLTEAVQKIAMESIVIWGKTPKFRLPIQKETWETWWTDYWQATW
IPEWEFVNTPPLVKLWYQLEKDPIAGVETFYV**A**GAANRETKLGKAGYVTDRGRQKIV
SLTETTNQKT**A**LQAIYLALQDSGSEVNIVT**A**SQYALGIIQAQPDKSESELVNQIIEQ
LIKKERVYLSWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKAQEEHEKYHSNWRA
MASDFNLPPVVAKEIVASCDQCQLKGEAMHGQVDCSPGIWQL**A**CTHLEGKIILVAVH
VASGYIEAEVIPAETGQETAYFILKLAGRWPVKVIHT**A**NGSNFTSAAVKAACWWAGI
QQEFGIPYNPQSQGVV**A**SMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIG
GYSAGERIIDIIATDIQTKELQKQIIKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAV
VIQDNSDIKVVPRRKVKIIKDYGKQMAGADCVAGRQDED

MOSAIC GAGPOL2 V3 (AA SEQUENCE)

SEQ ID NO:22

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLVWASRELERFALNPGLLETSEGC
KQIIKQLQPALQTGTEELRSLFNTVATLYCVHAEIEVRDTKEALDKIEEEQNKSQQK

TQQAKEADGKVSQNYPIVQNLQGQMVHQPISPRTLNAWVKVIEEKAFSPEVIPMFTA
LSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGPVAPGQMREP
RGSDIAGTTSNLQEQIAWMTSNPPIPVGDIYKRWIILGLNKIVRMYSPTSILDIKQG
PKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNANPDCKTILRALGPGATLEE
MMTACQGVGGPSHKARVLAEAMSQTNSTILMQRSNFKGSKRIVKCFNCGKEGHIARN
CRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPAE
SFRFEETTPAPKQEPKDREPLTSLRSLFGSDPLSQ**MA**PISPIETVPVKLKPGMDGPK
VKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPIFAIKKKDSTKWRKLVDF
RELNKRTQDFWEVQLGIPHPAGLKKKKSVTVL**A**VGDAYFSVPLDEDFRKYTAFTIPS
INNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYM**AA**LYVGS
DLEIGQHRTKIEELRQHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPE
KDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRGTKALTEVVPLTEEAELELAE
NREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHT
NDVKQLTEAVQKIATESIVIWGKTPKFKLPIQKETWEAWWTEYWQATWIPEWEFVNT
PPLVKLWYQLEKEPIVGAETFYV**A**GAANRETKLGKAGYVTDRGRQKVVSLTDTTNQK
T**A**LQAIHLALQDSGLEVNIVT**A**SQYALGIIQAQPDKSESELVSQIIEQLIKKEKVYL
AWVPAHKGIGGNEQVDKLVSRGIRKVLFLDGIDKAQEEHEKYHSNWRAMASEFNLPP
IVAKEIVASCDKCQLKGEAIHGQVDCSPGIWQL**A**CTHLEGKVILVAVHVASGYIEAE
VIPAETGQETAYFLLKLAGRWPVKTIHT**A**NGSNFTSATVKAACWWAGIKQEFGIPYN
PQSQGVV**A**SINKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGEYSAGERIV
DIIASDIQTKELQKQITKIQNFRVYYRDSRDPLWKGPAKLLWKGEGAVVIQDNSDIK
VVPRRKAKIIRDYGKQMAGDDCVASRQDED

MOS3 GAG-POL V3 (AA SEQUENCES)

1359 aa - GAG-POL FUSION WITH COMPLETE GAG AND MODIFIED POL

SEQ ID NO:23

MGARASVLSGGKLDAWEKIRLRPGGKKKYKLKHIVWASRELDRFALNPGLLETAEGC
QQIIEQLQPALQTGSEELKSLYNTVAVLYCVHQRIDVKDTKEALDKIEEIQNKSKQK
TQQAAADTGSSSKVSQNYPIVQNAQGQMVHQALSPRTLNAWVKVVEEKGFNPEVIPM
FSALAEGATPQDLNMMLNIVGGHQAAMQILKDTINEEAADWDRLHPVHAGPIPPGQM
REPRGSDIAGTTSTPQEQIGWMTSNPPVPVGEIYKRWIIMGLNKIVRMYSPVSILDI
KQGPKESFRDYVDRFFKVLRAEQATQEVKNWMTETLLIQNANPDCKSILRALGPGAS
LEEMMTACQGVGGPSHKARILAEAMSQANNTNIMMQRGNFKGQKRIKCFNCGKEGHL
ARNCRAPRKRGCWKCGREGHQMKDCNERQANFLGKIWPSSKGRPGNFPQSRPEPTAP
LEPTAPPAEPTAPPAESFGFGEEITPSPKQEQKDREPLTSLKSLFGSDPLLQ**MA**PIS
PIDTVPVTLKPGMDGPKIKQWPLTEEKIKALTEICTEMEKEGKISRIGPENPYNTPV
FAIKKKNSTRWRKLVDFRELNKKTQDFWEVQLGIPHPAGLKKKRSVTVL**A**VGDAYFS
VPLDKDFRKYTAFTIPSVNNETPGVRYQYNVLPQGWKGSPAIFQCSMTKILEPFRAQ
NPEIVIYQYV**AA**LYVGSDLEIEQHRTKIEELRAHLLSWGFTTPDKKHQREPPFLWMG
YELHPDRWTVQPIELPEKESWTVNDIQKLVGKLNWASQIYPGIKVKQLCRLLRGAKA
LTEVIPLTKEAELELAENREILREPVHGVYYDPSKDLVAEIQKQGQDQWTYQIYQEP
YKNLKTGKYARKRSAHTNDVRQLTEAVQKIALESIVIWGKIPKFRLPIQRETWETWW
TEYWQATWIPDWEFVNTPPLVKLWYQLEKEPIAGAETFYV**A**GASNRETKIGKAGYVT
DKGRQKVVSLTETTNQKA**AL**QAIQLALQDSGPEVNIVT**A**SQYVLGIIQAQPDRSESE
LVNQIIEELIKKEKVYLSWVPAHKGIGGNEQVDKLVSAGIRKILFLDGIDKAQEEHE
RYHSNWRTMASDFNLPPIVAKEIVANCDKCQLKGEAMHGQVDCSPGMWQL**A**CTHLEG
KIIIVAVHVASGYMEAEVIPAETGQETAYYILKLAGRWPVKVVHT**A**NGSNFTSTTVK

AACWWANVTQEFGIPYNPQSQGVI**A**SMNKELKKIIGQVREQAEHLKTAVQMAVLIHN
FKRRGGIGGYSAGERIVDIIATDIQTRELQKQIIKIQNFRVYFRDSRDPVWKGPAKL
LWKGEGAVVIQDNSEIKVVPRRKVKIIRDYGKQMAGDDCVAGRQDEDQ

VIII. 2-VALENT M MOSAIC GAGPOL FUSION SEQUENCES (VERSION 4; POL MINIMALLY INACTIVATED, COMPLETE PR-RT-IN)

MOSAIC GAGPOL1 V4 (AA SEQUENCE)

SEQ ID NO:24

MGARASVLSGGELDRWEKIRLRPGGKKKYRLKHIVWASRELERFAVNPGLLETSEGC
RQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALEKIEEEQNKSKKK
AQQAAADTGNSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPM
FSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQM
REPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPVSILDI
RQGPKEPFRDYVDRFYKTLRAEQASQDVKNWMTETLLVQNANPDCKTILKALGPAAT
LEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKTVKCFNCGKEG
HIAKNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSNKGRPGNFLQNRPEPT
APPEESFRFGEETTTPSQKQEPIDKEMYPLASLKSLFGNDPSSQRENLAFQQGEARE
FPSEQTRANSPTSRELQVRGDNPHSEAGAERQGTLNFPQITLWQRPLVSIKVGGQIR
EALL**A**TGADDTVLEDINLPGKWKPKMIGGIGGFIKV**G**QYDQILIEICGKKAIGTVLV
GPTPVNIIGRNMLTQLGCTLNFPISPIETVPVKLKPGMDGPRVKQWPLTEEKIKALT
AICEEMEKEGKITKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQL
GIPHPAGLKKKKSVTVLDVGDAYFSVPLDEGFRKYTAFTIPSTNNETPGIRYQYNVL
PQGWKGSPAIFQCSMTRILEPFRAKNPEIVIYQYMD**H**LYVGSDLEIGQHRAKIEELR
EHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQLPEKDSWTVNDIQKLVGK
LNWASQIYPGIKVRQLCKLLRGAKALTDIVPLTEEAELELAENREILKEPVHGVYYD
PSKDLIAEIQKQGHDQWTYQIYQEPFKNLKTGKYAKMRTAHTNDVKQLTEAVQKIAM
ESIVIWGKTPKFRLPIQKETWETWWTDYWQATWIPEWEFVNTPPLVKLWYQLEKDPI
AGVETFYVDGAANRETKLGKAGYVTDRGRQKIVSLTETTNQKTELQAIYLALQDSGS
EVNIVTDSQYALGIIQAQPDKSESELVNQIIEQLIKKERVYLSWVPAHKGIGGNEQV
DKLVSSGIRKVLFLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIVASCDQCQL
KGEAMHGQVDCSPGIWQL**A**CTHLEGKIILVAVHVASGYIEAEVIPAETGQETAYFIL
KLAGRWPVKVIHTDNGSNFTSAAVKAACWWAGIQQEFGIPYNPQSQGVVESMNKELK
KIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIIDIIATDIQTKELQKQ
IIKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKVKIIKDYGK
QMAGADCVAGRQDED

MOSAIC GAGPOL2 V4 (AA SEQUENCE)

SEQ ID NO:25

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLVWASRELERFALNPGLLETSEGC
KQIIKQLQPALQTGTEELRSLFNTVATLYCVHAEIEVRDTKEALDKIEEEQNKSQQK
TQQAKEADGKVSQNYPIVQNLQGQMVHQPISPRTLNAWVKVIEEKAFSPEVIPMFTA
LSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGPVAPGQMREP
RGSDIAGTTSNLQEQIAWMTSNPPIPVGDIYKRWIILGLNKIVRMYSPTSILDIKQG
PKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNANPDCKTILRALGPGATLEE

MMTACQGVGGPSHKARVLAEAMSQTNSTILMQRSNFKGSKRIVKCFNCGKEGHIARN
CRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPAE
SFRFEETTPAPKQEPKDREPLTSLRSLFGSDPLSQRENLAFPQGKAREFSSEQTRAN
SPTRRELQVWGRDNNSLSEAGADRQGTVSFSFPQITLWQRPLVTIKIGGQLKEALL**A**
TGADDTVLEEMNLPGRWKPKMIGGIGGFIKV**G**QYDQIPIEICGHKAIGTVLVGPTPV
NIIGRNLLTQIGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTE
MEKEGKISKIGPENPYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHP
AGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWK
GSPAIFQSSMTKILEPFRKQNPDIVIYQYMD**H**LYVGSDLEIGQHRTKIEELRQHLLR
WGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWAS
QIYAGIKVKQLCKLLRGTKALTEVVPLTEEAELELAENREILKEPVHGVYYDPSKDL
IAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKIATESIVI
WGKTPKFKLPIQKETWEAWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAET
FYVDGAANRETKLGKAGYVTDRGRQKVVSLTDTTNQKTELQAIHLALQDSGLEVNIV
TDSQYALGIIQAQPDKSESELVSQIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVS
RGIRKVLFLDGIDKAQEEHEKYHSNWRAMASEFNLPPIVAKEIVASCDKCQLKGEAI
HGQVDCSPGIWQL**A**CTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLAGR
WPVKTIHTDNGSNFTSATVKAACWWAGIKQEFGIPYNPQSQGVVESINKELKKIIGQ
VRDQAEHLKTAVQMAVFIHNFKRKGGIGEYSAGERIVDIIASDIQTKELQKQITKIQ
NFRVYYRDSRDPLWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGD
DCVASRQDED

IX. 2-VALENT M MOSAIC GAGPOL FUSION SEQUENCES (VERSION 5; POL MINIMALLY INACTIVATED, PR-DELETED)

MOSAIC GAGPOL1 V5 (AA SEQUENCE)

SEQ ID NO:26

MGARASVLSGGELDRWEKIRLRPGGKKKYRLKHIVWASRELERFAVNPGLLETSEGC
RQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALEKIEEEQNKSKKK
AQQAAADTGNSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPM
FSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQM
REPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPVSILDI
RQGPKEPFRDYVDRFYKTLRAEQASQDVKNWMTETLLVQNANPDCKTILKALGPAAT
LEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKTVKCFNCGKEG
HIAKNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSNKGRPGNFLQNRPEPT
APPEESFRFGEETTTPSQKQEPIDKEMYPLASLKSLFGNDPSSQMAPISPIETVPVK
LKPGMDGPRVKQWPLTEEKIKALTAICEEMEKEGKITKIGPENPYNTPVFAIKKKDS
TKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEGFR
KYTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQCSMTRILEPFRAKNPEIVIYQ
YMD**H**LYVGSDLEIGQHRAKIEELREHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKW
TVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGAKALTDIVPLT
EEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGHDQWTYQIYQEPFKNLKTGK
YAKMRTAHTNDVKQLTEAVQKIAMESIVIWGKTPKFRLPIQKETWETWWTDYWQATW
IPEWEFVNTPPLVKLWYQLEKDPIAGVETFYVDGAANRETKLGKAGYVTDRGRQKIV
SLTETTNQKTELQAIYLALQDSGSEVNIVTDSQYALGIIQAQPDKSESELVNQIIEQ
LIKKERVYLSWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKAQEEHEKYHSNWRA
MASDFNLPPVVAKEIVASCDQCQLKGEAMHGQVDCSPGIWQL**A**CTHLEGKIILVAVH

VASGYIEAEVIPAETGQETAYFILKLAGRWPVKVIHTDNGSNFTSAAVKAACWWAGI
QQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIG
GYSAGERIIDIIATDIQTKELQKQIIKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAV
VIQDNSDIKVVPRRKVKIIKDYGKQMAGADCVAGRQDED

MOSAIC GAGPOL2 V5 (AA SEQUENCE)

SEQ ID NO:27

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLVWASRELERFALNPGLLETSEGC
KQIIKQLQPALQTGTEELRSLFNTVATLYCVHAEIEVRDTKEALDKIEEEQNKSQQK
TQQAKEADGKVSQNYPIVQNLQGQMVHQPISPRTLNAWVKVIEEKAFSPEVIPMFTA
LSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGPVAPGQMREP
RGSDIAGTTSNLQEQIAWMTSNPPIPVGDIYKRWIILGLNKIVRMYSPTSILDIKQG
PKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNANPDCKTILRALGPGATLEE
MMTACQGVGGPSHKARVLAEAMSQTNSTILMQRSNFKGSKRIVKCFNCGKEGHIARN
CRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPAE
SFRFEETTPAPKQEPKDREPLTSLRSLFGSDPLSQMAPISPIETVPVKLKPGMDGPK
VKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPIFAIKKKDSTKWRKLVDF
RELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPS
INNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMD**H**LYVGS
DLEIGQHRTKIEELRQHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPE
KDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRGTKALTEVVPLTEEAELELAE
NREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHT
NDVKQLTEAVQKIATESIVIWGKTPKFKLPIQKETWEAWWTEYWQATWIPEWEFVNT
PPLVKLWYQLEKEPIVGAETFYVDGAANRETKLGKAGYVTDRGRQKVVSLTDTTNQK
TELQAIHLALQDSGLEVNIVTDSQYALGIIQAQPDKSESELVSQIIEQLIKKEKVYL
AWVPAHKGIGGNEQVDKLVSRGIRKVLFLDGIDKAQEEHEKYHSNWRAMASEFNLPP
IVAKEIVASCDKCQLKGEAIHGQVDCSPGIWQL**A**CTHLEGKVILVAVHVASGYIEAE
VIPAETGQETAYFLLKLAGRWPVKTIHTDNGSNFTSATVKAACWWAGIKQEFGIPYN
PQSQGVVESINKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGEYSAGERIV
DIIASDIQTKELQKQITKIQNFRVYYRDSRDPLWKGPAKLLWKGEGAVVIQDNSDIK
VVPRRKAKIIRDYGKQMAGDDCVASRQDED

X. 2-VALENT M MOSAIC GAGPOLNEF FUSION SEQUENCES (POL EXTENSIVELY INACTIVATED, PR-DELETED)

MOSAIC GAGPOLNEF1 (AA SEQUENCE)

SEQ ID NO:28

MGARASVLSGGELDRWEKIRLRPGGKKKYRLKHIVWASRELERFAVNPGLLETSEGC
RQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALEKIEEEQNKSKKK
AQQAAADTGNSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPM
FSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQM
REPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPVSILDI
RQGPKEPFRDYVDRFYKTLRAEQASQDVKNWMTETLLVQNANPDCKTILKALGPAAT
LEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKTVKCFNCGKEG
HIAKNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSNKGRPGNFLQNRPEPT

APPEESFRFGEETTTPSQKQEPIDKEMYPLASLKSLFGNDPSSQ**MA**PISPIETVPVK
LKPGMDGPRVKQWPLTEEKIKALTAICEEMEKEGKITKIGPENPYNTPVFAIKKKDS
TKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVL**A**VGDAYFSVPLDEGFR
KYTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQCSMTRILEPFRAKNPEIVIYQ
YM**AA**LYVGSDLEIGQHRAKIEELREHLLKWGFTTPDKKHQKEPPFLWMGYELHPDKW
TVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGAKALTDIVPLT
EEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGHDQWTYQIYQEPFKNLKTGK
YAKMRTAHTNDVKQLTEAVQKIAMESIVIWGKTPKFRLPIQKETWETWWTDYWQATW
IPEWEFVNTPPLVKLWYQLEKDPIAGVETFYV**A**GAANRETKLGKAGYVTDRGRQKIV
SLTETTNQKT**A**LQAIYLALQDSGSEVNIVT**A**SQYALGIIQAQPDKSESELVNQIIEQ
LIKKERVYLSWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKAQEEHEKYHSNWRA
MASDFNLPPVVAKEIVASCDQCQLKGEAMHGQVDCSPGIWQL**A**CTHLEGKIILVAVH
VASGYIEAEVIPAETGQETAYFILKLAGRWPVKVIHT**A**NGSNFTSAAVKAACWWAGI
QQEFGIPYNPQSQGVV**A**SMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIG
GYSAGERIIDIIATDIQTKELQKQIIKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAV
VIQDNSDIKVVPRRKVKIIKDYGKQMAGADCVAGRQDEDM**A**GKWSKSSVVGWPAIRE
RMRRAEPAADGVGAVSRDLEKHGAITSSNTAANNADCAWLEAQEEEEVGFPVRPQVP
LRPMTYKGALDLSHFLKEKGGLEGLIYSQKRQDILDLWVYHTQGYFPDWQNYTPGPG
IRYPLTFGWCFKLVPVEPEKIEEANEGENNSLLHPMSQHGMDDPEKEVLMWKFDSRL
AFHHMARELHPEYYKDC

MOSAIC GAGPOLNEF2 (AA SEQUENCE)

SEQ ID NO:29

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLVWASRELERFALNPGLLETSEGC
KQIIKQLQPALQTGTEELRSLFNTVATLYCVHAEIEVRDTKEALDKIEEEQNKSQQK
TQQAKEADGKVSQNYPIVQNLQGQMVHQPISPRTLNAWVKVIEEKAFSPEVIPMFTA
LSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGPVAPGQMREP
RGSDIAGTTSNLQEQIAWMTSNPPIPVGDIYKRWIILGLNKIVRMYSPTSILDIKQG
PKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNANPDCKTILRALGPGATLEE
MMTACQGVGGPSHKARVLAEAMSQTNSTILMQRSNFKGSKRIVKCFNCGKEGHIARN
CRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPAE
SFRFEETTPAPKQEPKDREPLTSLRSLFGSDPLSQ**MA**PISPIETVPVKLKPGMDGPK
VKQWPLTEEKIKALVEICTEMEKEGKISKIGPENPYNTPIFAIKKKDSTKWRKLVDF
RELNKRTQDFWEVQLGIPHPAGLKKKKSVTVL**A**VGDAYFSVPLDEDFRKYTAFTIPS
INNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYM**AA**LYVGS
DLEIGQHRTKIEELRQHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPE
KDSWTVNDIQKLVGKLNWASQIYAGIKVKQLCKLLRGTKALTEVVPLTEEAELELAE
NREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHT
NDVKQLTEAVQKIATESIVIWGKTPKFKLPIQKETWEAWWTEYWQATWIPEWEFVNT
PPLVKLWYQLEKEPIVGAETFYV**A**GAANRETKLGKAGYVTDRGRQKVVSLTDTTNQK
T**A**LQAIHLALQDSGLEVNIVT**A**SQYALGIIQAQPDKSESELVSQIIEQLIKKEKVYL
AWVPAHKGIGGNEQVDKLVSRGIRKVLFLDGIDKAQEEHEKYHSNWRAMASEFNLPP
IVAKEIVASCDKCQLKGEAIHGQVDCSPGIWQL**A**CTHLEGKVILVAVHVASGYIEAE
VIPAETGQETAYFLLKLAGRWPVKTIHT**A**NGSNFTSATVKAACWWAGIKQEFGIPYN
PQSQGVV**A**SINKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGEYSAGERIV
DIIASDIQTKELQKQITKIQNFRVYYRDSRDPLWKGPAKLLWKGEGAVVIQDNSDIK
VVPRRKAKIIRDYGKQMAGDDCVASRQDEDM**A**GKWSKSSIVGWPAVRERIRRAEPAA
EGVGAASQDLDKYGALTSSNTAATNADCAWLEAQEDEEVGFPVKPQVPLRPMTYKAA

FDLSFFLKEKGGLDGLIYSKKRQEILDLWVYNTQGFFPDWQNYTPGPGVRYPLTFGW
CFKLVPVDPREVEEANKGENNCLLHPMNLHGMDDPEREVLVWRFDSRLAFHHMAREK
HPEYYKNC

XI. OPTIMAL CLADE C ENV GP160, GAG, POL, NEF SEQUENCES

OPTIMAL CLADE C ENV GP160 (SN90.90.SE364) (AA SEQUENCE)

SEQ ID NO:30

MRVTGMLRNCQPWWIWGILGFWMLLIYNVGGNLWVTVYYGVPVWKEAKTTLFCASDA
KAYEKEVHNVWATHACVPTDPNPQEMVLENVTEYFNMWKNDMVDQMHEDIISLWDQS
LKPCVKLTPLCVTLNCRNVTTSNNATSNDNPNGEIKNCSFNITTELRDKRRNEYALF
YRLDIVPLSGSKNSSNSSEYRLINCNTSAITQACPKVSFDPIPIHYCAPAGYAILKC
NNKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLLNGSLAEGEIIIRSENLTNNAKTII
VHLNESIEIVCARPNNNTRKSMRIGPGQTFYATGDIIGDIRQAHCNISGNWNATLEK
VKGKLQEHFPGKNISFEPSSGGDLEITTHSFNCRGEFFYCDTSKLFNGTTHTANSSI
TIQCRIKQIINMWQGVGRAIYAPPIAGNITCKSNITGLLLTRDGGTLNNDTEKFRPG
GGDMRDNWRSELYKYKVVEIKPLGIAPTKAKRRVVE**REKR**AVGIGAVFLGFLGAAGS
TMGAASITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQTRVLAIER
YLKDQQLLGIWGCSGKIICTTAVPWNTSWSNKSLEDIWDNMTWMQWDREINNYTSII
YSLLEESQNQQEKNEKDLLALDSWNNLWNWFNITKWLWYIKIFIMIVGGLIGLRIIF
AVLSIVNRVRQGYSPLSFQTLIPNPRGPDRLGRIEEEGGEQDRDRSIRLVNGFLAIA
WDDLRSLCLFSYRRLRDFILIVARAVELLIQRGWETLKYLGSL?QYWGLELKKSAIS
LLDTIAITVAEGTDRIIELVQRICRAISNIPRRIRQGFEAALQ

