

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2007256717 B2**

(54) Title
HIV-1 Clade A consensus sequences, antigens, and transgenes

(51) International Patent Classification(s)
A61K 39/21 (2006.01) **A61K 39/38** (2006.01)
A61K 39/12 (2006.01)

(21) Application No: **2007256717** (22) Date of Filing: **2007.06.04**

(87) WIPO No: **WO07/143606**

(30) Priority Data

(31) Number	(32) Date	(33) Country
60/810,816	2006.06.02	US

(43) Publication Date: **2007.12.13**

(44) Accepted Journal Date: **2013.06.20**

(71) Applicant(s)
International Aids Vaccine Initiative

(72) Inventor(s)
Gupta, Kalpana; Jackson, Nicholas

(74) Agent / Attorney
FB Rice, L 14 90 Collins ST, Melbourne, VIC, 3000

(56) Related Art
Nkolola, J. et al. 2004 GENE THER., Vol. 11, pp. 1068-1080
Howard, T. et al. 1996 AIDS RES. HUM RETRO., Vol. 12, No. 15, pp. 1413-1425
Malm, M. et al. 2005 VIRAL IMMUNOL., Vol. 18, No.4, pp. 678-688
WO 2001/47955 A2

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
13 December 2007 (13.12.2007)

PCT

(10) International Publication Number
WO 2007/143606 A3

(51) International Patent Classification:

A61K 39/21 (2006.01) A61K 39/38 (2006.01)
A61K 39/12 (2006.01)

(21) International Application Number:

PCT/US2007/070321

(22) International Filing Date: 4 June 2007 (04.06.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/810,816 2 June 2006 (02.06.2006) US

(71) Applicant (for all designated States except US): **INTERNATIONAL AIDS VACCINE INITIATIVE** [US/US];
110 Williams Street, 27th Floor, New York, NY 10038-
3901 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GUPTA, Kalpana**
[IN/US]; 410 E. 60th Street, Apt. 36c, New York, NY
10022 (US). **JACKSON, Nicholas** [GB/GB]; 61 Quinton
Street, London SW18 3QR (GB).

(74) Agents: **KOWALSKI, Thomas** et al.; Frommer Lawrence
& Haug LLP, 745 Fifth Avenue, New York, NY 10151
(US).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,

AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH,
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, PG, PH, PL,
PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, 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, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL,
PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments

(88) Date of publication of the international search report:
13 November 2008

(15) Information about Correction:

Previous Correction:
see Notice of 14 February 2008

(54) Title: HIV-1 CLADE A CONSENSUS SEQUENCES, ANTIGENS, AND TRANSGENES

(57) Abstract: The present invention relates to consensus nucleotide and protein sequences for HIV- 1 Clade A antigens, and to nucleotide and protein sequences for Clade A antigens from circulating HIV-I field isolates wherein the antigen sequences are closely related to the these consensus sequences. In a preferred embodiment, the present invention relates to HIV-I Clade A transgenes that are derived from such sequences, and that encode either HIV-I Clade A Gag, Pol (RT and Int), and Nef (referred to as "GRIN"), HIV-I Clade A Gag, RT, and Nef (referred to as ("GRN"), or HIV-I Clade A Env. The invention also relates to vectors containing such transgenes, including in preferred embodiment, adenovirus vectors containing such transgenes. The invention also relates to immunogenic compositions comprising the HIV-I Clade A antigens, nucleotide sequences, vectors, or transgenes of the invention, and to methods of generating an immune response against HIV in a subject by administering an effective amount of such immuno-
genic compositions.



WO 2007/143606 A3

TITLE OF THE INVENTION

HIV-1 CLADE A CONSENSUS SEQUENCES, ANTIGENS, AND TRANSGENES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Patent Application No. 60/810,816
5 filed June 2, 2006.

The foregoing applications, and all documents cited therein or during their
prosecution (“appln cited documents”) and all documents cited or referenced in the appln
cited documents, and all documents cited or referenced herein (“herein cited documents”),
and all documents cited or referenced in herein cited documents, together with any
10 manufacturer’s instructions, descriptions, product specifications, and product sheets for any
products mentioned herein or in any document incorporated by reference herein, are hereby
incorporated herein by reference, and may be employed in the practice of the invention.

FIELD OF THE INVENTION

The present invention relates to consensus nucleotide and protein sequences for HIV-
15 1 Clade A antigens, and to nucleotide and protein sequences for Clade A antigens from
circulating HIV-1 field isolates wherein the antigen sequences are closely related to the these
consensus sequences. In a preferred embodiment, the present invention relates to HIV-1
Clade A transgenes that are derived from such sequences, and that encode either HIV-1 Clade
A Gag, Pol (RT and Int), and Nef (referred to as “GRIN”), HIV-1 Clade A Gag, RT, and Nef
20 (referred to as (“GRN”), or HIV-1 Clade A Env. The invention also relates to vectors
containing such transgenes, including in a preferred embodiment, adenovirus vectors
containing such transgenes. The invention also relates to immunogenic compositions
comprising the HIV-1 Clade A antigens, nucleotide sequences, vectors, or transgenes of the
invention, and to methods of generating an immune response against HIV-1 in a subject by
25 administering an effective amount of such immunogenic compositions.

BACKGROUND OF THE INVENTION

AIDS, or Acquired Immunodeficiency Syndrome, is caused by human
immunodeficiency virus (HIV) and is characterized by several clinical features including
wasting syndromes, central nervous system degeneration and profound immunosuppression
30 that results in opportunistic infections and malignancies. HIV is a member of the lentivirus
family of animal retroviruses, which include the visna virus of sheep and the bovine, feline,
and simian immunodeficiency viruses (SIV). Two closely related types of HIV, designated
HIV-1 and HIV-2, have been identified thus far, of which HIV-1 is by far the most common

cause of AIDS. However, HIV-2, which differs in genomic structure and antigenicity, causes a similar clinical syndrome.

An infectious HIV particle consists of two identical strands of RNA, each approximately 9.2 kb long, packaged within a core of viral proteins. This core structure is surrounded by a phospholipid bilayer envelope derived from the host cell membrane that also includes virally-encoded membrane proteins (Abbas et al., Cellular and Molecular Immunology, 4th edition, W.B. Saunders Company, 2000, p. 454). The HIV genome has the characteristic 5'-LTR-Gag-Pol-Env-LTR-3' organization of the retrovirus family. Long terminal repeats (LTRs) at each end of the viral genome serve as binding sites for transcriptional regulatory proteins from the host and regulate viral integration into the host genome, viral gene expression, and viral replication.

The HIV genome encodes several structural proteins. The Gag gene encodes core structural proteins of the nucleocapsid core and matrix. The Pol gene encodes reverse transcriptase (RT), integrase (Int), and viral protease enzymes required for viral replication. The tat gene encodes a protein that is required for elongation of viral transcripts. The rev gene encodes a protein that promotes the nuclear export of incompletely spliced or unspliced viral RNAs. The Vif gene product enhances the infectivity of viral particles. The vpr gene product promotes the nuclear import of viral DNA and regulates G2 cell cycle arrest. The vpu and nef genes encode proteins that down regulate host cell CD4 expression and enhance release of virus from infected cells. The Env gene encodes the viral envelope glycoprotein that is translated as a 160-kilodalton (kDa) precursor (gp160) and cleaved by a cellular protease to yield the external 120-kDa envelope glycoprotein (gp120) and the transmembrane 41-kDa envelope glycoprotein (gp41), which are required for the infection of cells (Abbas, pp. 454-456). Gp140 is a modified form of the env glycoprotein which contains the external 120-kDa envelope glycoprotein portion and a part of the gp41 portion of env and has characteristics of both gp120 and gp41. The Nef gene is conserved among primate lentiviruses and is one of the first viral genes that is transcribed following infection. *In vitro*, several functions have been described, including down regulation of CD4 and MHC class I surface expression, altered T-cell signaling and activation, and enhanced viral infectivity.

HIV infection initiates with gp120 on the viral particle binding to the CD4 and chemokine receptor molecules (e.g., CXCR4, CCR5) on the cell membrane of target cells such as CD4⁺ T-cells, macrophages and dendritic cells. The bound virus fuses with the target cell and reverse transcribes the RNA genome. The resulting viral DNA integrates into the cellular genome, where it directs the production of new viral RNA, and thereby viral

proteins and new virions. These virions bud from the infected cell membrane and establish productive infections in other cells. This process also kills the originally infected cell. HIV can also kill cells indirectly because the CD4 receptor on uninfected T-cells has a strong affinity for gp120 expressed on the surface of infected cells. In this case, the uninfected cells bind, via the CD4 receptor-gp120 interaction, to infected cells and fuse to form a syncytium, which cannot survive. Destruction of CD4+ T-lymphocytes, which are critical to immune defense, is a major cause of the progressive immune dysfunction that is the hallmark of AIDS disease progression. The loss of CD4+ T cells seriously impairs the body's ability to fight most invaders, but it has a particularly severe impact on the defenses against viruses, fungi, parasites and certain bacteria, including mycobacteria.

The different isolates of HIV-1 have been classified into three groups: M (main), O (outlier) and N (non-M, non-O). The HIV-1 M group dominates the global HIV pandemic (Gaschen et al., (2002) Science 296: 2354-2360). Since the HIV-1 M group began its expansion in humans roughly 70 years ago (Korber et al., Retroviral Immunology, Pantaleo et al., eds., Humana Press, Totowa, NJ, 2001, pp. 1-31), it has diversified rapidly (Jung et al., (2002) Nature 418: 144). The HIV-1 M group consists of a number of different clades (also known as subtypes) as well as variants resulting from the combination of two or more clades, known as circulating recombinant forms (CRFs). Subtypes are defined as having genomes that are at least 25% unique (AIDS epidemic update, December 2002). Eleven clades have been identified and a letter designates each subtype. When clades combine with each other and are successfully established in the environment, as can occur when an individual is infected with two different HIV subtypes, the resulting virus is known as a CRF. Thus far, roughly 13 CRFs have been identified. HIV-1 clades also exhibit geographical preference. For example, Clade A, the second-most prevalent clade, is prevalent in East Africa, while Clade B is common in Europe, the Americas and Australia. Clade C, the most common subtype, is widespread in southern Africa, India and Ethiopia (AIDS epidemic update, December 2002). Even within Clades there is variability in the virus between different strains and viral isolates.

This genetic variability of HIV creates a scientific challenge to vaccine development. One approach that has been suggested is to develop consensus sequences based on the sequences of multiple different HIV strains, and to develop vaccines based on these consensus sequences. The rationale behind such approaches is that the consensus sequences will encode antigens that are conserved among different HIV strains and that such antigens are therefore likely to be useful in generating immune responses against multiple different

strains of HIV. HIV-1 clade A consensus sequences have been generated by others. See for example, Nkolola et al. (2004) Gene Ther. 2004. Jul. 11 (13): 1068-80, and Korber B (eds) et al. Human Retroviruses and AIDS: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences. Los Alamos National Laboratory: Los Alamos, New Mexico, USA, (1997) which involve transgene RENTA and HIVA derived from consensus clade A sequences. However, the consensus sequences described in these articles appear to have been derived from the HIV-1 clade A consensus sequence obtained from the Los Alamos laboratory, and were not generated in the same way as the consensus sequences of the present invention. In addition, these references do not teach use of sequences from actual recently circulating HIV strains which closely match the consensus sequence. Instead they involve using the consensus sequences themselves.

Citation or identification of any document in this application is not an admission that such document is available as prior art to the present application.

SUMMARY OF THE INVENTION

The present invention provides new and improved consensus sequences for HIV-1 Clade A antigens and methods for producing such new and improved consensus sequences. The consensus sequences of the present invention are particularly advantageous because they are based on the antigen sequences of a large number of different HIV-1 Clade A strains, and also because they are based on the sequences of antigens from recently isolated HIV-1 Clade A strains. Accordingly, the consensus sequences of the present invention have superior biological relevance as compared to previously generated HIV-1 Clade A consensus sequences.

Another major advantage of the present invention is that it provides HIV-1 Clade A antigens, and strategies for producing such antigens, that are derived from naturally occurring HIV-1 Clade A strains. These antigens are selected such that they are closely related to, or have a small "protein distance" from, the consensus sequences of the present invention. An advantage of using these naturally occurring sequences with the closest match to the consensus sequences, as opposed to the artificially generated consensus sequences, is that less genetic manipulations are needed to generate these sequences and importantly biological relevance is assured.

In a first aspect the present invention is directed to a consensus amino acid sequence for an HIV-1 Clade A antigen. In one embodiment the invention relates to consensus amino acid sequences for the HIV-1 Clade A antigens Gag, Pol (comprising RT and Int), Nef and Env. In preferred embodiments, the invention relates to the consensus Gag amino acid

sequence of FIG. 1, the consensus Pol amino acid sequence of FIG. 3, to the consensus Env amino acid sequence of FIG. 5, and/or the consensus Nef amino acid sequence of FIG. 7.

5 In another embodiment, the present invention is directed to a non-naturally occurring consensus amino acid sequence for an HIV-1 Clade A Gag protein, wherein the Gag protein has the amino acid sequence of SEQ ID NO: 2.

In another embodiment, the present invention is directed to a non-naturally occurring consensus amino acid sequence for an HIV-1 Clade A Pol protein, wherein the Pol protein has the amino acid sequence of SEQ ID NO: 3.

10 In another embodiment, the present invention is directed to a non-naturally occurring consensus amino acid sequence for an HIV-1 Nef protein, wherein the Nef protein has the amino acid sequence of SEQ ID NO: 5.

In another embodiment, the present invention is directed to a non-naturally occurring consensus amino acid sequence for an HIV-1 Clade A antigen comprising amino acid sequences for HIV-1 Clade A Gag, Pol and Nef proteins, wherein the Gag protein has the amino acid sequence of SEQ ID NO: 2, the Pol protein has the amino acid sequence of SEQ ID NO: 3 and the Nef protein has the amino acid sequence of SEQ ID NO: 5.

20 In another embodiment, the present invention is directed to an HIV-1 Clade A fusion protein, wherein the fusion protein comprises HIV-1 Clade A Gag, Pol and Nef and wherein the fusion protein comprises the amino acid sequence of SEQ ID NO: 13.

In a further aspect the present invention is directed to a method of identifying a consensus amino acid sequence for an HIV-1 Clade A antigen of interest comprising determining the amino acid sequence of the antigen of interest in several circulating HIV-1 strains or field isolates, aligning such sequences, and determining the consensus sequence for that antigen.

30 In another aspect, the invention relates to a method of identifying an HIV-1 Clade A antigen from a circulating strain or field isolate of HIV-1 Clade A that has an amino acid sequence that is similar to the consensus amino acid sequence for that HIV-1 Clade A antigen. In a preferred embodiment the HIV-1 Clade A antigen is selected based on the degree of similarity to the consensus sequence, with sequences having the highest degree of similarity to, or the smallest "protein distance" from, the consensus sequence being preferred. In a further preferred embodiment the HIV-1 Clade A antigen is selected from a recently circulating strain or field isolate of HIV-1 Clade A.

35 In a further embodiment the invention relates to HIV-1 Clade A antigens identified using such methods.

In another aspect, the invention relates to a method of identifying an HIV-1 Clade A antigen from a circulating strain or field isolate of HIV-1 Clade A that has an amino acid sequence that is similar to the consensus amino acid sequence for that HIV-1 Clade A antigen, and then making mutations in that sequence to abrogate the biological functions of the sequences. It is preferred that a minimalist approach is used, i.e. that the number of mutations is kept to a minimum so that only those mutations necessary to abrogate function and facilitate obtaining regulatory authority approval are made and un-necessary alteration of the original HIV-1 gene sequences are avoided. For example, in one embodiment the Nef component of GRIN is not altered but rather fusion of the Nef N-terminus to the Int C-terminus abrogates nef function while retaining all the original nucleotide sequences of Nef.

In yet another aspect, the invention relates to a method of improving genetic stability of the HIV-1 Clade A transgene for insertion into viral vector technologies. The PR (protease) component is removed from Gag-full-length Pol-Nef (full length Pol contains PR, and Int and RT) so that only the Int and RT portions of Pol are left. This has the advantage of improved genetic stability and improved cloning and virus rescue properties, particularly using Ad35 and/or Ad11. Removing PR in this way is a minimalist approach in that only the smallest functional subunit of POL is removed, thereby preserving the larger IN & RT functional subunits. The invention also relates to HIV-1 Clade A antigens selected and produced using such methods.

In one embodiment the antigen is a Gag antigen from one of the strains listed in Table 1 and FIG. 2. Preferably the Gag antigen is selected from a strain in which the "protein distance" from the consensus Gag sequence is less than 0.07%, or more preferably less than 0.06%, or more preferably still less than 0.05%. In a preferred embodiment the Gag antigen is from HIV-1 Clade A strain TZA173, strain 97TZ02, strain KHNH1144 or strain SE7535UG.

In another embodiment the antigen is a Pol antigen from one of the strains listed in Table 2 and FIG. 4. Preferably the Pol antigen is selected from a strain in which the "protein distance" from the consensus Pol sequence is less than 0.03%, or more preferably less than 0.025%. In a preferred embodiment the Pol antigen is from HIV-1 Clade A strain MSA4070, strain SE7245SO, or strain SE8538.

In a further embodiment the antigen is an Env antigen from one of the strains listed in Table 3 and FIG. 6. Preferably the Env antigen is selected from a strain in which the "protein distance" from the consensus Gag sequence is less than 0.1, or more preferably less than 0.08%, or more preferably less than 0.07%, or more preferably still

less than 0.065%. In a preferred embodiment the Env antigen is from HIV-1 Clade A strain KEQ23, strain TZA341, or strain KNH1088.

In another embodiment the antigen is a Nef antigen from one of the strains listed in Table 4 and FIG. 8. Preferably the Nef antigen is selected from a strain in which the "protein distance" from the consensus Gag sequence is less than 0.1%, or more preferably less than 0.08%, or more preferably less than 0.07%, or more preferably less than 0.06, or more preferably still, less than 0.05%. In a preferred embodiment the Nef antigen is from HIV-1 Clade A strain MSA4070, or strain KNH1211, or strain 97TZ03, or strain 99UGA070, or strain SE8891UG.

In yet another aspect, the present invention is directed to the nucleotide sequences that encode the HIV-1 Clade A antigens of the invention. The invention also relates to vectors comprising these nucleotide sequences. The nucleotide sequences of the invention, and the vectors that comprise them, and also the antigens encoded by the nucleotide sequences of the invention, are useful in generating an immune response against HIV Clade A antigens *in vivo* and are useful in the production of vaccines against HIV-1 Clade A strains. The nucleotide sequences of the invention may also be useful for expressing and producing the HIV-1 Clade A antigens that they encode in cells or *in vitro*, for example, so that the antigens may be produced, isolated, and/or purified.

In one embodiment, the present invention is directed to a polynucleotide encoding the amino acid sequence of the invention or the fusion protein of the invention.

The nucleotides of the invention may be altered as compared to the consensus nucleotide sequences, or as compared to the sequences from circulating HIV-1 isolates that are closely related to such consensus sequences. For example, in one embodiment the nucleotide sequences may be mutated such that the activity of the encoded proteins *in vivo* is abrogated. In another embodiment the nucleotide sequences may be codon optimized, for example the codons may be optimized for human use. In preferred embodiments the nucleotide sequences of the invention are both mutated to abrogate the normal *in vivo* function of the encoded proteins, and codon optimized for human use. For example, each of the Gag, Pol, Env, Nef, RT, and Int sequences of the invention may be altered in these ways.

In a preferred embodiment, a single nucleotide sequence encodes a fusion protein comprising the Gag, RT (part of Pol) and Nef antigens of the invention. As used herein the abbreviations "GRN" and "GRtN" are used interchangeably to refer to HIV-1 Clade A fusion proteins comprising the Gag, RT and Nef antigens and to refer to the

nucleotide sequences that encode these fusion proteins. In a still more preferred embodiment the nucleotide sequence encoding GRN is inserted into a vector suitable for allowing expression of the GRN fusion protein. Preferably the vector is an adenovirus vector selected from the group consisting of Ad5, Ad35, Ad11, C6, and C7.

5 In another preferred embodiment a single nucleotide sequence encodes a fusion protein comprising the Gag, Pol (includes RT and Int) and Nef antigens of the invention. As used herein the abbreviations "GRIN" and "GRtIN" are used interchangeably to refer to HIV-1 Clade A fusion proteins comprising the Gag, Pol and Nef antigens and to refer to the nucleotide sequences that encode these fusion proteins.

10 In even more preferred embodiments GRIN has the amino acid sequence illustrated in FIGS. 16A-16J and is encoded by the nucleotide sequence illustrated in FIGS. 16A-16J. In a still more preferred embodiment the nucleotide sequence encoding GRIN is inserted into a vector suitable for allowing expression of the GRIN fusion protein. Preferably the vector is an adenovirus vector, more preferably and adenovirus vector

15 selected from the group consisting of Ad5, Ad35, Ad11, C6, and C7.

In yet another embodiment a single nucleotide sequence of the invention encodes an HIV-1 Clade A Env antigen according to the invention. In a preferred embodiment the Env antigen has the amino acid sequence illustrated in FIGS. 17A-17D and is encoded by the nucleotide sequence illustrated in FIGS. 17A-17D. In a still more

20 preferred embodiment the nucleotide sequence encoding Env is inserted into a vector suitable for allowing expression of the Env protein. Preferably the vector is an adenovirus vector, more preferably and adenovirus vector selected from the group consisting of Ad5, Ad35, Ad11, C6, and C7.

In yet another aspect, the present invention is directed to an immunogenic

25 composition comprising the amino acid sequence of the invention, the fusion protein of the invention, or the polynucleotide of the invention.

In another embodiment, the present invention provides methods of generating an immune response against HIV-1 Clade A antigens comprising administering to a subject a nucleotide sequence or antigen according to the invention. In preferred

30 embodiments the method of generating an immune response against HIV-1 Clade A comprises administering a nucleotide sequence encoding either GRIN or GRN wherein the nucleotide sequence is contained in an adenovirus vector selected from the group consisting of Ad5, Ad35, Ad11, C6, and C7. In further preferred embodiments, the vectors comprising GRIN or GRN are co-administered with a vector comprising a

35 nucleotide sequence encoding an Env antigen of the invention.

In another the embodiment, the present invention is directed to a method of generating an immune response against HIV-1 comprising administering to a subject the amino acid sequence of the invention, the fusion protein of the invention, the polynucleotide of the invention, or the composition of the invention.

5 In another the embodiment, the present invention is directed use of the amino acid sequence of the invention, the fusion protein of the invention, the polynucleotide the invention, or the composition of the invention, in the manufacture of a medicament for generating an immune response against HIV-1 in a subject.

10 In a further embodiment, the present invention provides immunogenic compositions or vaccine compositions comprising the nucleotide sequences of the invention.

It should be noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as "comprises", "comprised", "comprising" and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean "includes",
15 "included", "including", and the like; and that terms such as "consisting essentially of" and "consists essentially of" have the meaning ascribed to them in U.S. Patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention.

20 Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present disclosure as it existed before the priority date of each claim of this application.

25 These and other embodiments are disclosed or are obvious from and encompassed by, the following Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

30 The following Detailed Description, given by way of example, but not intended to limit the invention to the specific embodiments described, may be best understood in conjunction with the accompanying Figures.

FIG. 1 is a consensus amino acid sequence of the Gag protein of HIV-1 Clade A (SEQ ID NO: 2).

35 FIG. 2 is a graph illustrating the "distance" of the Gag protein sequences of circulating HIV-Clade A strains to that of the consensus HIV-1 Clade A Gag protein sequence.

FIG. 3 is a consensus amino acid sequence of the Pol protein of HIV-1 Clade A (SEQ ID NO: 3).

FIG. 4 is a graph illustrating the "distance" of the Pol protein sequences of circulating HIV-Clade A strains to that of the consensus HIV-1 Clade A Pol protein sequence.

FIG. 5 is a consensus amino acid sequence of the Env protein of HIV-1 Clade A (SEQ ID NO: 4).

FIG. 6 is a graph illustrating the "distance" of the Env protein sequences of circulating HIV-Clade A strains to that of the consensus HIV-1 Clade A Env protein sequence.

FIG. 7 is a consensus amino acid sequence of the Nef protein of HIV-1 Clade A (SEQ ID NO: 5).

FIG. 8 is a graph illustrating the "distance" of the Nef protein sequences of circulating HIV-Clade A strains to that of the consensus HIV-1 Clade A Nef protein sequence.

FIG. 9 is a schematic representation of the GRIN and GRN transgenes.

FIG. 10 illustrates the amino acid sequence of the Gag protein from HIV-1 Clade A strain TZA173 (SEQ ID NO: 6) having Genbank accession number AY253305.

FIG. 11 illustrates the amino acid sequence of the Pol protein from HIV-1 Clade A strain MSA4070 (SEQ ID NO: 7) having Genbank accession number AF457081.

FIG. 12 illustrates the amino acid sequence of the Nef protein from HIV-1 Clade A strain MSA4070 (SEQ ID NO: 8) having Genbank accession number AF457081.

FIG. 13 illustrates the amino acid sequence of the Env protein from HIV-1 Clade A strain TZA341 (SEQ ID NO: 9) having Genbank accession number AY253314.

FIGS. 14A-14C provide a sequence of GRIN as inserted into the Ad35 vector (SEQ ID NO: 10).

FIGS. 15A-15B provide a sequence of Env as inserted into the Ad35 vector (SEQ ID NO: 11).

FIGS. 16A-16J provide nucleotide (SEQ ID NO: 12) and amino acid (SEQ ID NO: 13) sequences of the codon optimized GRIN transgene.

FIGS. 17A-17D provide nucleotide (SEQ ID NO: 14) and amino acid (SEQ ID NO: 15) sequences of the codon optimized Env transgene.

FIG. 18 illustrates graphically the immunogenicity of Ad5-GRIN and Ad5-GRN in mice as measured by IFN-gamma ELIspot assay.

FIG. 19 illustrates graphically the immunogenicity of C7-GRIN and C7-GRN in mice as measured by IFN-gamma ELISpot assay.

FIG. 20 illustrates graphically the immunogenicity of C7-GRIN and C7-GRN in mice as measured by IL-2 ELISpot assay.

5 FIG. 21 illustrates graphically the immunogenicity of C6-GRIN and C6-GRN in mice as measured by IFN-gamma ELISpot assay.

FIG. 22 illustrates graphically the immunogenicity of C6-GRIN and C6-GRN in mice as measured by IL-2 ELISpot assay.

10 FIG. 23A illustrates IFN- γ ELISpot immunogenicity of Ad35-GRIN/ENV at the 10^{10} vp dose following a month 0-6 immunization schedule in rhesus macaques.

Definition of Positive Response: For a single peptide pool from a single sample:

Response = (mean peptide

count – mean no-peptide count). To be positive, a single peptide response must satisfy: 1. Mean peptide count > 4x mean no-peptide count from same plate; 2. Coefficient of variation amongst replicate counts $\leq 70\%$ & 3. Response > 55 SFC/106. Geometric mean responses for Spot Forming Cells (SFC) per million PBMCs to each antigen component (Gag, RT, IN and ENV) are shown on the y-axis and bleed timepoints in weeks on the x-axis.

FIG. 23B illustrates IFN- γ ELISpot immunogenicity of Ad35-GRIN/ENV at the 10^{11} vp dose following a month 0-6 immunization schedule in rhesus macaques. Definition of Positive Response: For a single peptide pool from a single sample: Response = (mean peptide count – mean no-peptide count). To be positive, a single peptide response must satisfy: 1. Mean peptide count > 4x mean no-peptide count from same plate; 2. Coefficient of variation amongst replicate counts $\leq 70\%$ & 3. Response > 55 SFC/106. Geometric mean responses for Spot Forming Cells (SFC) per million PBMCs to each antigen component (Gag, RT, IN and ENV) are shown on the y-axis and bleed timepoints in weeks on the x-axis.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to consensus nucleotide and protein sequences for HIV-1 clade A antigens, and to circulating HIV-1 field isolates that closely match these consensus sequences. The invention also relates to altered version of these sequences, which may be altered such that the function of the gene products *in vivo* is abrogated, to constructs and vectors comprising the sequences of the invention, and to immunogens, immunogenic compositions, and vaccines made using the sequences of the invention. The invention also relates to methods of generating an immune response against HIV-1 Clade A antigens in a subject and to methods of inducing protective immunity against challenge with HIV-1. The various embodiments of the invention are summarized above in the section entitled “Summary of the Invention.” Further details of the invention are provided in the Detailed Description and Examples that follow, and also in the Drawings.

As described in the above “Summary of the Invention” and the “Examples” below, the present invention provides HIV-1 Clade A consensus antigens, and also antigens from circulating HIV-1 Clade A strains that are closely related to these consensus sequences. The invention also provides HIV-1 transgenes and antigens encoded by these transgenes. These transgenes comprise sequences encoding the HIV-1 Clade A antigens of the invention, for example the Gag, Pol, Env, Nef, RT, and Int antigens of the invention. For example, in one preferred embodiment the present invention provides a GRIN (also referred to as GRtIN) transgene which comprises Gag, Pol (both RT and Int) and Nef antigens of the invention. In another preferred embodiment the present invention provides a GRN (also referred to as

GRtN) transgene which comprises the Gag, RT and Nef antigens of the invention. In another embodiment the present invention provides an Env transgene which comprises and Env antigens of the invention.

The terms "protein", "peptide", "polypeptide", and "amino acid sequence" are used interchangeably herein to refer to polymers of amino acid residues of any length. The polymer may be linear or branched, it may comprise modified amino acids or amino acid analogs, and it may be interrupted by chemical moieties other than amino acids. The terms also encompass an amino acid polymer that has been modified naturally or by intervention; for example disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation or modification, such as conjugation with a labeling or bioactive component.

As used herein, the terms "antigen" or "immunogen" are used interchangeably to refer to a substance, typically a protein, which is capable of inducing an immune response in a subject. The term also refers to proteins that are immunologically active in the sense that once administered to a subject (either directly or by administering to the subject a nucleotide sequence or vector that encodes the protein) is able to evoke an immune response of the humoral and/or cellular type directed against that protein.

It should be understood that the proteins and antigens of the invention may differ from the exact sequences illustrated and described herein. Thus, the invention contemplates deletions, additions and substitutions to the sequences shown, so long as the sequences function in accordance with the methods of the invention. In this regard, particularly preferred substitutions will generally be conservative in nature, *i.e.*, those substitutions that take place within a family of amino acids. For example, amino acids are generally divided into four families: (1) acidic--aspartate and glutamate; (2) basic--lysine, arginine, histidine; (3) non-polar--alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar--glycine, asparagine, glutamine, cystine, serine threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids. It is reasonably predictable that an isolated replacement of leucine with isoleucine or valine, or vice versa; an aspartate with a glutamate or vice versa; a threonine with a serine or vice versa; or a similar conservative replacement of an amino acid with a structurally related amino acid, will not have a major effect on the biological activity. Proteins having substantially the same amino acid sequence as the sequences illustrated and described but possessing minor amino acid substitutions that do not substantially affect the immunogenicity of the protein are, therefore, within the scope of the invention.

In one embodiment the present invention is directed to "consensus" amino acid sequences for an HIV-1 Clade A antigens. In one embodiment the invention relates to consensus amino acid sequences for the HIV-1 Clade A antigens Gag, Pol (comprising RT and Int), Nef and Env. In preferred embodiments, the invention relates to a consensus Gag amino acid sequence of FIG. 1, the consensus Pol amino acid sequence of FIG. 3, to a
5 consensus Env amino acid sequence of FIG. 5, and/or a consensus Nef amino acid sequence of FIG. 7. In a further aspect the present invention is directed to a method of identifying a consensus amino acid sequence for an HIV-1 Clade A antigen of interest comprising obtaining the amino acid sequence of the antigen of interest in several circulating HIV-1
10 strains or field isolates, aligning such sequences, and determining the consensus sequence for that antigen. For example, in one embodiment a database is generated using available sequences for HIV-1 Clade A non-recombinant circulating strains, and the individual HIV-1 genes (for example gag, pol, nef and env) from all the sequences in the database are then aligned, with dashes inserted to maintain alignment in regions with insertions or deletions in
15 the sequence, and a 50% consensus sequence can then be derived.

The present invention also relates to methods of identifying antigens from naturally occurring HIV-1 Clade A strains that have an amino acid sequence that has a small "protein distance" from the consensus amino acid sequence of that antigen. The "protein distance" is a measure of the level of similarity or difference between two amino acid sequences. Two
20 amino acid sequences that are very similar have a low protein distance. Two amino acid sequences that are very different have a high protein distance. Protein distances are preferably calculated using the Dayhoff PAM250 substitution matrix (M.O. Dayhoff, ed., 1978, Atlas of Protein Sequence and Structure, Vol. 5) which weights substitutions according to the degree of biochemical similarity. However, other methods for determining protein
25 distance can also be used.

As used herein the terms "nucleotide sequences" and "nucleic acid sequences" refer to deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) sequences, including, without limitation, messenger RNA (mRNA), DNA/RNA hybrids, or synthetic nucleic acids. The nucleic acid can be single-stranded, or partially or completely double-stranded (duplex).
30 Duplex nucleic acids can be homoduplex or heteroduplex.

As described in the above "Summary of the Invention" and the "Examples" below, the present invention provides HIV-1 Clade A consensus antigens and to the nucleotide sequences that encode these consensus antigen. The invention also relates to antigens from circulating HIV-1 Clade A strains that are closely related to these consensus sequences, and

to the nucleotide sequences that encode them. The invention also provides HIV-1 Clade A transgenes which comprise sequences encoding the HIV-1 Clade A antigens of the invention. As used herein the term “transgene” is used to refer to “recombinant” nucleotide sequences that are derived from either the HIV-1 Clade A consensus nucleotide sequences of the
5 invention, or from the nucleotide sequences that encode the antigens from recently circulating HIV-1 Clade A strains that have been identified as being closely matched to these consensus sequences. The term “recombinant” means a nucleotide sequence that has been manipulated “by man” and which does not occur in nature, or is linked to another nucleotide sequence or found in a different arrangement in nature. It is understood that manipulated “by man” means
10 manipulated by some artificial means, including by use of machines, codon optimization, restriction enzymes, etc. For example, in preferred embodiments the present invention provides the GRIN, GRN, and Env transgenes.

The nucleotides of the invention may be altered as compared to the consensus nucleotide sequences, or as compared to the sequences from circulating HIV-1 isolates that
15 are closely related to such consensus sequences. For example, in one embodiment the nucleotide sequences may be mutated such that the activity of the encoded proteins in vivo is abrogated. In another embodiment the nucleotide sequences may be codon optimized, for example the codons may be optimized for human use. In preferred embodiments the nucleotide sequences of the invention are both mutated to abrogate the normal in vivo
20 function of the encoded proteins, and codon optimized for human use. For example, each of the Gag, Pol, Env, Nef, RT, and Int sequences of the invention may be altered in these ways.

The types of mutations that can be made to abrogate the in vivo function of the antigens include, but are not limited to, the following which are also described in Example 7: Mutation of Gly2 to Ala in Gag to remove a myristylation site and prevent formation of
25 virus-like-particles (VLPs); Mutation of Gag to avoid slippage at the natural frame shift sequence to leave the conserved amino acid sequence (NFLG) intact and allow only the full-length GagPol protein product to be translated; Mutation of RT Asp185 to Ala and mutation of Asp186 to Ala to inactivate active enzyme residues. Mutation of Int Asp 64 to Ala, and mutation of Asp116 to Ala and mutation of Glu 152 to Ala to inactivate active enzyme
30 residues.