OPTIMAL CLADE C GAG (IN.70177) (AA SEQUENCE)

SEQ ID NO:31

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLVWASRELERFALNPGLLETSEGC
KQILKQLQPALQTGTEELRSLYNTVATLYCVHAGIEVRDTKEALDKIEEEQNKGQQK
TQQAKGADGKVSQNYPIVQNLQGQMVHQAISPRTLNAWVKVIEEKAFSPEVIPMFTA
LSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGPIAPGQMREP
RGSDIAGTTSTLQEQIAWMTNNPPVPVGDIYKRWIILGLNKIVRMYSPVSILDIKQG
PKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNANPDCKTILRALGPGATLEE
MMTACQGVGGPSHKARVLAEAMSQTGSTIMMQRSNFKGSKRIVKCFNCGKEGHIARN
CRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPAE
SFRFEETTPAPKQELKDREPLTSLKSLFGSDPLSQ

OPTIMAL CLADE C POL (ZA.04.04ZASK208B1) (AA SEQUENCE)

SEQ ID NO:32

FFRENLAFQQGEAREFPSEQARANSPTSREFQVRGDNPCSEAGVKGQGTLNFPQITL
WQRPLVSIKVGGQVKEALL**D**TGADDTVLEEINLPGKWKPKMIGGIGGFIKV**R**QYDQI
LIEICGKKAIGTVLVGPTPVNIIGRNMLTQLGCTLNFPISPIETVPVKLKPGMDGPK
IKQWPLTEEKIKALMAICEEMEKEGKITKIGPENPYNTPIFAIKKKDSTKWRKLVDF
RELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDESFRKYTAFTIPS
INNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILEPFRAKNPEIVIYQYMD**D**LYVGS

DLEIGQHRAKIEELREHLLRWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIQLPE
KDSWTVNDIQKLVGKLNWASQIYSGIKVRQLCKLLRGAKALTDIVPLTEEAELELAE
NREILKEPVHGVYYDPSKDLIAEIQKQGYDQWTYQIYQEPFKNLKTGKYAKMRTAHT
NDVKQLTEAVQKIALESIVIWGKTPKFRLPIQKETWEIWWTDYWQATWIPEWEFVNT
PPLVKLWYQLEKEPIAGAETFYVDGAANRETKIGKAGYVTDKGRQKIVTLTETTNQK
TELQAIQLALQDSGSEVNIVTDSQYALGIIQAQPDKSESELVNQIIEQLINKERVYL
SWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKAQEEHEKYHSNWRAMASEFNLPP
VVAKEIVASCDKCQLKGEAIHGQVDCSPGIWQL**D**CTHLEGKVILVAVHVASGYMEAE
VIPAETGQETAYYILKLAGRWPVKVIHTDNGSNFTSAAVKAACWWAGIQQEFGIPYN
PQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERII
DIIATDIQTKELQKQIIKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSDIK
VVPRRKVKIIKDYGKQMAGADCVAGRQDED

OPTIMAL CLADE C NEF (ZA00.1170MB) (AA SEQUENCE)

SEQ ID NO:33

MGGKWSKSSIVGWPDVRERMRRTEPAAEGVGAASQDLDKYGALTSSNTTHNNADCAW
LEAQEEGEVGFPVRPQVPLRPMTYKGAFDLSFFLKEKGGLDGLIYSKKRQEILDLWV
YHTQGFFPDWQNYTPGPGVRYPLTFGWCFKLVPVDPREVEEANKGENNCLLHPMSLH
GMEDEEREVLKWEFDSSLARRHLARELHPEYYKDC

XII. OPTIMAL CLADE C ENV GP140 SEQUENCE (CLEAVAGE/FUSION-DEFECTIVE)

OPTIMAL CLADE C ENV GP140 (SN90.90.SE364) (AA SEQUENCE)

SEQ ID NO:34

MRVTGMLRNCQPWWIWGILGFWMLLIYNVGGNLWVTVYYGVPVWKEAKTTLFCASDA
KAYEKEVHNVWATHACVPTDPNPQEMVLENVTEYFNMWKNDMVDQMHEDIISLWDQS
LKPCVKLTPLCVTLNCRNVTTSNNATSNDNPNGEIKNCSFNITTELRDKRRNEYALF
YRLDIVPLSGSKNSSNSSEYRLINCNTSAITQACPKVSFDPIPIHYCAPAGYAILKC
NNKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLLNGSLAEGEIIIRSENLTNNAKTII
VHLNESIEIVCARPNNNTRKSMRIGPGQTFYATGDIIGDIRQAHCNISGNWNATLEK
VKGKLQEHFPGKNISFEPSSGGDLEITTHSFNCRGEFFYCDTSKLFNGTTHTANSSI
TIQCRIKQIINMWQGVGRAIYAPPIAGNITCKSNITGLLLTRDGGTLNNDTEKFRPG
GGDMRDNWRSELYKYKVVEIKPLGIAPTKAKRRVVE**SEKS**AVGIGAVFLGFLGAAGS
TMGAASITLTVQARQLLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQTRVLAIER
YLKDQQLLGIWGCSGKIICTTAVPWNTSWSNKSLEDIWDNMTWMQWDREINNYTSII
YSLLEESQNQQEKNEKDLLALDSWNNLWNWFNITKWLW

XIII. OPTIMAL CLADE C POL SEQUENCE (EXTENSIVELY INACTIVATED, PR-DELETED)

OPTIMAL CLADE C POL (ZA.04.04ZASK208B1) (AA SEQUENCE)

SEQ ID NO:35

MAPISPIETVPVKLKPGMDGPKIKQWPLTEEKIKALMAICEEMEKEGKITKIGPENP
YNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVL**A**VG
DAYFSVPLDESFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKILE
PFRAKNPEIVIYQYM**AA**LYVGSDLEIGQHRAKIEELREHLLRWGFTTPDKKHQKEPP
FLWMGYELHPDKWTVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYSGIKVRQLCKLL
RGAKALTDIVPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGYDQWTYQ
IYQEPFKNLKTGKYAKMRTAHTNDVKQLTEAVQKIALESIVIWGKTPKFRLPIQKET
WEIWWTDYWQATWIPEWEFVNTPPLVKLWYQLEKEPIAGAETFYV**A**GAANRETKIGK
AGYVTDKGRQKIVTLTETTNQKT**A**LQAIQLALQDSGSEVNIVT**A**SQYALGIIQAQPD
KSESELVNQIIEQLINKERVYLSWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGIDKA
QEEHEKYHSNWRAMASEFNLPPVVAKEIVASCDKCQLKGEAIHGQVDCSPGIWQL**A**C
THLEGKVILVAVHVASGYMEAEVIPAETGQETAYYILKLAGRWPVKVIHT**A**NGSNFT
SAAVKAACWWAGIQQEFGIPYNPQSQGVV**A**SMNKELKKIIGQVRDQAEHLKTAVQMA
VFIHNFKRKGGIGGYSAGERIIDIIATDIQTKELQKQIIKIQNFRVYYRDSRDPIWK
GPAKLLWKGEGAVVIQDNSDIKVVPRRKVKIIKDYGKQMAGADCVAGRQDED

XIV. OPTIMAL CLADE C GAGNEF FUSION SEQUENCE

OPTIMAL CLADE C GAGNEF (IN.70177-ZA00.1170MB) (AA SEQUENCE)

SEQ ID NO:36

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLVWASRELERFALNPGLLETSEGC
KQILKQLQPALQTGTEELRSLYNTVATLYCVHAGIEVRDTKEALDKIEEEQNKGQQK
TQQAKGADGKVSQNYPIVQNLQGQMVHQAISPRTLNAWVKVIEEKAFSPEVIPMFTA
LSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVHAGPIAPGQMREP
RGSDIAGTTSTLQEQIAWMTNNPPVPVGDIYKRWIILGLNKIVRMYSPVSILDIKQG
PKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNANPDCKTILRALGPGATLEE
MMTACQGVGGPSHKARVLAEAMSQTGSTIMMQRSNFKGSKRIVKCFNCGKEGHIARN
CRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPAE
SFRFEETTPAPKQELKDREPLTSLKSLFGSDPLSQAGKWSKSSIVGWPDVRERMRRT
EPAAEGVGAASQDLDKYGALTSSNTTHNNADCAWLEAQEEGEVGFPVRPQVPLRPMT
YKGAFDLSFFLKEKGGLDGLIYSKKRQEILDLWVYHTQGFFPDWQNYTPGPGVRYPL
TFGWCFKLVPVDPREVEEANKGENNCLLHPMSLHGMEDEEREVLKWEFDSSLARRHL
ARELHPEYYKDC

XV. CONSENSUS SEQUENCES

M CONSENSUS ENV

SEQ ID NO:37

MRVRGIQRNCQHLWRWGTLILGMLMICSAAENLWVTVYYGVPVWKEANTTLF
CASDAKAYDTEVHNVWATHACVPTDPNPQEIVLENVTENFNMWKNNMVEQM
HEDIISLWDQSLKPCVKLTPLCVTLNCTNVNVTNTTNNTEEKGEIKNCSFNITTEI
RDKKQKVYALFYRLDVVPIDDNNNNSSNYRLINCNTSAITQACPKVSFEPIPIHYC
APAGFAILKCNDKKFNGTGPCKNVSTVQCTHGIKPVVSTQLLLNGSLAEEEIIIRS
ENITNNAKTIIVQLNESVEINCTRPNNNTRKSIRIGPGQAFYATGDIIGDIRQAHCN
I
SGTKWNKTLQQVAKKLREHFNNKTIIFKPSSGGDLEITTHSFNCRGEFFYCNTSG
LFNSTWIGNGTKNNNNTNDTITLPCRIKQIINMWQGVGQAMYAPPIEGKITCKSNI
TGLLLTRDGGNNNTNETEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAK
RRVVESEKSAVGIGAVFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLR
AIEAQQHLLQLTVWGIKQLQARVLAVERYLKDQQLLGIWGCSGKLICTTTVPWN
SSWSNKSQDEIWDNMTWMEWEREINNYTDIIYSLIEESQNQQEKNEQELLALDK
WASLWNWFDITNWLW

M CONSENSUS GAG

SEQ ID NO:38

MGARASVLSGGKLDAWEKIRLRPGGKKKYRLKHLVWASRELERFALNPGLLET
SEGCKQIIGQLQPALQTGSEELRSLYNTVATLYCVHQRIEVKDTKEALEKIEEEQN
KSQQKTQQAAADKGNSSKVSQNYPIVQNLQGQMVHQAISPRTLNAWVKVIEEK
AFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRL
HPVHAGPIPPGQMREPRGSDIAGTTSTLQEQIAWMTSNPPIPVGEIYKRWIILGLN
KIVRMYSPVSILDIRQGPKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQN
ANPDCKTILKALGPGATLEEMMTACQGVGGPGHKARVLAEAMSQVTNAAIMM
QRGNFKGQRRIIKCFNCGKEGHIARNCRAPRKKGCWKCGKEGHQMKDCTERQA
NFLGKIWPSNKGRPGNFLQSRPEPTAPPAESFGFGEEITPSPKQEPKDKEPPLTSLK
SLFGNDPLSQ

M CONSENSUS POL

SEQ ID NO:39

MAPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALTEICTEMEKEGKISKIGPEN
PYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTV
LDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSM
TKILEPFRTQNPEIVIYQYMDHLYVGSDLEIGQHRAKIEELREHLLRWGFTTPDKK
HQKEPPFLWMGYELHPDKWTVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYPGI
KVKQLCKLLRGAKALTDIVPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQ

KQGQDQWTYQIYQEPFKNLKTGKYAKMRSAHTNDVKQLTEAVQKIATESIVIW
GKTPKFRLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIAG
AETFYVDGAANRETKLGKAGYVTDRGRQKVVSLTETTNQKTELQAIHLALQDS
GSEVNIVTDSQYALGIIQAQPDKSESELVNQIIEQLIKKEKVYLSWVPAHKGIGGN
EQVDKLVSTGIRKVLFLDGIDKAQEEHEKYHSNWRAMASDFNLPPIVAKEIVASC
DKCQLKGEAMHGQVDCSPGIWQLACTHLEGKIILVAVHVASGYIEAEVIPAETG
QETAYFILKLAGRWPVKVIHTDNGSNFTSAAVKAACWWAGIQQEFGIPYNPQSQ
GVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIIDI
IATDIQTKELQKQITKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSDIK
VVPRRKAKIIRDYGKQMAGDDCVAGRQDED

Other Embodiments

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure that come within known or customary practice within the art to which the invention pertains and may be applied to the essential features hereinbefore set forth.

All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication or patent application was specifically and individually indicated to be incorporated by reference in their entirety.

What is claimed is:

CLAIMS

1. A vaccine comprising:

(a) an optimized viral polypeptide having the amino acid sequence set forth in SEQ ID N0:9, or

(b) a vector encoding an optimized viral polypeptide having the amino acid sequence set forth in SEQ ID NO:9.

2. The vaccine according to claim 1, wherein said vaccine comprises said optimized viral polypeptide having the amino acid sequence set forth in SEQ ID NO: 9, and wherein said vaccine further comprises:

(a) at least two distinct optimized gag polypeptides selected from any one or more of the groups: (i) SEQ ID NOs:3 and 4, (ii) SEQ ID NOs:3 and 15, and (iii) SEQ ID NOs:4 and 15; and/or

(b) at least two distinct optimized pol polypeptides selected from any one or more of the groups: (i) SEQ ID NOs:12 and 13, (ii) SEQ ID NOs:12 and 14, and (iii) SEQ ID NOs:13 and 14.

3. The vaccine according to claim 2, wherein said at least two distinct optimized gag polypeptides comprise the amino acid sequences set forth in SEQ ID NOs:3 and 4, respectively, and/or

wherein said at least two distinct optimized pol polypeptides comprise the amino acid sequences set forth in SEQ ID NOs:12 and 13, respectively.

4. The vaccine according to claim 1, wherein said vaccine comprises said vector encoding an optimized viral polypeptide having the amino acid sequence set forth in SEQ ID NO:9, and wherein said vaccine further comprises one or more vectors encoding:

at least two distinct optimized gag polypeptides selected from any one or more of the groups: (i) SEQ ID NOs:3 and 4, (ii) SEQ ID NOs:3 and 15, and (iii) SEQ ID NOs:4 and 15, and/or

at least two distinct optimized pol polypeptides selected from any one or more of the groups: (i) SEQ ID NOs:12 and 13, (ii) SEQ ID NOs:12 and 14, and (iii) SEQ ID NOs:13 and 14.

5. The vaccine according to claim 4, wherein said vaccine comprises one or more vectors encoding:

at least two distinct optimized gag polypeptides comprising the amino acid sequences set forth in SEQ ID NOs:3 and 4, respectively, and/or

at least two distinct optimized pol polypeptides comprising the amino acid sequences set forth in SEQ ID NOs:12 and 13, respectively.

6. The vaccine according to claim 5, wherein said vaccine comprises one or more vectors encoding:

at least two distinct optimized gag polypeptides comprising the amino acid sequences set forth in SEQ ID NOs:3 and 4, respectively, and

at least two distinct optimized pol polypeptides comprising the amino acid sequences set forth in SEQ ID NOs:12 and 13, respectively.

7. The vaccine according to claim 5 or 6, wherein said vaccine further comprises a vector encoding an optimized env polypeptide having the sequence set forth in SEQ ID NO:10.

8. The vaccine according to any one of claims 4 to 7, wherein said optimized gag, pol, and/or env polypeptides are encoded by (i) a single vector or (ii) multiple vectors.

9. The vaccine of any one of claims 1 and 4 to 8, wherein said vaccine comprises a viral vector selected from the group consisting of adenovirus serotype 26 (Ad26), adenovirus serotype 34 (Ad34), adenovirus serotype 35 (Ad35), adenovirus serotype 48 (Ad48), or adenovirus serotype 5 HVR48 (Ad5HVR48), poxvirus, and modified vaccinia virus Ankara (MVA).

10. The vaccine of claim 9, wherein said viral vector is an adenovirus serotype 26 (Ad26).

11. The vaccine of claim 9, wherein said viral vector is a modified vaccinia virus Ankara (MVA).

12. The vaccine of any one of claims 1 to 11, wherein said vaccine further comprises a pharmaceutically acceptable carrier, excipient, or diluent.

13. The vaccine of any one of claims 1 to 12 for use in a method of treating or reducing the risk of a human immunodeficiency virus type 1 (HIV-1) infection in a human.

14. A method of treating or reducing the risk of an HIV-1 infection in a human, said method comprising administering the vaccine of any one of claims 1 to 12 to said human.

15. A method of manufacturing a vaccine for treating or reducing the risk of an HIV-1 infection in a human, said method comprising synthesizing the vaccine of any one of claims 1 to 12.

16. A kit comprising:

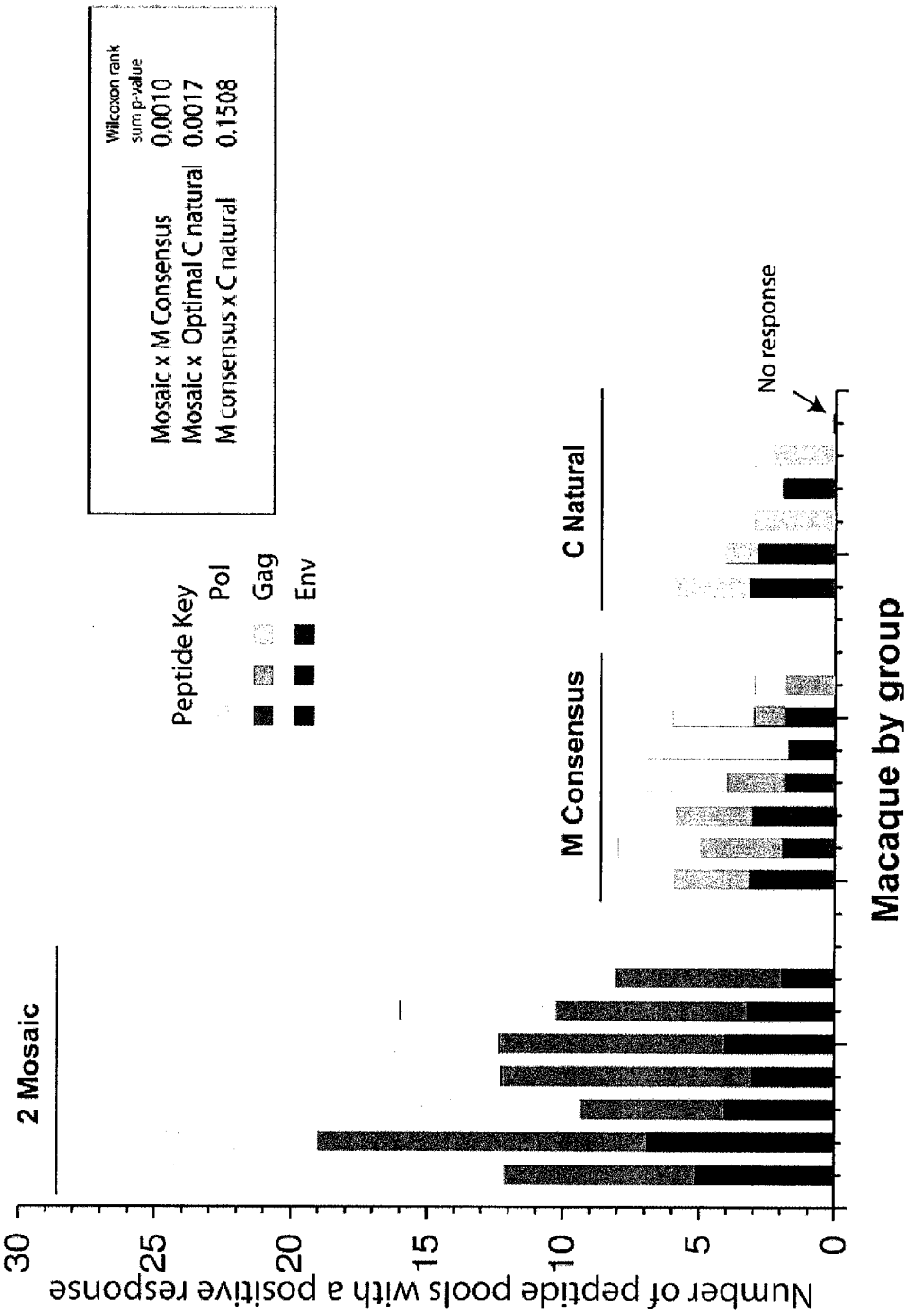
(a) the vaccine of any one of claims 1 to 11;

(b) a pharmaceutically acceptable carrier, excipient, or diluent;

(c) instructions for the use thereof; and, optionally,

(d) an adjuvant.