As regards codon optimization, the nucleic acid molecules of the invention have a nucleotide sequence that encodes the antigens of the invention and can be designed to employ codons that are used in the genes of the subject in which the antigen is to be produced. Many viruses, including HIV and other lentiviruses, use a large number of rare codons and, by

altering these codons to correspond to codons commonly used in the desired subject, enhanced expression of the antigens can be achieved. In a preferred embodiment, the codons used are "humanized" codons, i.e., the codons are those that appear frequently in highly expressed human genes (Andre et al., J. Virol. 72:1497-1503, 1998) instead of those codons that are frequently used by HIV. Such codon usage provides for efficient expression of the transgenic HIV proteins in human cells. Any suitable method of codon optimization may be used. For example, codons may be optimized for human usage as illustrated in Example 8. However, any other suitable methods of codon optimization may be used. Such methods, and the selection of such methods, are well known to those of skill in the art. In addition, there are several companies that will optimize codons of sequences, such as Geneart (geneart.com). Thus, the nucleotide sequences of the invention can readily be codon optimized.

The invention further encompasses nucleotide sequences encoding functionally and/or antigenically equivalent variants and derivatives of the antigens of the invention and functionally equivalent fragments thereof. These functionally equivalent variants, derivatives, and fragments display the ability to retain antigenic activity. For instance, changes in a DNA sequence that do not change the encoded amino acid sequence, as well as those that result in conservative substitutions of amino acid residues, one or a few amino acid deletions or additions, and substitution of amino acid residues by amino acid analogs are those which will not significantly affect properties of the encoded polypeptide. Conservative amino acid substitutions are glycine/alanine; valine/isoleucine/leucine; asparagine/glutamine; aspartic acid/glutamic acid; serine/threonine/methionine; lysine/arginine; and phenylalanine/tyrosine/tryptophan. In one embodiment, the variants have at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% homology or identity to the antigen, epitope, immunogen, peptide or polypeptide of interest.

For the purposes of the present invention, sequence identity or homology is determined by comparing the sequences when aligned so as to maximize overlap and identity while minimizing sequence gaps. In particular, sequence identity may be determined using any of a number of mathematical algorithms. A nonlimiting example of a mathematical algorithm used for comparison of two sequences is the algorithm of Karlin & Altschul, Proc. Natl. Acad. Sci. USA 1990; 87: 2264-2268, modified as in Karlin & Altschul, Proc. Natl. Acad. Sci. USA 1993;90: 5873-5877.

Another example of a mathematical algorithm used for comparison of sequences is the algorithm of Myers & Miller, CABIOS 1988;4: 11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a
5 PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Yet another useful algorithm for identifying regions of local sequence similarity and alignment is the FASTA algorithm as described in Pearson & Lipman, Proc. Natl. Acad. Sci. USA 1988; 85: 2444-2448.

Advantageous for use according to the present invention is the WU-BLAST
10 (Washington University BLAST) version 2.0 software. WU-BLAST version 2.0 executable programs for several UNIX platforms can be downloaded from <ftp://blast.wustl.edu/blast/executables>. This program is based on WU-BLAST version 1.4, which in turn is based on the public domain NCBI-BLAST version 1.4 (Altschul & Gish, 1996, Local alignment statistics, Doolittle ed., Methods in Enzymology 266: 460-480;
15 Altschul et al., Journal of Molecular Biology 1990; 215: 403-410; Gish & States, 1993; Nature Genetics 3: 266-272; Karlin & Altschul, 1993; Proc. Natl. Acad. Sci. USA 90: 5873-5877; all of which are incorporated by reference herein).

The various recombinant nucleotide sequences and transgenes of the invention are made using standard recombinant DNA and cloning techniques. Such techniques are well
20 known to those of skill in the art. See for example, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook et al. 1989).

The nucleotide sequences of the present invention may be inserted into "vectors." The term "vector" is widely used and understood by those of skill in the art, and as used herein the term "vector" is used consistent with its meaning to those of skill in the art. For
25 example, the term "vector" is commonly used by those skilled in the art to refer to a vehicle that allows or facilitates the transfer of nucleic acid molecules from one environment to another or that allows or facilitates the manipulation of a nucleic acid molecule.

Any vector that allows expression of the HIV-1 Clade A transgenes of the present invention may be used in accordance with the present invention. In certain embodiments, the
30 HIV-1 Clade A transgenes of the present invention may be used *in vitro* (such as using cell-free expression systems) and/or in cultured cells grown *in vitro* in order to produce the encoded HIV-1 antigens which may then be used for various applications such as in the production of proteinaceous vaccines. For such applications, any vector that allows expression of the HIV-1 Clade A transgenes *in vitro* and/or in cultured cells may be used.

For applications where it is desired that the transgenes be expressed *in vivo*, for example when the transgenes of the invention are used in DNA or DNA-containing vaccines, any vector that allows for the expression of the HIV-1 Clade A transgenes of the present invention and is safe for use *in vivo* may be used. In preferred embodiments the vectors used
5 are safe for use in humans, mammals and/or laboratory animals.

In order for the transgenes of the present invention to be expressed, the protein coding sequence should be "operably linked" to regulatory or nucleic acid control sequences that direct transcription and translation of the protein. As used herein, a coding sequence and a nucleic acid control sequence or promoter are said to be "operably linked" when they are
10 covalently linked in such a way as to place the expression or transcription and/or translation of the coding sequence under the influence or control of the nucleic acid control sequence. The "nucleic acid control sequence" can be any nucleic acid element, such as, but not limited to promoters, enhancers, IRES, introns, and other elements described herein that direct the expression of a nucleic acid sequence or coding sequence that is operably linked thereto. The
15 term "promoter" will be used herein to refer to a group of transcriptional control modules that are clustered around the initiation site for RNA polymerase II and that when operationally linked to the protein coding sequences of the invention lead to the expression of the encoded protein. The expression of the transgenes of the present invention can be under the control of a constitutive promoter or of an inducible promoter, which initiates transcription only when
20 exposed to some particular external stimulus, such as, without limitation, antibiotics such as tetracycline, hormones such as ecdysone, or heavy metals. The promoter can also be specific to a particular cell-type, tissue or organ. Many suitable promoters and enhancers are known in the art, and any such suitable promoter or enhancer may be used for expression of the transgenes of the invention. For example, suitable promoters and/or enhancers can be
25 selected from the Eukaryotic Promoter Database (EPDB).

The vectors used in accordance with the present invention should typically be chosen such that they contain a suitable gene regulatory region, such as a promoter or enhancer, such that the transgenes of the invention can be expressed.

For example, when the aim is to express the transgenes of the invention *in vitro*, or in
30 cultured cells, or in any prokaryotic or eukaryotic system for the purpose of producing the protein(s) encoded by that transgene, then any suitable vector can be used depending on the application. For example, plasmids, viral vectors, bacterial vectors, protozoal vectors, insect vectors, baculovirus expression vectors, yeast vectors, mammalian cell vectors, and the like, can be used. Suitable vectors can be selected by the skilled artisan taking into consideration

the characteristics of the vector and the requirements for expressing the transgenes under the identified circumstances.

When the aim is to express the transgenes of the invention *in vivo* in a subject, for example in order to generate an immune response against an HIV-1 antigen and/or protective immunity against HIV-1, expression vectors that are suitable for expression on that subject, and that are safe for use *in vivo*, should be chosen. For example, in some embodiments it may be desired to express the transgenes of the invention in a laboratory animal, such as for pre-clinical testing of the HIV-1 immunogenic compositions and vaccines of the invention. In other embodiments, it will be desirable to express the transgenes of the invention in human subjects, such as in clinical trials and for actual clinical use of the immunogenic compositions and vaccine of the invention. Any vectors that are suitable for such uses can be employed, and it is well within the capabilities of the skilled artisan to select a suitable vector. In some embodiments it may be preferred that the vectors used for these *in vivo* applications are attenuated to vector from amplifying in the subject. For example, if plasmid vectors are used, preferably they will lack an origin of replication that functions in the subject so as to enhance safety for *in vivo* use in the subject. If viral vectors are used, preferably they are attenuated or replication-defective in the subject, again, so as to enhance safety for *in vivo* use in the subject.

In preferred embodiments of the present invention viral vectors are used. Viral expression vectors are well known to those skilled in the art and include, for example, viruses such as adenoviruses, adeno-associated viruses (AAV), alphaviruses, herpesviruses, retroviruses and poxviruses, including avipox viruses, attenuated poxviruses, vaccinia viruses, and particularly, the modified vaccinia Ankara virus (MVA; ATCC Accession No. VR-1566). Such viruses, when used as expression vectors are innately non-pathogenic in the selected subjects such as humans or have been modified to render them non-pathogenic in the selected subjects. For example, replication-defective adenoviruses and alphaviruses are well known and can be used as gene delivery vectors.

In particularly preferred embodiments adenovirus vectors are used. Many adenovirus vectors are known in the art and any such suitable vector may be used. In preferred embodiments the adenovirus vector used is selected from the group consisting of the Ad5, Ad35, Ad11, C6, and C7 vectors.

The sequence of the Adenovirus 5 ("Ad5") genome has been published. (Chroboczek, J., Bieber, F., and Jacrot, B. (1992) The Sequence of the Genome of Adenovirus Type 5 and Its Comparison with the Genome of Adenovirus Type 2, *Virology* 186, 280-285; the contents

if which is hereby incorporated by reference). Ad35 vectors are described in U.S. Patent Nos. 6,974,695, 6,913,922, and 6,869,794. Ad11 vectors are described in U.S. Patent No. 6,913,922. C6 adenovirus vectors are described in U.S. Patent Nos. 6,780,407; 6,537,594; 6,309,647; 6,265,189; 6,156,567; 6,090,393; 5,942,235 and 5,833,975. C7 vectors are
5 described in U.S. Patent No. 6,277,558.

Adenovirus vectors that are E1-defective or deleted, E3-defective or deleted, and/or E4-defective or deleted may also be used. Certain adenoviruses having mutations in the E1 region have improved safety margin because E1-defective adenovirus mutants are replication-defective in non-permissive cells, or, at the very least, are highly attenuated.

10 Adenoviruses having mutations in the E3 region may have enhanced the immunogenicity by disrupting the mechanism whereby adenovirus down-regulates MHC class I molecules.

Adenoviruses having E4 mutations may have reduced immunogenicity of the adenovirus vector because of suppression of late gene expression. Such vectors may be particularly useful when repeated re-vaccination utilizing the same vector is desired. Adenovirus vectors
15 that are deleted or mutated in E1, E3, E4, E1 and E3, and E1 and E4 can be used in accordance with the present invention.

Furthermore, "gutless" adenovirus vectors, in which all viral genes are deleted, can also be used in accordance with the present invention. Such vectors require a helper virus for their replication and require a special human 293 cell line expressing both E1a and Cre, a
20 condition that does not exist in natural environment. Such "gutless" vectors are non-immunogenic and thus the vectors may be inoculated multiple times for re-vaccination. The "gutless" adenovirus vectors can be used for insertion of heterologous inserts/genes such as the transgenes of the present invention, and can even be used for co-delivery of a large number of heterologous inserts/genes.

25 The present invention also encompasses a design that puts the Env and GRIN on separate vectors to allow assessment of whether inclusion of Env is beneficial or detrimental in terms of cell-mediated immunity (CMI) and protective efficacy. The benefits and/or detriments of Env on CMI and protective efficacy remains an open question in the HIV vaccine field. Therefore, the present invention provides for the assessment of Env on CMI
30 and protective efficacy. It is within the purview of one of skill in the art to utilize the transgenes and vectors of the present invention to determine the effect of Env on CMI and protective efficacy.

The nucleotide sequences and vectors of the invention can be delivered to cells, for example if aim is to express and the HIV-1 antigens in cells in order to produce and isolate

the expressed proteins, such as from cells grown in culture. For expressing the transgenes in cells any suitable transfection, transformation, or gene delivery methods can be used. Such methods are well known by those skilled in the art, and one of skill in the art would readily be able to select a suitable method depending on the nature of the nucleotide sequences, vectors, and cell types used. For example, transfection, transformation, microinjection, infection, electroporation, lipofection, or liposome-mediated delivery could be used. Expression of the antigens can be carried out in any suitable type of host cells, such as bacterial cells, yeast, insect cells, and mammalian cells. The HIV-1 Clade A antigens of the invention can also be expressed using including *in vitro* transcription/translation systems. All of such methods are well known by those skilled in the art, and one of skill in the art would readily be able to select a suitable method depending on the nature of the nucleotide sequences, vectors, and cell types used.

Following expression, the antigens of the invention can be isolated and/or purified or concentrated using any suitable technique known in the art. For example, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, immuno-affinity chromatography, hydroxyapatite chromatography, lectin chromatography, molecular sieve chromatography, isoelectric focusing, gel electrophoresis, or any other suitable method or combination of methods can be used.

In preferred embodiments, the nucleotide sequences and/or antigens of the invention are administered *in vivo*, for example where the aim is to produce an immunogenic response in a subject. A "subject" in the context of the present invention may be any animal. For example, in some embodiments it may be desired to express the transgenes of the invention in a laboratory animal, such as for pre-clinical testing of the HIV-1 immunogenic compositions and vaccines of the invention. In other embodiments, it will be desirable to express the transgenes of the invention in human subjects, such as in clinical trials and for actual clinical use of the immunogenic compositions and vaccine of the invention. In preferred embodiments the subject is a human, for example a human that is infected with, or is at risk of infection with, HIV-1.

For such *in vivo* applications the nucleotide sequences and/or antigens of the invention are preferably administered as a component of an immunogenic composition comprising the nucleotide sequences and/or antigens of the invention in admixture with a pharmaceutically acceptable carrier. The immunogenic compositions of the invention are useful to stimulate an immune response against HIV-1 and may be used as one or more components of a

prophylactic or therapeutic vaccine against HIV-1 for the prevention, amelioration or treatment of AIDS. The nucleic acids and vectors of the invention are particularly useful for providing genetic vaccines, i.e. vaccines for delivering the nucleic acids encoding the HIV-1 Clade A antigens of the invention to a subject, such as a human, such that the HIV-1 Clade A antigens are then expressed in the subject to elicit an immune response.

The compositions of the invention may be injectable suspensions, solutions, sprays, lyophilized powders, syrups, elixirs and the like. Any suitable form of composition may be used. To prepare such a composition, a nucleic acid or vector of the invention, having the desired degree of purity, is mixed with one or more pharmaceutically acceptable carriers and/or excipients. The carriers and excipients must be "acceptable" in the sense of being compatible with the other ingredients of the composition. Acceptable carriers, excipients, or stabilizers are nontoxic to recipients at the dosages and concentrations employed, and include, but are not limited to, water, saline, phosphate buffered saline, dextrose, glycerol, ethanol, or combinations thereof, buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid and methionine; preservatives (such as octadecyldimethylbenzyl ammonium chloride; hexamethonium chloride; benzalkonium chloride, benzethonium chloride; phenol, butyl or benzyl alcohol; alkyl parabens such as methyl or propyl paraben; catechol; resorcinol; cyclohexanol; 3-pentanol; and m-cresol); low molecular weight (less than about 10 residues) polypeptide; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, histidine, arginine, or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrans; chelating agents such as EDTA; sugars such as sucrose, mannitol, trehalose or sorbitol; salt-forming counter-ions such as sodium; metal complexes (e.g., Zn-protein complexes); and/or non-ionic surfactants such as TWEENTM, PLURONICSTM or polyethylene glycol (PEG).

An immunogenic or immunological composition can also be formulated in the form of an oil-in-water emulsion. The oil-in-water emulsion can be based, for example, on light liquid paraffin oil (European Pharmacopea type); isoprenoid oil such as squalane, squalene, EICOSANETM or tetratetracontane; oil resulting from the oligomerization of alkene(s), e.g., isobutene or decene; esters of acids or of alcohols containing a linear alkyl group, such as plant oils, ethyl oleate, propylene glycol di(caprylate/caprates), glyceryl tri(caprylate/caprates) or propylene glycol dioleate; esters of branched fatty acids or alcohols, e.g., isostearic acid esters. The oil advantageously is used in combination with emulsifiers to form the emulsion.

The emulsifiers can be nonionic surfactants, such as esters of sorbitan, mannide (*e.g.*, anhydromannitol oleate), glycerol, polyglycerol, propylene glycol, and oleic, isostearic, ricinoleic, or hydroxystearic acid, which are optionally ethoxylated, and polyoxypropylene-polyoxyethylene copolymer blocks, such as the Pluronic® products, *e.g.*, L121. The adjuvant
 5 can be a mixture of emulsifier(s), micelle-forming agent, and oil such as that which is commercially available under the name Provax® (IDEC Pharmaceuticals, San Diego, CA).

The immunogenic compositions of the invention can contain additional substances, such as wetting or emulsifying agents, buffering agents, or adjuvants to enhance the effectiveness of the vaccines (Remington's Pharmaceutical Sciences, 18th edition, Mack
 10 Publishing Company, (ed.) 1980).

Adjuvants may also be included. Adjuvants include, but are not limited to, mineral salts (*e.g.*, $\text{AlK}(\text{SO}_4)_2$, $\text{AlNa}(\text{SO}_4)_2$, $\text{AlNH}(\text{SO}_4)_2$, silica, alum, $\text{Al}(\text{OH})_3$, $\text{Ca}_3(\text{PO}_4)_2$, kaolin, or carbon), polynucleotides with or without immune stimulating complexes (ISCOMs) (*e.g.*, CpG oligonucleotides, such as those described in Chuang, T.H. et al, (2002) J. Leuk. Biol.
 15 71(3): 538-44; Ahmad-Nejad, P. et al (2002) Eur. J. Immunol. 32(7): 1958-68; poly IC or poly AU acids, polyarginine with or without CpG (also known in the art as IC31; see Schellack, C. et al (2003) Proceedings of the 34th Annual Meeting of the German Society of Immunology; Lingnau, K. et al (2002) Vaccine 20(29-30): 3498-508), JuvaVax™ (U.S. Patent No. 6,693,086), certain natural substances (*e.g.*, wax D from *Mycobacterium*
 20 *tuberculosis*, substances found in *Cornyebacterium parvum*, *Bordetella pertussis*, or members of the genus *Brucella*), flagellin (Toll-like receptor 5 ligand; see McSorley, S.J. et al (2002) J. Immunol. 169(7): 3914-9), saponins such as QS21, QS17, and QS7 (U.S. Patent Nos. 5,057,540; 5,650,398; 6,524,584; 6,645,495), monophosphoryl lipid A, in particular, 3-de-O-acylated monophosphoryl lipid A (3D-MPL), imiquimod (also known in the art as IQM and
 25 commercially available as Aldara®; U.S. Patent Nos. 4,689,338; 5,238,944; Zuber, A.K. et al (2004) 22(13-14): 1791-8), and the CCR5 inhibitor CMPD167 (see Veazey, R.S. et al (2003) J. Exp. Med. 198: 1551-1562).

Aluminum hydroxide or phosphate (alum) are commonly used at 0.05 to 0.1% solution in phosphate buffered saline. Other adjuvants that can be used, especially with DNA
 30 vaccines, are cholera toxin, especially CTA1-DD/ISCOMs (see Mowat, A.M. et al (2001) J. Immunol. 167(6): 3398-405), polyphosphazenes (Allcock, H.R. (1998) App. Organometallic Chem. 12(10-11): 659-666; Payne, L.G. et al (1995) Pharm. Biotechnol. 6: 473-93), cytokines such as, but not limited to, IL-2, IL-4, GM-CSF, IL-12, IL-15 IGF-1, IFN- α , IFN-

β, and IFN-γ (Boyer et al., (2002) J. Liposome Res. 121:137-142; WO01/095919), immunoregulatory proteins such as CD40L (ADX40; see, for example, WO03/063899), and the CD1a ligand of natural killer cells (also known as CRONY or α-galactosyl ceramide; see Green, T.D. et al, (2003) J. Virol. 77(3): 2046-2055), immunostimulatory fusion proteins
5 such as IL-2 fused to the Fc fragment of immunoglobulins (Barouch et al., Science 290:486-492, 2000) and co-stimulatory molecules B7.1 and B7.2 (Boyer), all of which can be administered either as proteins or in the form of DNA, on the same expression vectors as those encoding the antigens of the invention or on separate expression vectors.

The immunogenic compositions can be designed to introduce the HIV-1 Clade A
10 antigens, nucleic acids or expression vectors to a desired site of action and release it at an appropriate and controllable rate. Methods of preparing controlled-release formulations are known in the art. For example, controlled release preparations can be produced by the use of polymers to complex or absorb the immunogen and/or immunogenic composition. A controlled-release formulations can be prepared using appropriate macromolecules (for
15 example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine sulfate) known to provide the desired controlled release characteristics or release profile. Another possible method to control the duration of action by a controlled-release preparation is to incorporate the active ingredients into particles of a polymeric material such as, for example, polyesters, polyamino acids,
20 hydrogels, polylactic acid, polyglycolic acid, copolymers of these acids, or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these active ingredients into polymeric particles, it is possible to entrap these materials into microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsule and poly-(methylmethacrylate)
25 microcapsule, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles and nanocapsules) or in macroemulsions. Such techniques are disclosed in New Trends and Developments in Vaccines, Voller et al. (eds.), University Park Press, Baltimore, Md., 1978 and Remington's Pharmaceutical Sciences, 16th edition.

30 Suitable dosages of the HIV-1 Clade A antigens, nucleic acids and expression vectors of the invention (collectively, the immunogens) in the immunogenic composition of the invention can be readily determined by those of skill in the art. For example, the dosage of the immunogens can vary depending on the route of administration and the size of the

subject. Suitable doses can be determined by those of skill in the art, for example by measuring the immune response of a subject, such as a laboratory animal, using conventional immunological techniques, and adjusting the dosages as appropriate. Such techniques for measuring the immune response of the subject include but are not limited to, chromium
5 release assays, tetramer binding assays, IFN- γ ELISPOT assays, IL-2 ELISPOT assays, intracellular cytokine assays, and other immunological detection assays, e.g., as detailed in the text "Antibodies: A Laboratory Manual" by Ed Harlow and David Lane.

When provided prophylactically, the immunogenic compositions of the invention are ideally administered to a subject in advance of HIV infection, or evidence of HIV infection,
10 or in advance of any symptom due to AIDS, especially in high-risk subjects. The prophylactic administration of the immunogenic compositions can serve to provide protective immunity of a subject against HIV-1 infection or to prevent or attenuate the progression of AIDS in a subject already infected with HIV-1. When provided therapeutically, the immunogenic compositions can serve to ameliorate and treat AIDS symptoms and are
15 advantageously used as soon after infection as possible, preferably before appearance of any symptoms of AIDS but may also be used at (or after) the onset of the disease symptoms.

The immunogenic compositions can be administered using any suitable delivery method including, but not limited to, intramuscular, intravenous, intradermal, mucosal, and topical delivery. Such techniques are well known to those of skill in the art. More specific
20 examples of delivery methods are intramuscular injection, intradermal injection, and subcutaneous injection. However, delivery need not be limited to injection methods. Further, delivery of DNA to animal tissue has been achieved by cationic liposomes (Watanabe et al., (1994) Mol. Reprod. Dev. 38:268-274; and WO 96/20013), direct injection of naked DNA into animal muscle tissue (Robinson et al., (1993) Vaccine 11:957-960;
25 Hoffman et al., (1994) Vaccine 12: 1529-1533; Xiang et al., (1994) Virology 199: 132-140; Webster et al., (1994) Vaccine 12: 1495-1498; Davis et al., (1994) Vaccine 12: 1503-1509; and Davis et al., (1993) Hum. Mol. Gen. 2: 1847-1851), or intradermal injection of DNA using "gene gun" technology (Johnston et al., (1994) Meth. Cell Biol. 43:353-365). Alternatively, delivery routes can be oral, intranasal or by any other suitable route. Delivery
30 also be accomplished via a mucosal surface such as the anal, vaginal or oral mucosa.

Immunization schedules (or regimens) are well known for animals (including humans) and can be readily determined for the particular subject and immunogenic composition. Hence, the immunogens can be administered one or more times to the subject.

Preferably, there is a set time interval between separate administrations of the immunogenic composition. While this interval varies for every subject, typically it ranges from 10 days to several weeks, and is often 2, 4, 6 or 8 weeks. For humans, the interval is typically from 2 to 6 weeks. The immunization regimes typically have from 1 to 6 administrations of the immunogenic composition, but may have as few as one or two or four. The methods of inducing an immune response can also include administration of an adjuvant with the immunogens. In some instances, annual, biannual or other long interval (5-10 years) booster immunization can supplement the initial immunization protocol.

The present methods also include a variety of prime-boost regimens, especially DNA prime-Adenovirus boost regimens. In these methods, one or more priming immunizations are followed by one or more boosting immunizations. The actual immunogenic composition can be the same or different for each immunization and the type of immunogenic composition (e.g., containing protein or expression vector), the route, and formulation of the immunogens can also be varied. For example, if an expression vector is used for the priming and boosting steps, it can either be of the same or different type (e.g., DNA or bacterial or viral expression vector). One useful prime-boost regimen provides for two priming immunizations, four weeks apart, followed by two boosting immunizations at 4 and 8 weeks after the last priming immunization. It should also be readily apparent to one of skill in the art that there are several permutations and combinations that are encompassed using the DNA, bacterial and viral expression vectors of the invention to provide priming and boosting regimens.

A specific embodiment of the invention provides methods of inducing an immune response against HIV in a subject by administering an immunogenic composition of the invention, preferably comprising an adenovirus vector containing DNA encoding one or more of the HIV-1 Clade A antigens of the invention, (preferably GRIN, GRN, or Env, or a combination thereof), one or more times to a subject wherein the HIV-1 Clade A antigen(s) are expressed at a level sufficient to induce a specific immune response in the subject. Such immunizations can be repeated multiple times at time intervals of at least 2, 4 or 6 weeks (or more) in accordance with a desired immunization regime.

The immunogenic compositions of the invention can be administered alone, or can be co-administered, or sequentially administered, with other HIV immunogens and/or HIV immunogenic compositions, e.g., with "other" immunological, antigenic or vaccine or therapeutic compositions thereby providing multivalent or "cocktail" or combination compositions of the invention and methods of employing them. Again, the ingredients and manner (sequential or co-administration) of administration, as well as dosages can be

determined taking into consideration such factors as the age, sex, weight, species and condition of the particular subject, and the route of administration.

When used in combination, the other HIV immunogens can be administered at the same time or at different times as part of an overall immunization regime, e.g., as part of a prime-boost regimen or other immunization protocol. Many other HIV immunogens are known in the art, one such preferred immunogen is HIVA (described in WO 01/47955), which can be administered as a protein, on a plasmid (e.g., pTHr.HIVA) or in a viral vector (e.g., MVA.HIVA). Another such HIV immunogen is RENTA (described in PCT/US2004/037699), which can also be administered as a protein, on a plasmid (e.g., pTHr.RENTA) or in a viral vector (e.g., MVA.RENTA).

For example, one method of inducing an immune response against HIV in a human subject comprises administering at least one priming dose of an HIV immunogen and at least one boosting dose of an HIV immunogen, wherein the immunogen in each dose can be the same or different, provided that at least one of the immunogens is an HIV-1 Clade A antigen of the invention, a nucleic acid encoding an HIV-1 Clade A antigen of the invention or an expression vector, preferably an adenovirus vector, encoding an HIV-1 Clade A antigen of the invention, and wherein the immunogens are administered in an amount or expressed at a level sufficient to induce an HIV-specific immune response in the subject. The HIV-specific immune response can include an HIV-specific T-cell immune response or an HIV-specific B-cell immune response. Such immunizations can be done at intervals, preferably of at least 2-6 or more weeks.

It is to be understood and expected that variations in the principles of invention as described above, and as described in the below example, may be made by one skilled in the art and it is intended that such modifications, changes, and substitutions are to be included within the scope of the present invention.

The following non-limiting examples are given for the purpose of illustrating various embodiments of the invention.

EXAMPLES**Example 1: Consensus sequence for Gag of HIV Clade A**Table 1

	Distance from		
	consensus	Country	Year
5	A_consensu	0	
	A_97TZ02_1	0.04081	TZ
	A_TZA173_1	0.0425	TZ
	A_KNH1144	0.04259	KE
	A_SE7535UG	0.04303	UG
	A_KNH1211	0.04463	KE
	A_KSM4024	0.04684	KE
10	A_KNH1207	0.04701	KE
	A_SE6594UG	0.04709	UG
	A_92UG037	0.05079	UG
	A_TZA195_1	0.05127	TZ
	A_MSA4079	0.05279	KE
	A_TZA341_1	0.05523	TZ
	A_MSA4072	0.05583	KE
	A_MSA4076	0.056	KE
15	A_KNH1199	0.05687	KE
	A_MSA4070	0.05947	KE
	A_98UG5713	0.06038	UG
	A_KEQ23-17	0.06072	KE
	A_KNH1209	0.06101	KE
	A_NKU3005	0.06108	KE
	A_SE7253SO	0.06113	SO
	A_98UG5713	0.06119	UG
20	A_SE8538TZ	0.06137	TZ
	A_KNH1088	0.06262	KE
	A_KER2008	0.065	KE
	A_99UGA070	0.06531	UG
	A_KER2012-	0.06654	KE
	A_KER2009	0.0674	KE
	A_99UGG033	0.06871	UG
	A_KSM4030-	0.07026	KE
25	A_KSM4021-	0.07145	KE
	A_98UG5713	0.07189	UG
	A_SE8891UG	0.07197	UG
	A_SE8131UG	0.07462	UG
	A_97TZ03_1	0.07653	TZ
	A_KNH1135	0.07687	KE
	A_98UG5714	0.0781	UG
	A_UGU455_1	0.08349	UG
30	A_MSA4069	0.08867	KE

The amino acid sequences of the Gag proteins of 39 non-recombinant HIV Clade A strains were analyzed. Table 1 lists the 39 strains used, and refers to each by its Genbank accession number. Table 1 also identifies the country and year of isolation of each of these 39 strains. 20 of the strains were from Kenya, 12 from Uganda, 6 from Tanzania, and 1 from

Somalia. 20 of the strains were isolated between 2000 and 2002, 10 were isolated between 1997 and 1999, 6 were isolated between 1994 and 1996 and 3 were isolated before 1993.

The Gag protein sequences were aligned with spaces added to preserve alignment in regions with insertions or deletions. A 50% consensus sequence was derived. The consensus amino acid sequence is shown FIG. 1. In FIG. 1 the spaces that were added to preserve alignment in regions with insertions or deletions are represented by dashes, and the positions for which a 50% consensus was not attained are represented by an "X".

For each of the 39 sequences used to generate the consensus sequence, the "distance" of that sequence from the consensus sequence was calculated using the Dayhoff PAM250 substitution matrix, which weights substitutions according to the degree of biochemical similarity. As shown in Table 1, the distance of each strain's sequence from the consensus sequence ranged from 4 to 9%.

FIG. 2 illustrates the distance of each strain's amino acid sequence from the consensus amino acid sequence in graphical form, and identifies the four strains having sequences that are closest to the consensus sequences. These four strains are strain 97TZ02 which from a low-risk individual in the Mbeya region of southwest Tanzania in 1997 which has Genbank accession number AF361872, strain TZA173 collected from an anonymous blood donor in the Mbeya region of southwest Tanzania in 2001 which has Genbank accession number AY253305, strain KNH1144 collected from an anonymous blood donor in southern Kenya in 2000 which has Genbank accession number AF4587006, and strain SE7535 collected in 1994 in Sweden from an individual thought to have been infected in Uganda which has Genbank accession number AF069671.

Example 2: Consensus sequence for Pol of HIV Clade A

The amino acid sequences of the Pol proteins of 36 non-recombinant HIV Clade A strains were analyzed. Table 2 lists the 36 strains used, and refers to each by its Genbank accession number. Table 2 also identifies the country and year of isolation of each of these 36 strains. 20 of the strains were from Kenya, 9 from Uganda, 6 from Tanzania, and 1 from Somalia. 19 of the strains were isolated between 2000 and 2002, 10 were isolated between 1997 and 1999, 4 were isolated between 1994 and 1996 and 3 were isolated before 1993.

The Pol protein sequences were aligned. There were no insertions or deletions. A 50% consensus sequence was derived. The consensus amino acid sequence is shown FIG. 3. In FIG. 3 the positions for which a 50% consensus was not attained are represented by an "X". There were 4 such positions out of 947 amino acid residues. For each of the 36 sequences used to generate the consensus sequence, the "distance" of that sequence from the consensus

sequence was calculated using the Dayhoff PAM250 substitution matrix, which weights substitutions according to the degree of biochemical similarity. As shown in Table 2, the distance of each strain's sequence from the consensus sequence ranged from 1.5 to 4.8%.

Table 2

	Distance from consensus	Country	Year
A_pol.cons	0		
A_MSA4070	0.01479	KE	2000
A_SE7253SO	0.01582	SO	1994
A_SE8538TZ	0.01898	TZ	1995
A_KER2012-	0.02329	KE	2000
A_97TZ02_3	0.0235	TZ	1997
A_KEQ23-17	0.02445	KE	1994
A_KNH1211	0.02449	KE	2000
A_TZA341_3	0.0246	TZ	2001
A_KSM4024	0.02528	KE	2000
A_97TZ03_3	0.02544	TZ	1997
A_KNH1088	0.02544	KE	1999
A_MSA4076	0.02564	KE	2000
A_KNH1207	0.0265	KE	2000
A_NKU3005	0.02661	KE	2000
A_TZA173_3	0.02756	TZ	2001
A_MSA4079	0.02762	KE	2000
A_KER2009	0.02765	KE	2000
A_TZA195_3	0.02881	TZ	2001
A_KSM4021-	0.02881	KE	1999
A_SE7535UG	0.02883	UG	1994
A_MSA4069	0.02886	KE	2000
A_SE6594UG	0.02889	UG	1993
A_98UG5713	0.02975	UG	1998
A_KNH1135	0.0299	KE	1999
A_92UG037	0.02993	UG	1992
A_KNH1209	0.03202	KE	2000
A_99UGG033	0.03291	UG	1999
A_KER2008	0.03294	KE	2000
A_KSM4030-	0.0343	KE	2000
A_KNH1199	0.03439	KE	2000
A_99UGA070	0.03537	UG	1999
A_MSA4072	0.03625	KE	2000
A_KNH1144	0.03863	KE	2000
A_98UG5713	0.04178	UG	1998
A_UGU455_3	0.04294	UG	1985
A_98UG5713	0.04808	UG	1998

FIG. 4 illustrates the distance of each strain's amino acid sequence from the consensus amino acid sequence in graphical form, and identifies the three strains having sequences that are closest to the consensus sequences. These three strains are strain MSA4070 from an anonymous blood donor in Southern Kenya in 2000, strain SE7235SO which was collected in 1994 from an individual in Sweden thought to have been infected in

Somalia, and strain SE8538 which was collected in 1995 from an individual in Sweden thought to have been infected in Tanzania.