Figure 1


Number of Positive Pools: 4x background
2 Mosaic
Peptide Key
Pol
Gag
Env
Wilcoxon rank sum p-value
Mosaic x M Consensus 0.0010
Mosaic x Optimal C natural 0.0017
M consensus x C natural 0.1508
M Consensus
C Natural
No response
Number of peptide pools with a positive response
0
5
10
15
20
25
30
Macaque by group

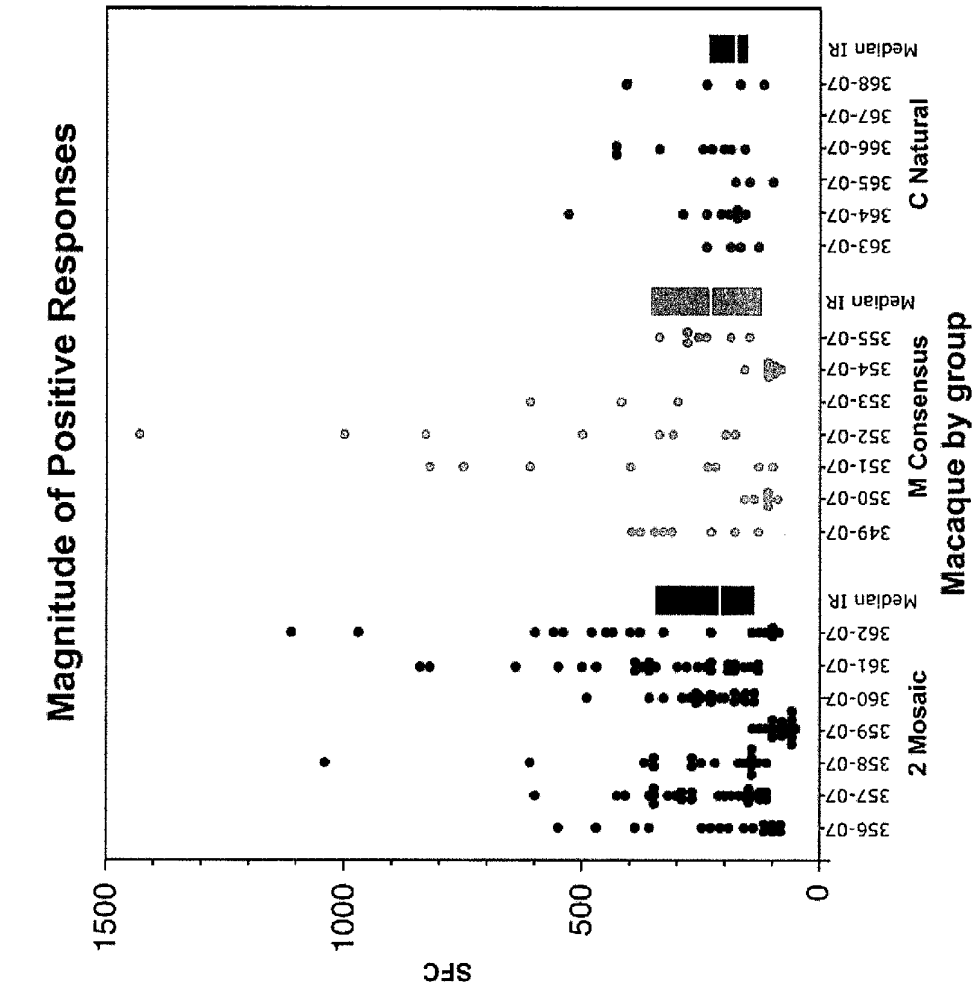

Magnitude of Positive Responses
SFC
1500
1000
500
0
356-07
357-07
358-07
359-07
360-07
361-07
362-07
Median IR
2 Mosaic
349-07
350-07
351-07
352-07
353-07
354-07
355-07
Median IR
M Consensus
363-07
364-07
365-07
366-07
367-07
368-07
Median IR
C Natural
Macaque by group
Mos x Con, p = 0.56
Mos x C natural, p = 0.95
Con x C natural, p = 0.71

Figure 2

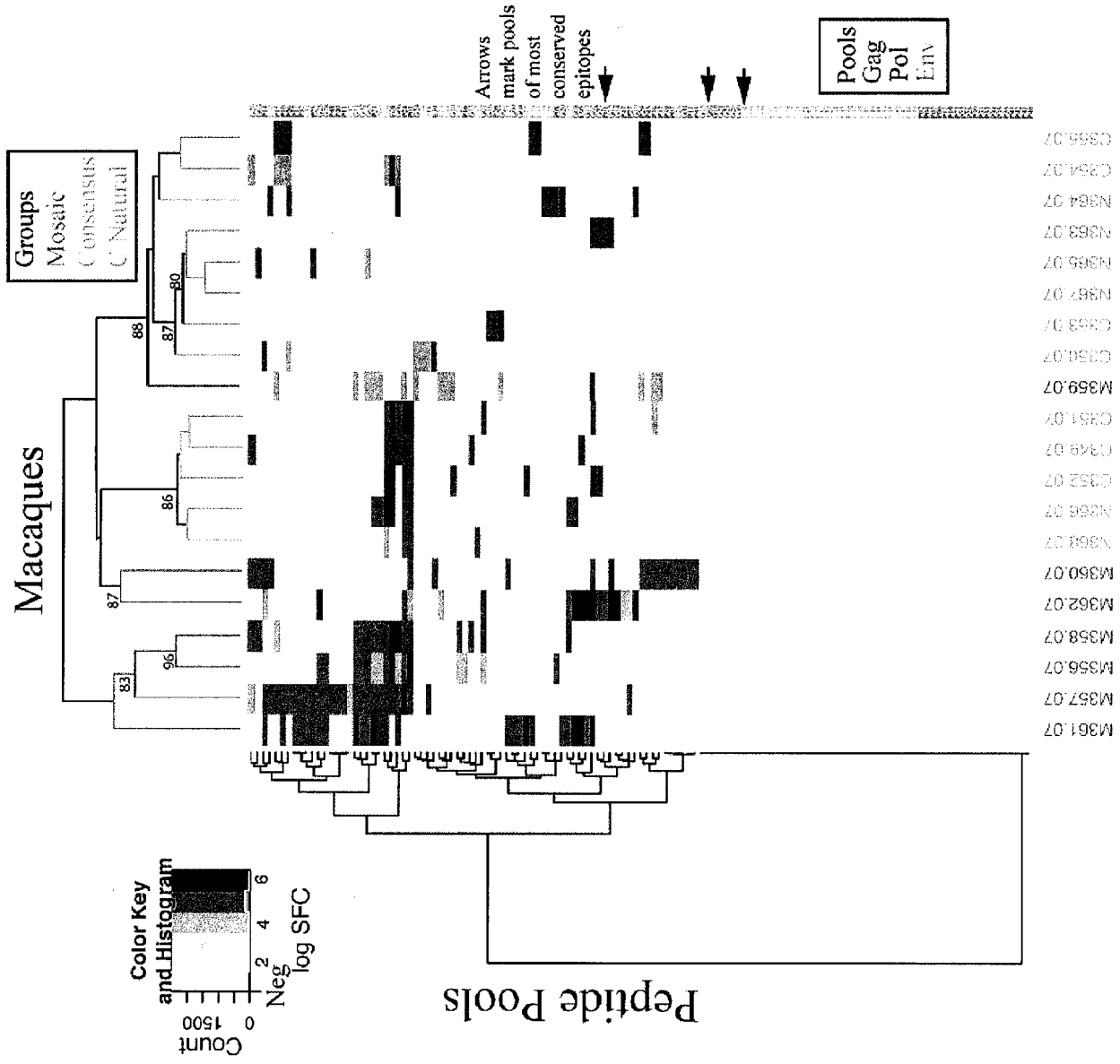

Groups
Mosaic
Consensus
Natural
Arrows
mark pools
of most
conserved
epitopes
Pools
Gag
Pol
Env
Macaques
Peptide Pools
Color Key
and Histogram
Count
0
1500
Neg
2
4
6
log SFC
C0556.07
C0541.07
N0364.07
N0363.07
N0366.07
N0367.07
C0358.07
C0350.07
M0359.07
C0351.07
C0349.07
C0352.07
N0366.07
N0368.07
M0360.07
M0362.07
M0358.07
M0356.07
M0357.07
M0361.07
88
87
80
86
87
96
83

Figure 3

Figure 4

PTE and M group 9-mer coverage of different vaccine candidates

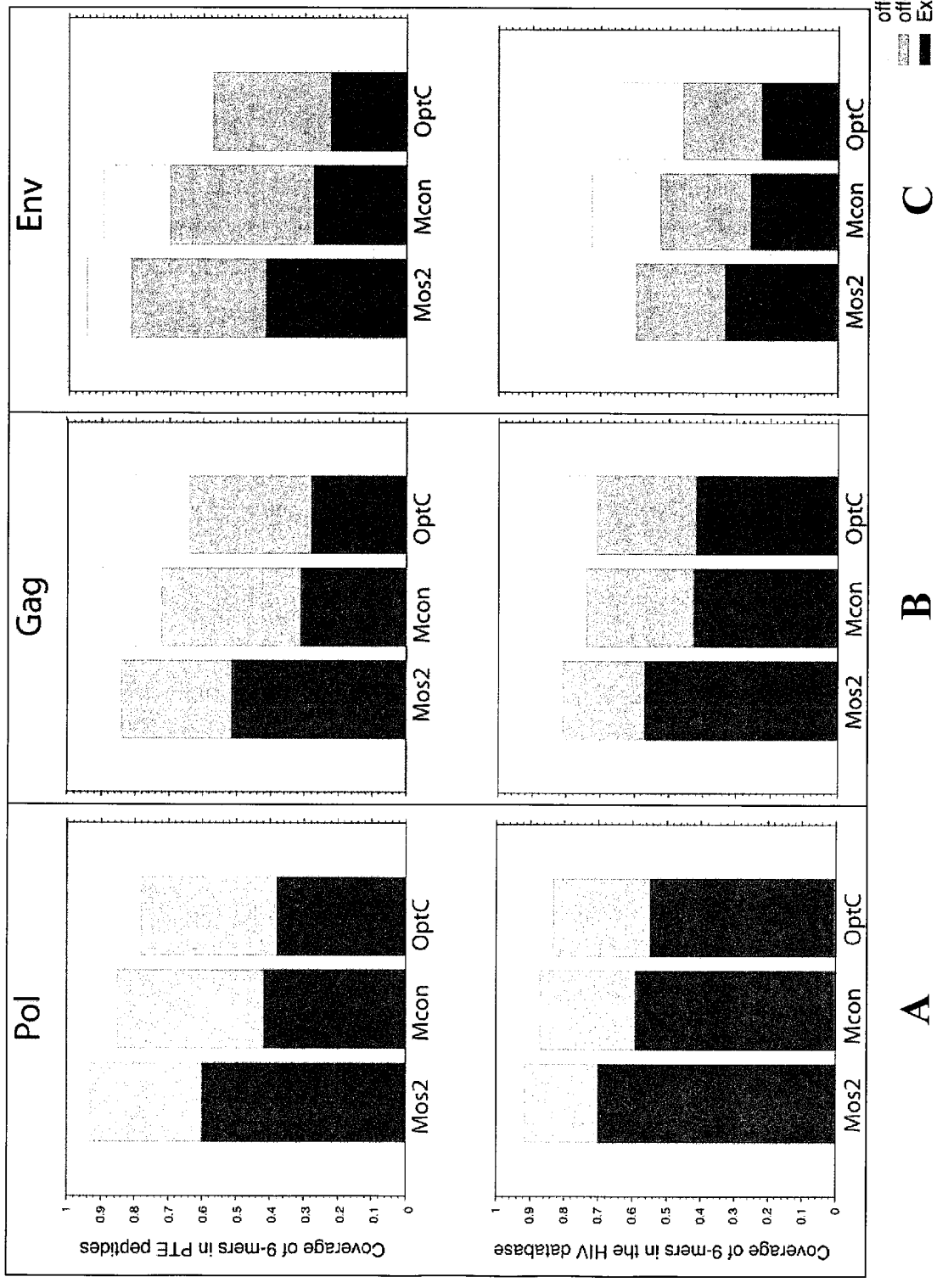


Figure 5

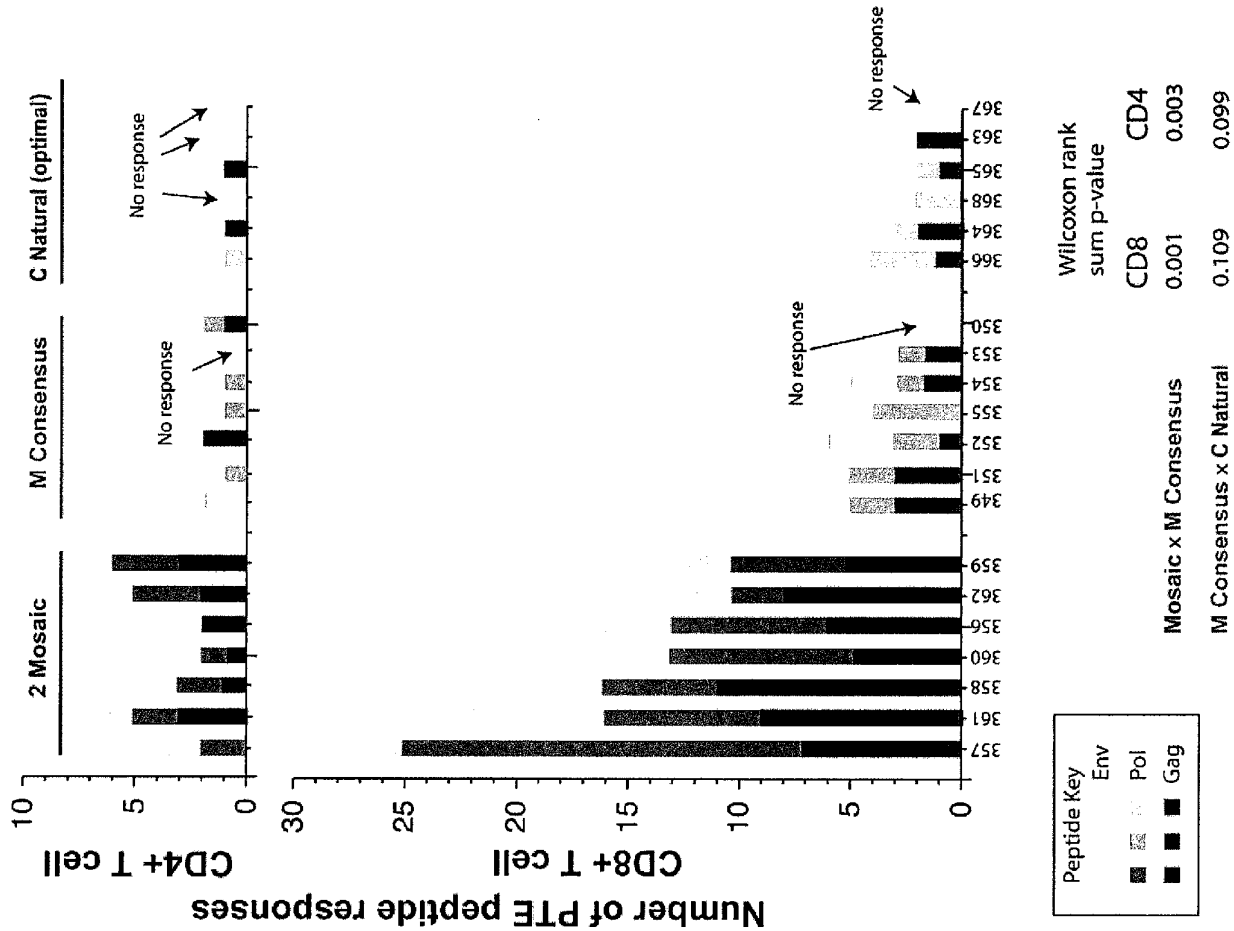

Number of PTE peptide responses
CD8+ T cell
CD4+ T cell
2 Mosaic
M Consensus
C Natural (optimal)
No response
No response
No response
No response
No response
Peptide Key
Env
Pol
Gag
Wilcoxon rank sum p-value
CD8 CD4
Mosaic x M Consensus 0.001 0.003
M Consensus x C Natural 0.109 0.099

Figure 6

- CD8 T cells: median (range)
  - 2 Mosaic: 16 (12-29)
  - Mcon: 6 (0-7)
  - OptC: 3 (0-7)
- CD4 T cells:
  - 2 Mosaic: 4 (2-6)
  - Mcon: 1 (0-2)
  - OptC: 0.5 (0-2)

Figure 7

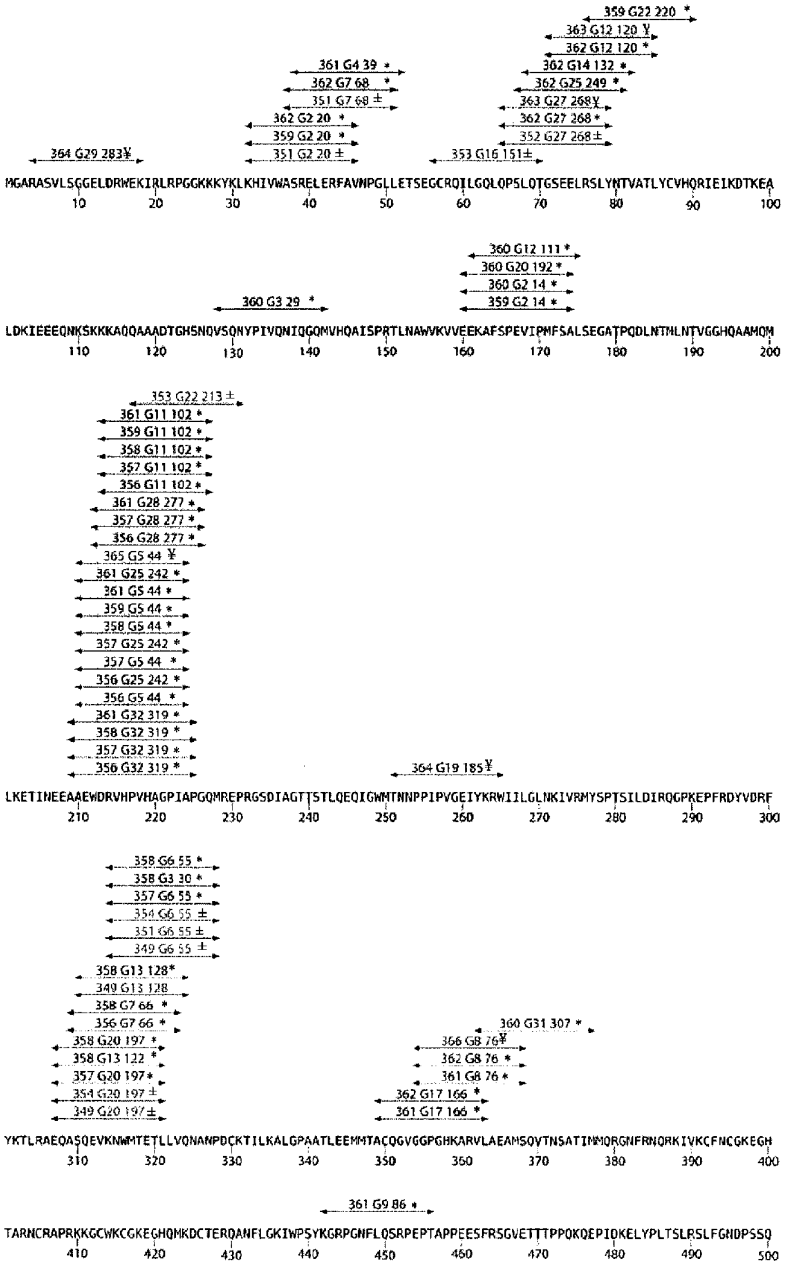

359 G22 220 *
363 G12 120 ¥
362 G12 120 *
362 G14 132 *
362 G25 249 *
363 G27 268¥
362 G27 268 *
352 G27 268±
361 G4 39 *
362 G7 68 *
351 G7 68 ±
362 G2 20 *
359 G2 20 *
351 G2 20 ±
364 G29 283¥
353 G16 151±
MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEA
10 20 30 40 50 60 70 80 90 100
360 G12 111 *
360 G20 192 *
360 G2 14 *
359 G2 14 *
360 G3 29 *
LDKIEEEQNKSKKKAQQAAADTGHSNQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQM
110 120 130 140 150 160 170 180 190 200
353 G22 213 ±
361 G11 102 *
359 G11 102 *
358 G11 102 *
357 G11 102 *
356 G11 102 *
361 G28 277 *
357 G28 277 *
356 G28 277 *
365 G5 44 ¥
361 G25 242 *
361 G5 44 *
359 G5 44 *
358 G5 44 *
357 G25 242 *
357 G5 44 *
356 G25 242 *
356 G5 44 *
361 G32 319 *
358 G32 319 *
357 G32 319 *
356 G32 319 *
364 G19 185¥
LKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
210 220 230 240 250 260 270 280 290 300
358 G6 55 *
358 G3 30 *
357 G6 55 *
354 G6 55 ±
351 G6 55 ±
349 G6 55 ±
358 G13 128*
349 G13 128
358 G7 66 *
356 G7 66 *
358 G20 197 *
358 G13 122 *
357 G20 197 *
354 G20 197 ±
349 G20 197 ±
360 G31 307 *
366 G8 76¥
362 G8 76 *
361 G8 76 *
362 G17 166 *
361 G17 166 *
YKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKIVKCFNCGKEGH
310 320 330 340 350 360 370 380 390 400
361 G9 86 *
TARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSYKGRPGNFLQSRPEPTAPPEESFRSGVETTTPPQKQEPIDKELYPLTSLRSLFGNDPSSQ
410 420 430 440 450 460 470 480 490 500

Figure 8

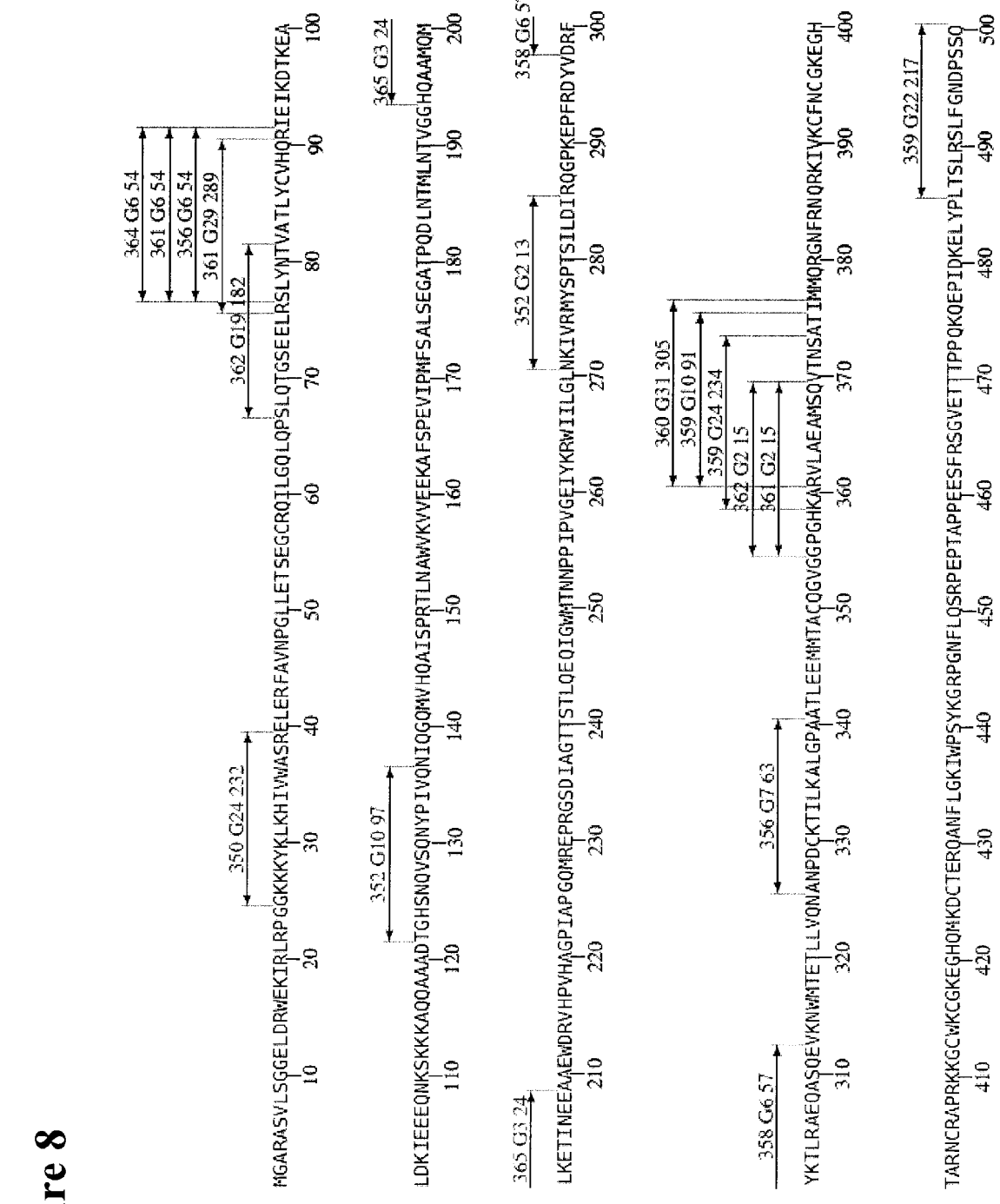

364 G6 54
361 G6 54
356 G6 54
361 G29 289
362 G19 182
350 G24 232
352 G10 97
365 G3 24
358 G6 57
352 G2 13
360 G31 305
359 G10 91
359 G24 234
362 G2 15
361 G2 15
356 G7 63
359 G22 217
MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEA
LDKIEEEQNKSKKKAQQAAADTGHSNQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQM
LKETIMEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
YKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKIVKCFNCGKEGH
TARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSYKGRPGNFLQSRPEPTAPPEESFRSGVETTTPPQKQEPIDKELYPLTSLRSLFGNDPSSQ

Figure 9

OptC 366-07:

5 CD8 responses:

OptC	I**VQQQSNLLRAIEAQQ**			
E54	VVQQQSNLLRAIEAQ	Env	548	562
E72	-VQQQNNLLRAIEAQH	Env	549	563
OptC	**AVFIHNFKRKGGIGGY**			
P22	AVFIHNFKRKGGIGG	Pol	894	908
P236	-VLIHNFKRKGGIGGY	Pol	895	909
OptC	**MAICEEMEKEGKITK**			
P224	TAICEEMEKEGKITK	Pol	190	204
OptC	**CIHGIKPVVSTQLLL**			
E15	CIHGIKPVVSTQLLL	Env	247	261
OptC	**GGPSHKARVLAEAMS**			
G76	GGPSHKARVLAEAMG	Gag	354	368