Example 3: Consensus sequence for Env of HIV Clade A

Table 3

	Dist from		
	A.cons	country	year
A.cons	0		
A_KEQ23-17	0.06307	KE	1994
A_TZA341_1	0.06413	TZ	2001
A_KNH1088	0.06524	KE	1999
A_KNH1209	0.0699	KE	2000
A_KNH1144	0.07088	KE	2000
A_99UGA070	0.07365	UG	1999
A_MSA4072	0.07516	KE	2000
A_KSM4021-	0.0778	KE	1999
A_97TZ02_1	0.07825	TZ	1997
A_KNH1199	0.07883	KE	2000
A_MSA4079	0.07944	KE	2000
A_SE7535UG	0.08375	UG	1994
A_SE8538TZ	0.08432	TZ	1995
A_98UG5713	0.08462	UG	1998
A_97TZ03_1	0.08541	TZ	1997
A_MSA4070	0.0874	KE	2000
A_NKU3005	0.0884	KE	2000
A_TZA173_1	0.09046	TZ	2001
A_KNH1207	0.09106	KE	2000
A_TZA195_1	0.09389	TZ	2001
A_MSA4076	0.09517	KE	2000
A_92UG037	0.098	UG	1992
A_98UG5714	0.09816	UG	1998
A_SE7253SO	0.09886	SO	1994
A_KER2012-	0.09984	KE	2000
A_98UG5713	0.10139	UG	1998
A_SE6594UG	0.10195	UG	1993
A_SE8891UG	0.10225	UG	1995
A_UGU455_1	0.10314	UG	1985
A_KER2009	0.10338	KE	2000
A_KNH1211	0.11319	KE	2000
A_SE8131UG	0.11321	UG	1995
A_MSA4069	0.11507	KE	2000
A_99UGG033	0.11653	UG	1999
A_KNH1135	0.11713	KE	1999
A_KER2008	0.12689	KE	2000

The amino acid sequences of the Env proteins of 36 non-recombinant HIV Clade A strains were analyzed. Table 3 lists the 36 strains used, and refers to each by its Genbank accession number. Table 3 also identifies the country and year of isolation of each of these 36 strains. 18 of the strains were from Kenya, 11 from Uganda, 6 from Tanzania, and 1 from

Somalia. 17 of the strains were isolated between 2000 and 2002, 10 were isolated between 1997 and 1999, 6 were isolated between 1994 and 1996 and 3 were isolated before 1993.

The Env protein sequences were aligned with spaces added to preserve alignment in regions with insertions or deletions. There were many regions with extensive heterogeneity in the length of insertions/deletions. A 50% consensus sequence was derived. The consensus amino acid sequence is shown FIG. 5. In FIG. 5 the spaces that were added to preserve alignment in regions with insertions or deletions are represented by dashes, and the positions for which a 50% consensus was not attained are represented by an "X". There were many amino acid positions for which a 50% consensus was not attained.

For each of the 36 sequences used to generate the consensus sequence, the "distance" of that sequence from the consensus sequence was calculated using the Dayhoff PAM250 substitution matrix, which weights substitutions according to the degree of biochemical similarity. As shown in Table 3, the distance of each strain's sequence from the consensus sequence ranged from 6.3 to 12.7%.

FIG. 6 illustrates the distance of each strain's amino acid sequence from the consensus amino acid sequence in graphical form, and identifies the three strains having sequences that are closest to the consensus sequences. These three strains were KEQ23 from a CSW in Kenya in 1994 (what is a CSW), TZA341 which was from an anonymous blood donor in Tanzania in 2002, and KNH1088 which was from an anonymous blood donor in Kenya in 1999.

Example 4: Consensus sequence for Nef of HIV Clade A

The amino acid sequences of the Nef proteins of 38 non-recombinant HIV Clade A strains were analyzed. Table 4 lists the 38 strains used, and refers to each by its Genbank accession number. The country and year of isolation of each of these 38 strains are described in Tables 1-3 in the previous Examples. More than half of the strains were from Kenya, with a substantial portion coming from Uganda, and a few strains coming from Tanzania. About half of the strains were isolated between 2000 and 2002.

Table 4

	A.cons
A_MSA4070	0.0318
A_KNH1211	0.04807
A_97TZ03_1	0.0535
A_99UGA070	0.05354
A_SE8891UG	0.05383
A_KEQ23-17	0.06476
A_98UG5713	0.07043

	A.cons
A NKU3005	0.0709
A SE7535UG	0.07117
A 98UG5714	0.07613
A SE6594UG	0.07634
A TZA341 1	0.0805
A MSA4069	0.08097
A KNH1199	0.08213
A 97TZ02 1	0.08276
A KSM4030-	0.08704
A KSM4021-	0.08795
A MSA4076	0.08873
A KNH1209	0.0899
A KER2012-	0.09224
A KNH1144	0.09577
A KER2008	0.09703
A MSA4072	0.09892
A 98UG5713	0.09892
A 99UGG033	0.09967
A KNH1088	0.10303
A 92UG037	0.10654
A SE8538TZ	0.10996
A KER2009	0.1102
A MSA4079	0.11083
A KSM4024	0.11126
A SE8131UG	0.11326
A SE7253SO	0.11453
A KNH1207	0.11549
A TZA173 1	0.13766
A 98UG5713	0.1399
A UGU455 1	0.15688
A KNH1135	0.16076
A.cons	0

The Nef protein sequences were aligned with spaces added to preserve alignment in regions with insertions or deletions. A 50% consensus sequence was derived. The consensus amino acid sequence is shown FIG. 7. In FIG. 7 the spaces that were added to preserve alignment in regions with insertions or deletions are represented by dashes, and the positions for which a 50% consensus was not attained are represented by an "X". There were six amino acid positions for which a 50% consensus was not attained.

For each of the 38 sequences used to generate the consensus sequence, the "distance" of that sequence from the consensus sequence was calculated using the Dayhoff PAM250 substitution matrix, which weights substitutions according to the degree of biochemical similarity. As shown in Table 4, the distance of each strain's sequence from the consensus sequence ranged from 3.2 to 16.1% with a mean distance of 9.3%.

FIG. 8 illustrates the distance of each strain's amino acid sequence from the consensus amino acid sequence in graphical form, and identifies the five strains having sequences that are closest to the consensus sequences. These five strains were MSA4070 and KNH1211, both of which were from anonymous donors in southern Kenya and were collected in 2000, 97TZ03 from a low-risk individual in the Mbeya region of southwest Tanzania which was collected in 1997, and UGA070 and SE8891 both of which were from individuals in Uganda and were collected in 1999 and 1995, respectively.

Example 5. Strains of HIV Clade A strains that are closest to the HIV clade A consensus sequences.

As described in Examples 1 to 4 above, and as summarized in Table 5, the strains of HIV Clade A having Gag, Pol, Env and Nef sequences that were most similar to the consensus sequences of each of these proteins were identified. In addition, the strains that were overall closest to the consensus sequence were identified by ranking each of the strains according to its closeness to the consensus sequence of a particular protein wherein the strain ranked number 1 was that whose sequence for that protein was closest to that of the consensus sequence, and then summing the rankings for each strain across all four of the proteins (i.e. Gag, Pol, Env, and Nef). The six strains that were overall closest to the consensus sequence across all four of the proteins studied are listed below in Table 6. It can be seen that strain 97TZ02 has a sequence which is overall closest to the consensus sequences of each of the Gag, Pol, Env and Nef genes.

Table 5

Gag	Pol	Env	Nef
97TZ02	MSA4070	KEQ23	MSA4070
TZA173	SE7245SO	TZA341	KNH1211
KNH1144	SE8538	KNH1088	97TZ03
SE7535UG			99UGA070
			SE8891UG

Table 6

	gag	pol	env	nef	sum
A_97TZ02_1	1	5	9	15	30
A_KEQ23-17	18	6	1	6	31
A_MSA4070	16	1	16	1	34
A_TZA341_1	12	8	2	12	34
A_SE7535UG	4	20	12	9	45
A_KNH1211_	5	7	31	2	45

Example 6. Construction of GRIN, GRN, and Env Transgenes.

Transgene constructs were made using HIV Clade A protein sequences derived from the most recently identified circulating HIV-1 field isolates that were the closest match to the HIV Clade A consensus sequence for each such protein. This strategy was developed in order to maximize the biological relevance of the HIV clade A sequences used. It should be understood that other sequences, i.e. sequences other than the specific sequences described in this example, can also be used in accordance with the invention. If other sequences are used it is preferred that the sequences are selected such that they are derived from recent field isolates and have sequences that are close to the HIV Clade A consensus sequences described herein, or to HIV Clade A consensus sequences that may be generated in the future.

Constructs referred to as GRIN and GRN were made. The GRIN construct contained HIV Clade A sequences encoding the Gag, Pol (RT and Integrase) and Nef proteins. The GRN transgene contained sequences encoding the Gag, RT and Nef proteins. The GRIN and GRN constructs are represented schematically in FIG. 9. The GRIN and GRN transgenes were made using the Gag protein sequence from strain TZA173 having Genbank accession number AY253305, the Pol (comprising both RT and Int sequences) sequence from strain MSA4070 having Genbank accession number AF457081, and the Nef sequence from strain MSA4070 having Genbank accession number AF457081. These sequences were selected because they were from the most recently identified circulating HIV-1 field isolates that had the closest match to the consensus sequence for each of Gag, Pol, and Nef, respectively. These sequences are illustrated in FIGS. 10, 11, and 12, respectively.

An Env construct was also made containing the Env coding sequence from the most recently identified circulating HIV-1 field isolate that had the closest match to the consensus Env sequence, i.e. the Env protein sequence from strain TZA341 having Genbank accession number AY253314. This sequence is illustrated in FIG. 13.

All sequences were then codon optimized for human expression by GeneArt (Germany). The optimized gene sequences allow high level and stable protein expression in humans or other mammalian cells. Further details of the codon optimization process are provided in Example 8. The transgenes were also engineered to incorporate specific mutations or arranged in a specific order to abrogate the normal function of the gene products *in vivo*. The details of these mutations, and the biological effects of each, are described in Example 7 below.

For the GRIN and GRN transgenes, the coding sequences for each of the Gag, Pol (RT & Int) or RT, and Nef proteins were joined in-frame such that each of the transgene

constructs (i.e. either GRIN or GRN) encoded a single fusion protein. Blast searches were performed to ensure that no neoepitopes were formed at the junctions. Although not used in the present example, it should be noted that it is also possible to insert spacer sequences between the sequences coding for the individual components of the final fusion protein to allow optimal protein domain folding, for example a spacer region may be added between Gag and Pol to allow the protein domains to fold in a more native conformation. Also, unique restriction sites were added at the 5' and 3' ends of each sequence in order to facilitate the joining together of each sequence (for example, the joining of the 5' end of Nef to the 3' end of Pol, etc.).

For use *in vivo* the GRIN, GRN and Env transgenes were inserted into either the Ad5, Ad35, Ad11, C6, or C7 adenovirus vectors. In order to facilitate cloning into these vectors unique restriction sites were added at the 5' and 3' ends of the GRIN, GRN, or Env constructs. FIGS. 14A-14C provides the sequence of GRIN as inserted into the Ad35 vector, and shows the restriction sites used to clone the GRIN sequence into the Ad35 vector (underlined and in bold typeface). The sequence shown also includes the CMV promoter sequence upstream of the GRIN sequence. FIGS. 15A-15B provides the sequence of Env as inserted into the Ad35 vector, and shows the restriction sites used to clone the Env sequence into the Ad35 vector (underlined and in bold typeface). The sequence shown also includes the CMV promoter sequence upstream of the Env sequence.

Standard recombinant DNA and cloning techniques were used to generate all of the above constructs. Such techniques are well known to those of skill in the art. See for example, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook et al. 1989).

Example 7: Mutations to abrogate normal *in vivo* function of HIV Clade A proteins

Table 7 summarizes the mutations engineered into the GRIN and GRN sequences to abrogate the *in vivo* function of their gene products. These mutations were made using standard recombinant DNA techniques. Such techniques are well known to those of skill in the art.

Table 7

Design	Gene	Mutation/Rationale
Mutation	gag	Gly2 → Ala : Removes myristylation site preventing VLP formation.
Mutation	gag	<i>To avoid slippage at the natural frame shift sequence, the DNA sequence was mutated in a manner that leaves the conserved amino acid sequence (NFLG) intact and allows only the full-length GagPol protein product to be translated.</i>
Mutation	RT	Asp185 → Ala & Asp186 → Ala: Inactivates active enzyme residues.
Mutations	Integrase (IN)	Asp 64 → Ala, Asp116 → Ala & Glu 152 → Ala: Inactivates active enzyme residues.

Design	Gene	Mutation/Rationale
No change	Nef	Fusion of nef N-terminus to IN C-terminus prevents myristylation and membrane targeting abrogating nef function.

Gag protein is expressed as a 55-kDa polyprotein precursor (Pr55^{gag}), and is cleaved by the HIV-1 viral protease. Four major viral proteins result from the cleavage; Matrix (MA), Capsid (CA), Nucleocapsid (NC), and p6; as well as two spacer polypeptides p2 and p1, which represent sequences between CA and NC and between NC and p6, respectively.

MA plays a key role in several steps in virus replication, including the critical mediation of viral particle assembly and budding from the cell plasma membrane through the formation of virus-like particles (VLPs) (See Gheysen, D., E. Jacobs, F. de Foresta, D. Thiriart, M. Francotte, D. Thines, and M. De Wilde. (1989). Assembly and release of HIV-1 precursor pr55gag virus-like particles from recombinant baculovirus-infected cells. Cell 59:103-112).

Both Pr55^{gag} and the MA (p17) are myristylated, i.e. amide bond formation to myristic acid. See Veronese di Marzo, F., Copeland, T. D., Oroszlan, S., Gallo, R. C. & Sarngadharan, M. G. (1988). J. Virol. 62, 795-801. See also section on Nef within Example 7 for a full description of myristylation process. Different HIV-1 isolates demonstrate that the myristyl-acceptor is the N-terminal glycine residue (Gly2). See Bryant & Ratner. (1990). Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. Proc. Natl. Acad. Sci. USA; 87: 523-527.

Bryant and Ratner (1990) demonstrated that substitution of Gly2 with Ala eliminated virus replication of an HIV-1 clone. The Pr55^{gag}, deficient of the myristyl-acceptor glycine, accumulated in infected Hela cells and was not processed into mature virion capsid. It was concluded that myristylation of the Gly2 is required for stable plasma membrane association and subsequent assembly of virions. Other groups have similarly demonstrated the importance of the myristylation of Gly2 in the MA. See Göttinger H G, Sodroski J G, Haseltine W A. (1989). Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. Proc Natl Acad Sci USA; 86:5781-5785, and Paul Spearman, Jaang-Jiun Wang, Nancy Vander Heyden and Lee Ratner. (1994). Identification of Human Immunodeficiency Virus Type 1 Gag Protein Domains Essential to Membrane Binding and Particle Assembly. J. Virol; 68 (5): 3232-3242.

If the myristyl-acceptor N-terminal glycine (Gly2) in MA is mutated, membrane binding is abrogated and particle assembly is prevented. Thus, clade A Gag is engineered to change Gly2 → Ala. This results in the loss of the Gag biological function.

Reverse transcriptase (RT) is a viral enzyme essential for replication. RT converts incoming viral RNA⁺ into dsDNA, catalyzed by the RNA- and DNA-dependent polymerase and RNase H activities of the enzyme. RT is a heterodimer composed of p66 and p51 subunit proteins. See Alfredo Jacobo-Molina et al. (1993). Crystal structure of human

5 immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. USA*; 90: 6320-6324. p66 has two domains, the polymerase and RNase H. p51 has the same polymerase domain.

The catalytically essential Asp-110, Asp-185, and Asp-186 residues are located in the highly conserved DNA polymerase active site. These three residues, termed the “the
10 catalytic triad” are thought to bind the divalent cations necessary for catalysis function. See Alfredo Jacobo-Molina et al. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. USA*; 90: 6320-6324.

The mutation of the aspartic acids at residues 185 and 186 into either asparagine or
15 glutamate have been demonstrated to result in mutant proteins which were catalytically inactive. See Lowe DM, Parmar V, Kemp SD, Larder BA. (1991). Mutational analysis of two conserved sequence motifs in HIV-1 reverse transcriptase. *FEBS Lett.*; 6;282(2):231-4.

Mutation of Asp185 → Ala & Asp186 → Ala in the Clade A RT will inactivate the RT polymerase enzyme by disrupting the “catalytic triad”. This will eliminate the biological
20 function of the Clade A RT.

Proviral cDNA generated by RT is integrated into the host cell genome through the action of the viral Integrase (Int) enzyme. Int contains a DNA recombinase domain that catalyzes two distinct endonucleolytic reactions. The first reaction, 3' processing, removes dinucleotides from each end of the cDNA producing two-nucleotide 5' extensions at both
25 ends. In the second reaction, Int non-specifically cleaves the host cell DNA and joins the free 3' groups of the cDNA termini to the 5' groups of the cleaved host cell DNA. Cellular enzymes repair gaps resulting in a fully integrated viral genome into the host cell DNA. See Coffin JM. *Retroviridae and their Replication*. Chapter 27. p645-708 & Wong-Staal F. *Human Immunodeficiency Viruses and Their Replication*. Chapter 28. p709-723. In Fields, BN. & Knipe DM. 2nd Edition *Fundamental Virology*. Raven Press. See also Engelman A,
30 Mizuuchi K, Craigie R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell*; 67(6):1211–1221. The catalytic domain, residues 50 to 212, contain a triad of residues Asp-64, Asp-116, and Glu-152 (termed the D,D-35-E motif) that

compromises the enzyme active site. See Esposito, D., and R. Craigie. (1999). HIV integrase structure and function. *Adv. Virus Res.* 52:319-333. See also Khan, E., J. P. G. Mack, R. A. Katz, J. Kulkosky, and A. M. Skalka. (1991). Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* 19:851-860.

5 Through a variety of techniques, groups have demonstrated the abrogation of endonuclease and/or integration function of IN through site directed mutation of Asp-64, Asp-116, and Glu-152 residues in the D,D-35-E motif. See Drelich M, Wilhelm R, Mous J. (1992). Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein in vitro. *Virology*;188(2):459-468. See also LaFemina RL, Schneider CL, Robbins HL, Callahan PL, LeGrow K, Roth E, Schleif WA, Emini EA. 10 (1992). Requirement of active human immunodeficiency virus type 1 integrase enzyme for productive infection of human T-lymphoid cells. *J Virol*; 66(12):7414-7419. See also Leavitt AD, Shiue L, Varmus HE. (1993). Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. *J Biol Chem*; 268(3):2113-15 2119.

 Mutation of Asp-64→Ala, Asp-116→Ala, and Glu-152→Ala in the Clade A Int will inactivate the Int active enzyme by disrupting the critical D,D-35-E motif. This will eliminate the biological function of Clade A Int.

 The Negative factor (Nef) protein (27-kDa) is the earliest viral protein to accumulate 20 in the newly infected cell. See Haseltine, W. (1991). *Molecular biology of the human immunodeficiency virus type 1*. FASEB. Vol 5. 2349-2360. Through myristylation, Nef is able to localize on the cytosol side of the cell membrane. See Yu G, Felsted RL. (1992). Effect of myristoylation on p27 nef subcellular distribution and suppression of HIV-LTR transcription. *Virology*. 187(1):46-55. See also Kaminchik, J., N. Bashan, A. Itach, N. Sarver, M. Gorecki, and A. Panet. (1991). Genetic characterization of human 25 immunodeficiency virus type 1 nef gene products translated in vitro and expressed in mammalian cells. *J. Virol.* 65:583-588. Myristylation of proteins is a co-translational event and involves the transfer of myristate from myristyl-Coenzyme A to the amino-terminal motif MGXXX of proteins by the enzyme N-myristyl transferase (NMT). See Towler, D. A., S. P. Adams, S. R. Eubanks, D. S. Towery, E. Jackson-Machelski, L. Glaser & J. I. Gordon (1987). 30 Purification and characterization of yeast myristoyl CoA:protein N- myristoyltransferase. *Proc Natl Acad Sci USA* 84:2708-2712. The lead methionine of the polypeptide is cleaved by the methionine amino peptidase during translation and NMT recognizes the newly

generated terminal amino group of glycine of the emerging peptide after approximately twenty residues are free of the ribosome. NMT transfers myristate to the glycine residue (the myristyl-acceptor) and myristylation is completed. Replacement of the penultimate glycine myristyl-acceptor with any other amino acid residue inhibits myristylation. See Towler, D.

- 5 A., S. R. Eubanks, D. S. Towery, S. P. Adams & L. Glaser (1987). Amino-terminal processing of proteins by N-myristoylation. Substrate specificity of N-myristoyl transferase. *J Biol Chem* 262:1030-1036.

Nef is a multifunctional protein able to modulate a number of surface molecules of the infected cell, such as CD4 (see Garcia, J. V., and A. D. Miller. (1991). Serine
10 phosphorylation-independent downregulation of cell-surface CD4 by nef. *Nature* 350:508-511; and Mariani R and Skowronski J. (1993). CD4 down-regulation by nef alleles isolated from human immunodeficiency virus type 1-infected individuals *Proc. Natl. Acad. Sci. USA*. Vol. 90, pp. 5549-5553; and Aiken C, Konner J, Landau NR, Lenburg ME, Trono D (1994). Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-
15 proximal CD4 cytoplasmic domain. *Cell*. 11;76(5):853-64), CD28 (see Swigut, T., N. Shohdy, and J. Skowronski. (2001). Mechanism for down-regulation of CD28 by Nef. *EMBO J*. 20:1593-1604), MHC-I (see Schwartz, O., V. Marechal, S. Le Gall, F. Lemonnier, and J. M. Heard. (1996). Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nat. Med*. 2:338-342), the macrophage-expressed MHC
20 1b protein HFE (see Drakesmith H, Chen N, Ledermann H, Screaton G, Townsend A, Xu XN. (2005). HIV-1 Nef down-regulates the hemochromatosis protein HFE, manipulating cellular iron homeostasis. *Proc Natl Acad Sci U S A*. 102(31):11017-22), MHC-II (see Stumptner-Cuvelette, P., S. Morchoisne, M. Dugast, S. Le Gall, G. Raposo, O. Schwartz, and P. Benaroch. (2001). HIV-1 Nef impairs MHC class II antigen presentation and surface
25 expression. *Proc. Natl. Acad. Sci. USA* 98:12144-12149), as well as disrupt signal transduction pathways (see Tolstrup, M., L. Ostergaard, A. L. Laursen, S. F. Pedersen, and M. Duch. (2004). HIV/SIV escape from immune surveillance: focus on Nef. *Curr. HIV Res*. 2:141-151) via association with multiple kinases and other cell surface proteins at the cell membrane. The mechanisms of these actions and the nef motifs involved remain to be fully
30 elucidated.

Specifically, a Nef mutant with deletion of the 19 N-terminal amino acids, including the N-terminus myristylation signal eliminated CD4 and MHC-1 down-regulation, while maintaining most CTL, T-helper and B-cell epitopes (see Peng B, Robert-Guroff M (2001). Deletion of N-terminal myristoylation site of HIV Nef abrogates both MHC-1 and CD4

down-regulation. Immunol Lett. 78(3):195-200). Other groups have demonstrated that mutation of the Nef amino-terminal glycine (Gly2) into alanine prevents myristylation (see Liang, X. et al. (2002). Development of HIV-1 Nef vaccine components: immunogenicity study of Nef mutants lacking myristylation and dileucine motif in mice. Vaccine 20: 3413-3421, and Kaminchik, J. et al. (1991). Genetic Characterization of Human Immunodeficiency Virus Type 1 nef Gene Products Translated in vitro and Expressed in Mammalian Cells. J. of Virol. 65(2): 583-588).

Since the amino-terminal motif MGXXX of the Clade A Nef is embedded within the GRIN fusion protein, there is no nascent methionine to be cleaved by the methionine amino peptidase during translation. Thus, no newly generated amino-terminal group of glycine occurs and NMT is unable to execute myristylation. In conclusion, the inability of Nef in GRIN to undergo myristylation abrogates the biological function of Nef.

Example 8: Codon optimization for GRIN (GagPolNef) and Env

The codon usage for each of GRIN and Env was adapted to the codon bias of human genes. The nucleotide and amino acid sequence of the codon optimized GRIN sequence is provided in FIGS. 16A-16J. The nucleotide and amino acid sequence of the codon optimized Env sequence is provided in FIGS. 17A-17D.

Regions of very high (greater than 80%) or very low (less than 30%) GC content were avoided where possible. During the optimization process the following cis-acting motifs were avoided: internal TATA boxes, chi-sites, ribosomal entry sites, AT-rich or GC-rich sequence stretches, ARE, INS, or CRS sequence elements, repeat sequences, RNA secondary structures, cryptic splice donor and acceptor sites, branch points, and HindIII, NcoI, BglII and BclI restriction sites except as indicated in the sequences provided in FIGS. 16 and 17. Also, a Kozak sequence was introduced upstream of the starting ATG for each of GRIN and Env to increase translation initiation, and two stop codons were added to each of GRIN and Env to ensure efficient termination. Restrictions sites to facilitate subcloning were also added, as indicated in FIGS. 16 and 17.

Example 9: Non-human primate study

A non-human primate (Chinese rhesus macaques) study was conducted with the primary objective to assess the immunogenicity of GRIN and ENV in a human adenovirus type 35 (Ad35) vector delivery system. Animals were given increasing doses of Ad35-GRIN/ENV (10^9 , 10^{10} and 10^{11} virus particles [vp]; intramuscular route) and received two immunizations at month 0 and month 6 (with 8 animals per group for the first immunization and 4 animals per group for the second immunization). At various timepoints (from week 0

through to week 50), animals were bled and immunogenicity measured by ELISpot for IFN-gamma (see FIGS. 23A and 23B for the 10^{10} and 10^{11} vp dosages, respectively).

A dose response was observed (data for 10^9 vp not shown), both in ELISpot intensity and frequency of responders following the prime (data not shown). Responses were seen to all vaccine antigen components of GRIN/ENV and IFN γ ELISPOT responses were boosted after the second immunization at month 6.

The invention is further described by the following numbered paragraphs:

1. A consensus nucleotide sequence for HIV-1 Clade A antigens, wherein the sequence comprises nucleotide sequences encoding HIV-1 Clade A Gag, Pol (RT and Int), and Nef ("GRIN), HIV-1 Clade A Gag, RT and Nef ("GRN") or HIV-1 Clade A Env.

2. A consensus nucleotide sequence according to paragraph 1 wherein the encoded Gag protein has the amino acid sequence of FIG. 1.

3. A consensus nucleotide sequence according to paragraph 1 wherein the encoded Pol protein has the amino acid sequence of FIG. 3.

4. A consensus nucleotide sequence according to paragraph 1 wherein the encoded Env protein has the amino acid sequence of FIG. 5.

5. A consensus nucleotide sequence according to paragraph 1 wherein the encoded Nef protein has the amino acid sequence of FIG. 7.

6. A method of identifying an HIV-1 Clade A antigen from a circulating strain or field isolate of HIV-1 that has an amino acid sequence that is similar to the consensus amino acid sequence for that HIV-1 Clade A antigen, comprising comparing the amino acid sequences of antigens from circulating strains or field isolates of HIV-1 to the consensus amino acid sequence for that protein, and selecting an antigen from the circulating strains or field isolates of HIV-1 that has a small protein distance from the consensus sequence.

7. An HIV-1 Clade A antigen identified using the method of paragraph 6.

8. A method of producing a transgenic HIV-1 Clade A antigen comprising selecting an HIV-1 Clade A antigen using the method of paragraph 6 and mutating the nucleotide sequence that encodes the antigen wherein the mutation abrogates the function of that antigen.

9. A method of generating an immune response against HIV-1 comprising administering to a subject a composition comprising a nucleotide sequence or antigen according to any of the previous paragraphs.

Having thus described in detail preferred embodiments of the present invention, it is to be understood that the invention defined by the appended claims is not to be limited by particular details set forth in the above description as many apparent variations thereof are possible without departing from the spirit or scope thereof.

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A non-naturally occurring consensus amino acid sequence for an HIV-1 Clade A
Gag protein, wherein the Gag protein has the amino acid sequence of SEQ ID NO: 2.
5
2. A non-naturally occurring consensus amino acid sequence for an HIV-1 Clade A
Pol protein, wherein the Pol protein has the amino acid sequence of SEQ ID NO: 3.
3. A non-naturally occurring consensus amino acid sequence for an HIV-1 Nef
10 protein, wherein the Nef protein has the amino acid sequence of SEQ ID NO: 5.
4. A non-naturally occurring consensus amino acid sequence for an HIV-1 Clade A
antigen comprising amino acid sequences for HIV-1 Clade A Gag, Pol and Nef
proteins, wherein the Gag protein has the amino acid sequence of SEQ ID NO: 2, the
15 Pol protein has the amino acid sequence of SEQ ID NO: 3 and the Nef protein has the
amino acid sequence of SEQ ID NO: 5.
5. An HIV-1 Clade A fusion protein, wherein the fusion protein comprises HIV-1
Clade A Gag, Pol and Nef and wherein the fusion protein comprises the amino acid
20 sequence of SEQ ID NO: 13.
6. A polynucleotide encoding the fusion protein of claim 5.
7. An immunogenic composition comprising the amino acid sequence of any one
25 of claims 1 to 4, the fusion protein of claim 5, or the polynucleotide of claim 6.
8. A method of generating an immune response against HIV-1 in a subject
comprising administering to the subject the amino acid sequence of any one of claims 1
to 4, the fusion protein of claim 5, the polynucleotide of claim 6, or the composition of
30 claim 7.
9. Use of the amino acid sequence of any one of claims 1 to 4, the fusion protein of
claim 5, the polynucleotide of claim 6, or the composition of claim 7, in the
manufacture of a medicament for generating an immune response against HIV-1 in a
35 subject.

10. The amino acid sequence of any one of claims 1 to 4, substantially as
hereinbefore described.
- 5 11. The fusion protein of claim 5, substantially as hereinbefore described.
12. The polynucleotide of claim 6, substantially as hereinbefore described.
13. The composition of claim 7, substantially as hereinbefore described.
- 10 14. The method of claim 8, substantially as hereinbefore described.

Consensus Gag

MGARASVLSGGKLDWEKIRLRPGGKKKYRLKHLVWASRELERFALNPSLLETAEGCQQIM
EQLQPALKTGTEELRSLFNTVATLYCVHQRIDVKTKEALDKIEEIQNKSKQK---TQQ--
AAADTGXSSKVS----
QNYPIVQNAQQQMIHQXLSRRTLNAWVKVIEEKAFSPEVIPMFALSEGATPQDLNMMMLNIVG
GHQAAMQMLKDTINEEAAEWDRLHPVHAGPIPPGQMREPRGSDIAGTTSTPQEQGAWMTG
NPPIPVGDIYKRWILGLNKIVRMYSVPSILDIKQGPKEPFRDYVDRFFKTLRAEQATQEVKGW
MTETLLVQNANPDCKSILRALGXGATLEEMMTACQGVGGPGHKKARVLAEAMSQVQQTN--
IMM-QRGNFRGQKR-
IKCFNCGKEGHLARNCRAPRKKGCKWCKGKEGHQMKDCTERQANFLGKIWPSSKGRPGNFP
QSRPEPTAPPAEI-FGMGEEIASPPKQEQK--DREQXXPPLVSLKSLFGNDPLSQ

FIG. 1

— 97TZ02 was from low-risk individual in Mbeya region of Southwest Tanzania in 1997 (Accession # AF361872)
 — TZA173 is from anonymous blood donor in Mbeya region, collected in 2001 (Accession # AY253305).
 — KNH1144 is from an anonymous blood donor in southern Kenya, collected in 2000 (Accession # AF457066).
 — SE7535 was collected in 1994 in Sweden from an individual thought to have been infected in Uganda (Accession # AF457066).

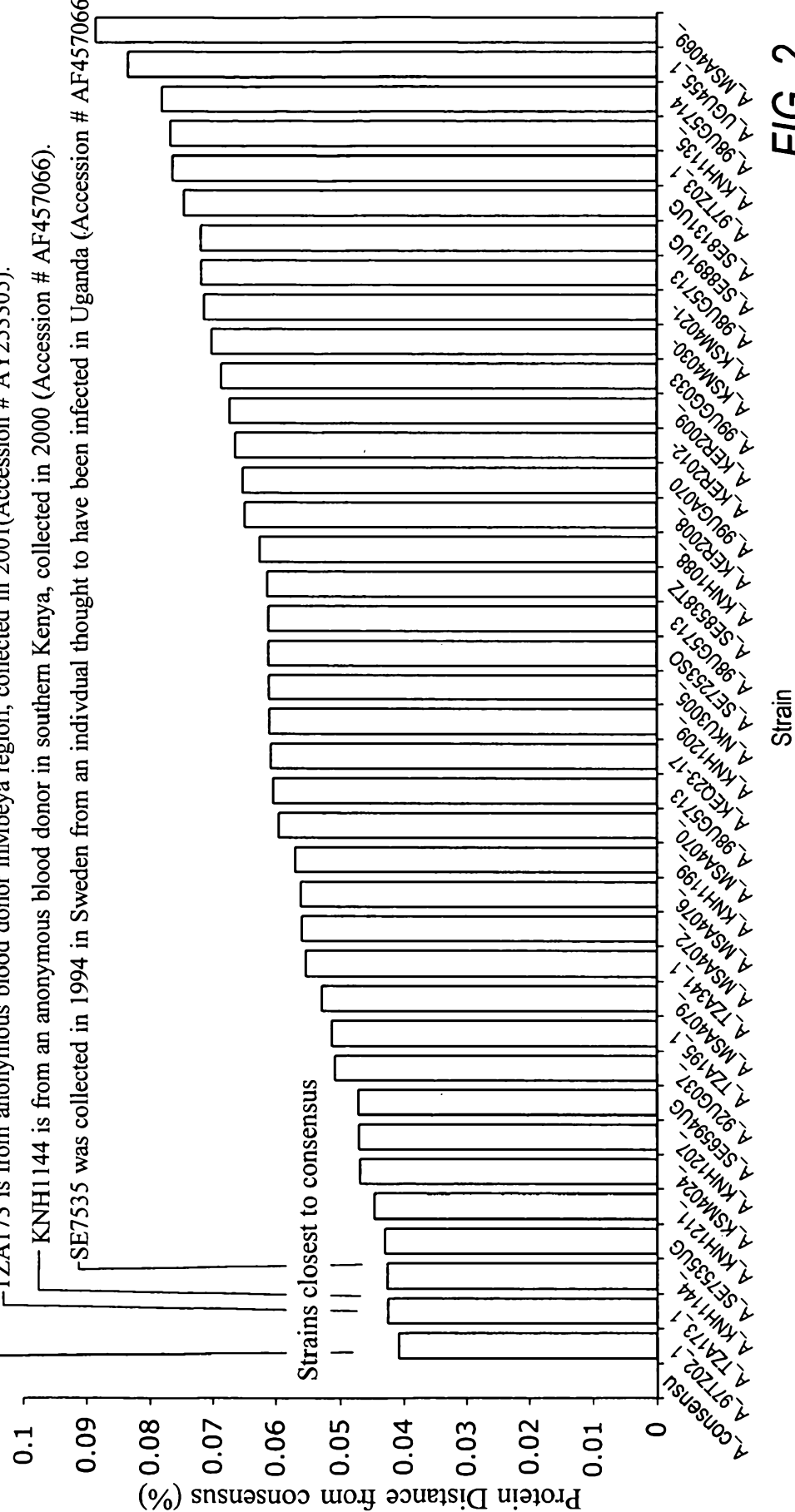


FIG. 2

Consensus Pol
 PQITLWQRPLVTVKIGGQLKEALLDTGADDTVLEDINLPGKWPKMIGGIGFIKVKQYD
 QILIEICGKKAIGTVLVGPTPVNIIGRNMLTQIGCTLNFPIPIETVPVKLKPMDGPKV
 KQWPLTEEKIKALTEICTEMEKEGKISKIGPENPYNTPIFAIKKKDSTKWRKLVDFRELN
 KRTQDFWEVQLGIPHPAGLKKKSVTVLDVGDAYFSVPLDESFRKYTAFTIPSTNNETPG
 IRYQYNVLPQGWKGSPAIFQSSMTKILEPFRSKNPEIIYQYMDDL YVGSDEIGQHRTK
 IEELRAHLLSWGFTTPDKKHQKEPPFLWMGYELHPDKWTVQPIXLPEKESWTVNDIQKLV
 GKLNWASQIYAGIKVKQLCKLLRGAKALTDIVTLTEEALELEAENREILKDPVHGVYYDP
 SKDLIAEIQKQGQDQWYQIYQEPFKNLKTGKYARKRSAHTNDVKQLAEVVQK VVMESIV
 IWGKTPKFKLPIQKETWETWMDYWQATWIPEWEFVNTPPLVKLWYQLEKDPXGAETFY
 VDGAANRETKLGKAGYVTDGRQKVVSLTETTNQKTELHAIXLALQDSGSEVNIVTDSQY
 ALGIIQAQPDRESELVNOIIEKLLIGKDKVYLSWVPAHKGIGGNEQVDKLVSSGIRKVLV
 LDGIDKAQEEHERYHSNWRXMASDFNLPPIVAKEIVASCDKCCQLKGEAMHGVQVDCSPGIW
 QLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLLKLAGRWPVKVVHTDNGSNF
 TSAAFKAAACWWANIQQEEFGIPYNPQSQGVVESMNKELKKIIGQVREQAEHLKTAVQMAVF
 IHNFKRKGIGGYSAGERIIDIIATDIQTKELQKQITKIQNFRVYYRDSRDPWKGPAKL
 LWKGEAVVIQDNSDIKVVPRRKAKIIRDYKGMAGDDCVAGRQDED

FIG. 3

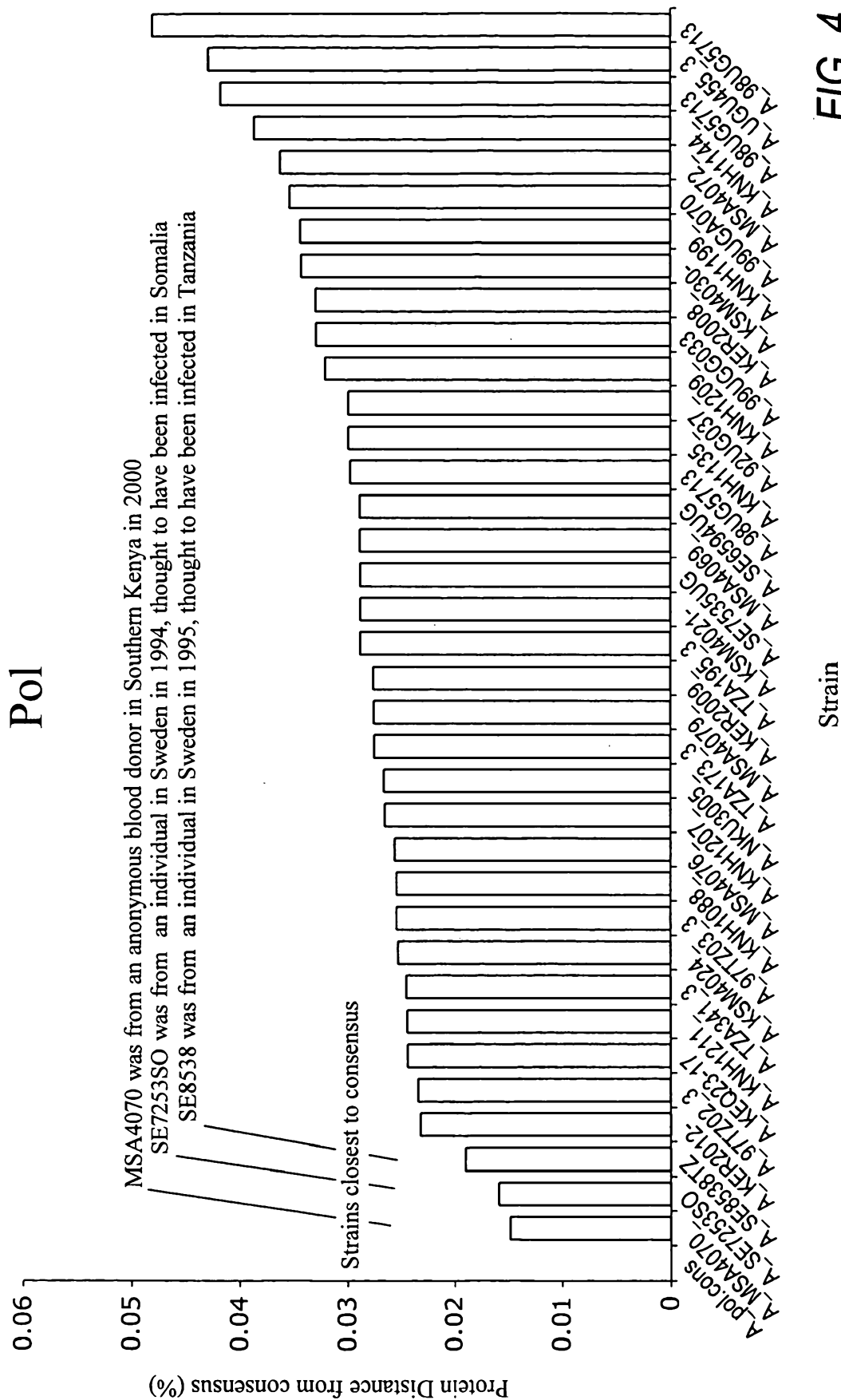


FIG. 4

Consensus Env
MRVMGIQRNCQHLLRWG-TMILGMIICS--XAENLWVTVYYGVPVWKDAETTLFCASDA
KAYTEXHNVWATHACVPTDPNPQEI XLXNVTEEFNMWKNDMVEQMHTDIISLWDQSLKP
CVKLTPLCVTLXCXXXXXXXXXXXXXXXXXXXXXXX--IKNC SFNMTTEL RDKKQ
KVYSLFYRLDVVQI-----XXXXXXXXXXSYRLINCNTSAITQACPKVSFEPIPIHY
CAPAGFALIKCXDXEFNGTGPCKNVSTVQCTHGIKPVVSTQLLNGSLAEXXVX-IRSEN
ITNNAKXIIVQLXXPVXINCTRPNNNTRK---SIRIGPGQAKYATGDIIGDIRQAHCNVS
RXXWNXTLQXVAXQLRXXXFXNKTIEXXSSGGDLEITHSFNCGGEFFYCN TSGLFNST
W-----XXXXXXXXXXSNDTITLXCRIKQIVNMWQRXGQAMYPPIQGVIRCES
NITGLILTRDGGXXXXXXXXXNETFRPGGDMRDNWRSEL YKYKVVKIEPLGVA PTRAKRR
VVEREKRAV-GIGAVFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQNLLRAIEAQQ
HLLKLT VWGIKQLQARVLAVERYLRDQQLGIWGCSGKLICTTNVPWNSSWSNKXXXEIW
DNMTWLQWDKEISNYTQIIYXLIIESNQKEKNEQD LALDKWANLWNWFDISNWL YWIK
IFIMIVGGLIGLRIVFAVLSINRVRQGYSPLSFQTHTPNPRGLDRPGRIEEEGEQGRD
RSIRLVSGFLALA WDDLRLSLCLFSYHRLRDFILIAARTVELLGHSSLKGLRGWEGLYL
WNLLXYWGRELKISAINLXDTIAIAVAGWTD R VIEIGQRIGRAILHIPRRIRQGLERALL

FIG. 5

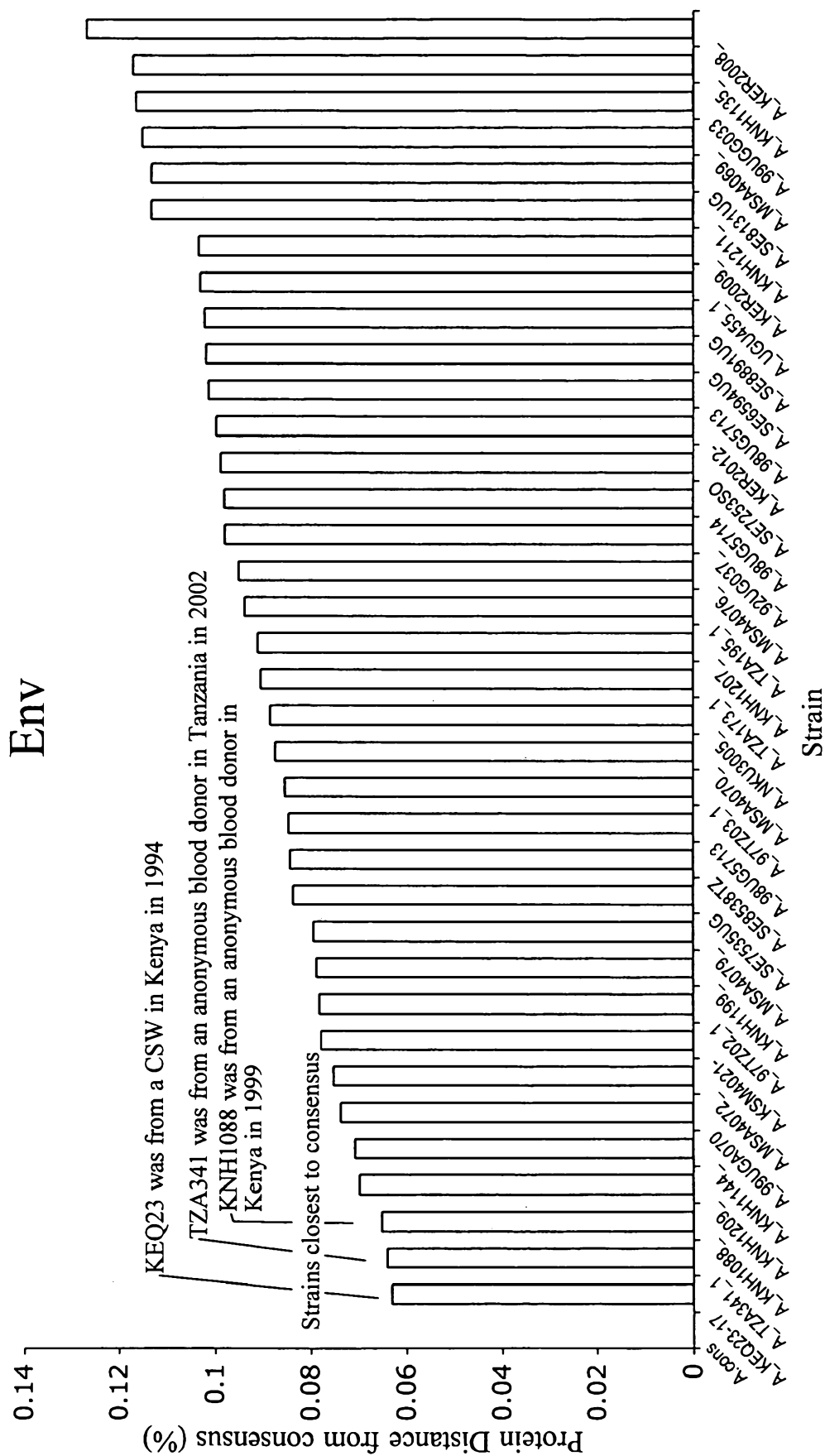


FIG. 6

Consensus Nef

MGGKWSKSSIVGWPEVRERMRRTPXAAX-----

GVGAVSQDLDKHGAITSSNIN

H--PSCVWLEAQEEEE--

VGFPVRPQVPLRPMTYKGAXDLSHFLKEKGGLDGLIYSRK

RQ

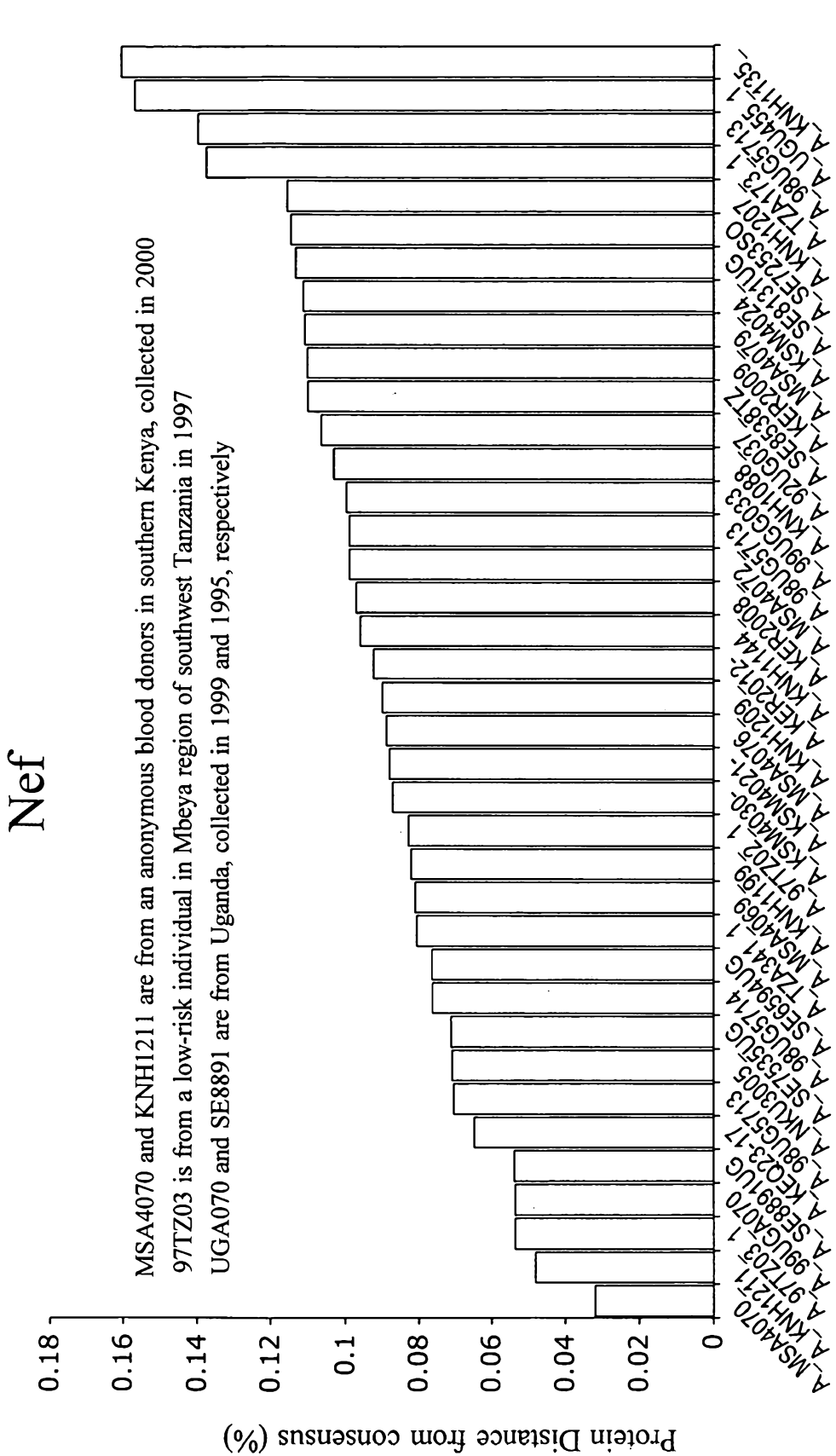
EILDLWVYHTQGYFPDWQNYTPGPGXRYPLTFGWCFKLV

PVDPDEVEKATEGENNSLLHP

ICQHGMDDEEREVLXWKFDSRLALKHRAXELHPEFYKD

The 50% consensus sequence is shown. There were 6 positions where a 50% consensus could not be reached. The mean protein distance in nef was 9.3%, range 3.2% -16.1%.

FIG. 7



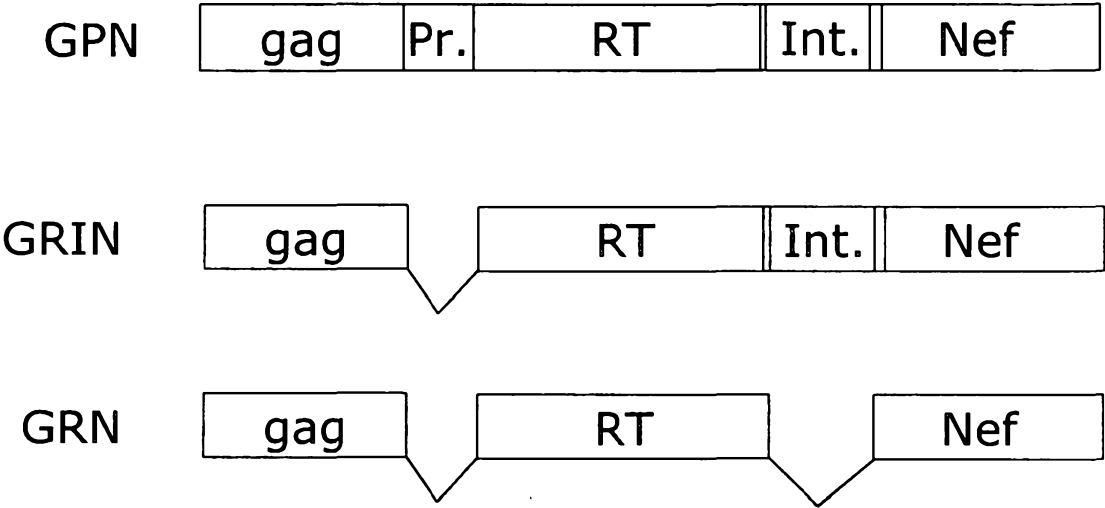


FIG. 9

GAG
AY253305 8766 bp DNA linear VRL 26-AUG-2004
 HIV-1 isolate 01TZA173 from Tanzania gag protein (gag) and pol protein (pol) genes, partial cds; and vif protein (vif), vpr protein (vpr), tat protein (tat), rev protein (rev), vpu protein (vpu), envelope glycoprotein (env), and nef protein (nef) genes, complete cds.

*MG ARASILSGGKLD AWEKIRLRPGGKKYRLKHL VWASRELDRLFAL
 NPSLLETTEGCQQIMNQLQPA VKTGTEEEKSLFNTVATLYCVHQRIDVKDTKEALDKI
 EEIQNKSKQKTQQAAADTGDSSKVSQNYPIQNAQQQMIHQNLSPRTLNAWVKVIEEK
 AFSPEVIPMFSA LSEGATPQDLNVMLNIVGGHQAAMQMLKDTINEEAAEWDRLHPVQA
 GPIPPQIREPRGSDIAGTTSTPQEQLQWMTGNPPIPVGNLYKR WILGLNKIVRMYS
 PVSILDIKQGPKEPRDYVDRFFKALRAEQATQDVKGWMTETLLVQNANPDCKSILKA
 LSGGATLEEMMTACQVGGPGHKARVLAEMSQAQQTNIMMQRGNGFRGQKRIKCFNCG
 KEGHLARNCRAPRKKGCVKCGKEGHQMKDCTERQANFLGKIWPSSKGRPGNFPQSRPE
 PTAPP AELFGMGEIASLPKQEQKDREQVPPLVSLKSLFGNDPLSQ

*MG missing in the genbank entry, artifact of amplicon primer.

FIG. 10

POL

AF457081 8827 bp DNA linear VRL 11-OCT-2002
HIV-1 isolate 00KE_MSA4070 from Kenya, partial genome

PQHILWQRPLVTVKIGGQLKEALLDTGADDTVLEDINLP GKWKPRM
IGGIGGFIKVKQYDQHLEICGKKAIGTVLVGPTPVNIGRNMLTQIGCTLNFPISPI
ETVPVTLKPGMDGPKVNQWPLTEEKIKALTEICTEMENEGKISKIGPENPYNTPIFAI
KKKDS TKWRKLVDERELNKRITQDFWEVQLGIPHPAGLKKKSVTVLDVGDAYFSVPLD
ENFRKYTAFTIPSTNNETPGVRYQYNVLPQGWKGPSAIFQSSMTKILEPFRSKNPEII
IYQYMDL YV'GSDLEIGQHRTKIEELRAHLLSWGFTTPDKKHQKEPPFLWMGYELHPD
KWTVQPIMLPKESWTVNDIQKLVGKLNWASQIVAGHKVKQLCRLLRGAKALTDIVTL
TEEALELELAENREILKDPVHG VYYDPSKDLVAEIQKQGDQWYQYQEPFKNLKTGK
YARKRSAHTNDVRQLAEVVQKVAMESIVWGKTPFKLPIQKETWETWMDY'WQATWT
PEWEFVNTPPLVKLWYQLEKDPILGAETFYVDGAANRETKLGKAGYVTDGRQKV'VSL
TETTNQKTELHAILLALQDSGSEVNIIVTDSQYALGHIQAQPDSESELVNQIEKLIIG
KDKIYLSWVPAHKGIGGNEQVDKLVSSGIRKVLFLDGDHKAQEDHERYHSNWR'TMASD
FNLPPIVAKEIVASCDKCOLKGEAMHGQVDCSPGIWQLDCTHLEGNVILVAVHVASGY
IEAEVIPAETGOETA YFLKLAGRWPVKV'VHTDNGSNFTSAAVKAAACWWANIQQEFGI
PYNPQSQGVV'ESMINKELKKIIGQVRDQAEHLKTA VQMAVFIHNFKRKGIGGYSAGER
IIDHIA'IDIQTKELQKQITKIQNFRVYYRDSRDPWKGPAKLLWKGEAVVIQDNSDI
KVVPRRKANILRDY'GKQMGAGDDCVAGRQDED

FIG. 11

NEF

AF457081 8827 bp DNA linear VRL 11-OCT-2002
HIV-1 isolate 00KE_MSA4070 from Kenya, partial genome.

MGGKWSKGSIVGWPEIRERMRRAPAAAPGVGAVSQDLDKHGAI
SSNNNPSCVWLEAQEEEEVGFPVRPQVPLRPMITYKGAFDLSHFLKEKGGLDGLIYSE
KRQEILDWVYHTQGYFPDWQNYTPGPGVRYPLTFGWCFKLVPMEDVEVEKATEGENN
SLLEHPICQHGMDDEEREVLWKFDSRLALNKHRAQELHPEFYKDC

FIG. 12

ENV**AY253314 8758 BP DNA LINEAR VRL 26-AUG-2004**

HIV-1 isolate 01TZA341 from Tanzania gag protein (gag) and pol protein (pol) genes, partial cds; and vif protein (vif), vpr protein (vpr), tat protein (tat), rev protein (rev), vpu protein (vpu), envelope glycoprotein (env), and nef protein (nef) genes, complete cds

The Env gp140 sequence below does NOT include the trans-membrane region

MRVMEIQRNCQHLLRWGIMILGMHICSTADNLWVTVVYGVVW
 RDAETTLFCASDAKAYSTEKHVWATHACVPTDENPQEIPLDNVTEEFNMWKNNMVDQ
 MHEDIISLWDQSLKPCVQLTPLCVTLNCSNARVNAFENSTEDREGMKNCSFNMITELR
 DKKQQVYSLFYRLDIEKINSNNSEYRLVNCNTSAITQACPKVTFEPIPIHYCAPAG
 FAILKCNDFTEFNGTGPCKNVSTVQCTHGKPPVSTQLLNGSLAEREVRIRSENIANN
 AKNIIVQFASPVKINCIRPNNTKSYRIGPGQTFYATDIVGDIRQAHCNVSRIDWNN
 TLRLVANQLRKYFSNKTIFINSSGGDLEITHSFNCGGEFFYCNTSGLFNSTWTINN
 MQESNDTSNGTITLPCRKQIRMWQRVQGAMYAPPIEGVRCESNTGLILTRDGGN
 NNSANETFRPGGGDIRDNWRSELYKYKVVKIEPLGVAPTRAKRRVVEREKRAVGIGAV
 FLGFLGAAGSTMGAASITLTVQARQLSGIVQQQNLLRAIEAQQLKLTWVGKQL
 QARVLAVERYLRDQQLGIWGCGLICTINVPWNSSWSNKSYYDDIWQNMITWLQWDKE
 ISNYTDHYSLIEESQNQKEKNEQDLLALDKWANLWNWFDISKWLWYI

FIG. 13

Assembled sequence GRIN insert (DZU36984)

AGTCTTCTGTTTTTACGTAGGTGTCAGCCTAGGTGGTCAATATTGGCCATTAG
CCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCCATTG
CATACTTGTATCCATATCATAATATGTACATTTATATTGGCTCATGTCCAAC
ATTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTA
CGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTACG
GTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA
TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC
ATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTG
GCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCT
ACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAAT
GGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTG
ACGTCAATGGGAGTTTGTGTTTGGCACCAAATCAACGGGACTTTCCAAAATG
TCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGG
GAGGTCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAGATCGCCTGGAGAC
GCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCT
CCGCGGCCGCGGAACGGTGCATTGGAAGCTTGCCGCCACCATGGCCGCCAGA
GCCAGCATCCTGAGCGGGGGCAAGCTGGACGCCTGGGAGAAGATCAGACTG
AGGCCTGGCGGCAAGAAGAAGTACCGGCTGAAGCACCTGGTGTGGGCCAGC
AGAGAGCTGGATCGCTTCGCCCTGAATCCTAGCCTGCTGGAGACCACCGAGG
GCTGCCAGCAGATCATGAACCAGCTGCAGCCCGCCGTGAAAACCGGCACCGA
GGAGATCAAGAGCCTGTTCAACACCGTGGCCACCCTGTACTGCGTGCACCAG
CGGATCGACGTGAAGGATACCAAGGAGGCCCTGGACAAGATCGAGGAGATC
CAGAACAAGAGCAAGCAGAAAACCCAGCAGGCCGCTGCCGACACCGGCGAC
AGCAGCAAAGTGAGCCAGAACTACCCCATCATCCAGAATGCCCAGGGCCAG
ATGATCCACCAGAACCTGAGCCCCAGAACCCTGAATGCCTGGGTGAAAGTGA
TCGAGGAAAAGGCCTTCAGCCCCGAAGTGATCCCTATGTTTCAGCGCCCTGAG
CGAGGGCGCCACCCCCCAGGACCTGAACGTGATGCTGAACATTGTGGGCGGA
CACCAGGCCGCCATGCAGATGCTGAAGGACACCATCAATGAGGAGGCCGCC
GAGTGGGACAGACTGCACCCCCGTGCAGGCCGGACCCATCCCCCTGGCCAGA
TCAGAGAGCCCAGAGGCAGCGACATCGCCGGCACCACTCCACCCCTCAAGA
ACAGCTGCAGTGGATGACCGGCAACCCTCCCATCCCTGTGGGCAACATCTAC
AAGCGGTGGATCATCCTGGGCCTGAACAAGATTGTGCGGATGTACAGCCCCG
TGTCCATCCTGGATATCAAGCAGGGCCCCAAGGAGCCCTTCAGAGACTACGT
GGACCGGTTCTTCAAGGCCCTGAGAGCCGAGCAGGCCACCCAGGACGTGAA
GGGCTGGATGACCGAGACCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAG
AGCATCCTGAAGGCCCTGGGCAGCGGCGCCACACTGGAGGAGATGATGACC
GCCTGCCAGGGAGTGGGCGGACCCGGCCACAAGGCCAGAGTGCTGGCCGAG
GCCATGAGCCAGGCCCAGCAGACCAACATCATGATGCAGCGGGGCAACTTCA
GAGGCCAGAAGCGGATCAAGTGCTTCAACTGCGGCAAGGAGGGGCCACCTGG
CCAGAACTGCAGAGCCCCCAGGAAGAAGGGCTGCTGGAAGTGTGGCAAGG
AAGGGCACCAGATGAAGGACTGCACCGAGAGGCAGGCCAATTTCTTGGGCA
AGATTTGGCCTAGCAGCAAGGGCAGACCCGGCAATTTCCCCCAGAGCAGACC

FIG. 14A

CAGCCCCATCGAGACCGTGCCCGTGACCCTGAAGCCCGGCATGGATGGCCCC
AAAGTGAAACAGTGGCCCCCTGACCGAGGAGAAGATTAAGGCCCTGACCGAA
ATCTGTACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAG
AACCCCTACAACACCCCCATCTTCGCCATCAAGAAGAAGGACAGCACCAAGT
GGCGGAAACTGGTGGACTTCCGGGAGCTGAACAAGAGGACCCAGGACTTCTG
GGAAGTGCAGCTGGGCATCCCCCACCCTGCCGGCCTGAAGAAGAAGAAGTCC
GTGACAGTGCTGGATGTGGGCGACGCCTACTTCAGCGTGCCCCCTGGACGAGA
ACTTCAGGAAGTACACCGCCTTCACCATCCCCAGCACCAACAACGAGACCCC
CGGAGTGAGATACCAGTACAACGTGCTGCCTCAGGGCTGGAAGGGCAGCCCC
GCCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGGAGCAAGA
ACCCCGAGATCATCATCTACCAGTACATGGCCGCCCTGTATGTGGGCAGCGA
TCTGGAGATCGGCCAGCACAGGACCAAGATCGAAGAGCTGAGGGCCACCT
GCTGAGCTGGGGCTTCACCACCCCCGATAAGAAGCACCAGAAGGAGCCCCCT
TTCCTGTGGATGGGCTACGAGCTGCACCCCGATAAGTGGACCGTGCAGCCCA
TCATGCTGCCCGATAAGGAGAGCTGGACCGTGAACGACATCCAGAAACTGGT
GGGCAAGCTGAATTGGGCCAGCCAAATCTACGCCGGCATTAAAGTGAAGCAG
CTGTGCAGGCTGCTGAGAGGCGCCAAAGCCCTGACAGACATCGTGACACTGA
CAGAGGAGGCCGAGCTGGAGCTGGCCGAGAACAGGGAGATCCTGAAGGACC
CCGTGCACGGCGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGATTCA
GAAGCAGGGCCAGGACCAGTGGACCTACCAAATCTACCAGGAGCCTTTCAAG
AACCTGAAAACCGGGAAGTACGCCAGGAAGAGAAGCGCCACACCAACGAT
GTGAGGCAGCTGGCCGAAGTGGTGCAGAAAGTGGCTATGGAGAGCATCGTG
ATCTGGGGCAAGACCCCCAAGTTCAAGCTGCCCATCCAGAAGGAGACCTGGG
AAACCTGGTGGATGGACTACTGGCAGGCCACCTGGATTCTGAGTGGGAGTT
CGTGAACACCCCCCTCTGGTGAAGCTGTGGTATCAGCTGGAGAAGGACCCC
ATCCTGGGCGCCGAGACCTTCTACGTGGACGGAGCCGCCAATAGAGAGACCA
AGCTGGGCAAGGCCGGCTACGTGACCGACAGAGGCAGACAGAAAGTGGTGT
CTCTGACCGAGACAACCAACCAGAAAACCGAGCTGCACGCCATCCTGCTGGC
CCTGCAGGACAGCGGCAGCGAAGTGAACATCGTGACCGACTCCAGTACGCC
CTGGGCATCATTACGGCCCAGCCCGATAGAAGCGAGAGCGAGCTGGTGAACC
AGATCATCGAGAAGCTGATCGGCAAGGACAAAATCTACCTGAGCTGGGTGCC
CGCCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGTCCAG
CGGCATCCGGAAAGTGCTGTTTCTGGACGGCATCGACAAGGCCCAGGAGGAC
CACGAGAGATACCACAGCAACTGGCGGACAATGGCCAGCGACTTCAACCTGC
CTCCCATCGTGGCCAAGGAGATCGTGGCCAGCTGCGATAAGTGTGAGCTGAA
GGGCGAGGCCATGCACGGCCAGGTGGACTGCAGCCCTGGCATCTGGCAGCTG
GCCTGCACCCACCTGGAGGGCAAAGTGATTCTGGTGGCCGTGCACGTGGCCA
GCGGCTACATCGAGGCCGAAGTGATTCCCGCCGAGACCGGCCAGGAGACCG
CCTACTTCCTGCTGAAGCTGGCCGGCAGATGGCCCGTGAAAGTGGTGCACAC
CGCCAACGGCAGCAACTTCACCTCTGCCGCCGTGAAGGCCGCCTGTTGGTGG
GCCAATATCCAGCAGGAGTTCGGCATCCCCTACAACCCTCAGAGCCAGGGCG
TGGTGGCCAGCATGAACAAGGAGCTGAAGAAGATCATCGGCCAGGTGAGGG
ACCAGGCCGAGCACCTGAAAACAGCCGTGCAGATGGCCGTGTTTCATCCCAA
CTTCAAGCGGAAGGGCGGCATTGGCGGCTACAGCGCCGGAGAGCGGATCAT
CGACATCATCGCCACCGATATCCAGACCAAGGAACTGCAGAAGCAGATCACC
AAGATTCAAGAACTTCAGAGTGTACTACCGGGACAGCAGGGACCCCATCTGGA

FIG. 14B

AGGATGAGGACAGATCTATGGGCGGCAAGTGGTCCAAGGGCAGCATTGTGG
GCTGGCCCGAGATCCGGGAGAGAATGAGAAGAGCCCCTGCCGCCGCTCCT
GGAGTGGGCGCCGTGTCTCAGGATCTGGATAAGCACGGCGCCATCACCAG
CAGCAACATCAACAACCCCAAGCTGTGTGTGGCTGGAGGCCCAAGAGGA
GGAAGTGGGCTTCCCTGTGAGACCCCAAGGTGCCCTGAGACCCATGACCTA
CAAGGGCGCCTTCGACCTGAGCCACTTCCTGAAGGAGAAGGGCGGCCTGG
ACGGCCTGATCTACAGCCGGAAGCGGCAGGAGATCCTGGATCTGTGGGTGT
ACCACACCCAGGGCTACTTCCCCGACTGGCAGAATTACACCCCTGGCCCTG
GAGTGCGGTATCCCCTGACCTTCGGCTGGTGCTTCAAGCTGGTGCCTATGG
AGCCCGACGAAGTGGAGAAGGCCACAGAGGGCGAGAACAACAGCCTGCTG
CACCTATCTGCCAGCACGGCATGGACGATGAGGAGCGGGAAGTGCTGATC
TGGAAGTTCGACAGCAGGCTGGCCCTGAAGCACAGAGCCCAGGAAGTGCAC
CCAGAGTTCTACAAGGACTGCTGATGATCATAATAAT**CTAGAC**GAGATCCGA
ACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATT
TCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCA
TCAATGTATCTTATCATGTCTAGATCTGAGGTATGATGATACGAGATCGAGGG
TGCGCGCATGCGAATGCGGAGGCAAGCATGCCAGGTTCCAGC