1 CD4 response:

OptC	**IIGQVRDQAEHLKTA**			
P86	LIGQVRDQAEHLKTA	Pol	876	890

Figure 10

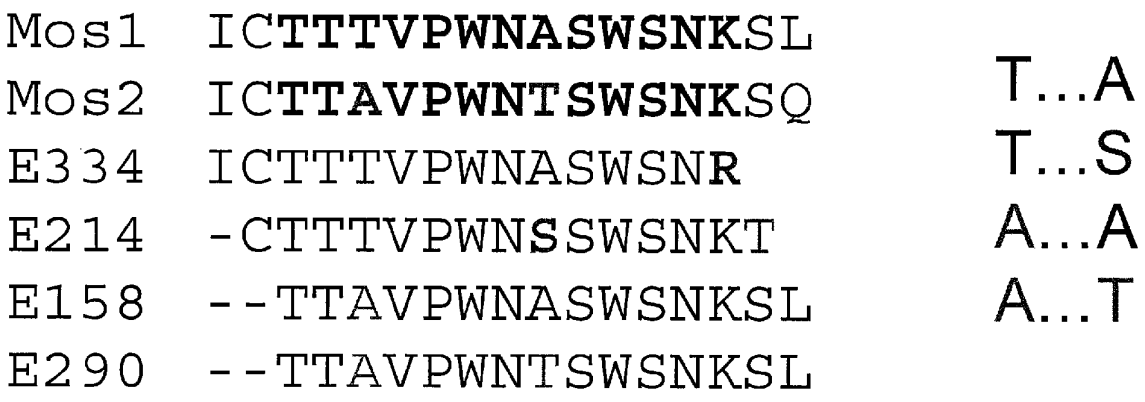

Mos1 ICTTTVPWNASWSNKSL
Mos2 ICTTAVPWNTSWSNKSQ
E334 ICTTTVPWNASWSNR
E214 -CTTTVPWNSSWSNKT
E158 --TTAVPWNASWSNKSL
E290 --TTAVPWNTSWSNKSL
T...A
T...S
A...A
A...T

Figure 11

8 CD8 responses:

Mos1	EQLIKK**ERVYLSWVP**AHKGIG			
Mos2	EQLIKK**EKVYLAWVP**AHKGIG			
P221	EELIKKEKVYLAWVP	Pol	678	692
P250	EPLIKKEKVYLSWVP	Pol	678	692
P316	-KLIEKDKVYLSWVPA	Pol	679	693
P223	--LIKKERVYLSWVPAH	Pol	680	694
P169	------EKVYLAWVPAHKGIG	Pol	684	698
P83	------EKVYLSWVPAHKGIG	Pol	684	698
Mos1	**CTHGIRPVVSTQLLL**			
Mos2	**CTHGIKPVVSTQLLL**			
E15	CTHGIKPVVSTQLLL	Env	247	261
E47	CTHGIRPVVSTQLLL	Env	247	261
Mos1	IC**TTTVPWNASWSNK**SL			
Mos2	IC**TTAVPWNTSWSNK**SQ			
E334	ICTTTVPWNASWSNR	Env	603	617
E214	-CTTTVPWNSSWSNKT	Env	604	618
E158	--TTAVPWNASWSNKSL	Env	605	619
E290	--TTAVPWNTSWSNKSL	Env	605	619
Mos1	ACQGV**GGPGHKARVL**AEAMS			
Mos2	ACQGV**GGPSHKARVL**AEAMS			
G166	ACQEVGGPGHKARVL	Gag	349	363
G76	-----GGPSHKARVLAEAMG	Gag	354	368
Mos1	AAEW**DRVHPVHAGPIA**PGQ			
Mos2	AAEW**DRLHPVHAGPVA**PGQ			
G319	AAE-DRLHPVHAGPIP	Gag	209	225
G242	-ADWDRLHPVHAGPVA	Gag	210	224
G44	-AEWDRLHPVHAGPIA	Gag	210	224
G277	---WDRVHPVHAGPNPPG	Gag	212	226
G102	----DRVHPVHAGPIPPGQ	Gag	212	226
Mos1	**HSNWRAMASDFNLPP**			
Mos2	**HSNWRAMASEFNLPP**			
P93	HSNWRAMASDFNLPP	Pol	731	745
Mos1	**KGRPGNFLQNRPEPT**			
Mos2	**KGRPGNFLQSRPEPT**			
G86	KGRPGNFLQNRPEPT	Gag	442	456
Mos1	**SRELERFAVNPGLLE**			
Mos2	**SRELERFALNPGLLE**			
G39	SRELERFALNPGLLE	Gag	38	52

5 CD4 responses:

Mos1	R**SLYNTVATLYCVHQ**R			
Mos2	R**SLFNTVATLYCVHA**E			
G289	KSLFNTVATLYCVHA	Gag	76	90
G54	-SLYNTVATLYCVHQR	Gag	77	91
Mos1	**YVTDRGRQKIVSLTE**			
Mos2	**YVTDRGRQKVVSLTD**			
P134	YVTDRGRQKVVSLTE	Pol		612
	626			
Mos1	**CTTTVPWNASWSNKS**			
Mos2	**CTTAVPWNTSWSNKS**			
E53	CTTNVPWNSSWSNKS	Env		604
	618			
Mos1	**WWAGIQQEFGIPYNP**			
Mos2	**WWAGIKQEFGIPYNP**			
P135	WWAGIQQEFGIPYNP	Pol		846
	860			
Mos1	**GPGHKARVLAEAMSQ**			
Mos2	**GPSHKARVLAEAMSQ**			
G15	GPGHKARVLAEAMSQ	Gag		355
	369			

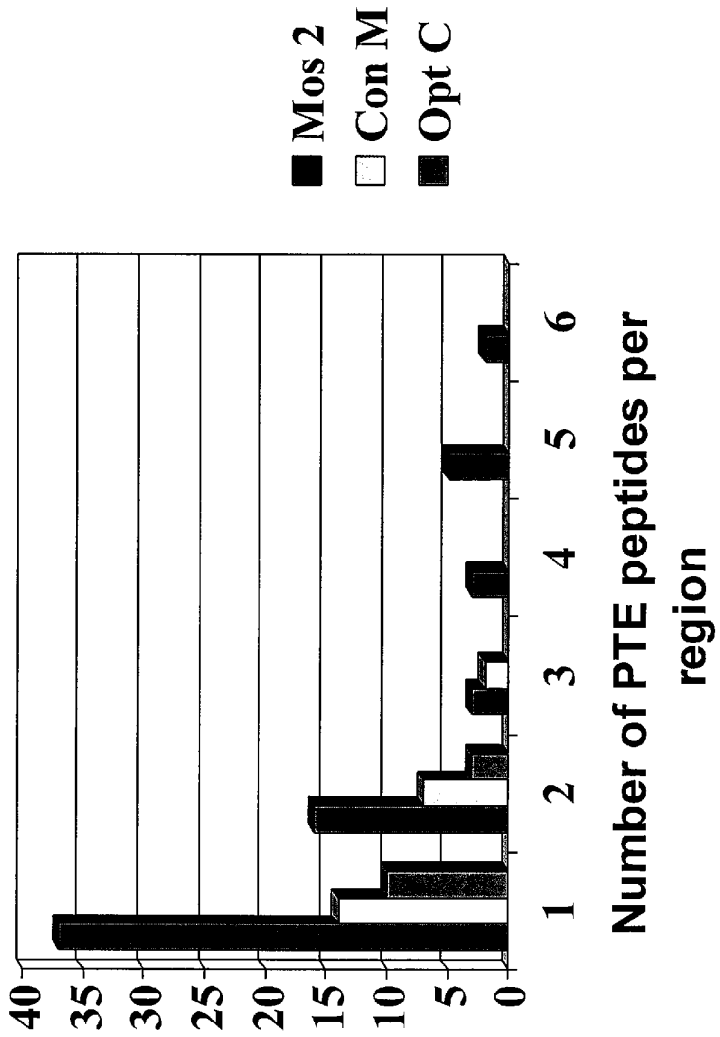

Mos 2
Con M
Opt C
Number of regions
40
35
30
25
20
15
10
5
0
1
2
3
4
5
6
Number of PTE peptides per region

Figure 12

Figure 13

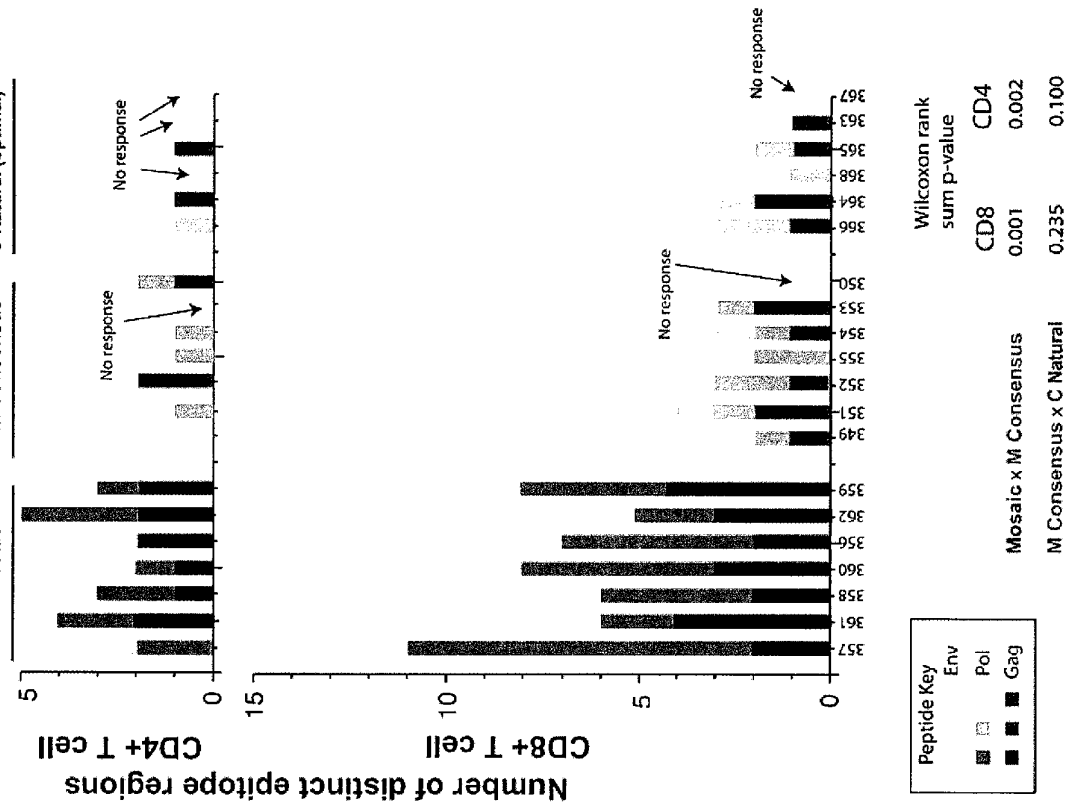

2 Mosaic
M Consensus
C Natural (optimal)
No response
No response
No response
No response
No response
CD4+ T cell
CD8+ T cell
Number of distinct epitope regions
Peptide Key
Env
Pol
Gag
Wilcoxon rank sum p-value
CD8
CD4
Mosaic x M Consensus
0.001
0.002
M Consensus x C Natural
0.235
0.100

Figure 14

- CD8 T cells: median (range)
  - 2 Mosaic: 8 (7-14)
  - Mcon: 3 (0-6)
  - OptC: 1.5 (0-5)
- CD4 T cells:
  - 2 Mosaic: 3 (2-5)
  - Mcon: 1 (0-2)
  - OptC: 0.5 (0-2)

Figure 15

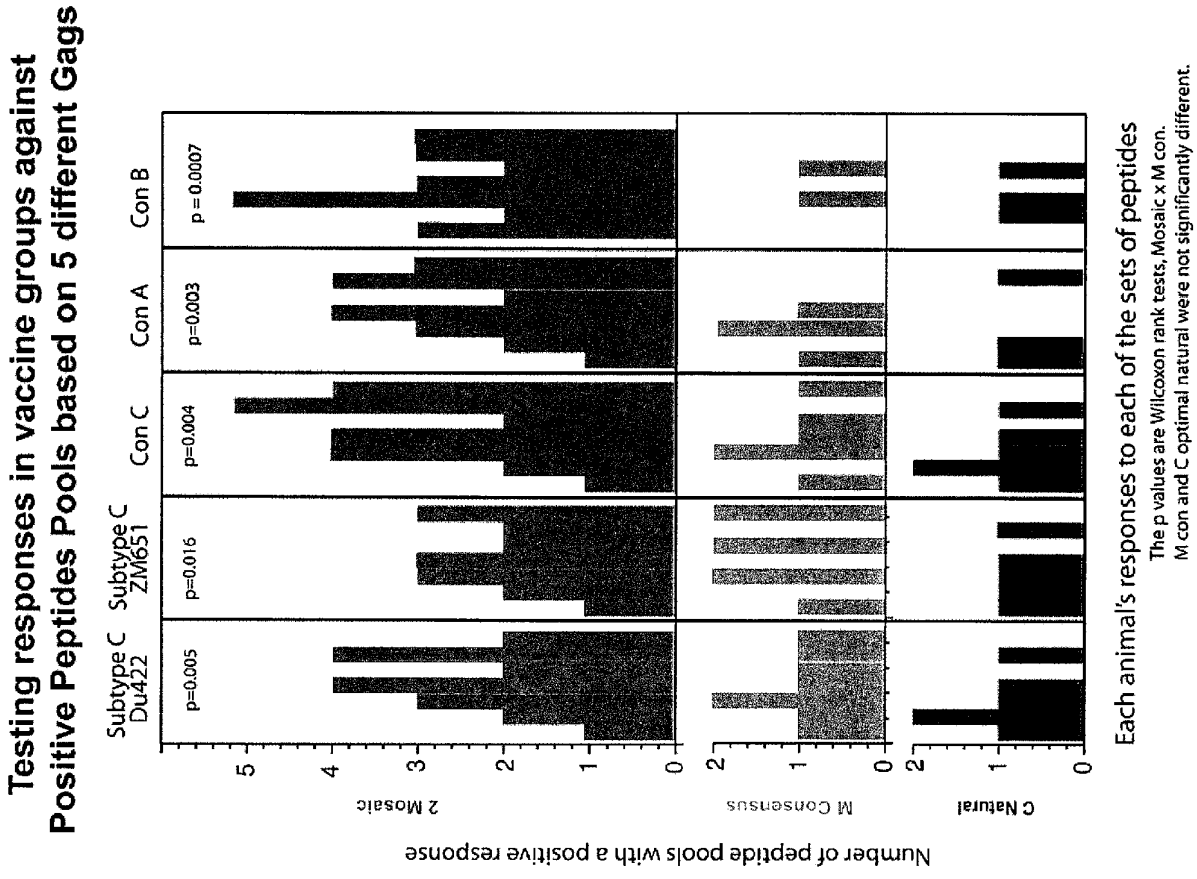

Testing responses in vaccine groups against
Positive Peptides Pools based on 5 different Gags
Subtype C Du422
Subtype C ZM651
Con C
Con A
Con B
p=0.005
p=0.016
p=0.004
p=0.003
p = 0.0007
2 Mosaic
M Consensus
C Natural
Number of peptide pools with a positive response
Each animal's responses to each of the sets of peptides
The p values are Wilcoxon rank tests, Mosaic x M con.
M con and C optimal natural were not significantly different.

Figure 16

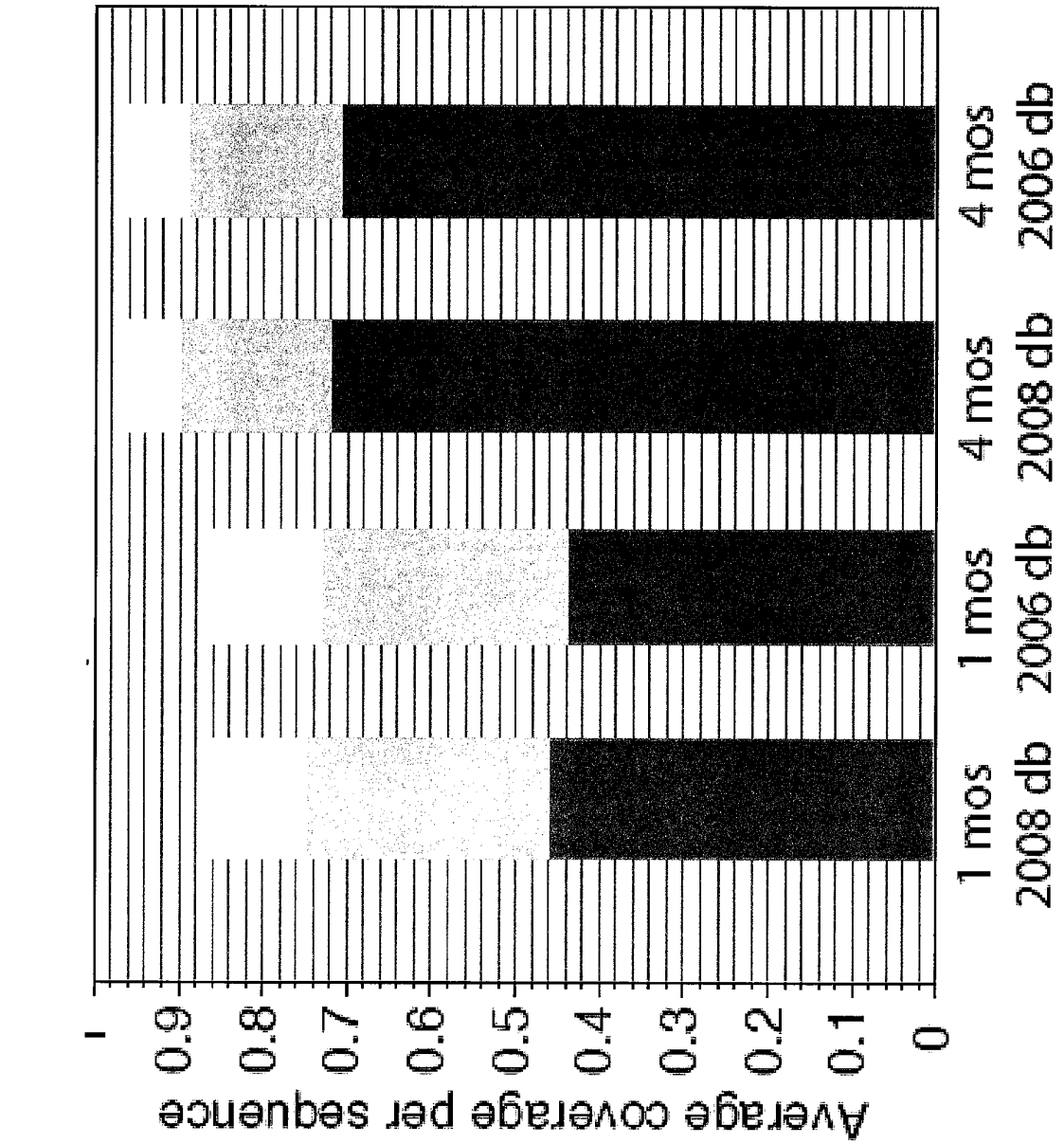
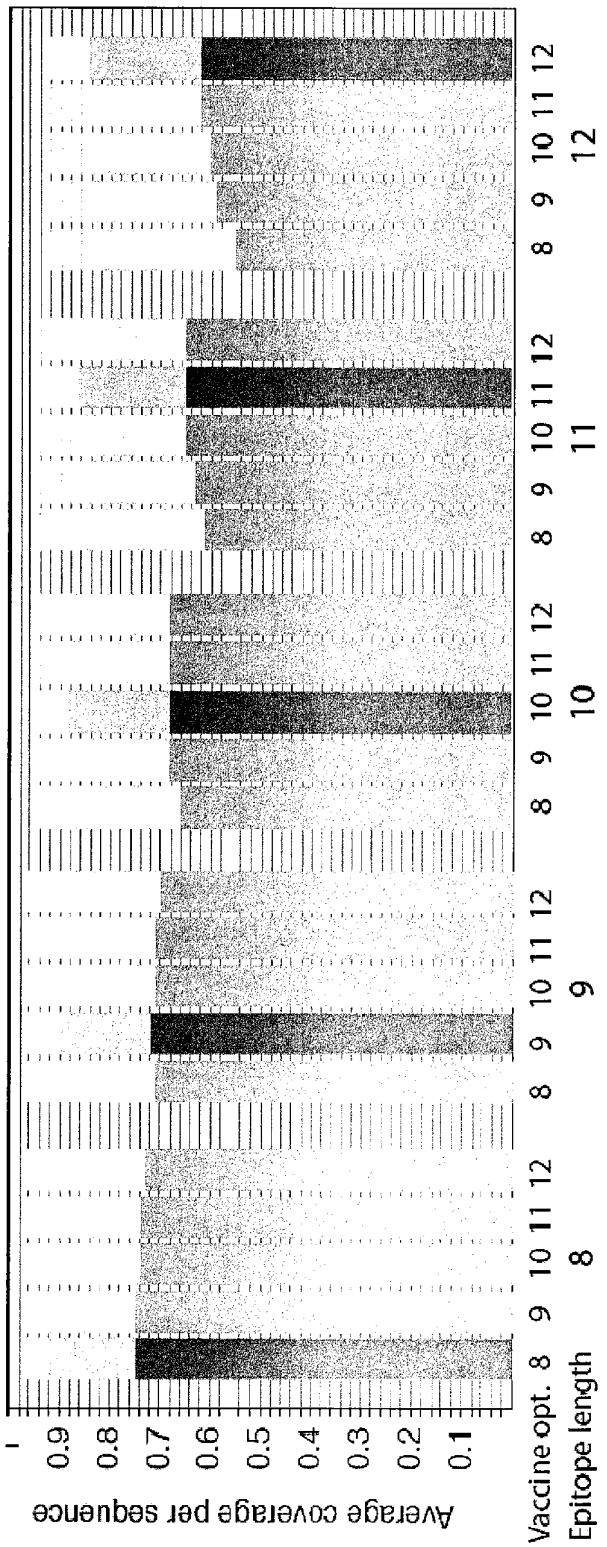

Number
Of Gags
2006
~ 550
2008
~1208
Average coverage per sequence
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 mos
2008 db
1 mos
2006 db
4 mos
2008 db
4 mos
2006 db

Figure 17


Average coverage per sequence
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Vaccine opt. 8 9 10 11 12 8 9 10 11 12 8 9 10 11 12 8 9 10 11 12
Epitope length 8 9 10 11 12