Note: cloning sites are underlined and bold

FIG. 14C

Appendix 2: Assembled sequence Env insert (DSP33447_01)

ATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAA
TGTATTTAgAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
TGCCACCTGACGTCTAAGAAACCATtATTATCATgACATTAACCTATAAAAAATA
GGCGtATCACGAGGCCctTTCGTCTTCAAGAATTGGTCGATGGCAAACAGCT
ATtATGGGTATTATGGGTTCTGAATTAATTAATCGACATCATCAATAATATACCT
TATAGATGGAATGGTGCCAATATGTAAATGAGGTGATTTTAAAAAGTGTGGG
CCGTGTGGTGATTGGCTGTGGGGTTAACGGTTAAAAGGGGCGGCGCGGCC
GTGGGAAAATGACGTTTTATGGGGGTGGAGTTTTTTTTGCAAGTTGTGCGCGG
GAAATGTTACGCATAAAAAGGCTTCTTTTCTCACGGAACACTAGTTTTCCC
ACGGTATTTAACAGGAAATGAGGTAGTTTTGACCGGATGCAAGTGAAAATTG
CTGATTTTCGCGCGAAAACCTGAATGAGGAAGTGTTTTTCTGAATAATGTGGT
ATTTATGGCAGGGTGGAGTATTTGTTCAGGGCCAGGTAGACTTTGACCCATT
ACGTGGAGGTTTCGATTACCGTGTTTTTTACCTGAATTTCCGCGTACCGTGT
CAAAGTCTTCTGTTTTTACGTAGGTGTCAGCCTAGGTGGTCAATATTGGCCA
TTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCTATTGGCC
ATTGCATACGTTGTATCCATATCATAATATGTACATTTATATTGGCTCATGTC
CAACATTACCGCCATGTTGACATTGATTATTGACTAGTTATTAATAGTAATCA
ATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGGTTACATAAC
TTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTG
ACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATT
GACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC
AAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATG
GCCCCGCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTTCTACTTG
GCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGG
CAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCOAAGT
CTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG
GACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGT
AGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTTAGTGAACCGT
CAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACAC
CGGGACCGATCCAGCCTCCGCGGCCCGGAACGGTGCATTGGAAGCTTGCC
GCCACCATGAGGGTGATGGAGATCCAGCGGAACCTGCCAGCACCTGCTGAG
ATGGGGCATCATGATCCTGGGCATGATTATCATCTGCAGCACCGCCGACAA
CCTGTGGGTGACCGTGTACTACGGCGTGCCTGTGTGGAGAGATGCCGAGA
CCACCCTGTTCTGCGCCAGCGACGCCAAGGCCTACAGCACCGAGAAGCAC
AATGTGTGGGCCACCCACGCCTGCGTGCCCTACCGATCCCAACCCTCAGGA
GATCCCCCTGGACAACGTGACCGAGGAGTTCAACATGTGGAAGAACAACAT
GGTGGACCAGATGCACGAGGACATCATCAGCCTGTGGGACCA

FIG. 15A

GAGCCTGAAGCCCTGCGTGCAGCTGACCCCCCTGTGCGTGACCCTGAACT
GCAGCAACGCCAGAGTGAACGCCACCTTCAACTCCACCGAGGACAGGGAG
GGCATGAAGAACTGCAGCTTCAACATGACCACCGAGCTGCGGGATAAGAAG
CAGCAGGTGTACAGCCTGTTCTACCGGCTGGACATCGAGAAGATCAACAGC
AGCAACAACAACAGCGAGTACCGGCTGGTGAAGTGAATACCAGCGCCATC
ACCCAGGCCTGCCCTAAGGTGACCTTCGAGCCCATCCCCATCCACTACTGC
GCCCCCTGCCGGCTTCGCCATCCTGAAGTGAACGACACCGAGTTCAATGG
CACCGGCCCCCTGCAAGAATGTGAGCACCGTGCAGTGCACCCACGGCATCA
AGCCCGTGGTGTCCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGAGA
GAAGTGCGGATCAGGAGCGAGAACATCGCCAACAACGCCAAGAACATCATC
GTGCAGTTCGCCAGCCCCGTGAAGATCAACTGCATCCGGCCCCAACAACAT
ACCCGGAAGAGCTACAGAATCGGCCCTGGCCAGACCTTCTACGCCACCGA
CATTGTGGGCGACATCAGACAGGCCCACTGCAACGTGTCCAGGACCGACT
GGAACAACACCCTGAGACTGGTGGCCAACCAGCTGCGGAAGTACTTCAGC
AACAAGACCATCATCTTCACCAACAGCAGCGGCGGAGACCTGGAGATCACC
ACCCACAGCTTCAATTGTGGCGGCGAGTTCTTCTACTGCAACACCTCCGGC
CTGTTCAATAGCACCTGGACCACCAACAACATGCAGGAGTCCAACGACACC
AGCAACGGCACCATCACCTGCCCTGCCGGATCAAGCAGATCATCCGGAT
GTGGCAGCGCGTGGGCCAGGCCATGTACGCCCTCCCATCGAGGGCGTG
ATTCGCTGCGAGAGCAACATCACCGGCCTGATCCTGACCAGAGATGGCGG
CAACAACAATTCGCCAACGAGACCTTCAGACCTGGCGGCGGAGATATCCG
GGACAACTGGCGGAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGC
CCCTGGGCGTGGCCCCCACCAGAGCCAAGAGAAGAGTGGTGGAGCGGGA
GAAGAGAGCCGTGGGCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCCG
CCGATCTACAATGGGAGCCGCCAGCATCACCTGACCGTGCAGGCCAGA
CAGCTGCTGAGCGGCATCGTGCAGCAGCAGAGCAATCTGCTGAGAGCCAT
CGAGGCCCAGCAGCAGCTGCTGAAGCTGACAGTGTGGGGCATCAAGCAGC
TGCAGGCCAGGGTGCTGGCCGTGGAGAGATACCTGAGGGACCAGCAGCTC
CTGGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCAACGTGCC
CTGGAATAGCAGCTGGAGCAACAAGAGCTACGACGACATCTGGCAGAACAT
GACCTGGCTGCAGTGGGACAAGGAGATCAGCAACTACACCGACATCATCTA
CAGCCTGATCGAGGAGAGCCAGAACCAGCAGGAGAAGAACGAGCAGGATC
TGCTGGCCCTGGACAAGTGGGCCAACCTGTGGAAGTGGTTCGACATCAGC
AAGTGGCTGTGGTACATCAGATCTTGATAAT**CTAGAC**GAGATCCGAACCTTGT
TTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACA
AATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAA
TGTATCTTATCATGTCTAGATCTGAGGTATGATGATACGAGATCGAGGGTGC
GCGCATGCGAATGCGGAGGCAAGCATGCCAGGTTCCAGCCGGTGTGTGTA
GATGTGACCGAAGATCTCAGACCGGATCATTTGGTTATTGCCCCGCACTGGA
GCAGAGTTCGGATCCAGTGGAGAAGAACTGACTAAGGTGAGTATTGGGAA
AACTTTG

FIG. 15B

GGGTGGGATTTTCAGATGGACAGATTGAGTAAAAATTTGTTTTTCTGTCTTG
CAGCTGACATGACT
GGAAATGCTTCTTTTAAGGGGGGGAGTCTTCAGCCCTTATCTGACAGGGCG
TCTCCCATCCTGGGCAGGAGTTCGT

Note: cloning sites are underlined and bold

FIG. 15C

SUBSTITUTE SHEET (RULE 26)

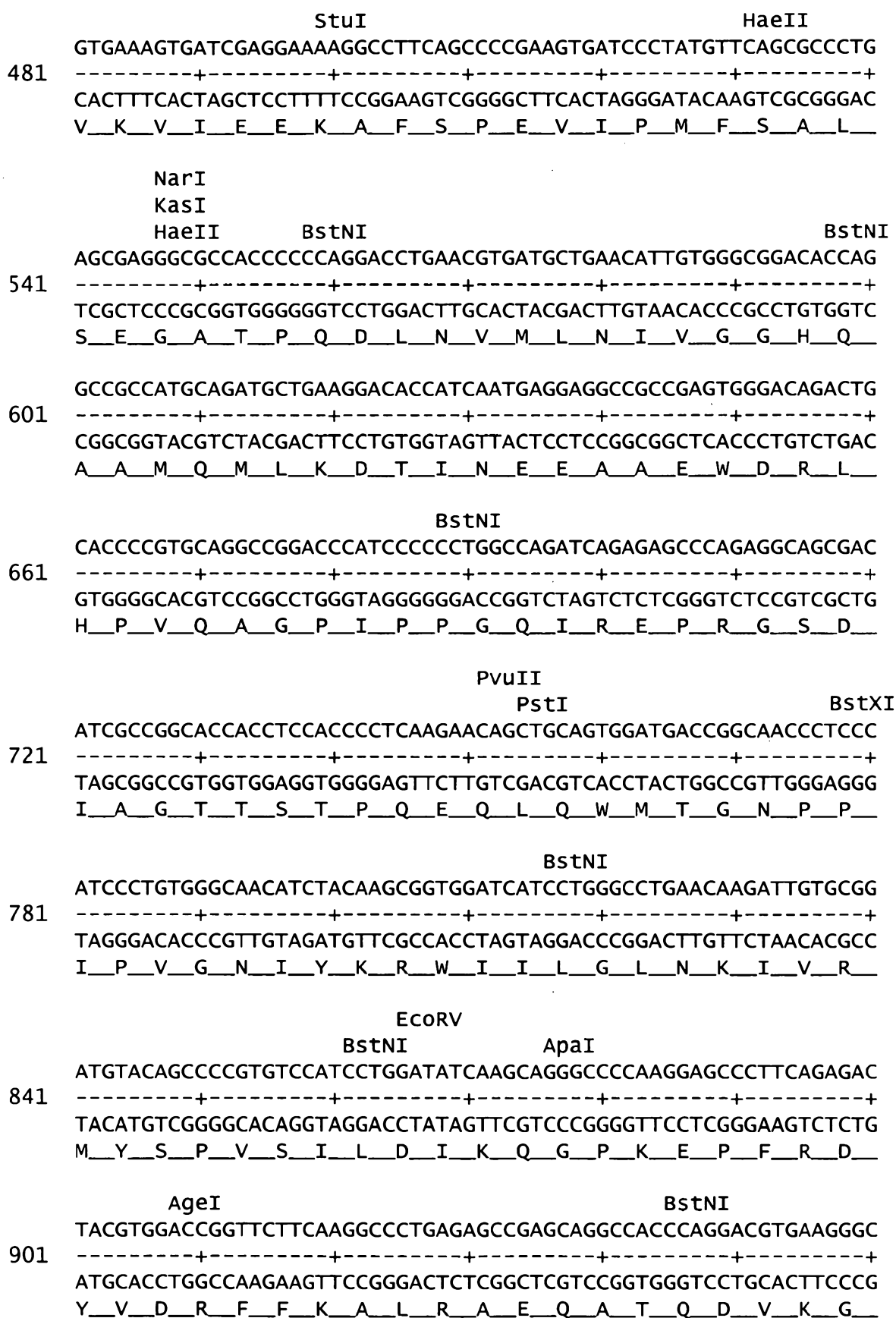


FIG. 16B

BsaI
961 TGGATGACCGAGACCCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAGAGCATCCTGAAG
-----+-----+-----+-----+-----+
ACCTACTGGCTCTGGGACGACCACGTCTTGCGGTTGGGGCTGACGTTCTCGTAGGACTTC
W_M_T_E_T_L_L_V_Q_N_A_N_P_D_C_K_S_I_L_K_

NarI
KasI
PflMI
BstNI
HaeII
1021 GCCCTGGGCAGCGGCCGACACTGGAGGAGATGATGACCGCCTGCCAGGGAGTGGGCGGA
-----+-----+-----+-----+-----+
CGGGACCCGTCGCCGCGGTGTGACCTCCTCTACTACTGGCGGACGGTCCCTCACCCGCT
A_L_G_S_G_A_T_L_E_E_M_M_T_A_C_Q_G_V_G_G_

BstXI
BstNI
1081 CCCGGCCACAAGGCCAGAGTGCTGGCCGAGGCCATGAGCCAGGCCAGCAGACCAACATC
-----+-----+-----+-----+-----+
GGGCCGGTGTTCCGGTCTCACGACCGGCTCCGGTACTCGGTCCGGGTCGTCTGGTTGTAG
P_G_H_K_A_R_V_L_A_E_A_M_S_Q_A_Q_Q_T_N_I_

1141 ATGATGCAGCGGGGCAACTTCAGAGGCCAGAAGCGGATCAAGTGCTTCAACTGCGGCAAG
-----+-----+-----+-----+-----+
TACTACGTCGCCCCGTTGAAGTCTCCGGTCTTCGCCTAGTTCACGAAGTTGACGCCGTTT
M_M_Q_R_G_N_F_R_G_Q_K_R_I_K_C_F_N_C_G_K_

BstNI
PstI
BstNI
1201 GAGGGCCACCTGGCCAGAACTGCAGAGCCCCAGGAAGAAGGGCTGCTGGAAGTGTGGC
-----+-----+-----+-----+-----+
CTCCCGGTGGACCGGTCTTTGACGTCTCGGGGGTCCTTCTTCCCGACGACCTTCACACCG
E_G_H_L_A_R_N_C_R_A_P_R_K_K_G_C_W_K_C_G_

BstXI
BstNI
1261 AAGGAAGGGCACCAGATGAAGGACTGCACCGAGAGGCAGGCCAATTTCTGGGCAAGATT
-----+-----+-----+-----+-----+
TTCCTTCCCGTGGTCTACTTCTGACGTGGCTCTCCGTCCGGTTAAAGGACCCGTTCTAA
K_E_G_H_Q_M_K_D_C_T_E_R_Q_A_N_F_L_G_K_I_

1321 TGGCCTAGCAGCAAGGGCAGACCCGGCAATTTCCCCCAGAGCAGACCCGAGCCACCGCC
-----+-----+-----+-----+-----+
ACCGGATCGTCGTTCCCGTCTGGGCCGTTAAAGGGGTCTCGTCTGGGCTCGGGTGGCGG
W_P_S_S_K_G_R_P_G_N_F_P_Q_S_R_P_E_P_T_A_

1381 CCTCCCGCCGAGCTGTTCGGCATGGGCGAGGGCATCGCCAGCCTGCCAAGCAGGAGCAG
-----+-----+-----+-----+-----+
GGAGGGCGGCTCGACAAGCCGTACCCGCTCCCGTAGCGGTCTCGACGGGTTCGTCTCTGTC
P_P_A_E_L_F_G_M_G_E_G_I_A_S_L_P_K_Q_E_Q_

BspMI
BstNI
1441 AAGGACAGAGAGCAGGTGCCCCCTGGTGTCCCTGAAGTCCCTGTTTCGGCAACGATCCT
-----+-----+-----+-----+-----+
TTCCTGTCTCTCGTCCACGGGGGGGACCACAGGGACTTCAGGGACAAGCCGTTGCTAGGA
K_D_R_E_Q_V_P_P_L_V_S_L_K_S_L_F_G_N_D_P_

FIG. 16C

BstNI NcoI
BamHI BsaI BstEII
CTGAGCCAGGGATCCATGGCCCCCAGATCACCTGTGGCAGAGACCCCTGGTGACCGTG
1501 -----+-----+-----+-----+-----+-----+
GACTCGGTCCCTAGGTACCGGGGGTCTAGTGGGACACCGTCTCTGGGGACCACTGGCAC
L_S_Q_G_S_M_A_P_Q_I_T_L_W_Q_R_P_L_V_T_V_

NarI
KasI
PvuII HaeII
AAGATCGGCGGCCAGCTGAAGGAAGCCCTGCTGGATACAGGCGCCGATGATACCGTGCTG
1561 -----+-----+-----+-----+-----+-----+
TTCTAGCCGCGGTCGACTTCCTTCGGGACGACCTATGTCCGCGGCTACTATGGCACGAC
K_I_G_G_Q_L_K_E_A_L_L_D_T_G_A_D_D_T_V_L_

BspMI
GAGGACATCAACCTGCCCCGCAAGTGAAGCCTAGAATGATCGGCGGCATCGGGGGCTTC
1621 -----+-----+-----+-----+-----+-----+
CTCCTGTAGTTGGACGGGCGTTCACCTTCGGATCTTACTAGCCGCGCTAGCCCCGAAG
E_D_I_N_L_P_G_K_W_K_P_R_M_I_G_G_I_G_G_F_

ATCAAAGTGAAGCAGTACGACCAGATCCTGATCGAGATTTGCGGAAGAAGGCCATCGGC
1681 -----+-----+-----+-----+-----+-----+
TAGTTTCACTTCGTATGCTGGTCTAGGACTAGCTCTAAACGCCCTTCTTCCGGTAGCCG
I_K_V_K_Q_Y_D_Q_I_L_I_E_I_C_G_K_K_A_I_G_

ApaI EagI
ACCGTGCTGGTGGGCCCCACCCCTGTGAATATCATCGGCCGAACATGCTGACCCAGATC
1741 -----+-----+-----+-----+-----+-----+
TGGCACGACCACCGGGGTGGGACACTTATAGTAGCCGGCCTTGTACGACTGGGTCTAG
T_V_L_V_G_P_T_P_V_N_I_I_G_R_N_M_L_T_Q_I_

BsaI
GGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACCGTGCCCGTGACCCTGAAGCCC
1801 -----+-----+-----+-----+-----+-----+
CCGACGTGGGACTTGAAGGGGTAGTCGGGGTAGCTCTGGCACGGGCACTGGGACTTCGGG
G_C_T_L_N_F_P_I_S_P_I_E_T_V_P_V_T_L_K_P_

GGCATGGATGGCCCCAAAGTGAAACAGTGGCCCCTGACCGAGGAGAAGATTAAGGCCCTG
1861 -----+-----+-----+-----+-----+-----+
CCGTACCTACCGGGGTTTCACTTTGTCACCGGGGACTGGCTCCTCTTCTAATTCCGGGAC
G_M_D_G_P_K_V_K_Q_W_P_L_T_E_E_K_I_K_A_L_

ACCGAAATCTGTACCGAGATGGAGAAGGAGGGCAAGATCAGCAAGATCGGCCCCGAGAAC
1921 -----+-----+-----+-----+-----+-----+
TGGCTTTAGACATGGCTCTACCTCTTCCTCCCGTTCTAGTCGTTCTAGCCGGGGCTCTTG
T_E_I_C_T_E_M_E_K_E_G_K_I_S_K_I_G_P_E_N_

CCCTACAACACCCCCATCTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGGAAACTG
1981 -----+-----+-----+-----+-----+-----+
GGGATGTTGTGGGGGTAGAAGCGGTAGTTCTTCTTCCTGTCGTGGTTACCGCCTTTGAC
P_Y_N_T_P_I_F_A_I_K_K_K_D_S_T_K_W_R_K_L_

FIG. 16D

2041 GTGGACTTCCGGGAGCTGAACAAGAGGACCCAGGACTTCTGGGAAGTGCAGCTGGGCATC
-----+-----+-----+-----+-----+-----+
CACCTGAAGGCCCTCGACTTGTTCTCCTGGGTCCTGAAGACCCTTCACGTCGACCCGTAG
V_D_F_R_E_L_N_K_R_T_Q_D_F_W_E_V_Q_L_G_I_

2101 CCCCACCTGCCGGCCTGAAGAAGAAGAAGTCCGTGACAGTGCTGGATGTGGGCGACGCC
-----+-----+-----+-----+-----+-----+
GGGGTGGGACGGCCGGACTTCTTCTTCTTCAGGCACTGTACGACCTACACCCGCTGCGG
P_H_P_A_G_L_K_K_K_K_S_V_T_V_L_D_V_G_D_A_

2161 BstNI
TACTTCAGCGTGCCCCTGGACGAGAACTTCAGGAAGTACACCGCCTTCACCATCCCCAGC
-----+-----+-----+-----+-----+-----+
ATGAAGTCGCACGGGGACCTGCTCTTGAAGTCCTTCATGTGGCGGAAGTGGTAGGGGTCG
Y_F_S_V_P_L_D_E_N_F_R_K_Y_T_A_F_T_I_P_S_

2221 BsaI
ACCAACAACGAGACCCCCGGAGTGAGATACAGTACAACGTGCTGCCTCAGGGCTGGAAG
-----+-----+-----+-----+-----+-----+
TGGTTGTTGCTCTGGGGCCTCACTCTATGGTCATGTTGCACGACGGAGTCCCCGACCTTC
T_N_N_E_T_P_G_V_R_Y_Q_Y_N_V_L_P_Q_G_W_K_

2281 BstXI BstNI
GGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGGAGCAAG
-----+-----+-----+-----+-----+-----+
CCGTCGGGGCGGTAGAAAGGTCTCGTCGTACTGGTTCTAGGACCTCGGGAAGGCCTCGTTC
G_S_P_A_I_F_Q_S_S_M_T_K_I_L_E_P_F_R_S_K_

2341 PflMI
AACCCCGAGATCATCATCTACAGTACATGGCCGCCCTGTATGTGGGACGCGATCTGGAG
-----+-----+-----+-----+-----+-----+
TTGGGGCTCTAGTAGTAGATGGTCATGTACCGGCGGGACATACACCCGTCGCTAGACCTC
N_P_E_I_I_I_Y_Q_Y_M_A_A_L_Y_V_G_S_D_L_E_

2401 ApaI BspMI
ATCGGCCAGCACAGGACCAAGATCGAAGAGCTGAGGGCCACCTGCTGAGCTGGGGCTTC
-----+-----+-----+-----+-----+-----+
TAGCCGGTCGTGTCCTGGTTCTAGCTTCTCGACTCCCGGGTGGACGACTCGACCCCGAAG
I_G_Q_H_R_T_K_I_E_E_L_R_A_H_L_L_S_W_G_F_

2461 ACCACCCCGATAAGAAGCACCAGAAGGAGCCCCCTTTCCTGTGGATGGGCTACGAGCTG
-----+-----+-----+-----+-----+-----+
TGGTGGGGGCTATTCTTCGTGGTCTTCCTCGGGGAAAGGACACCTACCCGATGCTCGAC
T_T_P_D_K_K_H_Q_K_E_P_P_F_L_W_M_G_Y_E_L_

2521 CACCCCGATAAGTGGACCGTGCAGCCCATCATGCTGCCCCGATAAGGAGAGCTGGACCGTG
-----+-----+-----+-----+-----+-----+
GTGGGGCTATTACCTGGCACGTGCGGTAGTACGACGGGCTATTCCTCTCGACCTGGCAC
H_P_D_K_W_T_V_Q_P_I_M_L_P_D_K_E_S_W_T_V_

FIG. 16E

Pf1MI
 AACGACATCCAGAACTGGTGGGCAAGCTGAATTGGGCCAGCCAAATCTACGCCGGCATT
 2581 -----+-----+-----+-----+-----+-----+
 TTGCTGTAGGTCTTTGACCACCCGTTTCGACTTAACCCGGTCGGTTTAGATGCGGCCGTAA
 N_D_I_Q_K_L_V_G_K_L_N_W_A_S_Q_I_Y_A_G_I_

NarI
 KasI
 PvuII
 HaeII
 AAAGTGAAGCAGCTGTGCAGGCTGCTGAGAGGCGCCAAAGCCCTGACAGACATCGTGACA
 2641 -----+-----+-----+-----+-----+-----+
 TTTCACTTCGTCGACACGTCCGACGACTCTCCGCGGTTTCGGGACTGTCTGTAGCACTGT
 K_V_K_Q_L_C_R_L_L_R_G_A_K_A_L_T_D_I_V_T_

CTGACAGAGGAGGCCGAGCTGGAGCTGGCCGAGAACAGGGAGATCCTGAAGGACCCCGTG
 2701 -----+-----+-----+-----+-----+-----+
 GACTGTCTCCTCCGGCTCGACCTCGACCGGCTCTTGTCCCTCTAGGACTTCTGGGGCAC
 L_T_E_E_A_E_L_E_L_A_E_N_R_E_I_L_K_D_P_V_

BstNI
 HinfI
 BstNI
 CACGGCGTGTACTACGACCCCAGCAAGGACCTGGTGGCCGAGATTCAGAAGCAGGGCCAG
 2761 -----+-----+-----+-----+-----+-----+
 GTGCCGCACATGATGCTGGGGTTCGTTCTGACCACCGGCTCTAAGTCTTCGTCCCGGTC
 H_G_V_Y_Y_D_P_S_K_D_L_V_A_E_I_Q_K_Q_G_Q_

BstNI
 GACCAGTGGACCTACCAAATCTACCAGGAGCCTTTCAAGAACCTGAAAACCGGGAAGTAC
 2821 -----+-----+-----+-----+-----+-----+
 CTGGTCACCTGGATGGTTTAGATGGTCCTCGGAAAGTTCTTGGACTTTTGGCCCTTCATG
 D_Q_W_T_Y_Q_I_Y_Q_E_P_F_K_N_L_K_T_G_K_Y_

BstNI
 HaeII
 PvuII
 GCCAGGAAGAGAAGCGCCACACCAACGATGTGAGGCAGCTGGCCGAAGTGGTGCAGAAA
 2881 -----+-----+-----+-----+-----+-----+
 CGGTCCTTCTCTTCGCGGGTGTGGTTGCTACACTCCGTGACCGGCTTCACCACGTCTTT
 A_R_K_R_S_A_H_T_N_D_V_R_Q_L_A_E_V_V_Q_K_

GTGGCTATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCAAGCTGCCCATCCAG
 2941 -----+-----+-----+-----+-----+-----+
 CACCGATACCTCTCGTAGCACTAGACCCGTTCTGGGGGTTCAAGTTCGACGGGTAGGTC
 V_A_M_E_S_I_V_I_W_G_K_T_P_K_F_K_L_P_I_Q_

BstNI
 BsaI
 BstNI
 HinfI
 BstNI
 AAGGAGACCTGGGAAACCTGGTGGATGGACTACTGGCAGGCCACCTGGATTCCTGAGTGG
 3001 -----+-----+-----+-----+-----+-----+
 TTCCTCTGGACCCTTTGGACCACCTACCTGATGACCGTCCGGTGGACCTAAGGACTCACC
 K_E_T_W_E_T_W_W_M_D_Y_W_Q_A_T_W_I_P_E_W_

FIG. 16F

3061 GAGTTCGTGAACACCCCCCTCTGGTGAAGCTGTGGTATCAGCTGGAGAAGGACCCCATC
-----+-----+-----+-----+-----+-----+
CTCAAGCACTTGTGGGGGGGAGACCACTTCGACACCATAGTCGACCTCTTCCTGGGGTAG
E_F_V_N_T_P_P_L_V_K_L_W_Y_Q_L_E_K_D_P_I_

NarI
KasI
HaeII BsaI BsaI
3121 CTGGGCGCCGAGACCTTCTACGTGGACGGAGCCGCAATAGAGAGACCAAGCTGGGCAAG
-----+-----+-----+-----+-----+-----+
GACCCGCGGCTCTGGAAGATGCACCTGCCTCGGCGGTTATCTCTCTGGTTCGACCCGTTCT
L_G_A_E_T_F_Y_V_D_G_A_A_N_R_E_T_K_L_G_K_

3181 GCCGGCTACGTGACCGACAGAGGCAGACAGAAAGTGGTGTCTCTGACCGAGACAACCAAC
-----+-----+-----+-----+-----+-----+
CGGCCGATGCACTGGCTGTCTCCGTCTGTCTTTCACCACAGAGACTGGCTCTGTTGGTTG
A_G_Y_V_T_D_R_G_R_Q_K_V_V_S_L_T_E_T_T_N_

BstXI PstI
3241 CAGAAAACCGAGCTGCACGCCATCCTGCTGGCCCTGCAGGACAGCGGCAGCGAAGTGAAC
-----+-----+-----+-----+-----+-----+
GTCTTTTGGCTCGACGTGCGGTAGGACGACCGGGACGTCCTGTCTGCGCGTCGCTTCACTTG
Q_K_T_E_L_H_A_I_L_L_A_L_Q_D_S_G_S_E_V_N_

HinfI BstNI
3301 ATCGTGACCGACTCCCAGTACGCCCTGGGCATCATTAGGCCCAGCCCGATAGAAGCGAG
-----+-----+-----+-----+-----+-----+
TAGCACTGGCTGAGGGTCATGCGGGACCCGTAGTAAGTCCGGGTCGGGCTATCTTCGCTC
I_V_T_D_S_Q_Y_A_L_G_I_I_Q_A_Q_P_D_R_S_E_

3361 AGCGAGCTGGTGAACCAGATCATCGAGAAGCTGATCGGCAAGGACAAAATCTACCTGAGC
-----+-----+-----+-----+-----+-----+
TCGCTCGACCACTTGGTCTAGTAGCTCTTCGACTAGCCGTTCTGTTTTAGATGGACTCG
S_E_L_V_N_Q_I_I_E_K_L_I_G_K_D_K_I_Y_L_S_

BspMI
3421 TGGGTGCCCCGCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGTCCAGC
-----+-----+-----+-----+-----+-----+
ACCCACGGGCGGGTGTTCCTCGTAGCCGCGGTTGCTCGTCCACCTGTTTCGACCACAGGTCTG
W_V_P_A_H_K_G_I_G_G_N_E_Q_V_D_K_L_V_S_S_

BstNI
3481 GGCATCCGGAAAGTGCTGTTTTCTGGACGGCATCGACAAGGCCAGGAGGACCACGAGAGA
-----+-----+-----+-----+-----+-----+
CCGTAGGCCTTTTACGACAAAGACCTGCCGTAGCTGTTCCGGGTCCTCCTGGTGCTCTCT
G_I_R_K_V_L_F_L_D_G_I_D_K_A_Q_E_D_H_E_R_

FIG. 16G

BspMI

3541 TACCACAGCAACTGGCGGACAATGGCCAGCGACTTCAACCTGCCTCCCATCGTGGCCAAG
-----+-----+-----+-----+-----+-----+
ATGGTGTCGTTGACCGCCTGTTACCGGTCGCTGAAGTTGGACGGAGGGTAGCACCGGTTTC
Y_H_S_N_W_R_T_M_A_S_D_F_N_L_P_P_I_V_A_K_

PvuII PvuII BstNI

3601 GAGATCGTGGCCAGCTGCGATAAGTGTGCTGAGCTGAAGGGCGAGGCCATGCACGGCCAGGTG
-----+-----+-----+-----+-----+-----+
CTCTAGCACCGGTCGACGCTATTACAGTCGACTTCCCCTCCGGTACGTGCCGGTCCAC
E_I_V_A_S_C_D_K_C_Q_L_K_G_E_A_M_H_G_Q_V_

PstI BstNI PvuII BstNI HinfI

3661 GACTGCAGCCCTGGCATCTGGCAGCTGGCCTGCACCCACCTGGAGGGCAAAGTGATTCTG
-----+-----+-----+-----+-----+-----+
CTGACGTCGGGACCGTAGACCGTCGACCGGACGTGGGTGGACCTCCCGTTTCACTAAGAC
D_C_S_P_G_I_W_Q_L_A_C_T_H_L_E_G_K_V_I_L_

HinfI BsaI BstNI

3721 GTGGCCGTGCACGTGGCCAGCGGCTACATCGAGGCCGAAGTGATTCCCGCCGAGACCGGC
-----+-----+-----+-----+-----+-----+
CACCGGCACGTGCACCGGTCGCCGATGTAGCTCCGGCTTCACTAAGGGCGGCTCTGGCCG
V_A_V_H_V_A_S_G_Y_I_E_A_E_V_I_P_A_E_T_G_

BsaI

3781 CAGGAGACCGCCTACTTCTGCTGAAGCTGGCCGGCAGATGGCCCGTGAAAGTGGTGCAC
-----+-----+-----+-----+-----+-----+
GTCCTCTGGCGGATGAAGGACGACTTCGACCGGCCGTCTACCGGGCACTTTCACCACGTG
Q_E_T_A_Y_F_L_L_K_L_A_G_R_W_P_V_K_V_V_H_

3841 ACCGCCAACGGCAGCAACTTCACCTCTGCCGCCGTGAAGGCCGCCTGTTGGTGGGCCAAT
-----+-----+-----+-----+-----+-----+
TGGCGGTTGCCGTCGTTGAAGTGGAGACGGCGGCACTTCCGGCGGACAACCACCGGTTA
T_A_N_G_S_N_F_T_S_A_A_V_K_A_A_C_W_W_A_N_

PflMI
BstNI

3901 ATCCAGCAGGAGTTTCGGCATCCCCTACAACCCTCAGAGCCAGGGCGTGGTGGCCAGCATG
-----+-----+-----+-----+-----+-----+
TAGGTCGTCCTCAAGCCGTAGGGGATGTTGGGAGTCTCGGTCCCGCACCACCGGTCGTAC
I_Q_Q_E_F_G_I_P_Y_N_P_Q_S_Q_G_V_V_A_S_M_

BstNI BstNI

3961 AACAAGGAGCTGAAGAAGATCATCGGCCAGGTGAGGGACCAGGCCGAGCACCTGAAAACA
-----+-----+-----+-----+-----+-----+
TTGTTCTCTGACTTCTTCTAGTAGCCGGTCCACTCCCTGGTCCGGCTCGTGGACTTTTGT
N_K_E_L_K_K_I_I_G_Q_V_R_D_Q_A_E_H_L_K_T_

4021 GCCGTGCAGATGGCCGTGTTTCATCCACAACCTTCAAGCGGAAGGGCGGCATTGGCGGCTAC
-----+-----+-----+-----+-----+-----+
CGGCACGTCTACCGGCACAAGTAGGTGTTGAAGTTCGCCTTCCCGCCGTAACCGCCGATG
A_V_Q_M_A_V_F_I_H_N_F_K_R_K_G_G_I_G_G_Y_

FIG. 16H

HaeII EcoRV PstI
4081 AGCGCCGGAGAGCGGATCATCGACATCATCGCCACCGATATCCAGACCAAGGAACTGCAG
-----+-----+-----+-----+-----+-----+
TCGCGGCCTCTCGCCTAGTAGCTGTAGTAGCGGTGGCTATAGGTCTGGTTCCTTGACGTC
S_A_G_E_R_I_I_D_I_I_A_T_D_I_Q_T_K_E_L_Q

HinfI
4141 AAGCAGATCACCAAGATTCAGAACTTCAGAGTGTACTACCGGGACAGCAGGGACCCCATC
-----+-----+-----+-----+-----+-----+
TTCGTCTAGTGGTTCTAAGTCTTGAAGTCTCACATGATGGCCCTGTCGTCCCTGGGGTAG
K_Q_I_T_K_I_Q_N_F_R_V_Y_Y_R_D_S_R_D_P_I

NarI
KasI
ApaI HaeII BstNI
4201 TGGAAGGGCCCTGCCAAGCTGCTGTGGAAGGGCGAAGGCGCCGTGGTGATCCAGGACAAC
-----+-----+-----+-----+-----+-----+
ACCTTCCCGGGACGGTTCGACGACACCTTCCCGCTTCCGCGGCACCACTAGGTCCTGTTG
W_K_G_P_A_K_L_L_W_K_G_E_G_A_V_V_I_Q_D_N

HinfI
4261 AGCGACATCAAAGTGGTGCCCCGGAGGAAGGCCAAGATTCTGCGGGACTACGGCAAACAG
-----+-----+-----+-----+-----+-----+
TCGCTGTAGTTTCACCACGGGGCCTCCTTCCGGTTCTAAGACGCCCTGATGCCGTTTGTC
S_D_I_K_V_V_P_R_R_K_A_K_I_L_R_D_Y_G_K_Q

BglII
4321 ATGGCCGGCGATGACTGCGTGGCCGGCAGGCAGGATGAGGACAGATCTATGGGCGGCAAG
-----+-----+-----+-----+-----+-----+
TACCGGCCGCTACTGACGCACCGGCCGTCCGTCTACTCCTGTCTAGATACCCGCCGTTT
M_A_G_D_D_C_V_A_G_R_Q_D_E_D_R_S_M_G_G_K

4381 TGGTCCAAGGGCAGCATTGTGGGCTGGCCCCGAGATCCGGGAGAGAATGAGAAGAGCCCCT
-----+-----+-----+-----+-----+-----+
ACCAGGTTCCCGTCGTAACACCCGACCGGGCTCTAGGCCCTCTCTTACTCTTCTCGGGGA
W_S_K_G_S_I_V_G_W_P_E_I_R_E_R_M_R_R_A_P

NarI NarI
KasI KasI
BstNI HaeII HaeII
4441 GCCGCCGCTCCTGGAGTGGGCGCCGTGTCTCAGGATCTGGATAAGCACGGCGCCATCACC
-----+-----+-----+-----+-----+-----+
CGGCGGCGAGGACCTCACCCGCGGCACAGAGTCCTAGACCTATTCGTGCCGCGGTAGTGG
A_A_A_P_G_V_G_A_V_S_Q_D_L_D_K_H_G_A_I_T

PvuII BstNI
4501 AGCAGCAACATCAACAACCCAGCTGTGTGTGGCTGGAGGCCAGGAAGAGGAGGAAGTG
-----+-----+-----+-----+-----+-----+
TCGTGTTGTAGTTGTTGGGGTGCACACACACCGACCTCCGGGTCCTTCTCCTCCTTAC
S_S_N_I_N_N_P_S_C_V_W_L_E_A_Q_E_E_E_V

FIG. 16I

NarI
KasI
HaeII

BsaI BstNI BsaI

4561 GGCTTCCCTGTGAGACCCCAGGTGCCCTGAGACCCATGACCTACAAGGGCGCCTTCGAC
-----+-----+-----+-----+-----+
CCGAAGGGACACTCTGGGGTCCACGGGACTCTGGGTACTGGATGTTCCCGCGGAAGCTG
G_F_P_V_R_P_Q_V_P_L_R_P_M_T_Y_K_G_A_F_D_

BstNI

4621 CTGAGCCACTTCCTGAAGGAGAAGGGCGGCCTGGACGGCCTGATCTACAGCCGGAAGCGG
-----+-----+-----+-----+-----+
GACTCGGTGAAGGACTTCCTCTTCCC GCCGACCTGCCGGA CTAGATGTCGGCCTTCGCC
L_S_H_F_L_K_E_K_G_G_L_D_G_L_I_Y_S_R_K_R_

BstNI BstNI

4681 CAGGAGATCCTGGATCTGTGGGTGTACCACACCCAGGGCTACTTCCCCGACTGGCAGAAT
-----+-----+-----+-----+-----+
GTCCTCTAGGACCTAGACACCCACATGGTGTGGGTCCCGATGAAGGGGCTGACCGTCTTA
Q_E_I_L_D_L_W_V_Y_H_T_Q_G_Y_F_P_D_W_Q_N_

BstNI BstNI

4741 TACACCCCTGGCCCTGGAGTGCGGTATCCCCTGACCTTCGGCTGGTGCTTCAAGCTGGTG
-----+-----+-----+-----+-----+
ATGTGGGGACCGGGACCTCACGCCATAGGGGACTGGAAGCCGACCACGAAGTTCGACCAC
Y_T_P_G_P_G_V_R_Y_P_L_T_F_G_W_C_F_K_L_V_

4801 CCTATGGAGCCCCGACGAAGTGGAGAAGGCCACAGAGGGCGAGAACAACAGCCTGCTGCAC
-----+-----+-----+-----+-----+
GGATACCTCGGGCTGCTTACCTCTTCCGGTGTCTCCCGCTCTTGTTGTCGGACGACGTG
P_M_E_P_D_E_V_E_K_A_T_E_G_E_N_N_S_L_L_H_

4861 CCTATCTGCCAGCACGGCATGGACGATGAGGAGCGGGAAGTGCTGATCTGGAAGTTCGAC
-----+-----+-----+-----+-----+
GGATAGACGGTCGTGCCGTACCTGCTACTCCTCGCCCTTCACGACTAGACCTTCAAGCTG
P_I_C_Q_H_G_M_D_D_E_E_R_E_V_L_I_W_K_F_D_

BstNI

4921 AGCAGGCTGGCCCTGAAGCACAGAGCCCAGGAACTGCACCCAGAGTTCTACAAGGACTGC
-----+-----+-----+-----+-----+
TCGTCCGACCGGGACTTCGTGTCTCGGGTCTTGACGTGGGTCTCAAGATGTTCTGACG
S_R_L_A_L_K_H_R_A_Q_E_L_H_P_E_F_Y_K_D_C_

BclI XbaI

4981 TGATGATCATAATAATCTAGAA
-----+-----+
ACTACTAGTATTATTAGATCTT
*
—

FIG. 16J

HindIII BspMI
1 AAGCTTGCCCGCCACCATGAGGGTGATGGAGATCCAGCGGAAGTCCAGCACCTGCTGAGA
-----+-----+-----+-----+-----+-----+
TTCGAACGGCGGTGGTACTCCCACTACCTCTAGGTCGCCTTGACGGTCGTGGACGACTCT
M_R_V_M_E_I_Q_R_N_C_Q_H_L_L_R

BstNI PstI BstEII
61 TGGGGCATCATGATCCTGGGCATGATTATCATCTGCAGCACCGCCGACAACCTGTGGGTG
-----+-----+-----+-----+-----+-----+
ACCCCGTAGTACTAGGACCCGTACTAATAGTAGACGTCGTGGCGGCTGTTGGACACCCAC
W_G_I_M_I_L_G_M_I_I_I_C_S_T_A_D_N_L_W_V

BsaI
121 ACCGTGTACTACGGCGTGCCTGTGTGGAGAGATGCCGAGACCACCCTGTTCTGCGCCAGC
-----+-----+-----+-----+-----+-----+
TGGCACATGATGCCGCACGGACACACCTCTCTACGGCTCTGGTGGGACAAGACGCGGTCTG
T_V_Y_Y_G_V_P_V_W_R_D_A_E_T_T_L_F_C_A_S

StuI
181 GACGCCAAGGCCTACAGCACCGAGAAGCACAAATGTGTGGGCCACCCACGCCTGCGTGCCT
-----+-----+-----+-----+-----+-----+
CTGCGGTTCCGGATGTCGTGGCTCTTCGTGTTACACACCCGGTGGGTGCGGACGCACGGA
D_A_K_A_Y_S_T_E_K_H_N_V_W_A_T_H_A_C_V_P

BstNI
241 ACCGATCCCAACCCTCAGGAGATCCCCCTGGACAACGTGACCGAGGAGTTCAACATGTGG
-----+-----+-----+-----+-----+-----+
TGGCTAGGGTTGGGAGTCCTCTAGGGGGACCTGTTGCACTGGCTCCTCAAGTTGTACACC
T_D_P_N_P_Q_E_I_P_L_D_N_V_T_E_E_F_N_M_W

301 AAGAACAACATGGTGGACCAGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTG
-----+-----+-----+-----+-----+-----+
TTCTTGTTGTACCACCTGGTCTACGTGCTCCTGTAGTAGTCGGACACCCTGGTCTCGGAC
K_N_N_M_V_D_Q_M_H_E_D_I_I_S_L_W_D_Q_S_L

PvuII PstI
361 AAGCCCTGCGTGCAGCTGACCCCCCTGTGCGTGACCCTGAACTGCAGCAACGCCAGAGTG
-----+-----+-----+-----+-----+-----+
TTCGGGACGCACGTCGACTGGGGGGACACGCACTGGGACTTGACGTCGTTGCGGTCTCAC
K_P_C_V_Q_L_T_P_L_C_V_T_L_N_C_S_N_A_R_V

PstI
421 AACGCCACCTTCAACTCCACCGAGGACAGGGAGGGCATGAAGAACTGCAGCTTCAACATG
-----+-----+-----+-----+-----+-----+
TTGCGGTGGAAGTTGAGGTGGCTCCTGTCCCTCCCGTACTTCTTGACGTCGAAGTTGTAC
N_A_T_F_N_S_T_E_D_R_E_G_M_K_N_C_S_F_N_M

BspMI
481 ACCACCGAGCTGCGGGATAAGAAGCAGCAGGTGTACAGCCTGTTCTACCGGCTGGACATC
-----+-----+-----+-----+-----+-----+
TGGTGGCTCGACGCCCTATTCTTCGTGCTCCACATGTCGGACAAGATGGCCGACCTGTAG
T_T_E_L_R_D_K_K_Q_Q_V_Y_S_L_F_Y_R_L_D_I

FIG. 17A

HaeII

541 GAGAAGATCAACAGCAGCAACAACAGCGAGTACCGGCTGGTGAAGTCAATACCAGC
-----+-----+-----+-----+-----+-----+
CTCTTCTAGTTGTCGTCGTTGTTGTTGTCGCTCATGGCCGACCACTTGACGTTATGGTCG
E_K_I_N_S_S_N_N_N_S_E_Y_R_L_V_N_C_N_T_S_

StuI
BstNI BstEII

601 GCCATCACCCAGGCCTGCCCTAAGGTGACCTTCGAGCCCATCCCCATCCACTACTGCGCC
-----+-----+-----+-----+-----+-----+
CGGTAGTGGGTCCGGACGGGATTCCACTGGAAGCTCGGGTAGGGGTAGGTGATGACGCGG
A_I_T_Q_A_C_P_K_V_T_F_E_P_I_P_I_H_Y_C_A_

661 CCTGCCGGCTTCGCCATCCTGAAGTGCAACGACACCGAGTTCAATGGCACCGGGCCCTGC
-----+-----+-----+-----+-----+-----+
GGACGGCCGAAGCGGTAGGACTTCACGTTGCTGTGGCTCAAGTTACCGTGGCCGGGGACG
P_A_G_F_A_I_L_K_C_N_D_T_E_F_N_G_T_G_P_C_

PvuII

721 AAGAATGTGAGCACCGTGCAGTGCACCCACGGCATCAAGCCCGTGGTGTCCACCCAGCTG
-----+-----+-----+-----+-----+-----+
TTCTTACACTCGTGGCACGTACGTGGGTGCCGTAGTTCGGGCACCACAGGTGGGTGCGAC
K_N_V_S_T_V_Q_C_T_H_G_I_K_P_V_V_S_T_Q_L_

BstNI

781 CTGCTGAACGGCAGCCTGGCCGAGAGAGAAGTGCGGATCAGGAGCGAGAACATCGCCAAC
-----+-----+-----+-----+-----+-----+
GACGACTTGCCGTCGGACCGGCTCTCTCTTCACGCCTAGTCCTCGCTCTTGTAGCGGTTG
L_L_N_G_S_L_A_E_R_E_V_R_I_R_S_E_N_I_A_N_

841 AACGCCAAGAACATCATCGTGCAGTTCGCCAGCCCCGTGAAGATCAACTGCATCCGGCCCC
-----+-----+-----+-----+-----+-----+
TTGCGGTTCTTGTAGTAGCACGTCAAGCGGTGCGGGCACTTCTAGTTGACGTAGGCCGGG
N_A_K_N_I_I_V_Q_F_A_S_P_V_K_I_N_C_I_R_P_

HinfI BstNI

901 AACACAATACCCGGAAGAGCTACAGAATCGGCCCTGGCCAGACCTTCTACGCCACCGAC
-----+-----+-----+-----+-----+-----+
TTGTTGTTATGGGCCTTCTCGATGTCTTAGCCGGGACCGGTCTGGAAGATGCGGTGGCTG
N_N_N_T_R_K_S_Y_R_I_G_P_G_Q_T_F_Y_A_T_D_

BstNI

961 ATTGTGGGCGACATCAGACAGGCCCACTGCAACGTGTCCAGGACCGACTGGAACAACACC
-----+-----+-----+-----+-----+-----+
TAACACCCGCTGTAGTCTGTCCGGGTGACGTTGCACAGGTCCTGGCTGACCTTGTTGTGG
I_V_G_D_I_R_Q_A_H_C_N_V_S_R_T_D_W_N_N_T_

PvuII ScaI

1021 CTGAGACTGGTGGCCAACCAGCTGCGGAAGTACTTCAGCAACAAGACCATCATCTTCACC
-----+-----+-----+-----+-----+-----+
GACTCTGACCACCGGTTGGTTCGACGCCTTCATGAAGTCGTTGTTCTGGTAGTAGAAGTGG
L_R_L_V_A_N_Q_L_R_K_Y_F_S_N_K_T_I_I_F_T_

FIG. 17B

BstNI

BsaI

1081 AACAGCAGCGGCGGAGACCTGGAGATCACCACCCACAGCTTCAATTGTGGCGGCGAGTTC
 -----+-----+-----+-----+-----+-----+
 TTGTCGTCGCCGCTCTGGACCTCTAGTGGTGGGTGTCGAAGTTAACACCGCCGCTCAAG
 N_S_S_G_G_D_L_E_I_T_T_H_S_F_N_C_G_G_E_F_

BstNI

HinfI

1141 TTCTACTGCAACACCTCCGGCCTGTTCAATAGCACCTGGACCACCAACAACATGCAGGAG
 -----+-----+-----+-----+-----+-----+
 AAGATGACGTTGTGGAGGCCGGACAAGTTATCGTGGACCTGGTGGTTGTTGTACGTCCTC
 F_Y_C_N_T_S_G_L_F_N_S_T_W_T_T_N_N_M_Q_E_

1201 TCCAACGACACCAGCAACGGCACCATCACCTGCCCTGCCGGATCAAGCAGATCATCCGG
 -----+-----+-----+-----+-----+-----+
 AGGTTGCTGTGGTCGTTGCCGTGGTAGTGGGACGGGACGGCCTAGTTCGTCTAGTAGGCC
 S_N_D_T_S_N_G_T_I_T_L_P_C_R_I_K_Q_I_I_R_

BstNI

HinfI

1261 ATGTGGCAGCGCTGGGCCAGGCCATGTACGCCCTCCCATCGAGGGCGTGATTCGCTGC
 -----+-----+-----+-----+-----+-----+
 TACACCGTCGCGCACCCGGTCCGGTACATGCGGGGAGGGTAGCTCCCGCACTAAGCGACG
 M_W_Q_R_V_G_Q_A_M_Y_A_P_P_I_E_G_V_I_R_C_

1321 GAGAGCAACATCACCGGCCTGATCCTGACCAGAGATGGCGGCAACAACAATTCCGCCAAC
 -----+-----+-----+-----+-----+-----+
 CTCTCGTTGTAGTGGCCGGACTAGGACTGGTCTCTACCGCCGTTGTTGTTAAGGCGGTTG
 E_S_N_I_T_G_L_I_L_T_R_D_G_G_N_N_N_S_A_N_

BsaI

BstNI

EcoRV

1381 GAGACCTTCAGACCTGGCGGCGGAGATATCCGGGACAACCTGGCGGAGCGAGCTGTACAAG
 -----+-----+-----+-----+-----+-----+
 CTCTGGAAGTCTGGACCGCCGCTCTATAGGCCCTGTTGACCGCCTCGCTCGACATGTTCT
 E_T_F_R_P_G_G_G_D_I_R_D_N_W_R_S_E_L_Y_K_

BstNI

1441 TACAAGGTGGTGAAGATCGAGCCCCTGGGCGTGGCCCCCACCAGAGCCAAGAGAAGAGTG
 -----+-----+-----+-----+-----+-----+
 ATGTTCCACCACTTCTAGCTCGGGGACCCGCACCGGGGGTGGTCTCGGTTCTCTTCTCAC
 Y_K_V_V_K_I_E_P_L_G_V_A_P_T_R_A_K_R_R_V_

NarI

KasI

HaeII

BstNI

1501 GTGGAGCGGGAGAAGAGAGCCGTGGGCATCGGCGCCGTGTTTCTGGGCTTCCTGGGAGCC
 -----+-----+-----+-----+-----+-----+
 CACCTCGCCCTCTTCTCTCGGCACCCGTAGCCGCGGCACAAAGACCCGAAGGACCCCTCGG
 V_E_R_E_K_R_A_V_G_I_G_A_V_F_L_G_F_L_G_A_

FIG. 17C

1561 PvuII
GCCGGATCTACAATGGGAGCCGCCAGCATCACCTGACCGTGCAGGCCAGACAGCTGCTG
-----+-----+-----+-----+-----+-----+
CGGCCTAGATGTTACCCTCGGCGGTCGTAGTGGGACTGGCACGTCCGGTCTGTGCGACGAC
A_G_S_T_M_G_A_A_S_I_T_L_T_V_Q_A_R_Q_L_L_

1621 PvuII
AGCGGCATCGTGCAGCAGCAGAGCAATCTGCTGAGAGCCATCGAGGCCAGCAGCAGCTG
-----+-----+-----+-----+-----+-----+
TCGCCGTAGCACGTCTGCTCTCGTTAGACGACTCTCGGTAGCTCCGGGTCGTCTGCGAC
S_G_I_V_Q_Q_Q_S_N_L_L_R_A_I_E_A_Q_Q_Q_L_

1681 PvuII BstXI
PstI BstNI
CTGAAGCTGACAGTGTGGGGCATCAAGCAGCTGCAGGCCAGGGTGCTGGCCGTGGAGAGA
-----+-----+-----+-----+-----+-----+
GACTTCGACTGTACACCCCGTAGTTCGTGCGACGTCCGGTCCCACGACCGGCACCTCTCT
L_K_L_T_V_W_G_I_K_Q_L_Q_A_R_V_L_A_V_E_R_

1741 BstNI PstI
TACCTGAGGGACCAGCAGCTCCTGGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACC
-----+-----+-----+-----+-----+-----+
ATGGACTCCCTGGTCGTGAGGACCCGTAGACCCGACGTGCGCCGTTGACTAGACGTGG
Y_L_R_D_Q_Q_L_L_G_I_W_G_C_S_G_K_L_I_C_T_

1801 BstNI PvuII
ACCAACGTGCCCTGGAATAGCAGCTGGAGCAACAAGAGCTACGACGACATCTGGCAGAAC
-----+-----+-----+-----+-----+-----+
TGGTTGCACGGGACCTTATCGTCGACCTCGTTGTTCTCGATGCTGCTGTAGACCGTCTTG
T_N_V_P_W_N_S_S_W_S_N_K_S_Y_D_D_I_W_Q_N_

1861 PstI
BstNI
ATGACCTGGCTGCAGTGGGACAAGGAGATCAGCAACTACACCGACATCATCTACAGCCTG
-----+-----+-----+-----+-----+-----+
TACTGGACCGACGTACCCCTGTTCTCTAGTCGTTGATGTGGCTGTAGTAGATGTGCGGAC
M_T_W_L_Q_W_D_K_E_I_S_N_Y_T_D_I_I_Y_S_L_

1921 BstNI
ATCGAGGAGAGCCAGAACCAGCAGGAGAAGAACGAGCAGGATCTGCTGGCCCTGGACAAG
-----+-----+-----+-----+-----+-----+
TAGCTCCTCTCGGTCTTGGTCGTCTTCTTGTCTGCTCCTAGACGACCGGGACCTGTTT
I_E_E_S_Q_N_Q_Q_E_K_N_E_Q_D_L_L_A_L_D_K_

1981 PflMI BglII
TGGGCCAACCTGTGGAAGTGGTTGACATCAGCAAGTGGCTGTGGTACATCAGATCTTGA
-----+-----+-----+-----+-----+-----+
ACCCGGTTGGACACCTTGACCAAGCTGTAGTCGTTACCGACACCATGTAGTCTAGAAGT
W_A_N_L_W_N_W_F_D_I_S_K_W_L_W_Y_I_R_S_*_

2041 XbaI
TAATCTAGAA
-----+
ATTAGATCTT

FIG. 17D

IAVA- Eval. of cloned AdH5
constructs in CB6F1; DAY 10-Elispot

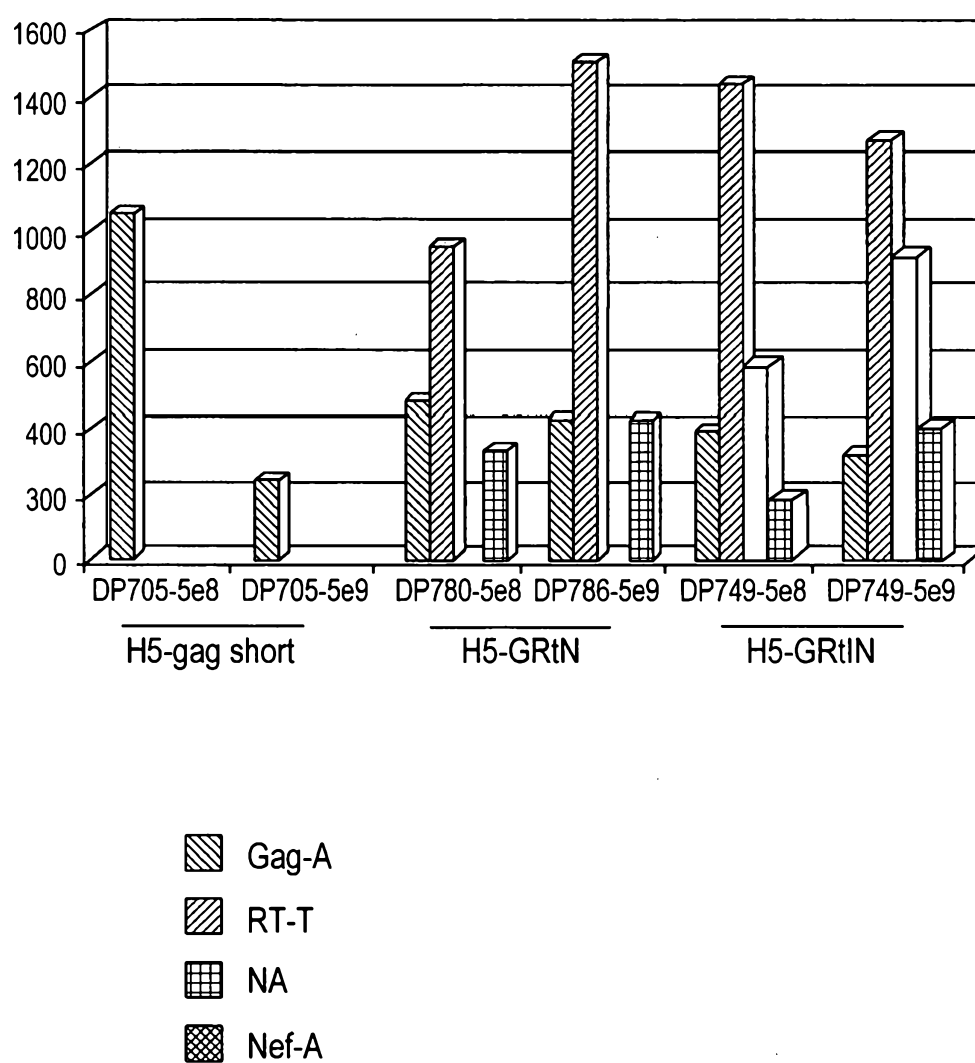


FIG. 18

**FCC/05.002: Mean responses to HIV Clade A peptide library pools by IFN-gamma ELIspt
at 10 days post-primary**
Each bar represents Mean of 4 or 8 samples (each tested in triplicate) +/- 1 Standard Error

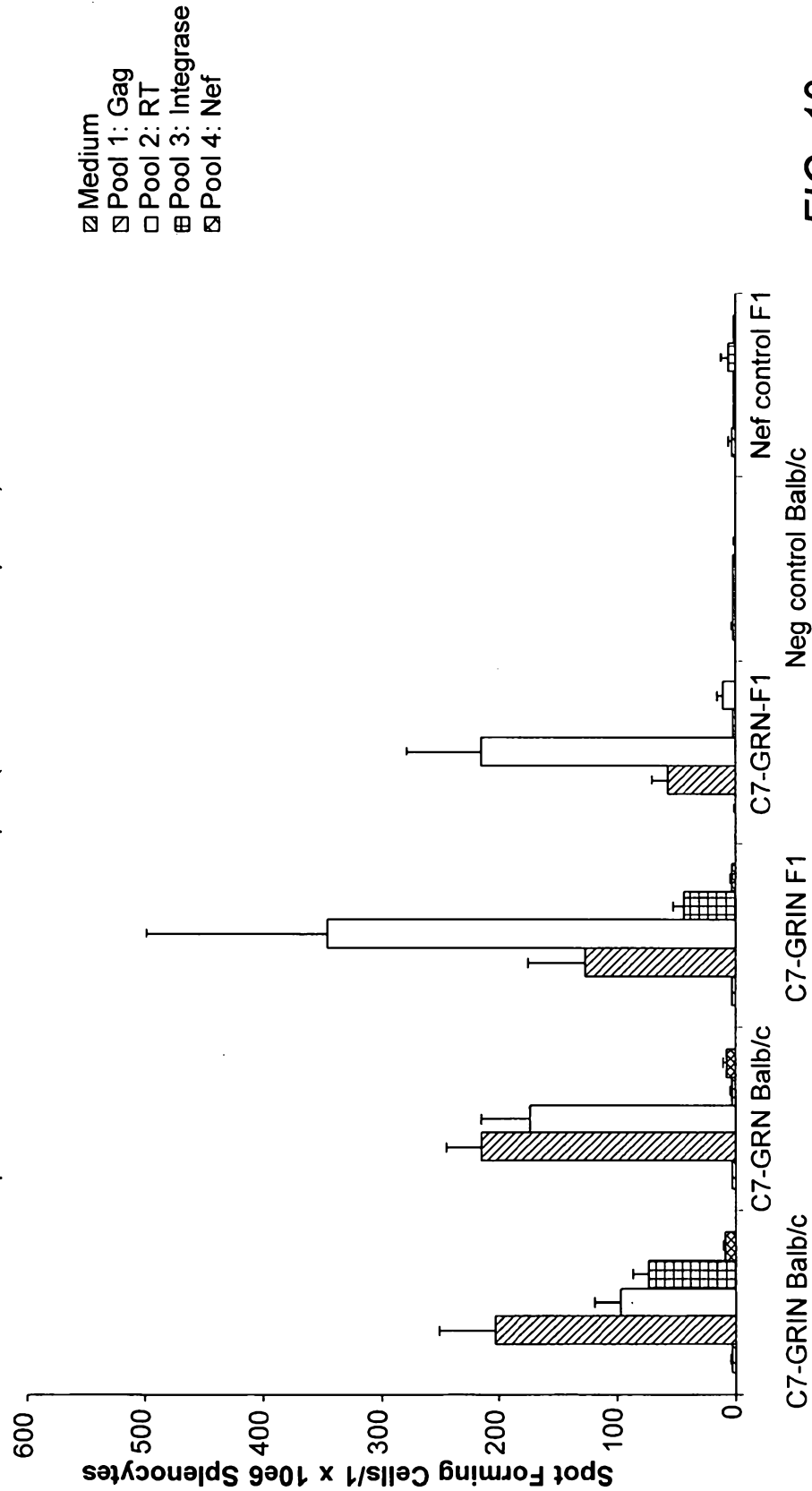


FIG. 19

**FCC/02.005: Mean responses to HIV Clade A peptide library pools by IL-2 ELISpot
at 10 days post-primary**

Each bar represents Mean of 4 or 8 samples (each tested in triplicate) +/- 1 Standard Error

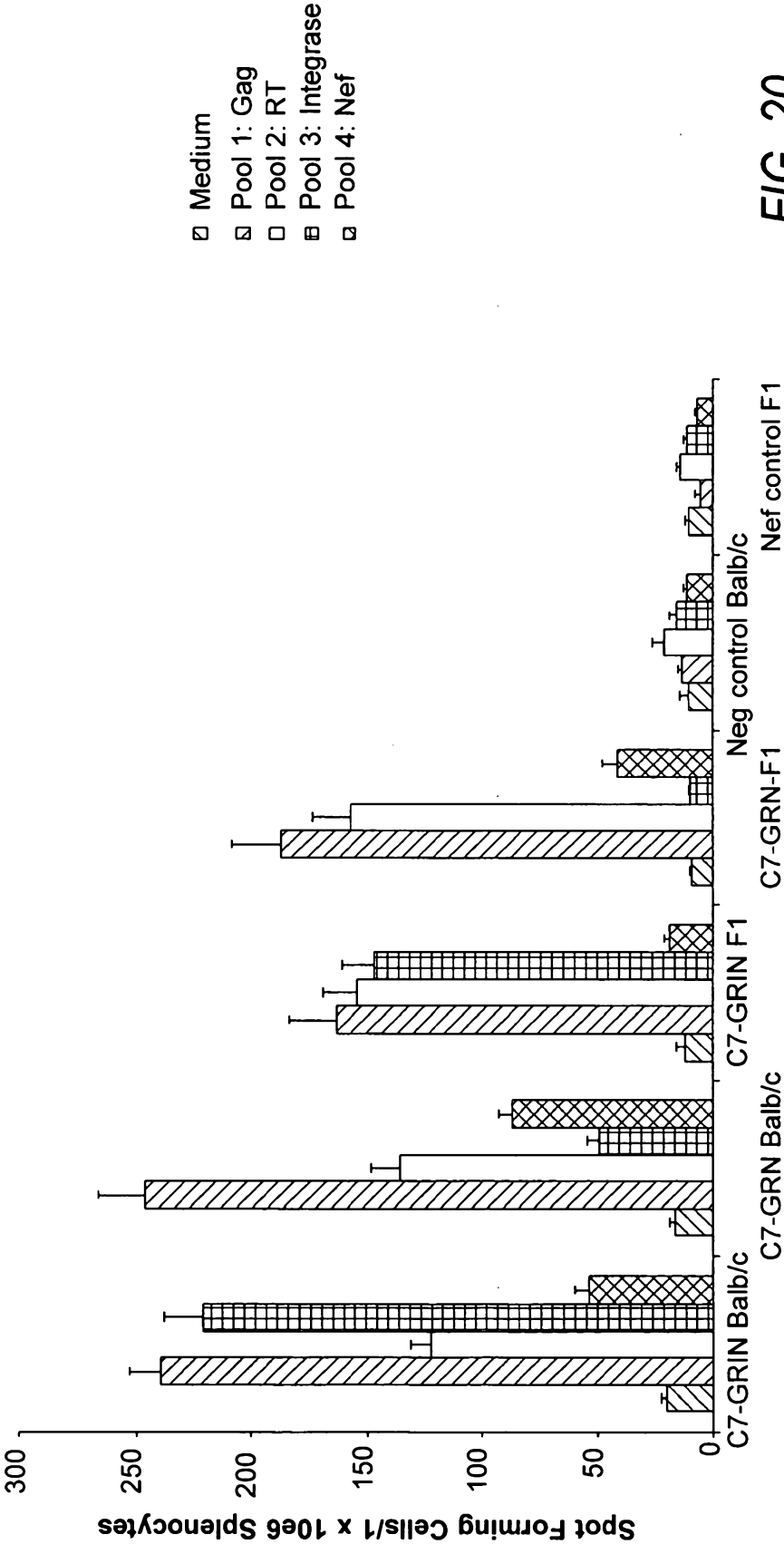


FIG. 20

FCC/05.003: Mean responses to HIV Clade A peptide library pools by IFN-gamma ELISpot at 10 days post-primary
Each bar represents Mean of 8 samples (each tested in triplicate) +/- 1 Standard Error

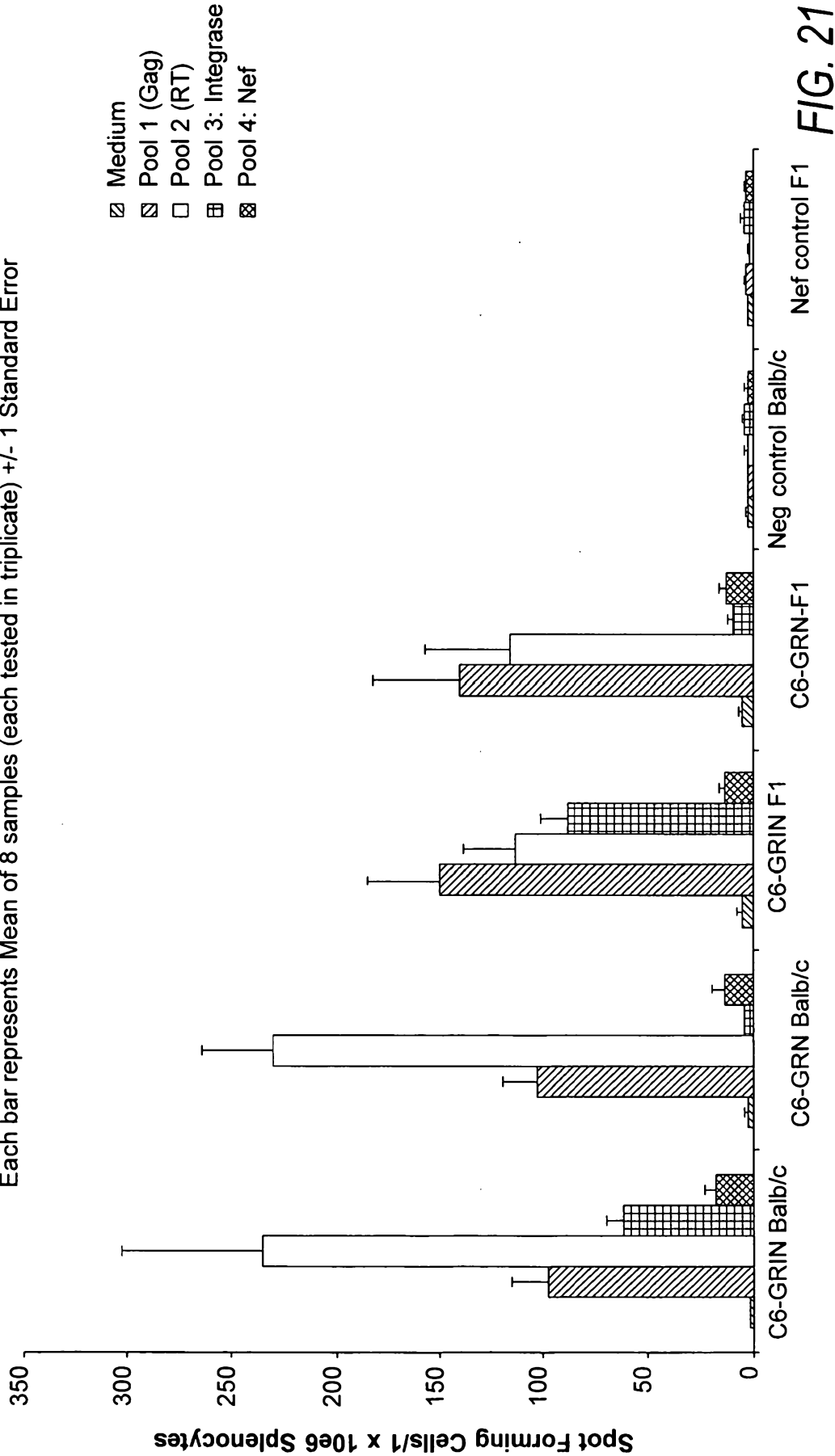


FIG. 21

FCC/05.003: Mean responses to HIV Clade A peptide library pools by IL-2 ELISpot at 10 days post-primary