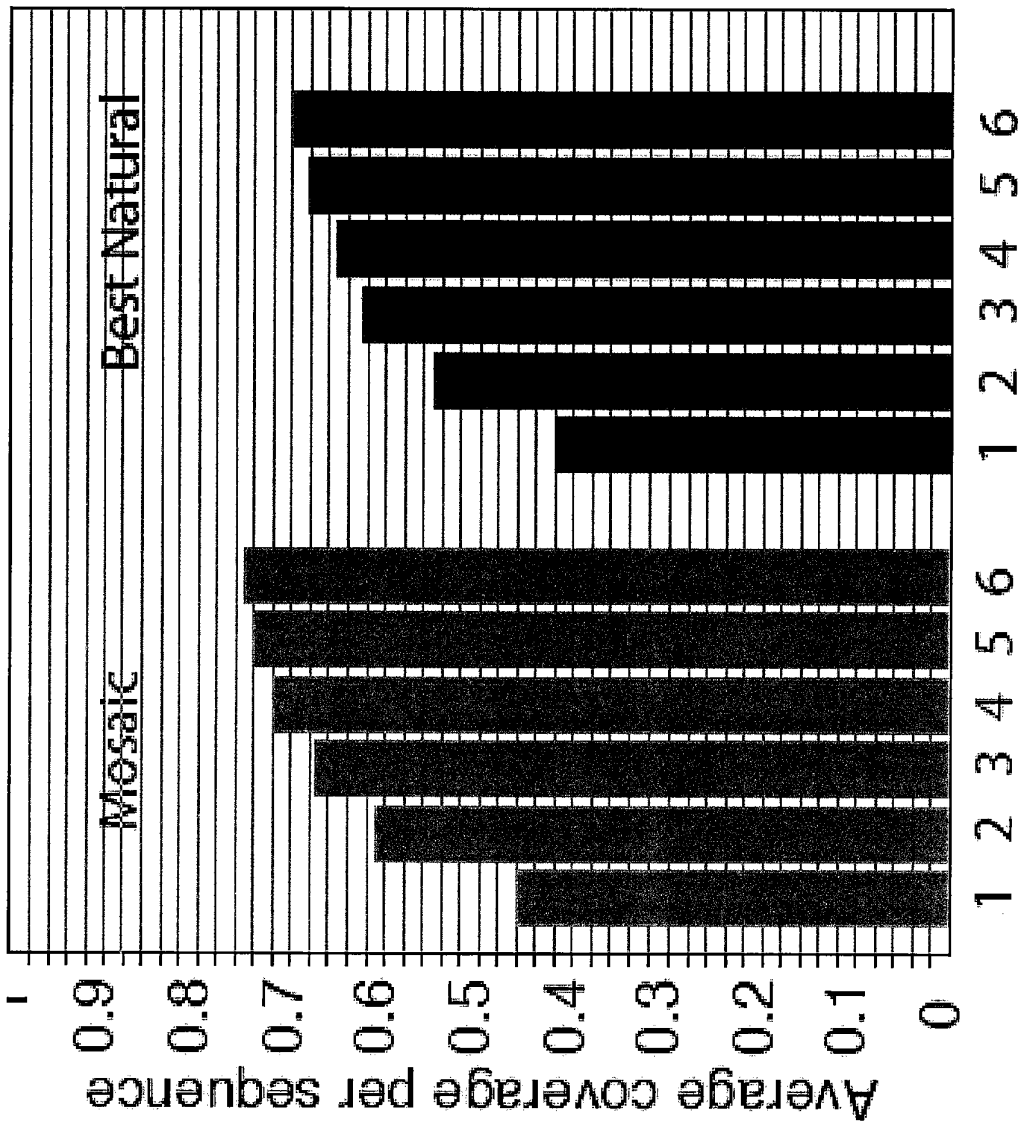

Average coverage per sequence
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1 2 3 4 5 6
1 2 3 4 5 6
Mosaic
Best Natural

Figure 18

Figure 19

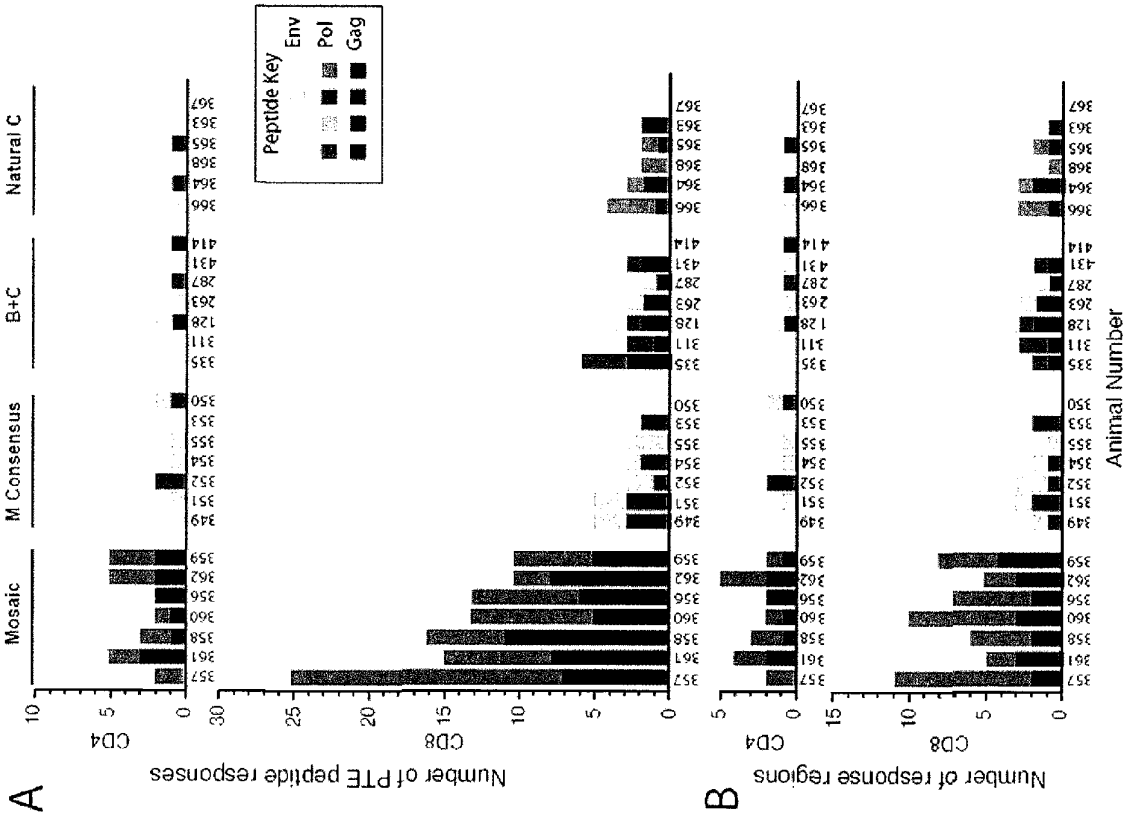

A
B
Mosaic
M Consensus
B+C
Natural C
Peptide Key
Env
Pol
Gag
CD4
CD8
Number of PTE peptide responses
Number of response regions
Animal Number

Figure 20A

CD8 Gag

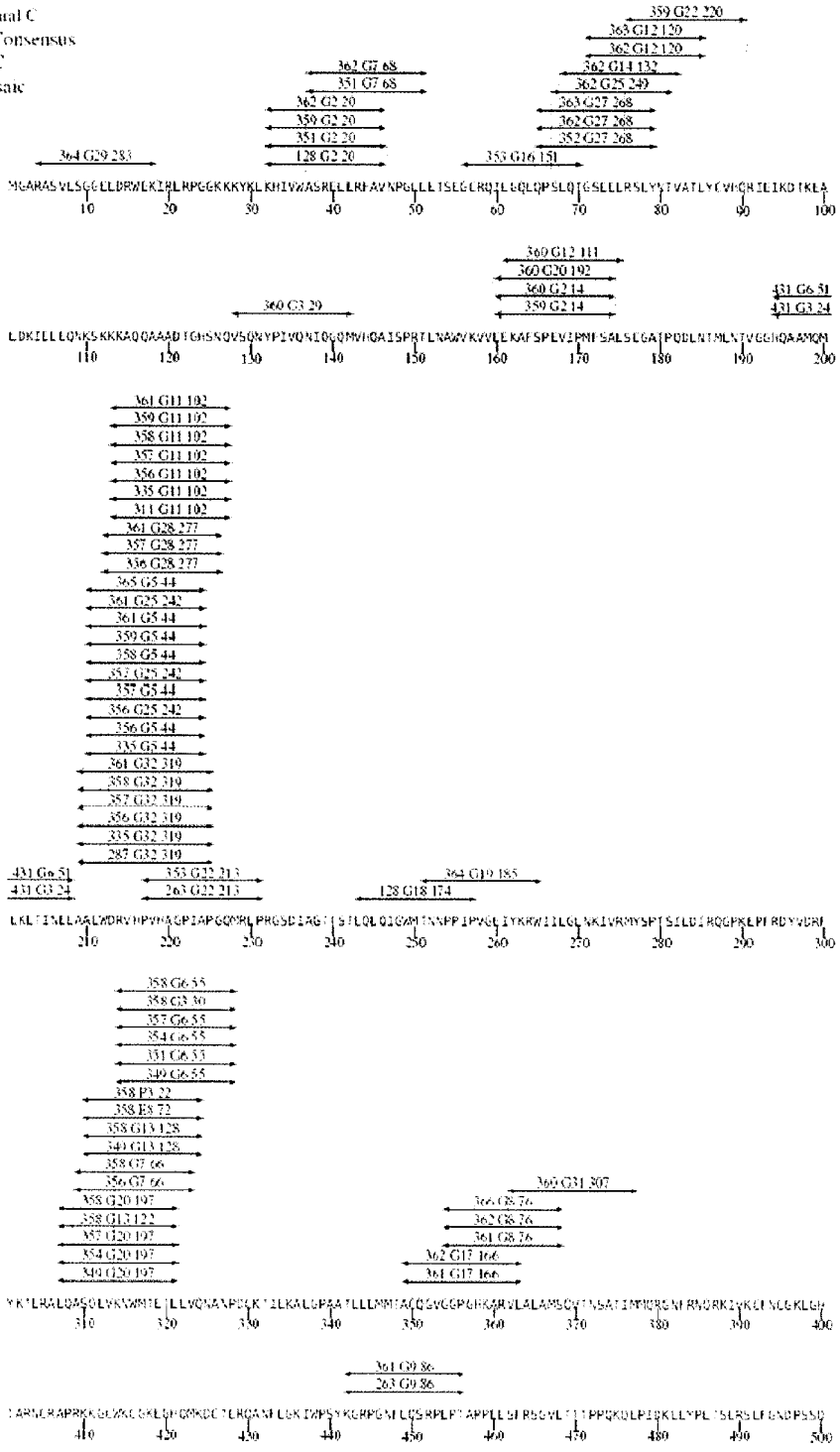


Figure 20B

CD8 Pol

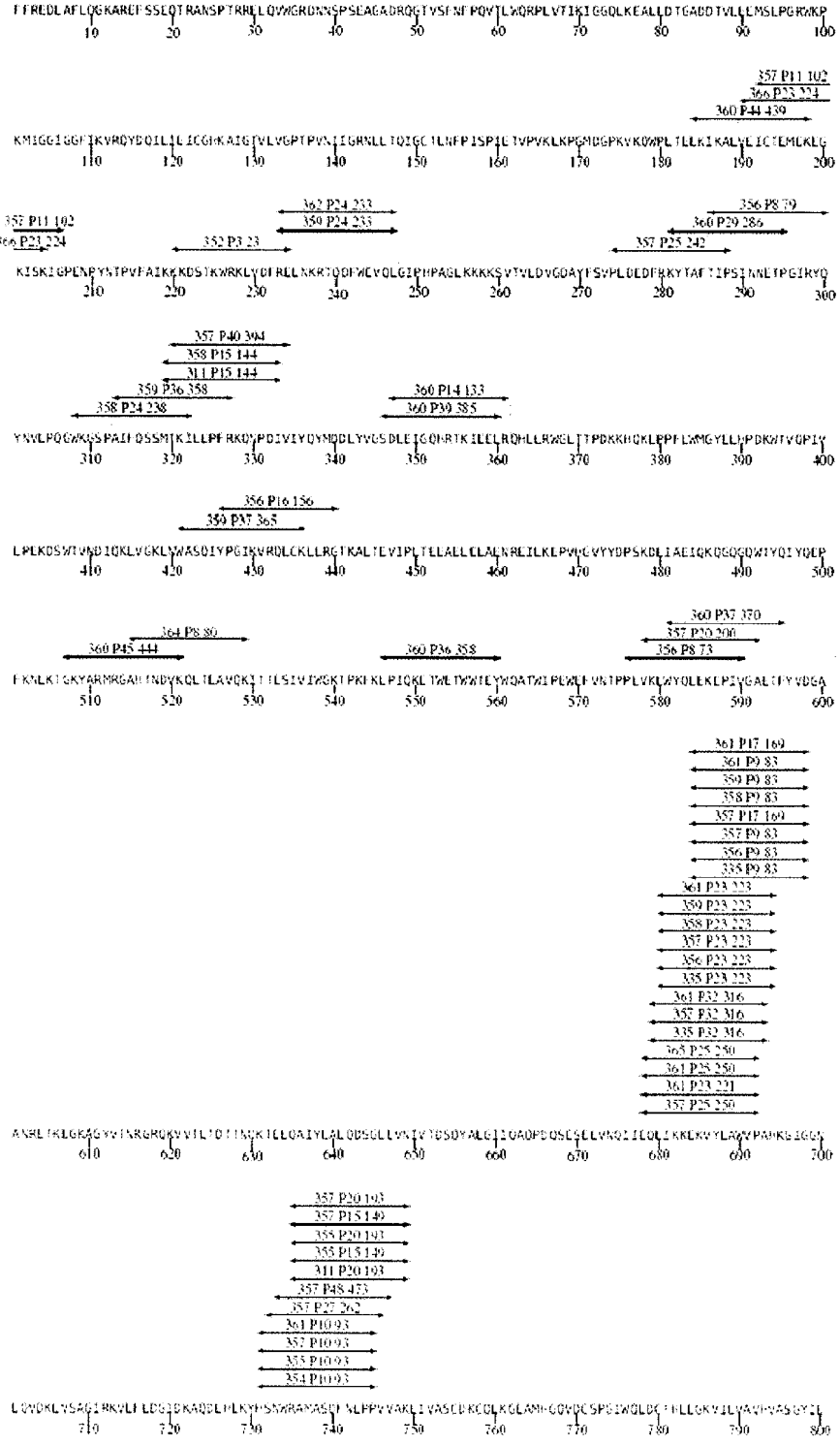


Figure 20B (con't)