Each bar represents Mean of 4 samples (each tested in triplicate) +/- Standard Error

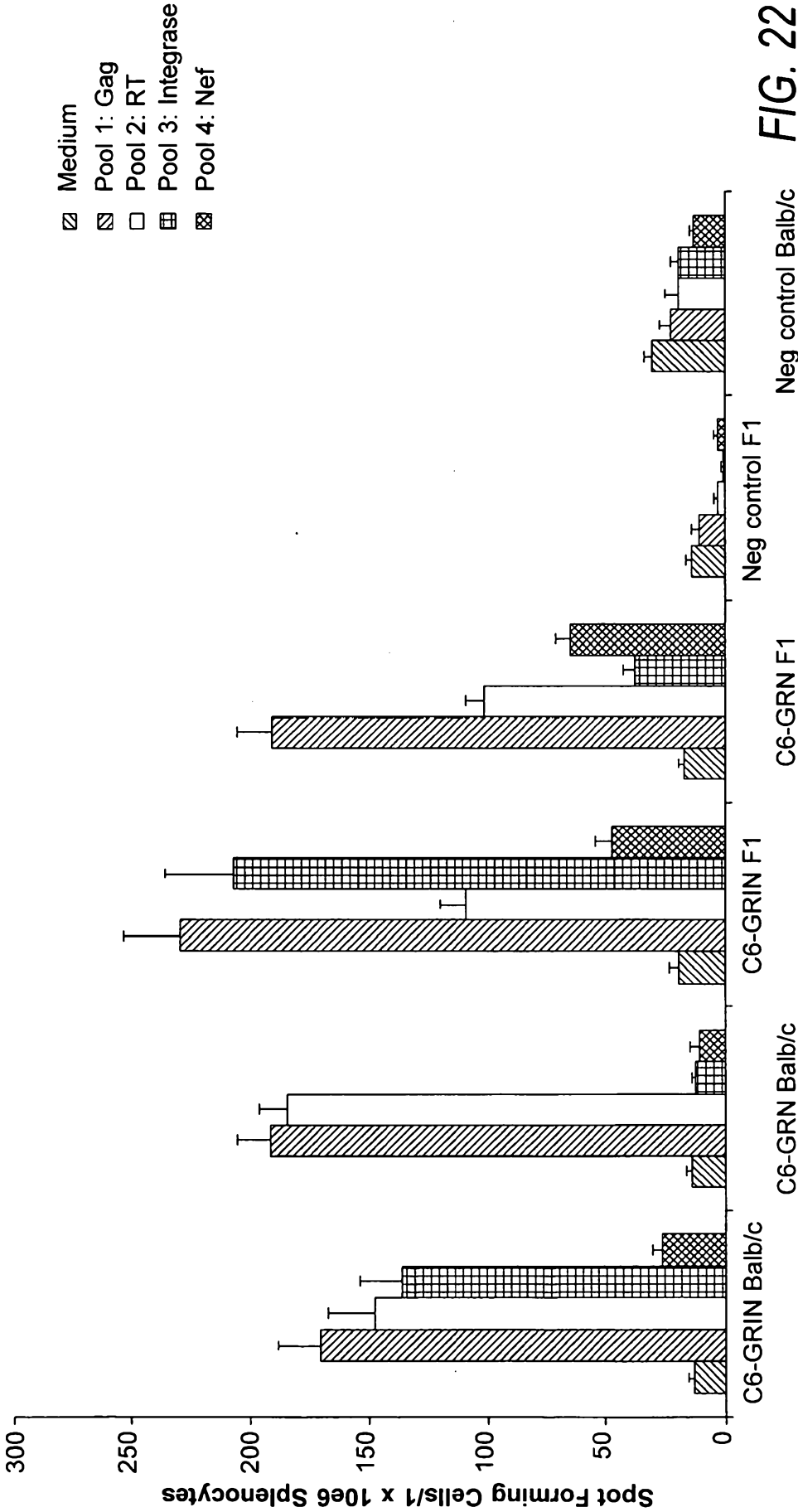


FIG. 22

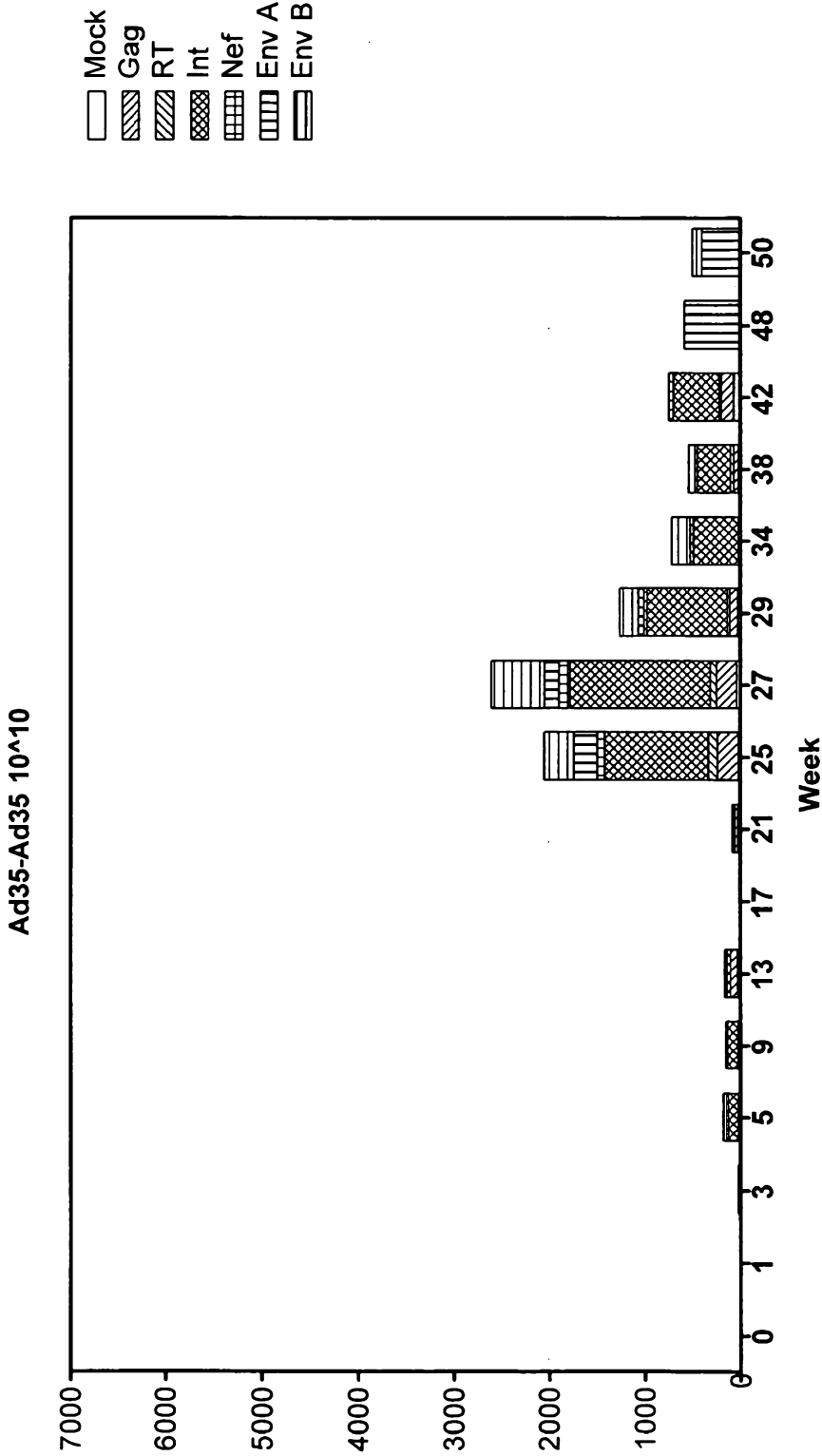


FIG. 23A

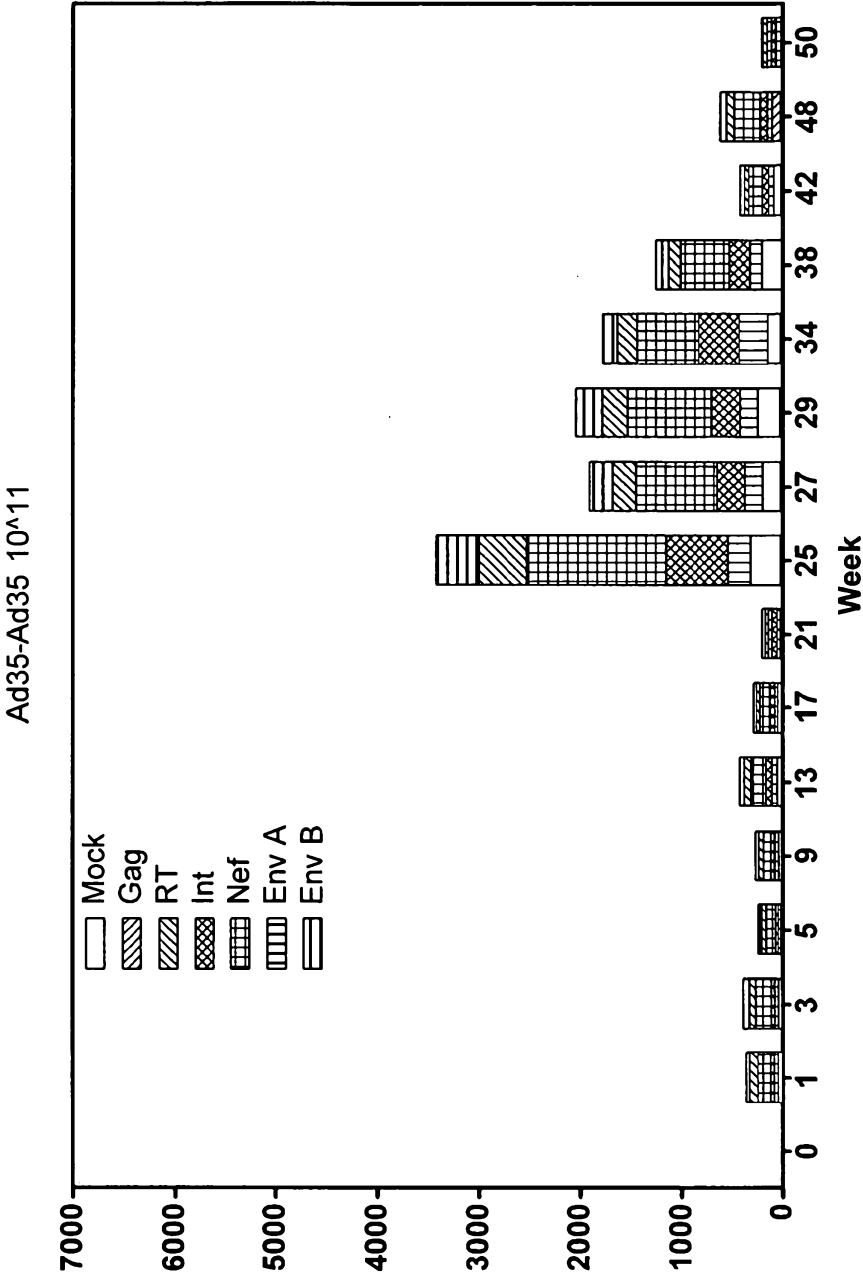


FIG. 23B

SEQUENCE LISTING

<110> International AIDS Vaccine Initiative
 <120> HIV-1 Clade A Consensus Sequences, Antigens, and Transgenes
 <130> 510500
 <140> AU2007256717
 <141> 2007-06-04
 <150> US60/810,816
 <151> 2006-06-02
 <160> 15
 <170> PatentIn version 3.5
 <210> 1
 <211> 4
 <212> PRT
 <213> Artificial Sequence
 <220>
 <223> peptide
 <400> 1
 Asn Phe Leu Gly
 1
 <210> 2
 <211> 499
 <212> PRT
 <213> Human immunodeficiency virus type 1
 <220>
 <221> MOD_RES
 <222> (124)..(124)
 <223> Variable amino acid
 <220>
 <221> MOD_RES
 <222> (146)..(146)
 <223> Variable amino acid
 <220>
 <221> MOD_RES
 <222> (339)..(339)
 <223> Variable amino acid
 <220>
 <221> MOD_RES
 <222> (481)..(482)
 <223> Variable amino acid
 <400> 2
 Met Gly Ala Arg Ala Ser Val Leu Ser Gly Gly Lys Leu Asp Ala Trp
 1 5 10 15
 Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys Lys Tyr Arg Leu Lys
 20 25 30
 His Leu Val Trp Ala Ser Arg Glu Leu Glu Arg Phe Ala Leu Asn Pro
 35 40 45
 Ser Leu Leu Glu Thr Ala Glu Gly Cys Gln Gln Ile Met Glu Gln Leu
 50 55 60
 Gln Pro Ala Leu Lys Thr Gly Thr Glu Glu Leu Arg Ser Leu Phe Asn
 65 70 75 80
 Thr Val Ala Thr Leu Tyr Cys Val His Gln Arg Ile Asp Val Lys Asp
 85 90 95
 Thr Lys Glu Ala Leu Asp Lys Ile Glu Glu Ile Gln Asn Lys Ser Lys
 100 105 110
 Gln Lys Thr Gln Gln Ala Ala Ala Asp Thr Gly Xaa Ser Ser Lys Val
 115 120 125
 Ser Gln Asn Tyr Pro Ile Val Gln Asn Ala Gln Gly Gln Met Ile His
 130 135 140
 Gln Xaa Leu Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Ile Glu
 145 150 155 160
 Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Ser Ala Leu Ser
 165 170 175

2007256717 29 Apr 2013

2007256717 29 Apr 2013

Glu Gly Ala Thr Pro Gln Asp Leu Asn Met Met Leu Asn Ile Val Gly
 180 185 190
 Gly His Gln Ala Ala Met Gln Met Leu Lys Asp Thr Ile Asn Glu Glu
 195 200 205
 Ala Ala Glu Trp Asp Arg Leu His Pro Val His Ala Gly Pro Ile Pro
 210 215 220
 Pro Gly Gln Met Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr
 225 230 235 240
 Ser Thr Pro Gln Glu Gln Gly Ala Trp Met Thr Gly Asn Pro Pro Ile
 245 250 255
 Pro Val Gly Asp Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys
 260 265 270
 Ile Val Arg Met Tyr Ser Pro Val Ser Ile Leu Asp Ile Lys Gln Gly
 275 280 285
 Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Phe Lys Thr Leu
 290 295 300
 Arg Ala Glu Gln Ala Thr Gln Glu Val Lys Gly Trp Met Thr Glu Thr
 305 310 315 320
 Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Ser Ile Leu Arg Ala
 325 330 335
 Leu Gly Xaa Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly
 340 345 350
 Val Gly Gly Pro Gly His Lys Ala Arg Val Leu Ala Glu Ala Met Ser
 355 360 365
 Gln Val Gln Gln Thr Asn Ile Met Met Gln Arg Gly Asn Phe Arg Gly
 370 375 380
 Gln Lys Arg Ile Lys Cys Phe Asn Cys Gly Lys Glu Gly His Leu Ala
 385 390 395 400
 Arg Asn Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys Gly Lys
 405 410 415
 Glu Gly His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn Phe Leu
 420 425 430
 Gly Lys Ile Trp Pro Ser Ser Lys Gly Arg Pro Gly Asn Phe Pro Gln
 435 440 445
 Ser Arg Pro Glu Pro Thr Ala Pro Pro Ala Glu Ile Phe Gly Met Gly
 450 455 460
 Glu Glu Ile Ala Ser Pro Pro Lys Gln Glu Gln Lys Asp Arg Glu Gln
 465 470 475 480
 Xaa Xaa Pro Pro Leu Val Ser Leu Lys Ser Leu Phe Gly Asn Asp Pro
 485 490 495
 Leu Ser Gln

<210> 3
 <211> 947
 <212> PRT
 <213> Human immunodeficiency virus type 1

<220>
 <221> MOD_RES
 <222> (344)..(344)
 <223> variable amino acid

<220>
 <221> MOD_RES
 <222> (534)..(534)
 <223> variable amino acid

<220>
 <221> MOD_RES
 <222> (582)..(582)
 <223> variable amino acid

<220>
 <221> MOD_RES
 <222> (680)..(680)
 <223> variable amino acid

<400> 3

2007256717 29 Apr 2013

Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly
 1 5 10 15
 Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
 20 25 30
 Leu Glu Asp Ile Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
 35 40 45
 Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Ile Leu Ile
 50 55 60
 Glu Ile Cys Gly Lys Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65 70 75 80
 Pro Val Asn Ile Ile Gly Arg Asn Met Leu Thr Gln Ile Gly Cys Thr
 85 90 95
 Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
 100 105 110
 Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
 115 120 125
 Lys Ile Lys Ala Leu Thr Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
 130 135 140
 Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
 145 150 155 160
 Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
 165 170 175
 Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
 180 185 190
 Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu
 195 200 205
 Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Ser Phe Arg
 210 215 220
 Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
 225 230 235 240
 Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
 245 250 255
 Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Ser
 260 265 270
 Lys Asn Pro Glu Ile Ile Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
 275 280 285
 Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
 290 295 300
 Arg Ala His Leu Leu Ser Trp Gly Phe Thr Thr Pro Asp Lys Lys His
 305 310 315 320
 Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
 325 330 335
 Lys Trp Thr Val Gln Pro Ile Xaa Leu Pro Glu Lys Glu Ser Trp Thr
 340 345 350
 Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
 355 360 365
 Ile Tyr Ala Gly Ile Lys Val Lys Gln Leu Cys Lys Leu Leu Arg Gly
 370 375 380
 Ala Lys Ala Leu Thr Asp Ile Val Thr Leu Thr Glu Glu Ala Glu Leu
 385 390 395 400
 Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Asp Pro Val His Gly Val
 405 410 415
 Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly
 420 425 430
 Gln Asp Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu
 435 440 445
 Lys Thr Gly Lys Tyr Ala Arg Lys Arg Ser Ala His Thr Asn Asp Val
 450 455 460
 Lys Gln Leu Ala Glu Val Val Gln Lys Val Val Met Glu Ser Ile Val
 465 470 475 480

2007256717 29 Apr 2013

Ile Trp Gly Lys Thr 485 Pro Lys Phe Lys Leu 490 Pro Ile Gln Lys Glu Thr 495
 Trp Glu Thr Trp 500 Trp Met Asp Tyr Trp 505 Gln Ala Thr Trp Ile 510 Pro Glu 510
 Trp Glu Phe Val Asn Thr Pro 520 Leu Val Lys Leu Trp 525 Tyr Gln Leu 525
 Glu Lys 530 Asp Pro Ile xaa 535 Ala Glu Thr Phe Tyr 540 Val Asp Gly Ala 540
 Ala 545 Asn Arg Glu Thr Lys 550 Leu Gly Lys Ala Gly 555 Tyr Val Thr Asp Arg 560
 Gly Arg Gln Lys Val 565 Val Ser Leu Thr Glu 570 Thr Thr Asn Gln Lys Thr 575
 Glu Leu His Ala 580 Ile xaa Leu Ala Leu 585 Gln Asp Ser Gly Ser 590 Glu Val 590
 Asn Ile 595 Val Thr Asp Ser Gln Tyr 600 Ala Leu Gly Ile Ile 605 Gln Ala Gln 605
 Pro Asp 610 Arg Ser Glu Ser Glu 615 Leu Val Asn Gln Ile Ile 620 Glu Lys Leu 620
 Ile Gly Lys Asp Lys Val 630 Tyr Leu Ser Trp Val 635 Pro Ala His Lys Gly 640
 Ile Gly Gly Asn Glu 645 Gln Val Asp Lys Leu Val 650 Ser Ser Gly Ile Arg 655
 Lys Val Leu Phe 660 Leu Asp Gly Ile Asp 665 Lys Ala Gln Glu Glu 670 His Glu 670
 Arg Tyr His 675 Ser Asn Trp Arg xaa 680 Met Ala Ser Asp Phe 685 Asn Leu Pro 685
 Pro Ile 690 Val Ala Lys Glu Ile 695 Val Ala Ser Cys Asp 700 Lys Cys Gln Leu 700
 Lys Gly Glu Ala Met His 710 Gly Gln Val Asp Cys 715 Ser Pro Gly Ile Trp 720
 Gln Leu Asp Cys Thr 725 His Leu Glu Gly Lys 730 Val Ile Leu Val Ala Val 735
 His Val Ala Ser 740 Gly Tyr Ile Glu Ala 745 Glu Val Ile Pro Ala Glu Thr 750
 Gly Gln Glu Thr Ala Tyr Phe Leu 760 Leu Lys Leu Ala Gly 765 Arg Trp Pro 765
 Val Lys 770 Val Val His Thr Asp 775 Asn Gly Ser Asn Phe 780 Thr Ser Ala Ala 780
 Phe Lys Ala Ala Cys Trp 790 Trp Ala Asn Ile Gln 795 Gln Glu Phe Gly Ile 800
 Pro Tyr Asn Pro Gln 805 Ser Gln Gly Val Val 810 Glu Ser Met Asn Lys Glu 815
 Leu Lys Lys Ile 820 Ile Gly Gln Val Arg 825 Glu Gln Ala Glu His 830 Leu Lys 830
 Thr Ala Val 835 Gln Met Ala Val Phe 840 Ile His Asn Phe Lys 845 Arg Lys Gly 845
 Gly Ile 850 Gly Gly Tyr Ser Ala 855 Gly Glu Arg Ile Ile 860 Asp Ile Ile Ala 860
 Thr Asp Ile Gln Thr Lys 870 Glu Leu Gln Lys Gln 875 Ile Thr Lys Ile Gln 880
 Asn Phe Arg Val Tyr 885 Tyr Arg Asp Ser Arg 890 Asp Pro Ile Trp Lys Gly 895
 Pro Ala Lys Leu 900 Leu Trp Lys Gly Glu 905 Gly Ala Val Val Ile Gln Asp 910
 Asn Ser Asp 915 Ile Lys Val Val Pro 920 Arg Arg Lys Ala Lys 925 Ile Ile Arg 925
 Asp Tyr 930 Gly Lys Gln Met Ala 935 Gly Asp Asp Cys Val 940 Ala Gly Arg Gln 940
 Asp Glu Asp 945

2007256717 29 Apr 2013

<210> 4
 <211> 875
 <212> PRT
 <213> Human immunodeficiency virus type 1

<220>
 <221> MOD_RES
 <222> (29)..(29)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (61)..(61)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (64)..(64)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (84)..(84)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (86)..(86)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (129)..(129)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (131)..(158)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (190)..(198)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (200)..(200)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (239)..(239)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (241)..(241)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (278)..(279)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (281)..(281)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (293)..(293)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (299)..(300)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (303)..(303)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (345)..(346)
 <223> Variable amino acid

2007256717 29 Apr 2013

<220>
 <221> MOD_RES
 <222> (349)..(349)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (353)..(353)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (356)..(356)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (360)..(362)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (364)..(364)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (371)..(372)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (405)..(415)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (417)..(418)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (426)..(426)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (439)..(439)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (469)..(476)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (631)..(632)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (656)..(656)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (820)..(820)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (834)..(834)
 <223> Variable amino acid

<400> 4

Met Arg Val Met Gly Ile Gln Arg Asn Cys Gln His Leu Leu Arg Trp
 1 5 10 15

Gly Thr Met Ile Leu Gly Met Ile Ile Cys Ser Xaa Ala Glu Asn
 20 25 30

Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Lys Asp Ala Glu
 35 40 45

Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Xaa Thr Glu Xaa
 50 55 60

His 65	Asn	Val	Trp	Ala	Thr 70	His	Ala	Cys	Val	Pro 75	Thr	Asp	Pro	Asn	Pro 80
Gln	Glu	Ile	Xaa	Leu 85	Xaa	Asn	Val	Thr	Glu 90	Glu	Phe	Asn	Met	Trp 95	Lys
Asn	Asp	Met	Val 100	Glu	Gln	Met	His	Thr 105	Asp	Ile	Ile	Ser	Leu 110	Trp	Asp
Gln	Ser	Leu 115	Lys	Pro	Cys	Val	Lys 120	Leu	Thr	Pro	Leu	Cys 125	Val	Thr	Leu
Xaa	Cys 130	Xaa	Xaa	Xaa	Xaa	Xaa 135	Xaa	Xaa	Xaa	Xaa	Xaa 140	Xaa	Xaa	Xaa	Xaa
Xaa 145	Xaa	Xaa	Xaa	Xaa	Xaa 150	Xaa	Xaa	Xaa	Xaa	Xaa 155	Xaa	Xaa	Xaa	Ile	Lys 160
Asn	Cys	Ser	Phe	Asn 165	Met	Thr	Thr	Glu	Leu 170	Arg	Asp	Lys	Lys	Gln 175	Lys
Val	Tyr	Ser	Leu 180	Phe	Tyr	Arg	Leu	Asp 185	Val	Val	Gln	Ile	Xaa 190	Xaa	Xaa
Xaa	Xaa	Xaa 195	Xaa	Xaa	Xaa	Ser	Xaa 200	Tyr	Arg	Leu	Ile	Asn 205	Cys	Asn	Thr
Ser	Ala 210	Ile	Thr	Gln	Ala	Cys 215	Pro	Lys	Val	Ser	Phe 220	Glu	Pro	Ile	Pro
Ile 225	His	Tyr	Cys	Ala	Pro 230	Ala	Gly	Phe	Ala	Leu 235	Ile	Lys	Cys	Xaa	Asp 240
Xaa	Glu	Phe	Asn	Gly 245	Thr	Gly	Pro	Cys	Lys 250	Asn	Val	Ser	Thr	Val 255	Gln
Cys	Thr	His	Gly 260	Ile	Lys	Pro	Val	Val 265	Ser	Thr	Gln	Leu	Leu 270	Leu	Asn
Gly	Ser	Leu 275	Ala	Glu	Xaa	Xaa	Val 280	Xaa	Ile	Arg	Glu	Ser 285	Asn	Ile	Thr
Asn	Asn 290	Ala	Lys	Xaa	Ile	Ile 295	Val	Gln	Leu	Xaa	Xaa 300	Pro	Val	Xaa	Ile
Asn 305	Cys	Thr	Arg	Pro	Asn 310	Asn	Asn	Thr	Arg	Lys 315	Ser	Ile	Arg	Ile	Gly 320
Pro	Gly	Gln	Ala	Lys 325	Tyr	Ala	Thr	Gly	Asp 330	Ile	Ile	Gly	Asp	Ile 335	Arg
Gln	Ala	His	Cys 340	Asn	Val	Ser	Arg	Xaa 345	Xaa	Trp	Asn	Xaa	Thr 350	Leu	Gln
Xaa	Val	Ala 355	Xaa	Gln	Leu	Arg	Xaa 360	Xaa	Xaa	Phe	Xaa	Asn 365	Lys	Thr	Ile
Ile	Phe 370	Xaa	Xaa	Ser	Ser	Gly 375	Gly	Asp	Leu	Glu	Ile 380	Thr	Thr	His	Ser
Phe 385	Asn	Cys	Gly	Gly	Glu 390	Phe	Phe	Tyr	Cys	Asn 395	Thr	Ser	Gly	Leu	Phe 400
Asn	Ser	Thr	Trp	Xaa 405	Xaa	Xaa	Xaa	Xaa	Xaa 410	Xaa	Xaa	Xaa	Xaa	Xaa 415	Ser
Xaa	Xaa	Ser	Asn 420	Asp	Thr	Ile	Thr	Leu 425	Xaa	Cys	Arg	Ile	Lys 430	Gln	Ile
Val	Asn	Met 435	Trp	Gln	Arg	Xaa	Gly 440	Gln	Ala	Met	Tyr	Ala 445	Pro	Pro	Ile
Gln	Gly 450	Val	Ile	Arg	Cys	Glu 455	Ser	Asn	Ile	Thr	Gly 460	Leu	Ile	Leu	Thr
Arg 465	Asp	Gly	Gly	Xaa	Xaa 470	Xaa	Xaa	Xaa	Xaa	Xaa 475	Xaa	Asn	Glu	Thr	Phe 480
Arg	Pro	Gly	Gly	Gly 485	Asp	Met	Arg	Asp	Asn 490	Trp	Arg	Ser	Glu	Leu 495	Tyr
Lys	Tyr	Lys	Val 500	Val	Lys	Ile	Glu	Pro 505	Leu	Gly	Val	Ala	Pro 510	Thr	Arg
Ala	Lys	Arg 515	Arg	Val	Val	Glu	Arg 520	Glu	Lys	Arg	Ala	Val 525	Gly	Ile	Gly
Ala	Val 530	Phe	Leu	Gly	Phe	Leu 535	Gly	Ala	Ala	Gly	Ser 540	Thr	Met	Gly	Ala

2007256717 29 Apr 2013

Ala Ser Ile Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile
 545 550 555 560
 Val Gln Gln Gln Ser Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln His
 565 570 575
 Leu Leu Lys Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Val
 580 585 590
 Leu Ala Val Glu Arg Tyr Leu Arg Asp Gln Gln Leu Leu Gly Ile Trp
 595 600 605
 Gly Cys Ser Gly Lys Leu Ile Cys Thr Thr Asn Val Pro Trp Asn Ser
 610 615 620
 Ser Trp Ser Asn Lys Ser Xaa Xaa Glu Ile Trp Asp Asn Met Thr Trp
 625 630 635 640
 Leu Gln Trp Asp Lys Glu Ile Ser Asn Tyr Thr Gln Ile Ile Tyr Xaa
 645 650 655
 Leu Ile Glu Glu Ser Gln Asn Gln Gln Glu Lys Asn Glu Gln Asp Leu
 660 665 670
 Leu Ala Leu Asp Lys Trp Ala Asn Leu Trp Asn Trp Phe Asp Ile Ser
 675 680 685
 Asn Trp Leu Tyr Trp Ile Lys Ile Phe Ile Met Ile Val Gly Gly Leu
 690 695 700
 Ile Gly Leu Arg Ile Val Phe Ala Val Leu Ser Ile Ile Asn Arg Val
 705 710 715 720
 Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr His Thr Pro Asn Pro
 725 730 735
 Arg Gly Leu Asp Arg Pro Gly Arg Ile Glu Glu Glu Gly Gly Glu Gln
 740 745 750
 Gly Arg Asp Arg Ser Ile Arg Leu Val Ser Gly Phe Leu Ala Leu Ala
 755 760 765
 Trp Asp Asp Leu Arg Ser Leu Cys Leu Phe Ser Tyr His Arg Leu Arg
 770 775 780
 Asp Phe Ile Leu Ile Ala Ala Arg Thr Val Glu Leu Leu Gly His Ser
 785 790 795 800
 Ser Leu Lys Gly Leu Arg Leu Gly Trp Glu Gly Leu Lys Tyr Leu Trp
 805 810 815
 Asn Leu Leu Xaa Tyr Trp Gly Arg Glu Leu Lys Ile Ser Ala Ile Asn
 820 825 830
 Leu Xaa Asp Thr Ile Ala Ile Ala Val Ala Gly Trp Thr Asp Arg Val
 835 840 845
 Ile Glu Ile Gly Gln Arg Ile Gly Arg Ala Ile Leu His Ile Pro Arg
 850 855 860
 Arg Ile Arg Gln Gly Leu Glu Arg Ala Leu Leu
 865 870 875

<210> 5
 <211> 203
 <212> PRT
 <213> Human immunodeficiency virus type 1

<220>
 <221> MOD_RES
 <222> (25)..(25)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (28)..(28)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (83)..(83)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (131)..(131)
 <223> Variable amino acid

<220>
 <221> MOD_RES

2007256717 29 Apr 2013

<222> (180)..(180)
 <223> variable amino acid

<220>
 <221> MOD_RES
 <222> (194)..(194)
 <223> variable amino acid

<400> 5

Met Gly Gly Lys Trp Ser Lys Ser Ser Ile Val Gly Trp Pro Glu Val
 1 5 10 15
 Arg Glu Arg Met Arg Arg Thr Pro Xaa Ala Ala Xaa Gly Val Gly Ala
 20 25 30
 Val Ser Gln Asp Leu Asp Lys His Gly Ala Ile Thr Ser Ser Asn Ile
 35 40 45
 Asn His Pro Ser Cys Val Trp Leu Glu Ala Gln Glu Glu Glu Val
 50 55 60
 Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys
 65 70 75 80
 Gly Ala Xaa Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Asp
 85 90 95
 Gly Leu Ile Tyr Ser Arg Lys Arg Gln Glu Ile Leu Asp Leu Trp Val
 100 105 110
 Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly
 115 120 125
 Pro Gly Xaa Arg Tyr Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val
 130 135 140
 Pro Val Asp Pro Asp Glu Val Glu Lys Ala Thr Glu Gly Glu Asn Asn
 145 150 155 160
 Ser Leu Leu His Pro Ile Cys Gln His Gly Met Asp Asp Glu Glu Arg
 165 170 175
 Glu Val Leu Xaa Trp Lys Phe Asp Ser Arg Leu Ala Leu Lys His Arg
 180 185 190
 Ala Xaa Glu Leu His Pro Glu Phe Tyr Lys Asp
 195 200

<210> 6
 <211> 498
 <212> PRT
 <213> Human immunodeficiency virus type 1

<400> 6

Met Gly Ala Arg Ala Ser Ile Leu Ser Gly Gly Lys Leu Asp Ala Trp
 1 5 10 15
 Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys Lys Tyr Arg Leu Lys
 20 25 30
 His Leu Val Trp Ala Ser Arg Glu Leu Asp Arg Phe Ala Leu Asn Pro
 35 40 45
 Ser Leu Leu Glu Thr Thr Glu Gly Cys Gln Gln Ile Met Asn Gln Leu
 50 55 60
 Gln Pro Ala Val Lys Thr Gly Thr Glu Glu Ile Lys Ser Leu Phe Asn
 65 70 75 80
 Thr Val Ala Thr Leu Tyr Cys Val His Gln Arg Ile Asp Val Lys Asp
 85 90 95
 Thr Lys Glu Ala Leu Asp Lys Ile Glu Glu Ile Gln Asn Lys Ser Lys
 100 105 110
 Gln Lys Thr Gln Gln Ala Ala Ala Asp Thr Gly Asp Ser Ser Lys Val
 115 120 125
 Ser Gln Asn Tyr Pro Ile Ile Gln Asn Ala Gln Gly Gln Met Ile His
 130 135 140
 Gln Asn Leu Ser Pro Arg Thr Leu Asn Ala Trp Val Lys Val Ile Glu
 145 150 155 160
 Glu Lys Ala Phe Ser Pro Glu Val Ile Pro Met Phe Ser Ala Leu Ser
 165 170 175
 Glu Gly Ala Thr Pro Gln Asp Leu Asn Val Met Leu Asn Ile Val Gly