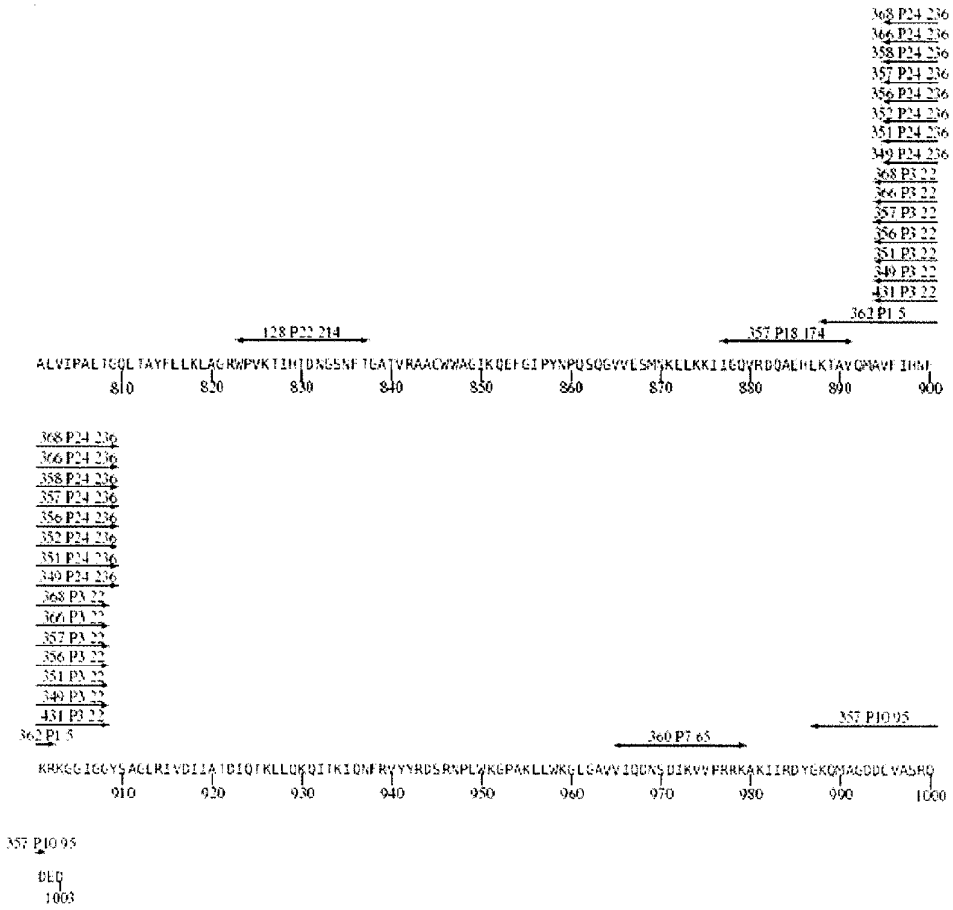


Figure 20C

CD8 Env

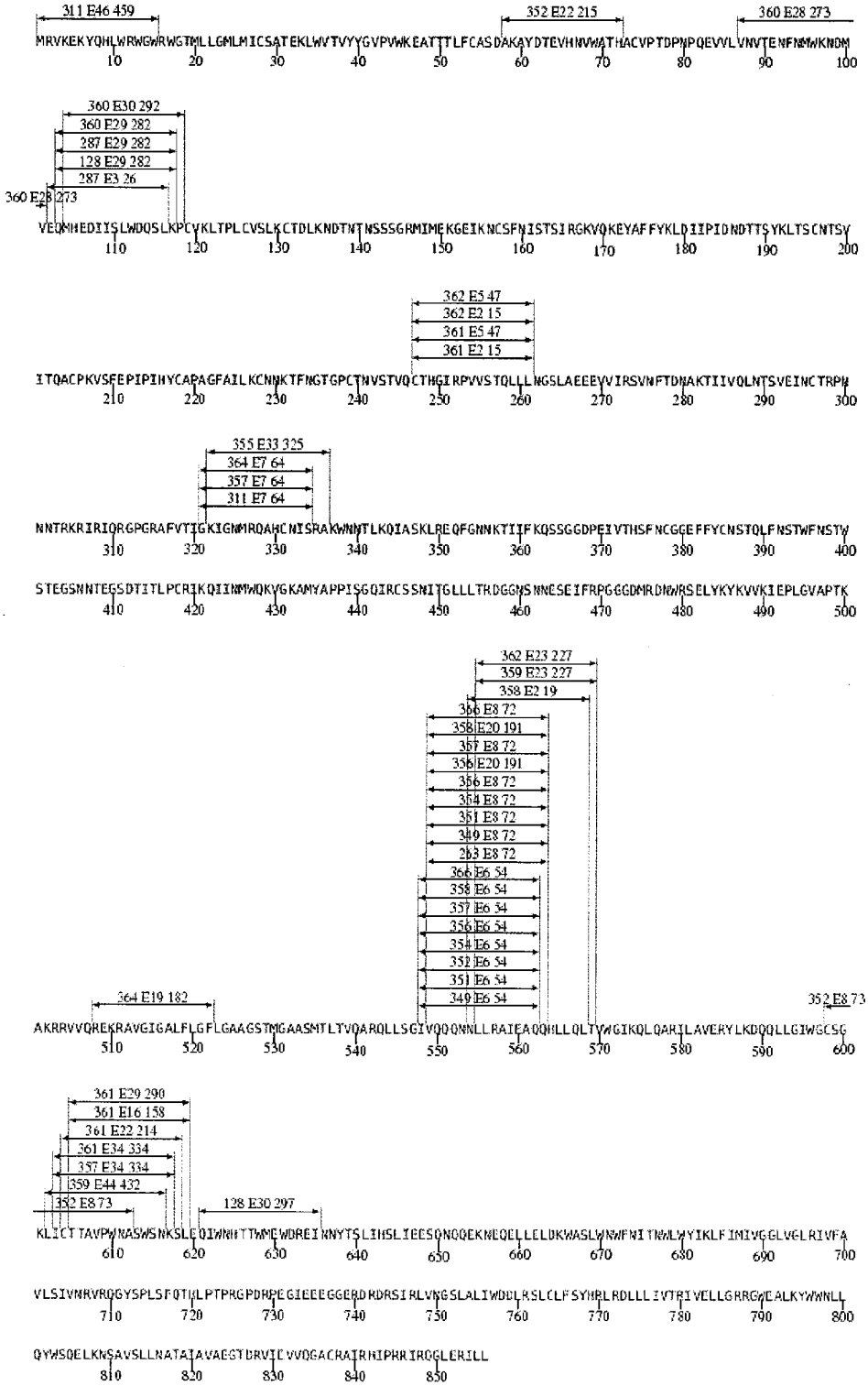


Figure 21A

CD4 Gag

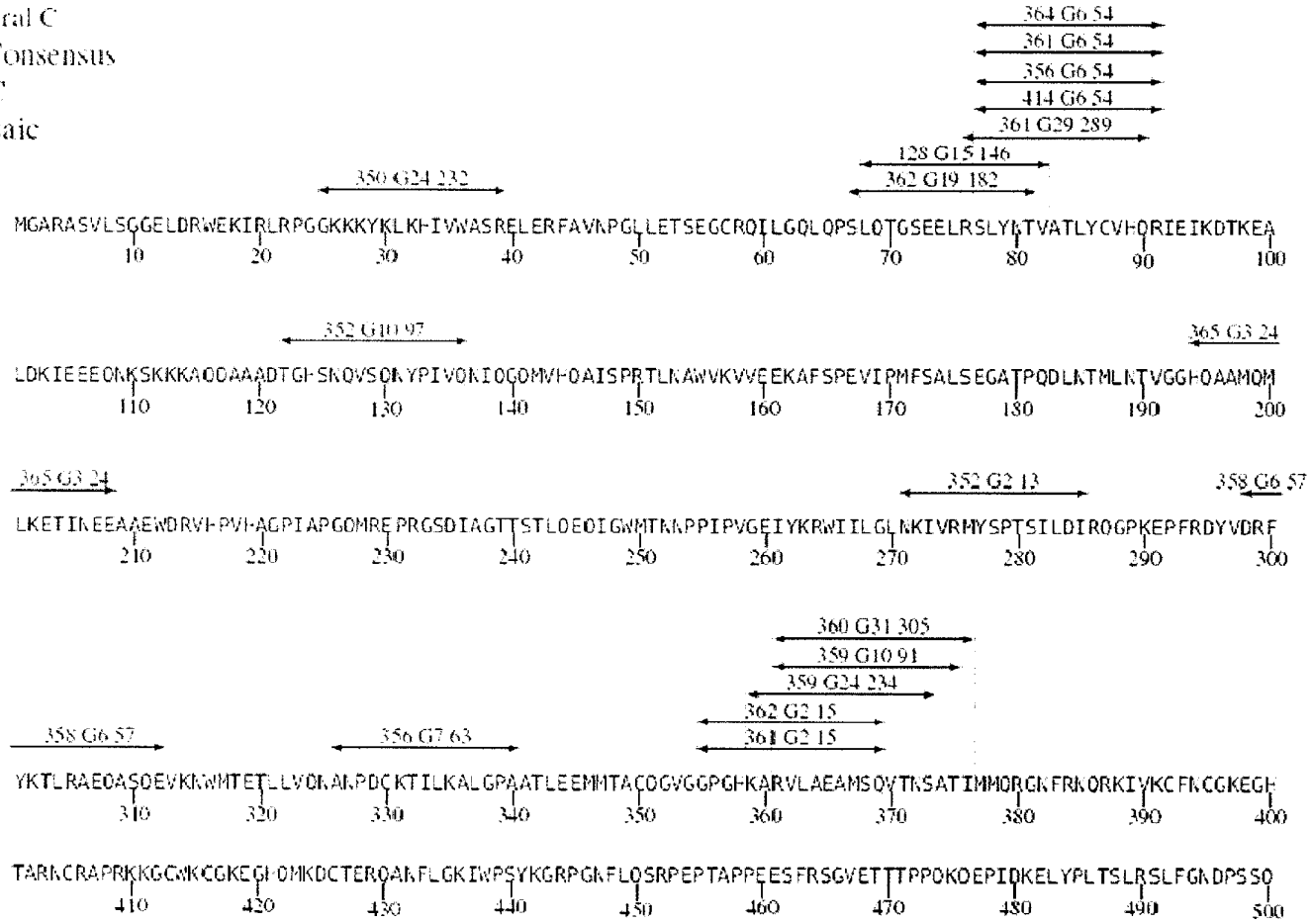


Figure 21B

CD4 Pol

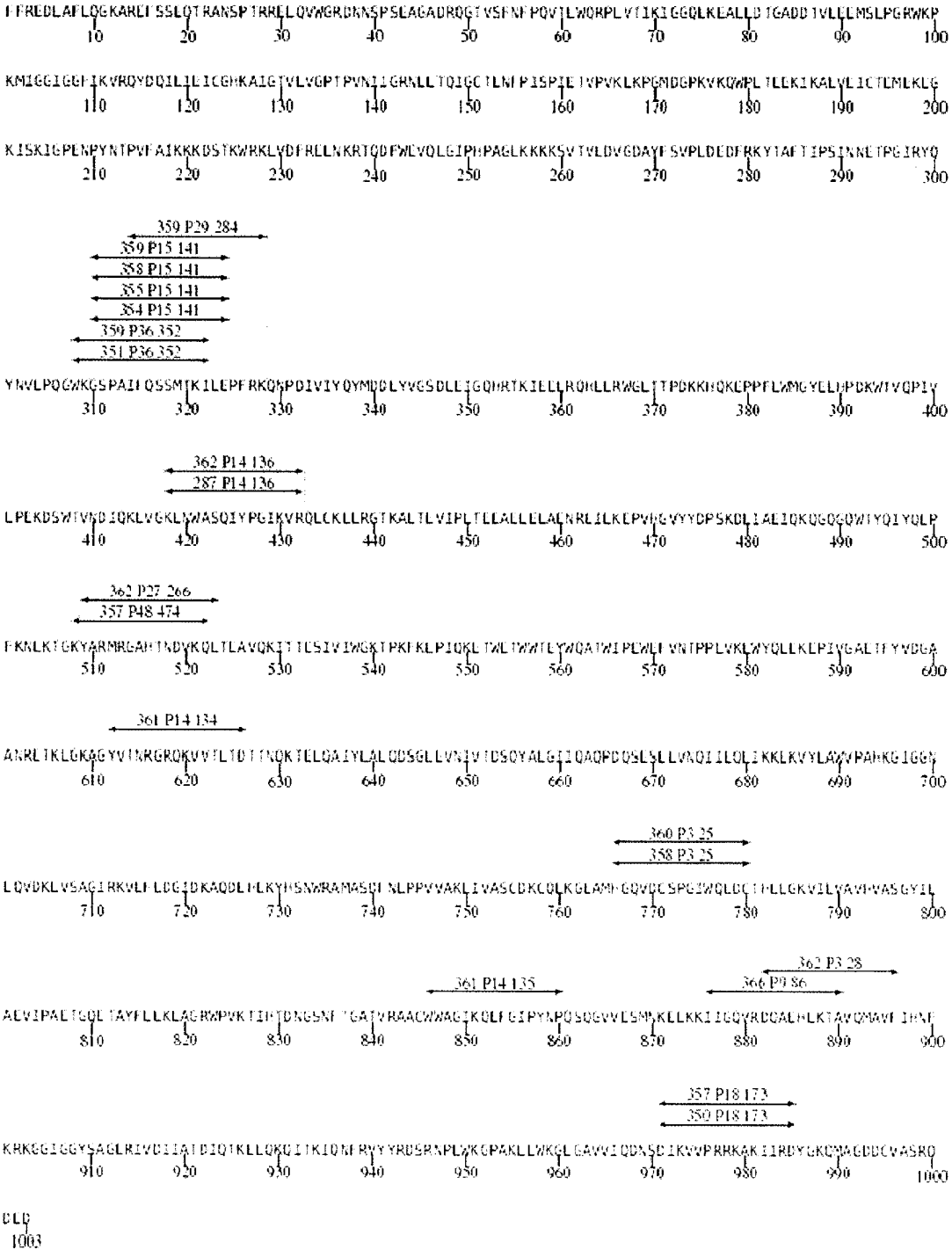


Figure 21C

CD4 Env

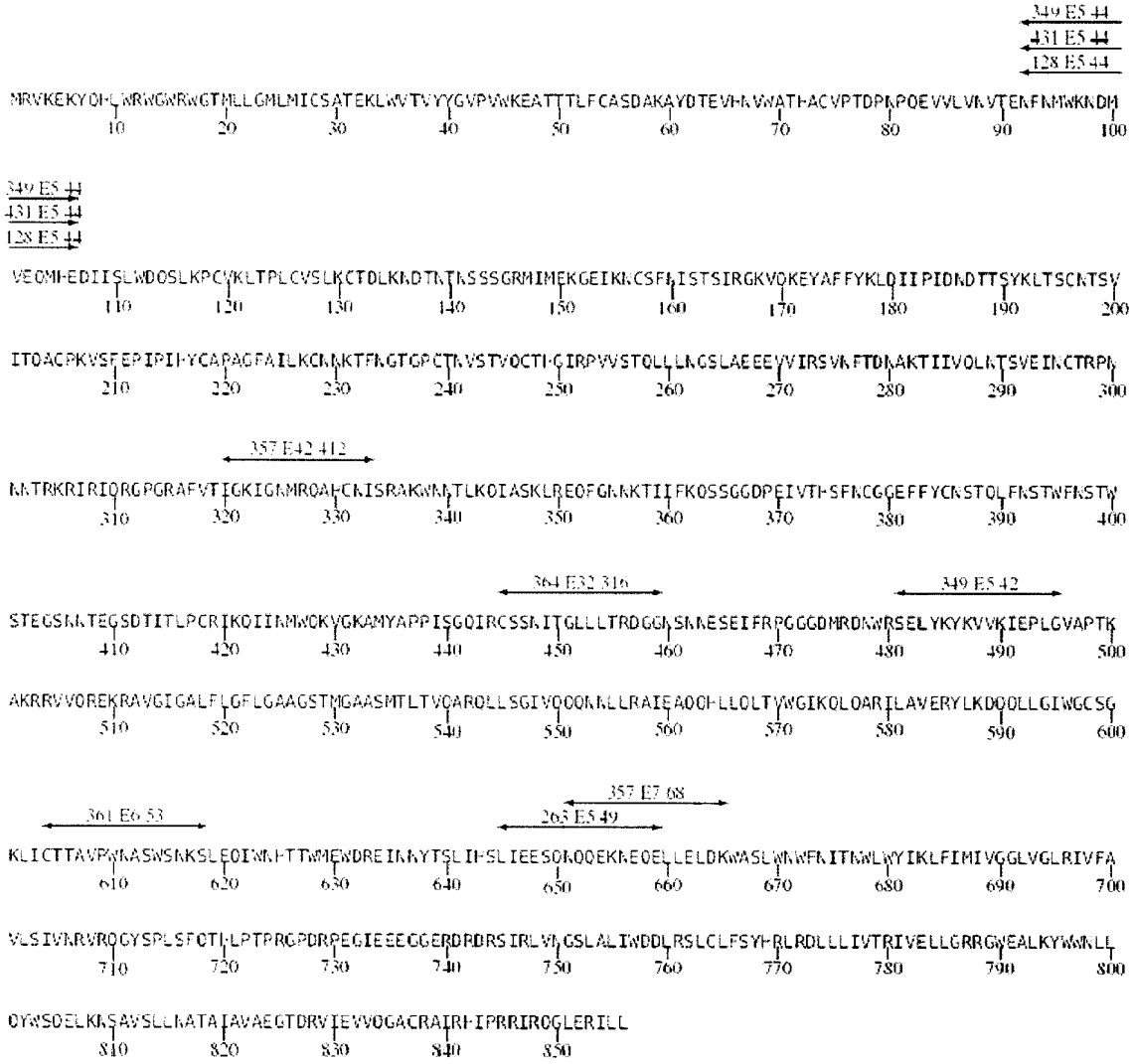


Figure 22

Reactive peptides and breadth and depth of responses: Clade B+C Vaccine

Clade B+C 287-95:
3 CD8 PTE+, 1 CD4 PTE+
2 CD8+ regions, 1 CD4+ region
Number of overlapping peptides per region: CD8: 1 2 CD4: 1

3 CD8 responses:

C	**AAEWDRLHPVHAGPIA**			
B	**AAEWDRLHPVHAGPIA**			
G319	AAE-DRLHPVHAGPIP	Gag	209	225
C	**DQMHEDIISLWDQSLK**			
B	**EQMHEDIISLWDQSLK**			
E26	EQMHEDIISLWDQSL	Env	102	116
E282	HMHEDIISLWDESLK	Env	103	117

1 CD4 response:

C	**KLNWASQIYSGIKVR**			
B	**KLNWASQIYPGIKVR**			
P136	KLNWASQIYAGIKVK	Pol	418	432

Clade B+C 128-92:
5 CD8 PTE+, 2 CD4 PTE+
5 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 1 1 1 1 1 CD4: 1 1

5 CD8 responses:

C	**KHLVWASRELERFAL**			
B	**KHIVWASRELERFAV**			
G20	KHLVWASRELERFAL	Gag	32	46
C	**LQEQIAWMTNNPPVP**			
B	**LQEQIGWMTNNPPIP**			
G174	LAEQIAWMTSNPPIP	Gag	243	257
C	**QMHEDIISLWDQSLK**			
B	**QMHEDIISLWDQSLK**			
E282	HMHEDIISLWDESLK	Env	103	117
C	**DIWDNMTWMQWDREI**			
B	**EIWDNMTWMEWEREI**			
E297	EIWNNMTWMEWEKEI	Env	621	635
C	**WPVKVIHTANGSNFT**			
B	**WPVKTIHTANGSNFT**			
P214	WPVKVVHIDNGSNFT	Pol	823	837

2 CD4 responses:

C	**AQTGTEELRSLYNTV**			
B	**LQTGSEELRSLYNTV**			
G146	LQTGSEELKSLFNTV	Gag	68	82
C	**YFNMWKNDMVDQMHE**			
B	**NFNMWKNNMVEQMHE**			
E544	NFNMWKNNMVEQMHE	Env	92	106

Figure 22 (con't)

Clade B+C 263-00:
3 CD8 PTE+, 1 CD4 PTE+
3 CD8+ regions, 1 CD4+ region
Number of overlapping peptides per region: CD8: 1 1 1 CD4: 1

3 CD8 responses:

C	**GRPGNFLQSRPEPT**			
B	**GRPGNFLQSRPEPT**			
G86	GRPGNFLQ**N**RPEPT	Gag	442	456
C	**PVHAGPIAPGQMREP**			
B	**PVHAGPIAPGQMREP**			
G213	PVQAGPIAPGQMREP	Gag	217	231
C	**VQQQSNLLRAIEAQQ**			
B	**VQQQNNLLRAIEAQQ**			
E72	VQQQNNLLRAIEAQH	Env	549	563

1 CD4 response:

C	**LLEESQNQQEKNEKD**			
B	**LIEESQNQQEKNEQE**			
E49	LIEESQNQQEKNEQD	Env	645	659

Clade B+C 311-00:
5 CD8 PTE+, 0 CD4 PTE+
5 CD8+ regions, 0 CD4+ region
Number of overlapping peptides per region: CD8: 1 1 1 1 1 CD4: 0

5 CD8 responses:

C	**DRLHPVHAGPIAPGQ**			
B	**DRLHPVHAGPIAPGQ**			
G102	DRVHPVHAGPI**P**PGQ	Gag	213	227
C	**GDIIGDIRQAHCNIS**			
B	**GEIIGDIRQAHCNIS**			
E64	GDIIGDIRQAHCNIS	Env	321	334
C	**TGMLRNCQPWWIWGI**			
B	**KGIRKNYQHLWRWGT**			
E459	**KE**IRKNYQHLWRWGT	Env	4	18
C	**MTKILEPFRAKNPEI**			
B	**MTKILEPFRKQNPDI**			
P144	MTKILEPFRKQNPDI	Pol	319	333
C	**RAMASEFNLPPVVAK**			
B	**RAMASDFNLPPVVAK**			
P193	KAMASDFNLPPIVAK	Pol	735	749

Figure 22 (con't)

Clade B+C 335-96:

6 CD8 PTE+, 0 CD4 PTE+
2 CD8+ regions, 0 CD4+ region
Number of overlapping peptides per region: CD8: 3 3 CD4: 0

6 CD8 responses:

C	AAEW**DRLHPVHAGPIA**PGQ			
B	AAEW**DRLHPVHAGPIA**PGQ			
G319	AAE-DRLHPVHAGPI**P**	Gag	209	225
G44	AEWDRLHPVHAGPIA	Gag	210	224
G102	DRVHPVHAGPI**P**PGQ	Gag	213	227
C	QLINKE**RVYLSWVPA**HKGIG			
B	QLIKKE**KVYLAWVPA**HKGIG			
P316	KLIEKDKVYLSWVPA	Pol	679	693
P223	LIKKE**R**VYLSWVPAH	Pol	680	694
P83	EKVYLSWVPAHKGIG	Pol	684	698

Clade B+C 414-95:

0 CD8 PTE+, 1 CD4 PTE+
0 CD8+ regions, 1 CD4+ region
Number of overlapping peptides per region: CD8: 0 CD4: 1

1 CD4 responses:

C	**SLYNTVATLYCVHAG**			
B	**SLYNTVATLYCVHQK**			
G54	SLYNTVATLYCVHQR	Gag	77	91

Clade B+C 431-01:

3 CD8 PTE+, 1 CD4 PTE+
2 CD8+ regions, 1 CD4+ region
Number of overlapping peptides per region: CD8: 2 1 CD4: 1

3 CD8 responses:

C	**HQAAMQMLKDTINEE**			
B	**HQAAMQMLKETINEE**			
G24	HQAAMQMLK**D**TINEE	Gag	194	208
G51	HQAAMQMLKEIINEE	Gag	194	208
C	**AVFIHNFKRKGGIGG**			
B	**AVFIHNFKRKGGIGG**			
P22	AVFIHNFKRKGGIGG	Pol	894	908

1 CD4 response:

C	**YFNMWKNDMVDQMHE**			
B	**NFNMWKNNMVEQMHE**			
E44	NFNMWKNNMVEQMHE	Env	92	106

Figure 22 (con't)

Reactive peptides and breadth and depth of responses: M Consensus Vaccine

M consensus 349-07:
7 CD8 FTE-, 2 CD4 FTE+
3 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 3 2 2 CD4: 1 1

3 CD8 responses:

Mcon	**EQATQDVKNWMTDTLLVQNANP**			
G197	EQSTQEVKNWMIDTL	Gag	307	321
G128	---TQDVKNWMIDTLLIQ	Gag	310	324
G55	-------KNWMIDTLLVQNANP	Gag	314	328
Mcon	**IVQQQSNLLRAIEAQQ**			
E54	IVQQQSNLLRAIEAQ	Env	548	562
E72	-VQQQNNLLRAIEAQH	Env	549	563
Mcon	**AVFIHNFKRKGGIGGY**			
P22	AVFIHNFKRKGGIGG	Pol	894	908
P236	-VLIHNFKRKGGIGGY	Pol	895	909

2 CD4 responses:

Mcon	**SELYKYKVVKIEPLG**			
E40	SELYKYKVVKIEPLG	Env	481	495
Mcon	**NFNMWKNNMVEQMHE**			
E44	NFNMWKNNMVEQMHE	Env	92	106

M consensus 350-07:
3 CD8 FTE-, 2 CD4 FTE+
3 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: none CD4: 1 1

2 CD4 responses:

Mcon	**GKKKYRLKHLVWASR**			
G232	GRKKYRLKHIVWASR	Gag	25	39
Mcon	**DIKVVPRRKAKIIRD**			
P173	DIKVVPRRKVKIIRD	Pol	971	985

M consensus 351-07:
7 CD8 FTE-, 1 CD4 FTE+
4 CD8+ regions, 1 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 2 1 CD4: 1

Figure 22 (con't)

4 CD8 responses:

Mcon	KHLVW**ASRELERFAL**NPGLL			
G20	KHLVWASRELERFAL	Gag	32	46
G68	-----ASRELERFAVNPGLL	Gag	37	5
Mcon	I**VQQQSNLLRAIEAQ**Q			
E6	VVQQQSNLLRAIEAQ	Env	548	562
E8	-VQQQ**N**NLLRAIEAQH	Env	549	563
Mcon	**AVFIHNFKRKGGIGG**Y			
P22	AVFIHNFKRKGGIGG	Pol	894	908
P236	-VLIHNFKRKGGIGGY	Pol	895	909
Mcon	**KNWMTDTLLVQNANP**			
G55	KNWMTDTLLVQNANP	Gag	314	328

1 CD4 response:

Mcon	**WKGSPAIFQSSMTKI**			
P352	WKGSPAIFQ**A**SMTKI	Pol	308	322

M consensus 352-07:

6 CD8 PTE+, 2 CD4 PTE+
6 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 1 1 1 1 1 1 CD4: 1

6 CD8 responses:

Mcon	**QPALQTGSEELRSLY**			
G268	**L**PALK**T**GSEELRSLY	Gag	65	79
Mcon	**IVQQQSNLLRAIEAQ**			
E54	VVQQQSNLLRAIEAQ	Env	548	562
Mcon	**CSGKLICTTTVPWNS**			
E73	CSGKLICTT**A**VPWNS	Env	598	612
Mcon	**AKAYDTEVHNVWATH**			
E215	AKAY**E**KEVHNVWATH	Env	58	72
Mcon	**KKDSTKWRKLVDFRE**			
P23	KKDSTKWRKLVDFRE	Pol	220	234
Mcon	**VFIHNFKRKGGIGGY**			
P236	VLIHNFKRKGGIGGY	Pol	895	909

2 CD4 responses:

Mcon	**NKIVRMYSPVSILDI**			
G13	NKIVRMYSPVSILDI	Gag	271	285
Mcon	**KGNSSKVSQNYPIVQ**			
G97	T**G**NSSQVSQNYPIVQ	Gag	122	136

Figure 22 (con't)

M consensus 353-07:
2 CD8 PTE+, 0 CD4 PTE+
2 CD8+ regions, 0 CD4+ regions
Number of overlapping peptides per region: CD8: 1 1 CD4: none

2 CD8 responses:

Mcon	**GCKQIIGQLQPALQT**			
G151	GC**RQILGQLQPSLQT**	Gag	56	70
Mcon	**PVHAGPIPPGQMREP**			
G213	PVQAGPIA**P**GQMREP	Gag	217	231

M consensus 354-07:
5 CD8 PTE+, 1 CD4 PTE+
3 CD8+ regions, 1 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 1 CD4: 1

3 CD8 responses:

Mcon	EQATQDV**KNWMTDTL**LVQNANP			
G197	EQSTQEVKNWMTDTL	Gag	307	321
G55	-------KNWMTDTLLVQNANP	Gag	314	328
Mcon	**IVQQQSNLLRAIEAQ**Q			
E54	VVQQQSNLLRAIEAQ	Env	548	562
E72	-VQQQNNLLRAIEAQH	Env	549	563
Mcon	**HSNWRAMASDFNLPP**			
P93	HSNWRAMASDFNLPP	Pol	731	745

1 CD4 response:

Mcon	**GSPAIFQSSMTKILE**			
P141	GSPAIFQSSMIKILD	Pol	310	324

M consensus 355-07:
3 CD8 PTE+, 1 CD4 PTE+
2 CD8+ regions, 1 CD4+ regions
Number of overlapping peptides per region: CD8: 1 2 CD4: 1

2 CD8 responses:

Mcon	**IIGDIRQAHCNISGT**			
E325	ITGDIRQAHCNVSRS	Env	322	336
Mcon	SNW**RAMASDFNLPP**IVAK			
P93	SNWRAMASDFNLPP	Pol	731	745
P149	---RTMASDFNLPPVVAK	Pol	735	749
P193	---KAMASDFNLPPIVAK	Pol	735	749

1 CD4 response:

Mcon	**GSPAIFQSSMTKILE**			
P141	GSPAIFQSSMIKILD	Pol	310	324

Figure 22 (con't)

Reactive peptides and breadth and depth of responses: Mosaic Vaccine

Mosaic 356-07:
15 CD8 PTE+, 2 CD4 PTE+
8 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 5 3 1 2 1 1 1 2 CD4: 1 1

8 CD8 responses:

Mos1	AAEW**DRVHPVHAGPIA**PGQ			
Mos2	AAEW**DRLHPVHAGPVA**PGQ			
G319	AAE-DRLHPVHAGP**IP**	Gag	209	225
G44	-AEWDRLHPVHAGP**IA**	Gag	210	224
G242	-ADWDRLHPVHAGP**VA**	Gag	210	224
G277	---WDRVHPVHAGP**NP**PG	Gag	212	226
G102	----DRVHPVHAGP**IP**PGQ	Gag	213	227

Mos1	**IVQQQNNLLRAIEAQQ**			
Mos2	**IVQQQSNLLRAIEAQQ**			
E54	VVQQQSNLLRAIEAQ	Env	548	562
E70	-VQQQ**NN**LLRAIEAQH	Env	549	563
E191	-VQQQSNLLKAIEAQQ	Env	549	563

Mos1	**ASQDVKNWMTETLLV**			
Mos2	**ATQDVKNWMTDTLLV**			
G66	ASQEVKNWMI**E**TLLI	Gag	309	323

Mos1	**AVFIHNFKRKGGIGGY**			
Mos2	**AVFIHNFKRKGGIGEY**			
P22	AV**F**IHNFKRKGGIG**G**	Pol	894	908
P236	-VLIHNFKRKGGIG**GY**	Pol	895	909

Mos1	**PLVKLWYQLEKDPIA**			
Mos2	**PLVKLWYQLEKEPIV**			
P73	PLVKLWYQLEKEPIV	Pol	576	590

Mos1	**TIPSTNNETPGIRYQ**			
Mos2	**TIPSINNETPGIRYQ**			
P79	TIPSINNETPGIRYQ	Pol	286	300

Mos1	**YPGIKVRQLCKLLRG**			
Mos2	**YAGIKVKQLCKLLRG**			
P156	YPGIKVRQLCKLLRG	Pol	426	440

Mos1	LIKK**ERVYLSWVPAH**KGIG			
Mos2	LIKK**EKVYLAWVPAH**KGIG			
P223	LIKKERVYLSWVPAH	Pol	680	694
P83	----EKVYLSWVPAHKGIG	Pol	684	698

2 CD4 responses:

Mos1	**SLYNTVATLYCVHQR**			
Mos2	**SLFNTVATLYCVHAE**			
G54	SLYNTVATLYCVH**QR**	Gag	77	91

Mos1	**ANPDCKTILKALGPA**			
Mos2	**ANPDCKTILRALGPG**			
G63	ANPDCKTILRALGPG	Gag	326	340

Figure 22 (con't)

Mosaic 357-07:
29 CD8 PTE+, 4 CD4 PTE+
14 CD8+ regions, 4 CD4+ regions
Number of overlapping peptides per region: CD8: 5 2 2 5 5 2 1 1 1 1 1 1 1 1 CD4: 1 1 1 1

14 CD8 responses:

Mos1	EQLIKK**ERVYLSWVPAHKGIG**			
Mos2	EQLIKK**EKVYLAWVPAHKGIG**			
P250	EPLIKKEKVYLSWVP	Pol	678	692
P316	-KLIEKDKVYLSWVPA	Pol	679	693
P223	--LIKKERVYLSWVPAH	Pol	680	694
P169	------EKVYLAWVPAHKGIG	Pol	684	698
P83	------EKVYLSWVPAHKGIG	Pol	684	698
Mos1	**IVQQQNNLLRAIEAQQ**			
Mos2	**IVQQQSNLLRAIEAQQ**			
E54	VVQQQSNLLRAIEAQ	Env	548	562
E72	-VQQQNNLLRAIEAQH	Env	549	563
Mos1	EQASQDV**KNWMTETLL**VQNANP			
Mos2	EQATQDV**KNWMTDTLL**VQNANP			
G197	EQSTQEVKNWMTDTL	Gag	307	321
G55	-------KNWMTDTLLVQNANP	Gag	314	328
Mos1	HSNW**RAMASDFNLPP**VVAK			
Mos2	HSNW**RAMASEFNLPP**IVAK			
P93	HSNWRAMASDFNLPP	Pol	731	745
P262	-SNWRAMASEFNLPPI	Pol	732	746
P473	--NWRTMASDFNLPPVI	Pol	733	747
P149	----RTMASDFNLPPVVAK	Pol	735	749
P193	----KAMASDFNLPPIVAK	Pol	735	749
Mos1	AEW**DRVHPVHAGPIA**PGQ			
Mos2	AEW**DRLHPVHAGPVA**PGQ			
G319	AAEDRLHPVHAGP**IP**	Gag	209	225
G242	ADWDRLHPVHAGP**VA**	Gag	210	224
G44	AEWDRLHPVHAGP**IA**	Gag	210	224
G277	--WDRVHPVHAGP**NP**PG	Gag	212	226
G102	---DRVHPVHAGP**IP**PGQ	Gag	213	227
Mos1	**AVFIHNFKRKGGIGGY**			
Mos2	**AVFIHNFKRKGGIGEY**			
P22	AVFIHNFKRKGGIGG	Pol	894	908
P236	-VLIHNFKRKGGIGGY	Pol	895	909
Mos1	**PLDEGFRKYTAFTIP**			
Mos2	**PLDEDFRKYTAFTIP**			
P242	PLDESFRKYTAFTIP	Pol	274	288
Mos1	**TRILEPFRAKNPEIV**			
Mos2	**TKILEPFRKQNPDIV**			
P394	IKILEPFR**A**QNP**E**LV	Pol	320	334
Mos1	**ICEEMEKEGKITKIG**			
Mos2	**ICTEMEKEGKISKIG**			
P112	ICTEMEKEGKISKIG	Pol	192	206
Mos1	**VKLWYQLEKDPIAGV**			
Mos2	**VKLWYQLEKEPIVGA**			
P200	VKLWYQLEKDPIVGA	Pol	578	592
Mos1	**GDIIGDIRQAHCNIS**			
Mos2	**GDIIGDIRQAHCNLS**			
E764	GDIIGDIRQAHCNIS	Env	321	334
Mos1	**ICTTTVPWNASWSNK**			
Mos2	**ICTTAVPWNTSWSNK**			
E334	ICTTTVPWN**A**SWSN**R**	Env	633	647

Figure 22 (con't)

```
Mos1  IGQVRDQAEHLKTAV
Mos2  IGQVRDQAEHLKTAV
P174  IGQVREQAEHLKTAV            Pol     877     891

Mos1  GKQMAGADCVAGRQD
Mos2  GKQMAGDDCVASRQD
P95   GKQMAGDDCVAGRQD            Pol     987     1001

4 CD4 responses:
Mos1  KYAKMRTAHTNDVKQ
Mos2  KYARMRGAHTNDVKQ
P474  KYAKMRTAHTNDVRQ            Pol   503   517

Mos1  AGDIIGDIRQAHCNI
Mos2  TGDIIGDIRQAHCNL
P412  TGEIIGDIRQAHCNL            Env   320   334

Mos1  NQQEKNEKDLLALDS
Mos2  NQQEKNEQELLELDK
P61   NQQEKNEQDLLALDK            Env   651   665

Mos1  DIKVVPRRKVKIIKD
Mos2  DIKVVPRRKAKIIRD
P179  DIKVVPRRKVKIIRQ            Pol   971   985

Mosaic 358-07:
19 CD8 PTE+, 3 CD4 PTE-
7 CD8+ regions, 3 CD4+ regions
Number of overlapping peptides per region: CD8: 2 4 2 6 3 1 1  CD4: 1 1 1

7 CD8 responses:
Mos1  LIKKERVYLSWVPAHKGIG
Mos2  LIKKEKVYLAWVPAHKGIG
P202  LIKKERVYLSWVPAK            Pol     680     694
P38   ----EKVYLSWVPAHKGIG        Pol     684     698

Mos1  IVQQQNNLLRAIEAQQHLLQL
Mos2  IVQQQNNLLRAIEAQQHVLQL
E54   IVQQQNLLRAIEAQ              Env   548     562
E191  -VQQQNLLKAIEAQQ             Env   549     563
P72   -VQQQNNLLRAIEAQH            Env   549     563
E19   ------NLLRAIEAQQHLLQL      Env   554     568

Mos1  AVFIHNFKRKGGIGGY
Mos2  AVFIHNFKRKGGIGEY
```

Figure 22 (con't)

```
P222  AVFIHNFKRKGGIGG            Pol     894     908
P236  -VLIHNFKRKGGIGGY           Pol     895     909

Mos1  EQASQDVKNWMTETLLVQNANP
Mos2  EQATQDVKNWMTDTLLVQNANP
G122  EQASQDVKNWMTETL            Gag     307     321
G197  EQSTQEVKNWMTDTL            Gag     307     321
G66   --ASQEVKNWMTETLLT          Gag     309     323
G128  ---TQDVKNWMTDTLLIQ         Gag     310     324
G30   -------KNWMTETLLVQNANP     Gag     314     328
G55   -------KNWMTDTLLVQNANP     Gag     314     328

Mos1  AEWDRVHPVHAGPIAPGQ
Mos2  AEWDRLHPVHAGPVAPGQ
G319  AASDRLHPVHAGPIP           Gag     209     225
G44   AEWDRLHPVHAGPIA           Gag     210     224
G11   ---DRVHPVHAGPIPPGQ        Gag     213     227

Mos1  WKGSPAIFQCSMTRI
Mos2  WKGSPAIFQSSMTKI
P238  WKGSPAIFQCSMTKI            Pol   308   322

Mos1  MTRILEPFRAKNPEI
Mos2  MTKILEPFRKQNPDI
P144  MTKILEPFRKQNPDI            Pol   319   333

3 CD4 responses:
Mos1  GSPAIFQCSMTRILE
Mos2  GSPAIFQSSMTKILE
P141  GSPAIFQSSMTKILD            Pol   310   324

Mos1  HGQVDCSPGIWQLAC
Mos2  HGQVDCSPGIWQLAC
P25   HGQVDCSPGIWQLDC            Pol   766   780

Mos1  DRFYKTLRAEQASQD
Mos2  DRFFKTLRAEQATQD
G57   DRFYKTLRAEQASQE            Gag   298   312
```

Mosaic 359-07:
12 CD8 PTE+, 5 CD4 PTE+
10 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 1 1 1 1 1 1 1 1 CD4: 3 2

10 CD8 responses:
Mos1 LIKKERVYLSWVPAHKGIG

Figure 22 (con't)

```
Model  LIKKEKVYLAWVPAHKGIG          Pol    660   694
E203   LIKKERVNISWVPAH               Pol    664   696
E13    ----EKVNISWVPAHKGIG          Pol    664   696

Model  AEWDRVHPVHAGPIAPGQ
Model  AEWDRLHPVHAGPVAPGQ
344    AEWDRLHPVHAGPIA              Gag    210   224
3103   ---DRVHPVHAGPIFPGQ           Gag    213   227

Model  AIFQCSMTRILEPFR
Model  AIFQSSMTKILEPFR
P358   AIFQSSMTRILEPER              Pol    313   327

Model  ASQIYPGIKVRQLCK
Model  ASQIYKGIKVKQLCK
P325   WSQLVAGIKVRQLCK              Pol    421   436

Model  RELNKRTQDFWEVQL
Model  RELNKRTQDFWEVQL
P338   RELNRRTQDFWEVQL              Pol    433   447

Model  LLRAIEAQQHLLQLT
Model  LLRAIEAQQHMLQLT
E207   LLWAIEAQQHLLQLT              Env    555   569

Model  LICTTVPWNASWSN
Model  LICTTAVPWNTSWSN
E430   RICTTNVPWNASWSN              Env    602   616

Model  KHIVWASRELERFAV
Model  KHIVWASRELERFAL
3207   KHLVWASRELERFAL              Gag    31    46

Model  RSLYNTVATLYCVHQ
Model  RSLFNTVATLYCVHA
3203   KSLYNVAVTLYCVHQ              Gag    76    90

Model  EEKAFSPEVIPMFSA
Model  EEKAFSPEVIPMFTA
314    EEKAFSPEVIPNFSA              Gag    160   174

2 CD4 responses:
Model  WKGSPAIFQCSMTRILEPFRA
Model  WKGSPAIFQSSMTKILEPFRA
P350   WKGSPRIFQASMTKI              Pol    308   322
P141   --SSPAIFQSSMTKIL             Pol    310   324
```

Figure 22 (con't)

P284	------IFQCSMTKILEPFRA	Pol	314	
Mos1	KARVLAEAMSQVTNSAT			
Mos2	KARVLAEAMSQTNSTIL			
G234	KAKVLAEAMSQVQQT	Gag	359	373
G91	--RVLAEAMSQVTNSAT	Gag	361	375

Mosaic 360-07:
16 CD8 PTE+, 2 CD4 PTE+
12 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 3 1 1 1 1 1 1 1 1 1 CD4: 1 1

12 CD8 responses:

Mos1	SDLEIGQHRAKIEELR			
Mos2	SDLEIGQHRTKIEELR			
P385	SDLEIGQHRIKIEEL	Pol	346	360
P133	-DLEIGQHRAKIEELR	Pol	347	361
Mos1	QMHEDIISLWDQSLKP			
Mos2	QMHEDIIRLWDQSLKP			
E282	MHEDIISLWDESLK	Env	103	117
E292	-MHEDVISLWDQSLKP	Env	104	118
Mos1	EEKAFSPEVIPMFSAL			
Mos2	EEKAFSPEVIPMFTAL			
G14	EEKAFSPEVIPMFSA	Gag	160	174
G192	EEKGFNPEVIPMFSA	Gag	160	174
G111	-EKGFSPEVIPMFTAL	Gag	161	175
Mos1	KYTAFTIPSTNNETP			
Mos2	KYTAFTIPSINNETP			
P286	KYTAFTIPSTNNETP	Pol	281	295
Mos1	EKIKALTAICEEMEK			
Mos2	EKIKALVEICTEMEK			
P439	EKIEALTAICEEMEK	Pol	184	198
Mos1	GKYAKMRTAHTNDVK			
Mos2	GKYARMRGAHTNDVK			
P444	GKYARMRGAHTNDVR	Pol	507	521
Mos1	LPIQKETWETWWTDY			
Mos2	LPIQKETWEAWWTEY			
P358	LPIQKETWEAWWTEY	Pol	546	560
Mos1	WYQLEKDPIAGVETF			

Figure 22 (con't)

```
Mos2  WYQLEKEPIVGAETF
P370  WYQLEKDPIAGAETF       Pol   581   595

Mos1  ENVTENFNMWKNNMV
Mos2  ENVTENFNMWKNDMV
E273  ENVTEEFNMWKNNMV       Env   87    101

Mos1  VIQDNSDIKVVPRRK
Mos2  VIQDNSDIKVVPRRK
P765  VIQDNSDIKVVPRRK       Pol   965   979

Mos1  VSQNYPIVQNIQGQM
Mos2  VSQNYPIVQNLQGQM
G29   VSQNYPIVQNLQGQM       Gag   128   142

Mos1  VLAEAMSQVTNSATIM
Mos2  VLAEAMSQ-TNSTILM
G307  VLAEAMSQ-AQQTNIM      Gag   362   377

2 CD4 responses:
Mos1  HGQVDCSPGIWQLAC
Mos2  HGQVDCSPGIWQLAC
P25   HGQVDCSPGIWQLDC       Pol   766   780

Mos1  RVLAEAMSQVTNSAT
Mos2  RVLAEAMSQTNSTIL
G305  RVLAEAMSQTNSAIL       Gag   361   376

_____________________________________________
Mosaic 361-07:
21 CD8 PTE+, 6 CD4 PTE-
7 CD8- regions, 5 CD4- regions
Number of overlapping peptides per region: CD8: 6 2 4 2 5 1 1 CD4: 2 1 1 1 1

7 CD8 responses:
Mos1  EQLIKKERVYLSWVPAHKGIG
Mos2  EQLIKKEKVYLAWVPAHKGIG
P221  EELIKKEKVYLAWVP             Pol     678     692
P250  EELIKKEKVYLSWVP             Pol     678     692
P31d  -KLIEKDKVYLSWVPA            Pol     679     693
P223  --LIKKERVYLSWVPAH           Pol     680     694
P169  ------EKVYLAWVPAHKGIG       Pol     684     698
P33   ------EKVYLSWVPAHKGIG       Pol     684     698

Mos1  CTHGIRPVVSTQLLL
Mos2  CTHGIKPVVSTQLLL
E18   CTHGIKPVVSTQLLL            Env     247     261
```

Figure 22 (con't)

```
E47   CIHGIRPVVSIQLLL          Env   247     261

Mos1  ICTTTVPWNASWSNKSL
Mos2  ICTTAVPWNTSWSNKSQ
E334  ICTTTVPWNASWSNR          Env   603     617
E214  -CTTTVPWNSSWSNKT         Env   604     618
E159  --TTAVPWNASWSNKSL        Env   605     619
E290  --TTAVPWNTSWSNKSL        Env   605     619

Mos1  ACQGVGGPGHKARVLAEAMS
Mos2  ACQGVGGPSHKARVLAEAMS
G166  ACQGVGGPGHKARVL          Gag     349     363
G76   -----GGPSHKARVLAEAMS     Gag     354     368

Mos1  AAEWDRVHPVHAGPIAPGQ
Mos2  AAEWDRLHPVHAGPVAPGQ
G319  AAE-DRLHPVHAGPIP        Gag     209     225
G242  -AEWDRLHPVHAGPVA        Gag     210     224
G44   -AEWDRLHPVHAGPIA        Gag     210     224
G277  ---WDRVHPVHAGPNPPG      Gag     212     226
G102  ----DRVHPVHAGPIPPGQ     Gag     212     226

Mos1  HSNWRAMASDFNLPP
Mos2  HSNWRAMASEFNLPP
P93   HSNWRAMASDFNLPP         Pol   731   745

Mos1  KGRPGNFLQNRPEPT
Mos2  KGRPGNFLQSRPEPT
G86   KGRPGNFLQNRPEPT         Gag   442   456

5 CD4 responses:
Mos1  RSLYNTVATLYCVHQR
Mos2  RSLFNTVATLYCVHAE
G219  RSLFNTVAILYCVHA         Gag     76      90
G54   -SLYNTVAILYCVHQR        Gag     77      91

Mos1  YVTDRGRQKIVSLTE
Mos2  YVTDRGRQKVVSLTD
P134  YVTDRGRQKVVSLTE         Pol   610   624

Mos1  CTTTVPWNASWSNKS
Mos2  CTTAVPWNTSWSNKS
E53   CITNVPWNSSWSNKS         Env   604   618

Mos1  WWAGIQQEFGIPYNP
Mos2  WWAGIKQEFGIPYNP
```

Figure 22 (con't)

P135	WWAGIQQEFGIPYNP	Pol	846	860
Mos1	**GPGHKARVLAEAMSQ**			
Mos2	**GPSHKARVLAEAMSQ**			
G15	GPGHKARVLAEAMSQ	Gag	355	369

Mosaic 362-07:
13 CD8 PTE+, 5 CD4 PTE+
7 CD8+ regions, 5 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 4 2 1 1 1 CD4: 1 1 1 1 1

7 CD8 responses:

Mos1	**CTHGIRPVVSTQLLL**			
Mos2	**CTHGIKPVVSTQLLL**			
E15	CTHGIKPVVSTQLLL	Env	247	261
E47	CTHGIRPVVSTQLLL	Env	247	261
Mos1	KHIVW**ASRELERFAV**NPGLL			
Mos2	KHLVW**ASRELERFAL**NPGLL			
G20	KHLVWASRELERFAL	Gag	32	46
G68	-----ASRELERFA**V**NPGLL	Gag	37	51
Mos1	QPSLQT**GSEELRSLY**NTVATL			
Mos2	QPALQT**GTEELRSLF**NTVATL			
G268	QPALKTG**SEELRSLY**	Gag	65	79
G249	--ALQTG**SEELRSLF**NT	Gag	67	81
G132	---LQTG**TEELRSLF**NTV	Gag	68	82
G120	------G**SEELRSLY**NTVATL	Gag	71	85
Mos1	ACQGV**GGPGHKARVL**AEAMS			
Mos2	ACQGV**GGPS HKARVL**AEAMS			
G166	ACQEVGGP**G**HKARVL	Gag	349	363
G76	-----GGPSHKARVLAEAMG	Gag	354	368
Mos1	**RELNKRTQDFWEVQL**			
Mos2	**RELNKRTQDFWEVQL**			
P233	RELNRRTQDFWEVQL	Pol	233	247
Mos1	**LLRAIEAQQHLLQLT**			
Mos2	**LLRAIEAQQHMLQLT**			
E227	LLMAIEAQQH**L**LQLT	Env	555	569
Mos1	**KTAVQMAVFIHNFKR**			
Mos2	**KTAVQMAVFIHNFKR**			
P5	KTAVQMAVFIHNFKR	Pol	888	902

Figure 22 (con't)

```
5 CD4 responses:
Mos1  KLNWASQIYPGIKVR
Mos2  KLNWASQIYAGIKVK
B136  KLNWASQIYAGIKVK          Pol  418  432

Mos1  YAKMRTAHTNDVKQL
Mos2  YARMRGAHTNDVKQL
B266  YARKRSAHINDVKQL          Pol  509  523

Mos1  DQAEHLKTAVQMAVF
Mos2  DQAEHLKTAVQMAVF
B28   DQAEHLKTAVQMAVF          Pol  882  896

Mos1  SLQTGSEELRSLYNT
Mos2  ALQTGTEELRSLFNT
G182  ALKDGTEELRSLYNT          Gag  67   81

Mos1  GPGHKARVLAEAMSQ
Mos2  GPSHKARVLAEAMSQ
G15   GPGHKARVLAEAMSQ          Gag  355  369
```

Reactive peptides and breadth and depth of responses: Optimal Natural Clade C Vaccine

```
Natural Clade C 363-07:
2 CD8 PTE+, 3 CD4 PTE+
1 CD8- regions, 0 CD4+ regions
Number of overlapping peptides per region: CD8: 2 CD4: none

1 CD8 responses:
0     QPALQTGTEELRSLYNTVAIL
G268  IPALKTGSEELRSLY
G130  ------GSEELRSLYNTVAIL
```

```
Natural Clade C 364-07:
5 CD8 PTE+, 2 CD4 PTE+
5 CD8- regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 1 1 1 1 1 CD4: 1 1

5 CD8 responses:
0     AHTNDVKQLTEAVQK
B10   AHTNDVKQLIEAVQK Pol  515  529

0     GDIIGDIRQAHCNIS
B64   GDIIGDIRQAHCNIS Env  321  334
```

Figure 22 (con't)

C SEKSAVGIGAVFLGF
E182 REKRAVGLGAVFLGF Env 508 522

C RASILRGGKLDKWEK
G283 SASILRGGKLDKWEK Gag 4 18

C TNNPPVPVGDIYKRW
G185 TSNPPVPVGDIYKRW Gag 231 245

2 CD4 responses:
C CKSNITGLLLTRDGG
E316 CTSNITGLLLVRDGG Env 445 459

C SLYNTVATLYCVHAG
G54 SLYNTVATLYCVHQR Gag 77 91

Natural Clade C 365-07:
2 CD8 PTE+, 1 CD4 PIE+
2 CD8+ regions, 1 CD4+ regions
Number of overlapping peptides per region: CD8: 1 1 CD4: 1

2 CD8 responses:
C EQLINKERVYLSWVP
P251 EPLIKKEKVYLSWVP Pol 678 692

C AEWDRLHPVHAGPIA
G44 AEWDRLHPVHAGPIA Gag 210 224

1 CD4 response:
C HQAAMQMLKDTINEE
G214 HQAAMQMLKDTINEE Gag 194 208

Natural Clade C 366-07:
6 CD8 PTE+, 1 CD4 PIE+
4 CD8+ regions, 1 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 1 1 CD4: 1

4 CD8 responses:
C IVQQQSNLLRAIEAQQ
E54 VVQQQSNLLRAIEAQ Env 548 562
E72 -VQQQNNLLRAIEAQH Env 549 563

C AVFIHNFKRKGGIGGY
P22 AVFIHNFKRKGGIGG Pol 894 908
P236 -VLIHNFKRKGGIGGY Pol 895 909

Figure 22 (con't)

```
C     MAICEEMEKEGKITK
P004  TAICEEMEKEGKITK        Pol   190   204

C     GGPSHKARVLAEAMS
G76   GGPSHKARVLAEAMG        Gag   354   368

1 CD4 response:
C     IIGQVRDQAEHLKTA
P96   LIGQVRDQAEHLKTA Pol   876   890

_____________________________________________
Natural Clade C 367-07:
0 CD8 PTE+, 0 CD4 PTE+
0 CD8- regions, 0 CD4+ regions
Number of overlapping peptides per region: none

0 CD8 responses

0 CD4 responses

_____________________________________________
Natural Clade C 368-07:
2 CD8 PTE+, 0 CD4 PTE+
1 CD8- regions, 0 CD4+ regions
Number of overlapping peptides per region: CD8: 2 CD4: 0

1 CD8 response
C     AVFIHNFKRKGGIGGY
P22   AVFIHNFKRKGGIGG       Pol     894     908
P236  -VLIHNFKRKGGIGGY      Pol     895     909
```

Figure 23

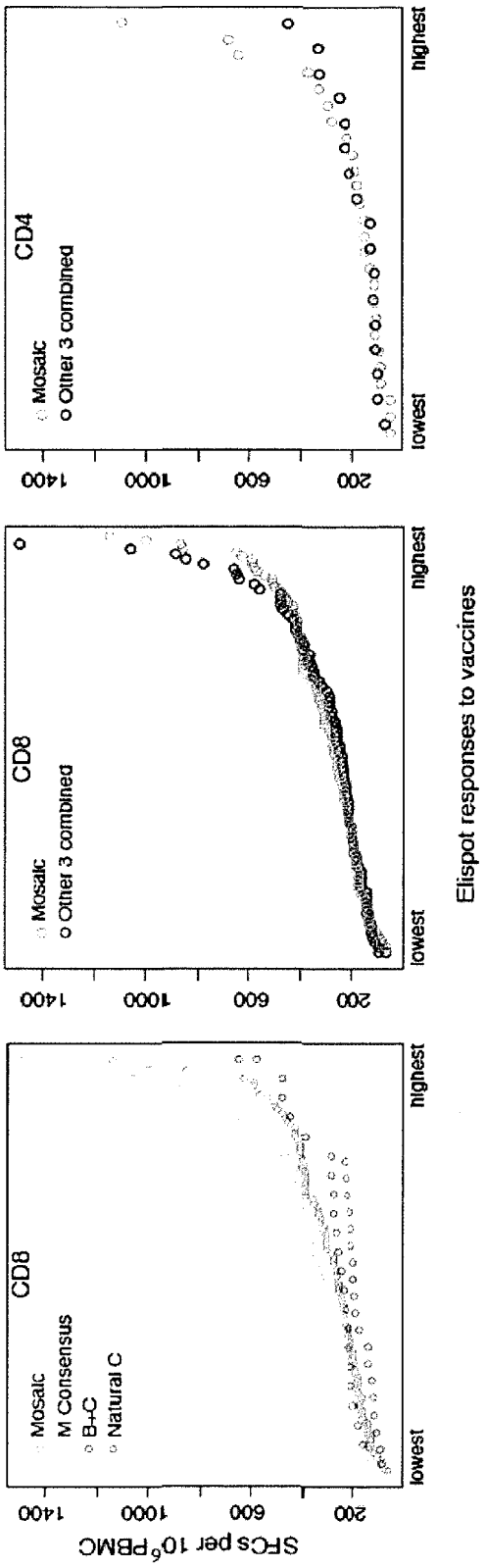

A
B
C
CD8
CD8
CD4
Mosaic
M Consensus
B-C
Natural C
Other 3 combined
SFCs per 10^6PBMC
200
600
1000
1400
lowest
highest
Elispot responses to vaccines

Figure 24

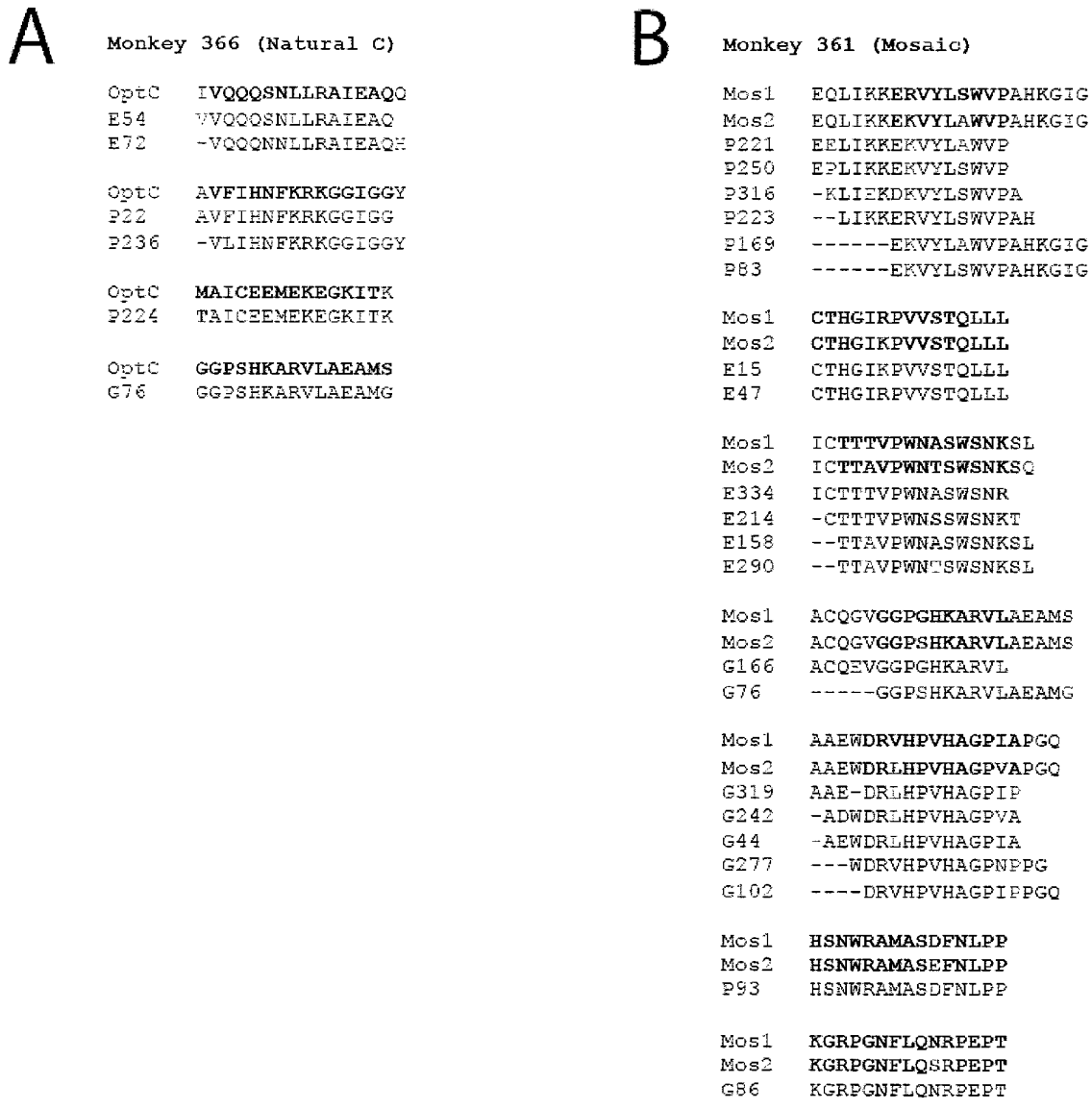

A
Monkey 366 (Natural C)
OptC IVQQQSNLLRAIEAQQ
E54 VVQQQSNLLRAIEAQ
E72 -VQQQNNLLRAIEAQH
OptC AVFIHNFKRKGGIGGY
P22 AVFIHNFKRKGGIGG
P236 -VLIHNFKRKGGIGGY
OptC MAICEEMEKEGKITK
P224 TAICEEMEKEGKITK
OptC GGPSHKARVLAEAMS
G76 GGPSHKARVLAEAMG
B
Monkey 361 (Mosaic)
Mos1 EQLIKKERVYLSWVPAHKGIG
Mos2 EQLIKKEKVYLAWVPAHKGIG
P221 EELIKKEKVYLAWVP
P250 EPLIKKEKVYLSWVP
P316 -KLIEKDKVYLSWVPA
P223 --LIKKERVYLSWVPAH
P169 ------EKVYLAWVPAHKGIG
P83 ------EKVYLSWVPAHKGIG
Mos1 CTHGIRPVVSTQLLL
Mos2 CTHGIKPVVSTQLLL
E15 CTHGIKPVVSTQLLL
E47 CTHGIRPVVSTQLLL
Mos1 ICTTTVPWNASWSNKSL
Mos2 ICTTAVPWNTSWSNKSQ
E334 ICTTTVPWNASWSNR
E214 -CTTTVPWNSSWSNKT
E158 --TTAVPWNASWSNKSL
E290 --TTAVPWNTSWSNKSL
Mos1 ACQGVGGPGHKARVLAEAMS
Mos2 ACQGVGGPSHKARVLAEAMS
G166 ACQEVGGPGHKARVL
G76 -----GGPSHKARVLAEAMG
Mos1 AAEWDRVHPVHAGPIAPGQ
Mos2 AAEWDRLHPVHAGPVAPGQ
G319 AAE-DRLHPVHAGPIP
G242 -ADWDRLHPVHAGPVA
G44 -AEWDRLHPVHAGPIA
G277 ---WDRVHPVHAGPNPPG
G102 ----DRVHPVHAGPIPPGQ
Mos1 HSNWRAMASDFNLPP
Mos2 HSNWRAMASEFNLPP
P93 HSNWRAMASDFNLPP
Mos1 KGRPGNFLQNRPEPT
Mos2 KGRPGNFLQSRPEPT
G86 KGRPGNFLQNRPEPT

Figure 24 (con't)

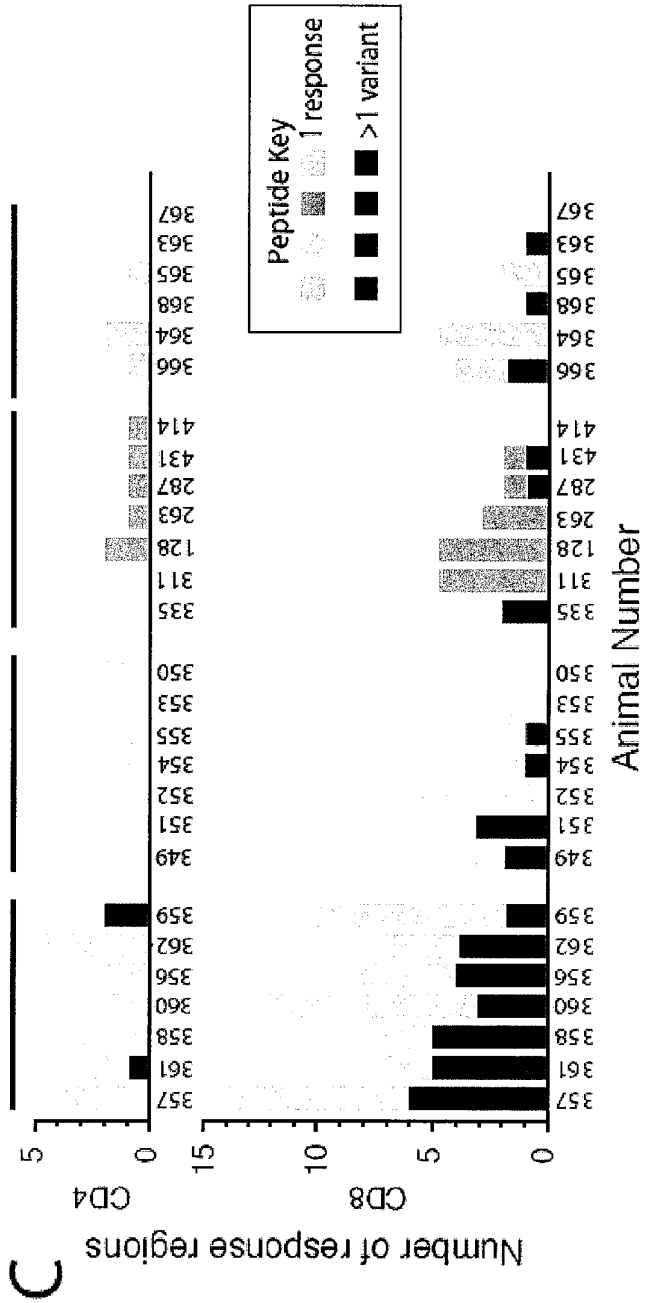


Figure 25

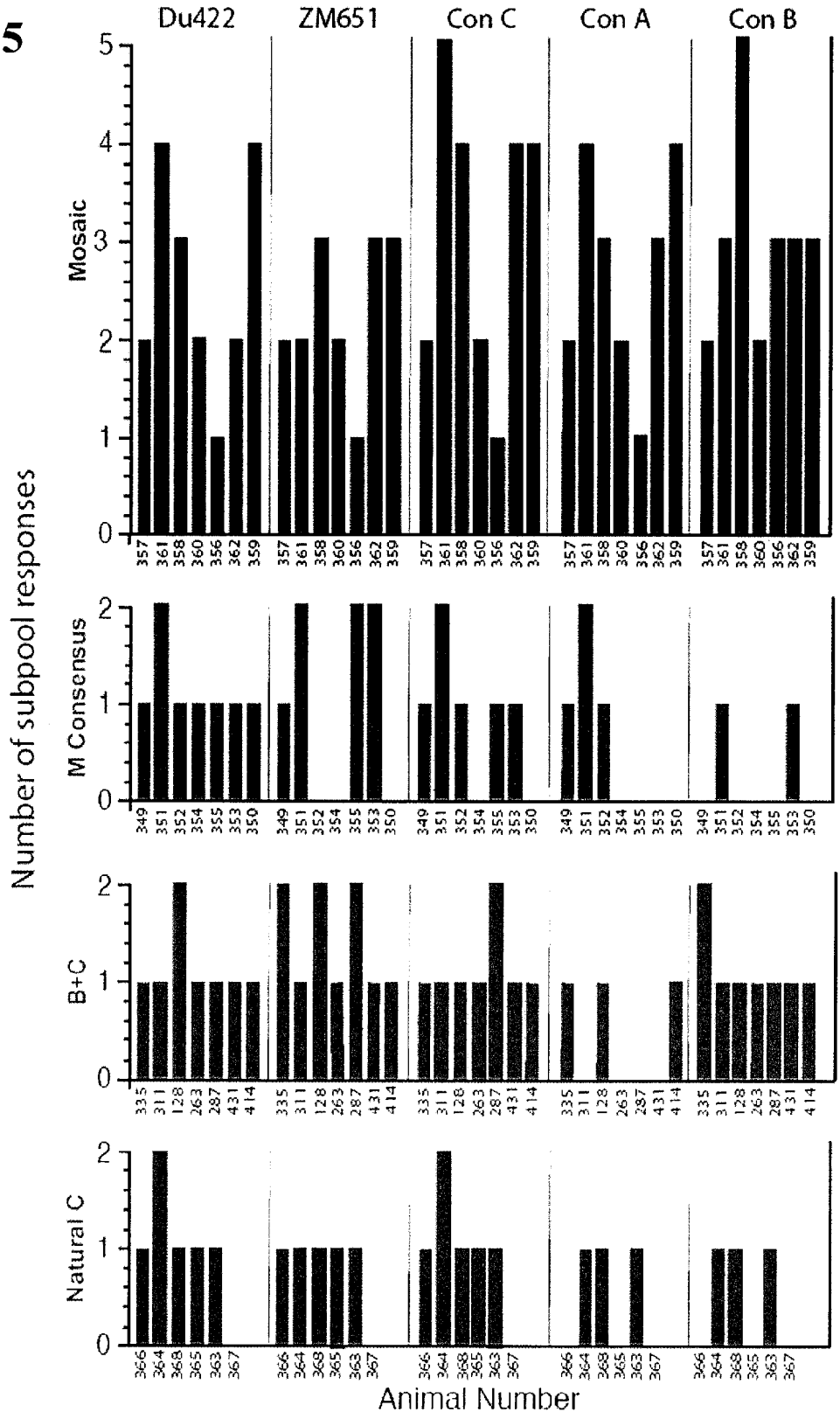

Du422
ZM651
Con C
Con A
Con B
Mosaic
M Consensus
B+C
Natural C
Number of subpool responses
Animal Number

Figure 26

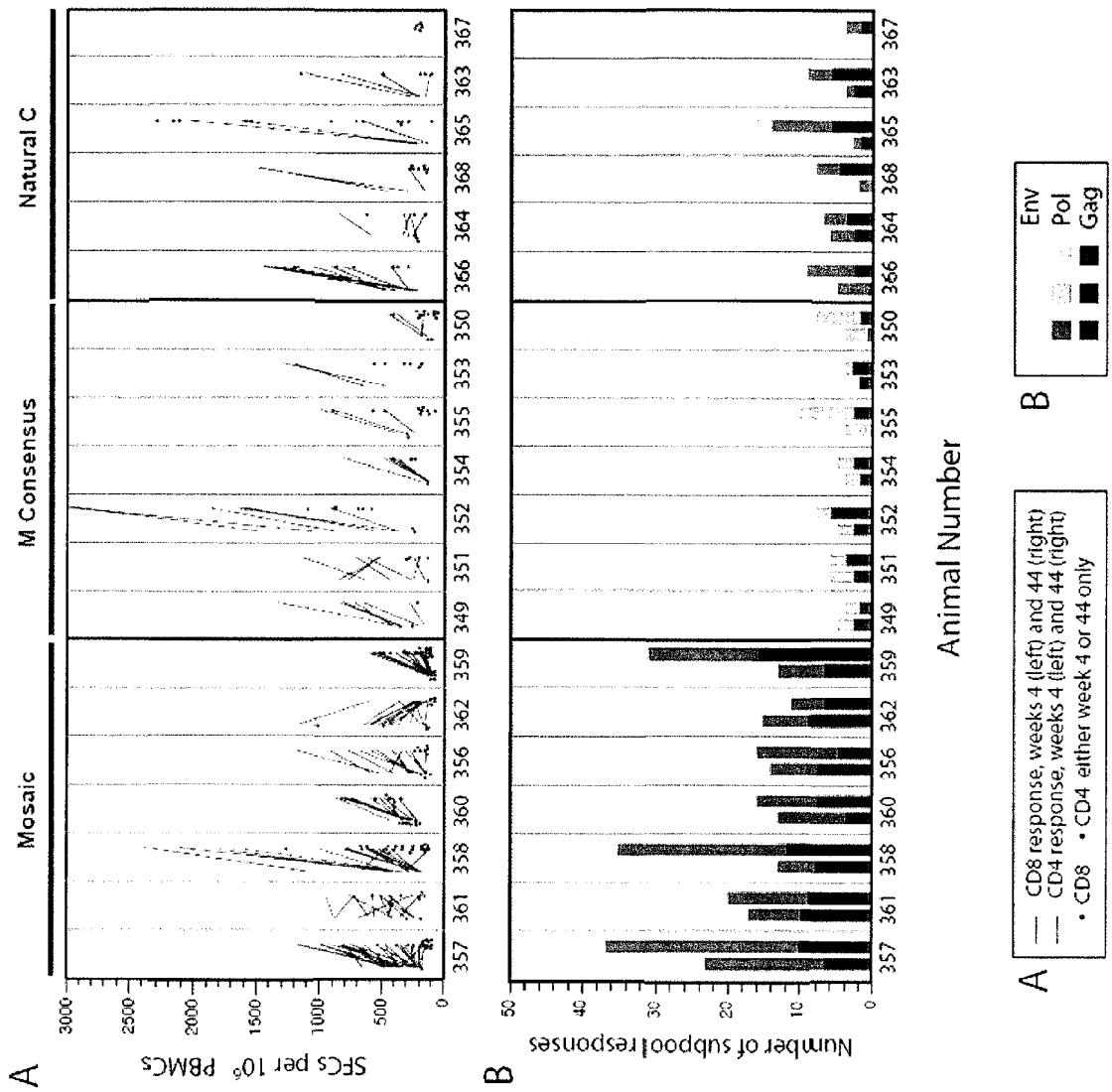
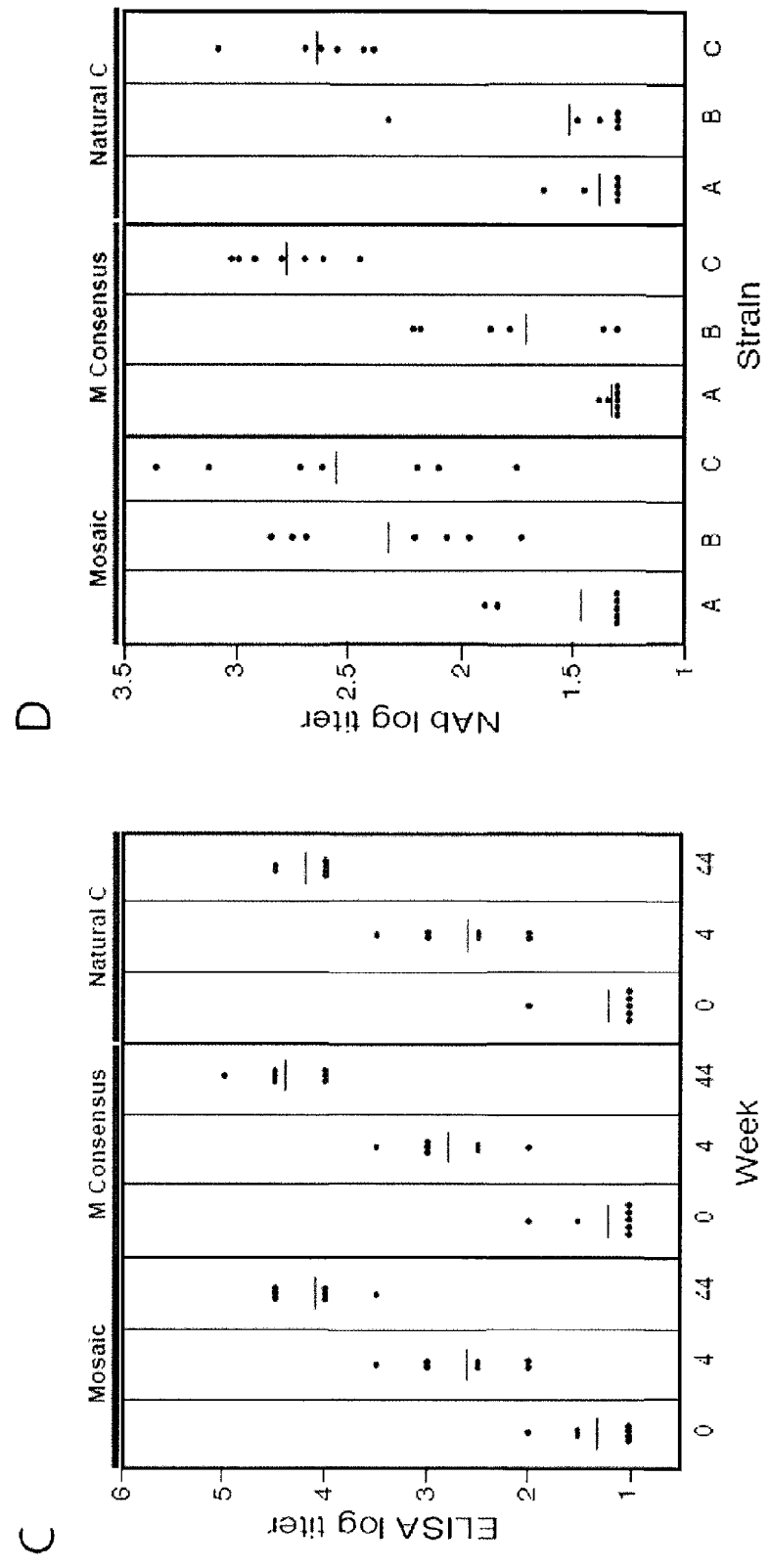

A
Mosaic
M Consensus
Natural C
SFCs per 10^6 PBMCs
3000
2500
2000
1500
1000
500
0
357 361 358 360 356 362 359 349 351 352 354 355 353 350 366 364 368 365 363 367
B
Number of subpool responses
50
40
30
20
10
0
Animal Number
A
CD8 response, weeks 4 (left) and 44 (right)
CD4 response, weeks 4 (left) and 44 (right)
CD8 • CD4 either week 4 or 44 only
B
Env
Pol
Gag

Figure 26 (con't)



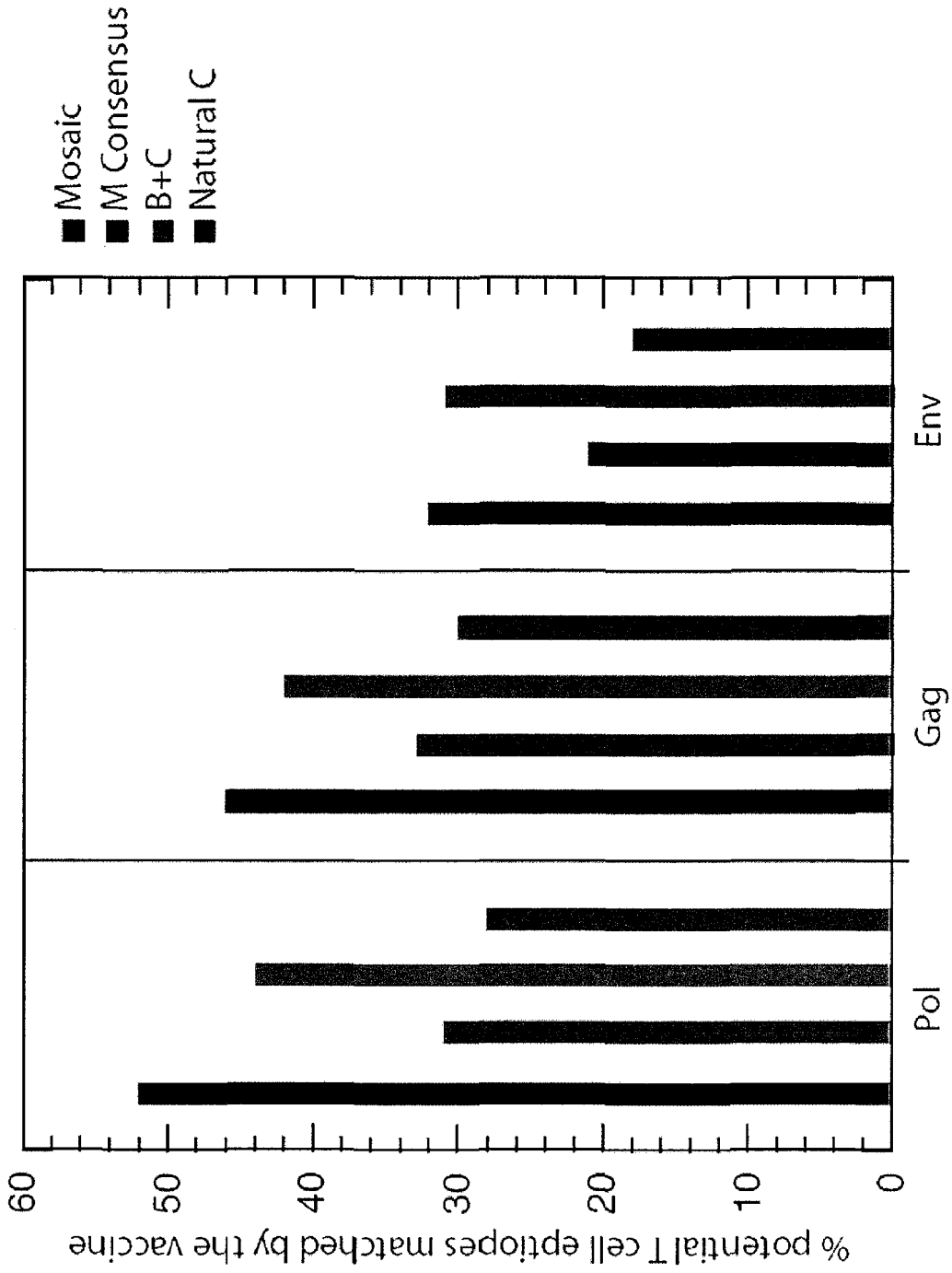

Mosaic
M Consensus
B+C
Natural C
% potential T cell epitopes matched by the vaccine
0
10
20
30
40
50
60
Pol
Gag
Env

Figure 27