2007256717 29 Apr 2013

180 185 190
 Gly His Gln Ala Ala Met Gln Met Leu Lys Asp Thr Ile Asn Glu Glu
 195 200 205
 Ala Ala Glu Trp Asp Arg Leu His Pro Val Gln Ala Gly Pro Ile Pro
 210 215 220
 Pro Gly Gln Ile Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr
 225 230 235 240
 Ser Thr Pro Gln Glu Gln Leu Gln Trp Met Thr Gly Asn Pro Pro Ile
 245 250 255
 Pro Val Gly Asn Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys
 260 265 270
 Ile Val Arg Met Tyr Ser Pro Val Ser Ile Leu Asp Ile Lys Gln Gly
 275 280 285
 Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Phe Lys Ala Leu
 290 295 300
 Arg Ala Glu Gln Ala Thr Gln Asp Val Lys Gly Trp Met Thr Glu Thr
 305 310 315 320
 Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Ser Ile Leu Lys Ala
 325 330 335
 Leu Gly Ser Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly
 340 345 350
 Val Gly Gly Pro Gly His Lys Ala Arg Val Leu Ala Glu Ala Met Ser
 355 360 365
 Gln Ala Gln Gln Thr Asn Ile Met Met Gln Arg Gly Asn Phe Arg Gly
 370 375 380
 Gln Lys Arg Ile Lys Cys Phe Asn Cys Gly Lys Glu Gly His Leu Ala
 385 390 395 400
 Arg Asn Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys Gly Lys
 405 410 415
 Glu Gly His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn Phe Leu
 420 425 430
 Gly Lys Ile Trp Pro Ser Ser Lys Gly Arg Pro Gly Asn Phe Pro Gln
 435 440 445
 Ser Arg Pro Glu Pro Thr Ala Pro Pro Ala Glu Leu Phe Gly Met Gly
 450 455 460
 Glu Gly Ile Ala Ser Leu Pro Lys Gln Glu Gln Lys Asp Arg Glu Gln
 465 470 475 480
 Val Pro Pro Leu Val Ser Leu Lys Ser Leu Phe Gly Asn Asp Pro Leu
 485 490 495
 Ser Gln

<210> 7
 <211> 947
 <212> PRT
 <213> Human immunodeficiency virus type 1

<400> 7

Pro Gln Ile Leu Thr Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly
 1 5 10 15
 Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
 20 25 30
 Leu Glu Asp Ile Asn Leu Pro Gly Lys Trp Lys Pro Arg Met Ile Gly
 35 40 45
 Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Ile Leu Ile
 50 55 60
 Glu Ile Cys Gly Lys Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
 65 70 75 80
 Pro Val Asn Ile Ile Gly Arg Asn Met Leu Thr Gln Ile Gly Cys Thr
 85 90 95
 Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Thr Leu Lys
 100 105 110

2007256717 29 Apr 2013

Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
 115 120 125
 Lys Ile Lys Ala Leu Thr Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
 130 135 140
 Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
 145 150 155 160
 Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
 165 170 175
 Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
 180 185 190
 Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu
 195 200 205
 Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asn Phe Arg
 210 215 220
 Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly
 225 230 235 240
 Val Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro
 245 250 255
 Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Ser
 260 265 270
 Lys Asn Pro Glu Ile Ile Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val
 275 280 285
 Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu
 290 295 300
 Arg Ala His Leu Leu Ser Trp Gly Phe Thr Thr Pro Asp Lys Lys His
 305 310 315 320
 Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp
 325 330 335
 Lys Trp Thr Val Gln Pro Ile Met Leu Pro Asp Lys Glu Ser Trp Thr
 340 345 350
 Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln
 355 360 365
 Ile Tyr Ala Gly Ile Lys Val Lys Gln Leu Cys Arg Leu Leu Arg Gly
 370 375 380
 Ala Lys Ala Leu Thr Asp Ile Val Thr Leu Thr Glu Glu Ala Glu Leu
 385 390 395 400
 Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Asp Pro Val His Gly Val
 405 410 415
 Tyr Tyr Asp Pro Ser Lys Asp Leu Val Ala Glu Ile Gln Lys Gln Gly
 420 425 430
 Gln Asp Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu
 435 440 445
 Lys Thr Gly Lys Tyr Ala Arg Lys Arg Ser Ala His Thr Asn Asp Val
 450 455 460
 Arg Gln Leu Ala Glu Val Val Gln Lys Val Ala Met Glu Ser Ile Val
 465 470 475 480
 Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr
 485 490 495
 Trp Glu Thr Trp Trp Met Asp Tyr Trp Gln Ala Thr Trp Ile Pro Glu
 500 505 510
 Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu
 515 520 525
 Glu Lys Asp Pro Ile Leu Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala
 530 535 540
 Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asp Arg
 545 550 555 560
 Gly Arg Gln Lys Val Val Ser Leu Thr Glu Thr Thr Asn Gln Lys Thr
 565 570 575
 Glu Leu His Ala Ile Leu Leu Ala Leu Gln Asp Ser Gly Ser Glu Val
 580 585 590

2007256717 29 Apr 2013

Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln
 595 600 605
 Pro Asp Arg Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Lys Leu
 610 615 620
 Ile Gly Lys Asp Lys Ile Tyr Leu Ser Trp Val Pro Ala His Lys Gly
 625 630 635 640
 Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser Ser Gly Ile Arg
 645 650 655
 Lys Val Leu Phe Leu Asp Gly Ile Asp Lys Ala Gln Glu Asp His Glu
 660 665 670
 Arg Tyr His Ser Asn Trp Arg Thr Met Ala Ser Asp Phe Asn Leu Pro
 675 680 685
 Pro Ile Val Ala Lys Glu Ile Val Ala Ser Cys Asp Lys Cys Gln Leu
 690 695 700
 Lys Gly Glu Ala Met His Gly Gln Val Asp Cys Ser Pro Gly Ile Trp
 705 710 715 720
 Gln Leu Asp Cys Thr His Leu Glu Gly Lys Val Ile Leu Val Ala Val
 725 730 735
 His Val Ala Ser Gly Tyr Ile Glu Ala Glu Val Ile Pro Ala Glu Thr
 740 745 750
 Gly Gln Glu Thr Ala Tyr Phe Leu Leu Lys Leu Ala Gly Arg Trp Pro
 755 760 765
 Val Lys Val Val His Thr Asp Asn Gly Ser Asn Phe Thr Ser Ala Ala
 770 775 780
 Val Lys Ala Ala Cys Trp Trp Ala Asn Ile Gln Gln Glu Phe Gly Ile
 785 790 795 800
 Pro Tyr Asn Pro Gln Ser Gln Gly Val Val Glu Ser Met Asn Lys Glu
 805 810 815
 Leu Lys Lys Ile Ile Gly Gln Val Arg Asp Gln Ala Glu His Leu Lys
 820 825 830
 Thr Ala Val Gln Met Ala Val Phe Ile His Asn Phe Lys Arg Lys Gly
 835 840 845
 Gly Ile Gly Gly Tyr Ser Ala Gly Glu Arg Ile Ile Asp Ile Ile Ala
 850 855 860
 Thr Asp Ile Gln Thr Lys Glu Leu Gln Lys Gln Ile Thr Lys Ile Gln
 865 870 875 880
 Asn Phe Arg Val Tyr Tyr Arg Asp Ser Arg Asp Pro Ile Trp Lys Gly
 885 890 895
 Pro Ala Lys Leu Leu Trp Lys Gly Glu Gly Ala Val Val Ile Gln Asp
 900 905 910
 Asn Ser Asp Ile Lys Val Val Pro Arg Arg Lys Ala Lys Ile Leu Arg
 915 920 925
 Asp Tyr Gly Lys Gln Met Ala Gly Asp Asp Cys Val Ala Gly Arg Gln
 930 935 940

Asp Glu Asp
 945

<210> 8
 <211> 204
 <212> PRT
 <213> Human immunodeficiency virus type 1

<400> 8

Met Gly Gly Lys Trp Ser Lys Gly Ser Ile Val Gly Trp Pro Glu Ile
 1 5 10 15
 Arg Glu Arg Met Arg Arg Ala Pro Ala Ala Ala Pro Gly Val Gly Ala
 20 25 30
 Val Ser Gln Asp Leu Asp Lys His Gly Ala Ile Thr Ser Ser Asn Ile
 35 40 45
 Asn Asn Pro Ser Cys Val Trp Leu Glu Ala Gln Glu Glu Glu Val
 50 55 60
 Gly Phe Pro Val Arg Pro Gln Val Pro Leu Arg Pro Met Thr Tyr Lys
 65 70 75 80

2007256717 29 Apr 2013

Gly Ala Phe Asp Leu Ser His Phe Leu Lys Glu Lys Gly Gly Leu Asp
 85 90 95
 Gly Leu Ile Tyr Ser Arg Lys Arg Gln Glu Ile Leu Asp Leu Trp Val
 100 105 110
 Tyr His Thr Gln Gly Tyr Phe Pro Asp Trp Gln Asn Tyr Thr Pro Gly
 115 120 125
 Pro Gly Val Arg Tyr Pro Leu Thr Phe Gly Trp Cys Phe Lys Leu Val
 130 135 140
 Pro Met Glu Pro Asp Glu Val Glu Lys Ala Thr Glu Gly Glu Asn Asn
 145 150 155 160
 Ser Leu Leu His Pro Ile Cys Gln His Gly Met Asp Asp Glu Glu Arg
 165 170 175
 Glu Val Leu Ile Trp Lys Phe Asp Ser Arg Leu Ala Leu Lys His Arg
 180 185 190
 Ala Gln Glu Leu His Pro Glu Phe Tyr Lys Asp Cys
 195 200

<210> 9
 <211> 672
 <212> PRT
 <213> Human immunodeficiency virus type 1
 <400> 9

Met Arg Val Met Glu Ile Gln Arg Asn Cys Gln His Leu Leu Arg Trp
 1 5 10 15
 Gly Ile Met Ile Leu Gly Met Ile Ile Ile Cys Ser Thr Ala Asp Asn
 20 25 30
 Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Arg Asp Ala Glu
 35 40 45
 Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Ser Thr Glu Lys
 50 55 60
 His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro
 65 70 75 80
 Gln Glu Ile Pro Leu Asp Asn Val Thr Glu Glu Phe Asn Met Trp Lys
 85 90 95
 Asn Asn Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp
 100 105 110
 Gln Ser Leu Lys Pro Cys Val Gln Leu Thr Pro Leu Cys Val Thr Leu
 115 120 125
 Asn Cys Ser Asn Ala Arg Val Asn Ala Thr Phe Asn Ser Thr Glu Asp
 130 135 140
 Arg Glu Gly Met Lys Asn Cys Ser Phe Asn Met Thr Thr Glu Leu Arg
 145 150 155 160
 Asp Lys Lys Gln Gln Val Tyr Ser Leu Phe Tyr Arg Leu Asp Ile Glu
 165 170 175
 Lys Ile Asn Ser Ser Asn Asn Asn Ser Glu Tyr Arg Leu Val Asn Cys
 180 185 190
 Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val Thr Phe Glu Pro
 195 200 205
 Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys
 210 215 220
 Asn Asp Thr Glu Phe Asn Gly Thr Gly Pro Cys Lys Asn Val Ser Thr
 225 230 235 240
 Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu
 245 250 255
 Leu Asn Gly Ser Leu Ala Glu Arg Glu Val Arg Ile Arg Ser Glu Asn
 260 265 270
 Ile Ala Asn Asn Ala Lys Asn Ile Ile Val Gln Phe Ala Ser Pro Val
 275 280 285
 Lys Ile Asn Cys Ile Arg Pro Asn Asn Asn Thr Arg Lys Ser Tyr Arg
 290 295 300
 Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Asp Ile Val Gly Asp Ile

2007256717 29 Apr 2013

305 310 315 320
 Arg Gln Ala His Cys Asn Val Ser Arg Thr Asp Trp Asn Asn Thr Leu
 325 330 335
 Arg Leu Val Ala Asn Gln Leu Arg Lys Tyr Phe Ser Asn Lys Thr Ile
 340 345 350
 Ile Phe Thr Asn Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His Ser
 355 360 365
 Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Thr Ser Gly Leu Phe
 370 375 380
 Asn Ser Thr Trp Thr Thr Asn Asn Met Gln Glu Ser Asn Asp Thr Ser
 385 390 395 400
 Asn Gly Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Arg Met
 405 410 415
 Trp Gln Arg Val Gly Gln Ala Met Tyr Ala Pro Pro Ile Glu Gly Val
 420 425 430
 Ile Arg Cys Glu Ser Asn Ile Thr Gly Leu Ile Leu Thr Arg Asp Gly
 435 440 445
 Gly Asn Asn Asn Ser Ala Asn Glu Thr Phe Arg Pro Gly Gly Gly Asp
 450 455 460
 Ile Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys
 465 470 475 480
 Ile Glu Pro Leu Gly Val Ala Pro Thr Arg Ala Lys Arg Arg Val Val
 485 490 495
 Glu Arg Glu Lys Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe
 500 505 510
 Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr
 515 520 525
 Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn
 530 535 540
 Leu Leu Arg Ala Ile Glu Ala Gln Gln Gln Leu Leu Lys Leu Thr Val
 545 550 555 560
 Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr
 565 570 575
 Leu Arg Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu
 580 585 590
 Ile Cys Thr Thr Asn Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser
 595 600 605
 Tyr Asp Asp Ile Trp Gln Asn Met Thr Trp Leu Gln Trp Asp Lys Glu
 610 615 620
 Ile Ser Asn Tyr Thr Asp Ile Ile Tyr Ser Leu Ile Glu Glu Ser Gln
 625 630 635 640
 Asn Gln Gln Glu Lys Asn Glu Gln Asp Leu Leu Ala Leu Asp Lys Trp
 645 650 655
 Ala Asn Leu Trp Asn Trp Phe Asp Ile Ser Lys Trp Leu Trp Tyr Ile
 660 665 670

<210> 10
 <211> 5775
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<400> 10
 agtcttctgt ttttacgtag gtgtcagcct agtggtgcaa tattggccat tagccatatt 60
 attcattggt tatatagcat aaatcaatat tggctattgg ccattgcata cgttgtatcc 120
 atatcataat atgtacattt atattggctc atgtccaaca ttaccgccat gttgacattg 180
 attattgact agttattaat agtaatcaat tacgggggtca ttagttcata gcccatatat 240
 ggagttccgc gttacataac ttacggtaaa tggcccgctt ggctgaccgc ccaacgaccc 300
 ccgcccattg acgtcaataa tgacgtatgt tcccatagta acgccaatag ggactttcca 360
 ttgacgtcaa tgggtggagt atttacggtta aactgcccac ttggcagtag atcaagtgtta 420

2007256717 29 Apr 2013

tcatatgcc	agtacgcccc	ctattgacgt	caatgacggt	aaatggcccc	cctggcatta	480
tgccagtag	atgaccttat	gggactttcc	tacttggcag	tacatctacg	tattagtcac	540
cgctattacc	atggtgatgc	ggttttggca	gtacatcaat	gggcgtggat	agcggtttga	600
ctcacgggga	tttccaagtc	tccaccccat	tgacgtcaat	gggagtttgt	tttggcacca	660
aaatcaacgg	gactttccaa	aatgtcgtaa	caactccgcc	ccattgacgc	aaatgggcgg	720
taggcgtgta	cggtagggag	tctatataag	cagagctcgt	ttagtgaacc	gtcagatcgc	780
ctggagacgc	catccacgct	gttttgacct	ccatagaaga	caccgggacc	gatccagcct	840
ccgcggccgg	gaacgggtga	ttggaagctt	gccgccacca	tggccgccag	agccagcatc	900
ctgagcgggg	gcaagctgga	cgcttgggag	aagatcagac	tgaggcctgg	cggaagaag	960
aagtaccggc	tgaagcacct	ggtgtgggcc	agcagagagc	tgatcgctt	cgccctgaat	1020
cctagcctgc	tggagaccac	cgagggtcgc	cagcagatca	tgaaccagct	gcagcccgcc	1080
gtgaaaaccg	gcaccgagga	gatcaagagc	ctgttcaaca	ccgtggccac	cctgtactgc	1140
gtgcaccagc	ggatcgacgt	gaaggatacc	aaggaggccc	tggacaagat	cgaggagatc	1200
cagaacaaga	gcaagcagaa	aaccagcag	gccgctgccg	acaccggcga	cagcagcaaa	1260
gtgagccaga	actaccccat	catccagaat	gcccagggcc	agatgatcca	ccagaacctg	1320
agccccagaa	ccctgaatgc	ctgggtgaaa	gtgatcgagg	aaaaggcctt	cagccccgaa	1380
gtgatcccta	tgctcagcgc	cctgagcgag	ggcgccaccc	cccaggacct	gaacgtgatg	1440
ctgaacattg	tgggcggaca	ccaggccgcc	atgcagatgc	tgaaggacac	catcaatgag	1500
gaggccgcgc	agtgggacag	actgcacccc	gtgcaggccg	gacccatccc	ccctggccag	1560
atcagagagc	ccagaggcag	cgacatcgcc	ggcaccacct	ccaccctca	agaacagctg	1620
cagtggatga	ccggcaaccc	tcccatccct	gtgggcaaca	tctacaagcg	gtggatcatc	1680
ctgggcctga	acaagattgt	gcggatgtac	agccccgtgt	ccatcctgga	tatcaagcag	1740
ggccccaagg	agcccttcag	agactacgtg	gaccggttct	tcaaggccct	gagagccgag	1800
caggccaccc	aggacgtgaa	gggctggatg	accgagaccc	tgctgggtga	gaacgccaac	1860
cccgactgca	agagcatcct	gaaggccctg	ggcagcggcg	ccacactgga	ggagatgatg	1920
accgcctgcc	agggagtggt	cggaaccggc	cacaaggcca	gagtgcctgg	cgaggccatg	1980
agccaggccc	agcagaccaa	catcatgatg	cagcggggca	acttcagagg	ccagaagcgg	2040
atcaagtgtc	tcaactgcgg	caaggagggc	cacctggcca	gaaactgcag	agcccccagg	2100
aagaagggct	gctggaagtg	tggcaaggaa	gggcaccaga	tgaaggactg	caccgagagg	2160
caggccaatt	tcctggggca	gatttggcct	agcagcaagg	gcagaccggg	caatttcccc	2220
cagagcagac	ccgagcccac	cgccctccc	gccgagctgt	tcggcatggg	cgagggcac	2280
gccagcctgc	ccaagcagga	gcagaaggac	agagagcagg	tgccccctct	ggtgtccctg	2340
aagtccctgt	tcggcaacga	tcctctgagc	cagggatccc	ccatcagccc	catcgagacc	2400
gtgcccgtga	ccctgaagcc	cgcatggat	ggccccaag	tgaaacagt	gcccctgacc	2460
gaggagaaga	ttaaggccct	gaccgaaatc	tgtaccgaga	tggagaagga	gggcaagatc	2520
agcaagatcg	gccccgagaa	cccctacaac	acccccatct	tcgccatcaa	gaagaaggac	2580
agcaccaagt	ggcggaaact	ggtggacttc	cgggagctga	acaagaggac	ccaggacttc	2640
tgggaagtgc	agctgggcat	ccccaccct	gccggcctga	agaagaagaa	gtccgtgaca	2700
gtgctggatg	tgggcgacgc	ctacttcagc	gtgcccctgg	acgagaactt	caggaagtac	2760
accgccttca	ccatccccag	caccaacaac	gagacccccg	gagtgcagata	ccagtacaac	2820
gtgctgcctc	agggctggaa	gggcagcccc	gccatcttcc	agagcagcat	gaccaagatc	2880
ctggagccct	tccggagcaa	gaaccccgag	atcatcatct	accagtacat	ggccgcccctg	2940
tatgtgggca	gcgatctgga	gatcgggcag	cacaggacca	agatcgaaga	gctgagggcc	3000
cacctgctga	gctggggcct	caccaccccc	gataagaagc	accagaagga	gccccctttc	3060
ctgtggatgg	gctacgagct	gcaccccgat	aagtggaccg	tgacgcccat	catgctgccc	3120

2007256717 29 Apr 2013

gataaggaga gctggaccgt gaacgacatc cagaaactgg tgggcaagct gaattgggcc	3180
agccaaatct acgccggcat taaagtgaag cagctgtgca ggctgctgag aggcgccaaa	3240
gccctgacag acatcgtgac actgacagag gagggcagc tggagctggc cgagaacagg	3300
gagatcctga aggaccccg gcacggcgtg tactacgacc ccagcaagga cctggtggcc	3360
gagattcaga agcagggccca ggaccagtgg acctaccaa tctaccagga gcctttcaag	3420
aacctgaaaa ccgggaagta cgccaggaag agaagcgccc acaccaacga tgtgaggcag	3480
ctggccgaag tgggtgcagaa agtggctatg gagagcatcg tgatctgggg caagacccc	3540
aagtccaagc tgcccatcca gaaggagacc tgggaaacct ggtgatgga ctactggcag	3600
gccacctgga ttctgtagt ggagttcgtg aacaccccc ctctggtgaa gctgtggtat	3660
cagctggaga aggaccccat cctgggcgcc gagaccttct acgtggacgg agccgccaat	3720
agagagacca agctgggcaa ggccggctac gtgaccgaca gaggcagaca gaaagtgggtg	3780
tctctgaccg agacaaccaa ccagaaaacc gagctgcacg ccacccctgct ggccctgcag	3840
gacagcggca gcgaagtga catcgtgacc gactcccagt acgccctggg catcattcag	3900
gccagcccg atagaagcga gagcgagctg gtgaaccaga tcatcgagaa gctgatcggc	3960
aaggacaaaa tctacctgag ctgggtgccc gccacaagg gcatcggcgg caacgagcag	4020
gtggacaagc tgggtgtccag cggcatccgg aaagtgtgt ttctggacgg catcgacaag	4080
gccaggagg accacgagag ataccacagc aactggcgga caatggccag cgacttcaac	4140
ctgcctccca tctgggccaa ggagatcgtg gccagctgcg ataagtgtca gctgaagggc	4200
gaggccatgc acggccaggt ggactgcagc cctggcatct ggcagctggc ctgcaccac	4260
ctggagggca aagtgattct ggtggcgtg cacgtggcca gcggctacat cgaggccgaa	4320
gtgattcccg ccgagaccgg ccaggagacc gcctacttcc tgctgaagct ggccggcaga	4380
tggcccgta aagtgggtga caccgccaac ggcagcaact tcacctctgc cgccgtgaag	4440
ggccctgtt ggtgggcaa tatccagcag gagttcggca tccccataa ccctcagagc	4500
cagggcgtgg tggccagcat gaacaaggag ctgaagaaga tcatcgcca ggtgaggac	4560
caggccgagc acctgaaaac agccgtgcag atggccgtgt tcatccaaa cttcaagcgg	4620
aaggcgggca ttggcggcta cagcgcgga gagcgatca tgcacatcat cgccaccgat	4680
atccagacca aggaactgca gaagcagatc accaagattc agaacttcag agtgtactac	4740
cgggacagca gggaccccat ctggaagggc cctgccaagc tgctgtggaa gggcgaaggc	4800
gccgtggtga tccaggacaa cagcgacatc aaagtgtgc cccggaggaa ggccaagatt	4860
ctgcgggact acggcaaaca gatggccggc gatgactgcg tggccggcag gcaggatgag	4920
gacagatcta tgggcggcaa gtggtccaag ggcagcattg tgggtggcc cgagatccgg	4980
gagagaatga gaagagcccc tgccgccgt cctggagtgg gcgccgtgtc tcaggatctg	5040
gataagcacg gcgccatcac cagcagcaac atcaacaacc ccagctgtgt gtggctggag	5100
gccaggaag aggaggaagt gggcttccct gtgagacccc aggtgcccct gagacccatg	5160
acctacaagg gcgccttcga cctgagccac ttcttgaagg agaagggcgg cctggacggc	5220
ctgatctaca gccggaagcg gcaggagatc ctggatctgt ggggtgtacca caccagggc	5280
tacttccccg actggcagaa ttacacccct ggccctggag tgcggtatcc cctgaccttc	5340
ggctgggtgct tcaagctggt gcctatggag ccgacgaag tggagaaggc cacagagggc	5400
gagaacaaca gcctgctgca ccctatctgc cagcacggca tggacgatga ggagcgggaa	5460
gtgctgatct ggaagtgcga cagcaggctg gccctgaagc acagagccca ggaactgcac	5520
ccagagttct acaaggactg ctgatgatca taataatcta gacgagatcc gaacttgttt	5580
attgcagctt ataatggtta caataaagc aatagcatca caaatttcac aaataaagca	5640
tttttttcac tgcatcttag ttgtggtttg tccaaactca tcaatgtatc ttatcatgtc	5700
tagatctgag gtatgatgat acgagatcga ggggtgcgcg atgcgaatgc ggaggcaagc	5760
atgccagggt ccagc	5775

2007256717 29 Apr 2013

<210> 11
 <211> 4071
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<400> 11
 attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt gaatgtattt 60
 agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcc cctgacgtct 120
 aagaaccat tattatcatg acattaacct ataaaaatag gcgtatcacg aggccctttc 180
 gtcttcaaga attggtcgat ggcaaacagc tattatgggt attatgggtt cgaattaatt 240
 aatcgacatc atcaataata taccttatag atggaatggg gccaatatgt aaatgaggtg 300
 attttaaaaa gtgtgggccg tgtggtgatt ggctgtgggg ttaacggtta aaagggcgcg 360
 cgcggccgtg ggaaaaatgac gttttatggg ggtggagttt ttttgcaagt tgcgcggga 420
 aatgttacgc ataaaaaggc ttcttttctc acggaactac ttagttttcc cacggtattt 480
 aacaggaaat gaggtagttt tgaccggatg caagtgaata ttgctgattt tcgcgcgaaa 540
 actgaatgag gaagtgtttt tctgaataat gtggtattta tggcaggggt gagtatttgt 600
 tcagggccag gtgacatttg acccattacg tggaggtttc gattaccgtg tttttacct 660
 gaatttccgc gtaccgtgtc aaagtcttct gtttttacgt aggtgtcagc ctagggtgtc 720
 aatattggcc attagccata ttattcattg gttatatagc ataatcaat attggctatt 780
 ggccattgca tacgttgtat ccatatcata atatgtacat ttatattggc tcatgtccaa 840
 cattaccgcc atgttgacat tgattattga ctagtatta atagtaatca attacggggt 900
 cattagtcca tagcccatat atggagttcc gcgttacata acttacggta aatggccgc 960
 ctggctgacc gcccaacgac ccccgcccat tgacgtcaat aatgacgtat gttcccatag 1020
 taacgccaat agggactttc cattgacgtc aatgggtgga gtatttacgg taaactgccc 1080
 acttggcagt acatcaagtg tatcatatgc caagtacgcc ccctattgac gtcaatgacg 1140
 gtaaattggc cgcttgcat tatgcccagt acatgacctt atgggacttt cctacttggc 1200
 agtacatcta cgtattagtc atcgctatta ccatggtgat gcggttttgg cagtacatca 1260
 atgggcgtgg atagcggttt gactcacggg gatttccaag tctccacccc attgacgtca 1320
 atgggagttt gttttggcac caaatcaac gggactttcc aaaatgtcgt aacaactccg 1380
 ccccatgac gcaaatgggc ggtaggcgtg tacggtggga ggtctatata agcagagctc 1440
 gtttagtgaa ccgtcagatc gcctggagac gccatccacg ctgttttgac ctccatagaa 1500
 gacaccggga ccgatccagc ctccgcggcc gggaacggtg cattggaagc ttgccgccac 1560
 catgaggggtg atggagatcc agcggaaactg ccagcacctg ctgagatggg gcatcatgat 1620
 cctgggcatg attatcatct gcagcaccgc cgacaacctg tgggtgaccg tgtactacgg 1680
 cgtgcctgtg tggagagatg ccgagaccac cctgttctgc gccagcgacg ccaaggccta 1740
 cagcaccgag aagcacaatg tgtgggccac ccacgcctgc gtgcctaccg atcccaacct 1800
 tcaggagatc cccctggaca acgtgaccga ggagttcaac atgtggaaga acaacatggt 1860
 ggaccagatg cacgaggaca tcatcagcct gtgggaccag agcctgaagc cctgcgtgca 1920
 gctgaccccc ctgtgcgtga ccctgaactg cagcaacgcc agagtgaagc ccaccttcaa 1980
 ctccaccgag gacagggagg gcatgaagaa ctgcagcttc aacatgacca ccgagctgcg 2040
 ggataagaag cagcaggtgt acagcctgtt ctaccggctg gacatcgaga agatcaacag 2100
 cagcaacaac aacagcgagt accggctggt gaactgcaat accagcgcca tcaccaggc 2160
 ctgcccctaag gtgaccttcg agcccatccc catccactac tgcgccccctg ccggcttcgc 2220
 catcctgaag tgcaacgaca ccgagttcaa tggcaccggc ccctgcaaga atgtgagcac 2280
 cgtgcagtgc acccaccgca tcaagccgtt ggtgtccacc cagctgctgc tgaacggcag 2340
 cctggccgag agagaagtgc ggatcaggag cgagaacatc gccacaacg ccaagaacat 2400

2007256717 29 Apr 2013

```

catcgtgcag ttcgccagcc cctgaagat caactgcac cggcccaaca acaatacccg 2460
gaagagctac agaatcggcc ctggccagac cttctacgcc accgacattg tgggcgacat 2520
cagacaggcc cactgcaacg tgtccaggac cgactggaac aacaccctga gactggtggc 2580
caaccagctg cggaagtact tcagcaacaa gaccatcatc ttcaccaaca gcagcggcgg 2640
agacctggag atcaccaccc acagcttcaa ttgtggcggc gagttcttct actgcaacac 2700
ctccggcctg ttcaatagca cctggaccac caacaacatg caggagtcca acgacaccag 2760
caacggcacc atcacctgc cctgcccggat caagcagatc atccggatgt ggcagcgcgt 2820
gggccaggcc atgtacgccc ctcccatcga gggcgtgatt cgctgcgaga gcaacatcac 2880
cggcctgac ctgaccagag atggcggcaa caacaattcc gccaacgaga ctttcagacc 2940
tggcggcgga gatatccggg acaactggcg gagcgagctg tacaagtaca aggtggtgaa 3000
gatcgagccc ctgggcgtgg cccccaccag agccaagaga agagtgtgtg agcgggagaa 3060
gagagccgtg ggcacgagg cctgttttct gggcttctct ggagccgccg gatctacaat 3120
gggagccgcc agcatcacc tgaccgtgca ggccagacag ctgctgagcg gcatcgtgca 3180
gcagcagagc aatctgctga gagccatcga ggcccagcag cagctgctga agctgacagt 3240
gtggggcatc aagcagctgc aggccagggt gctggccgtg gagagatacc tgaggggacca 3300
gcagctcctg ggcacgtggg gctgcagcgg caagctgatc tgcaccacca acgtgccctg 3360
gaatagcagc tggagcaaca agagctacga cgacatctgg cagaacatga cctggctgca 3420
gtgggacaag gagatcagca actacaccga catcatctac agcctgatcg aggagagcca 3480
gaaccagcag gagaagaacg agcaggatct gctggccctg gacaagtggg ccaacctgtg 3540
gaactgggtc gacatcagca agtggtgtg gtacatcaga tcttgataat ctagacgaga 3600
tccgaacttg tttattgcag cttataatgg ttacaataaa agcaatagca tcacaaattt 3660
cacaaataaa gcatTTTTTT cactgcattc tagttgtggt ttgtccaaac tcatcaatgt 3720
atcttatcat gtctagatct gaggtatgat gatacgagat cgagggtgcg cgcagcgaa 3780
tgcggaggca agcatgccag gttccagccg gtgtgtgtag atgtgaccga agatctcaga 3840
ccggatcatt tggttattgc ccgactgga gcagagtctg gatccagtgg agaagaaact 3900
gactaagggt agtattggga aaactttggg gtgggatttt cagatggaca gattgagtaa 3960
aaatttggtt tttctgtctt gcagctgaca tgactggaaa tgcttctttt aaggggggga 4020
gtcttcagcc cttatctgac agggcgtctc ccatcctggg caggagtctg t 4071

```

<210> 12

<211> 5002

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic construct

<220>

<221> CDS

<222> (16)..(4980)

<400> 12

```

aagcttgccg ccacc atg gcc gcc aga gcc agc atc ctg agc ggg ggc aag 51
                  Met Ala Ala Arg Ala Ser Ile Leu Ser Gly Gly Lys
                  1          5          10

ctg gac gcc tgg gag aag atc aga ctg agg cct ggc ggc aag aag aag 99
Leu Asp Ala Trp Glu Lys Ile Arg Leu Arg Pro Gly Gly Lys Lys Lys
15          20          25

tac cgg ctg aag cac ctg gtg tgg gcc agc aga gag ctg gat cgc ttc 147
Tyr Arg Leu Lys His Leu Val Trp Ala Ser Arg Glu Leu Asp Arg Phe
30          35          40

gcc ctg aat cct agc ctg ctg gag acc acc gag ggc tgc cag cag atc 195
Ala Leu Asn Pro Ser Leu Leu Glu Thr Thr Glu Gly Cys Gln Gln Ile
45          50          55          60

atg aac cag ctg cag ccc gcc gtg aaa acc ggc acc gag gag atc aag 243
Met Asn Gln Leu Gln Pro Ala Val Lys Thr Gly Thr Glu Glu Ile Lys
65          70          75

agc ctg ttc aac acc gtg gcc acc ctg tac tgc gtg cac cag cgg atc 291

```

2007256717 29 Apr 2013

Ser	Leu	Phe	Asn	Thr	Val	Ala	Thr	Leu	Tyr	Cys	Val	His	Gln	Arg	Ile	
			80					85					90			
gac	gtg	aag	gat	acc	aag	gag	gcc	ctg	gac	aag	atc	gag	gag	atc	cag	339
Asp	Val	Lys	Asp	Thr	Lys	Glu	Ala	Leu	Asp	Lys	Ile	Glu	Glu	Ile	Gln	
		95					100					105				
aac	aag	agc	aag	cag	aaa	acc	cag	cag	gcc	gct	gcc	gac	acc	ggc	gac	387
Asn	Lys	Ser	Lys	Gln	Lys	Thr	Gln	Gln	Ala	Ala	Ala	Asp	Thr	Gly	Asp	
	110					115					120					
agc	agc	aaa	gtg	agc	cag	aac	tac	ccc	atc	atc	cag	aat	gcc	cag	ggc	435
Ser	Ser	Lys	Val	Ser	Gln	Asn	Tyr	Pro	Ile	Ile	Gln	Asn	Ala	Gln	Gly	
125					130				135						140	
cag	atg	atc	cac	cag	aac	ctg	agc	ccc	aga	acc	ctg	aat	gcc	tgg	gtg	483
Gln	Met	Ile	His	Gln	Asn	Leu	Ser	Pro	Arg	Thr	Leu	Asn	Ala	Trp	Val	
				145					150					155		
aaa	gtg	atc	gag	gaa	aag	gcc	ttc	agc	ccc	gaa	gtg	atc	cct	atg	ttc	531
Lys	Val	Ile	Glu	Glu	Lys	Ala	Phe	Ser	Pro	Glu	Val	Ile	Pro	Met	Phe	
			160					165					170			
agc	gcc	ctg	agc	gag	ggc	gcc	acc	ccc	cag	gac	ctg	aac	gtg	atg	ctg	579
Ser	Ala	Leu	Ser	Glu	Gly	Ala	Thr	Pro	Gln	Asp	Leu	Asn	Val	Met	Leu	
		175					180					185				
aac	att	gtg	ggc	gga	cac	cag	gcc	gcc	atg	cag	atg	ctg	aag	gac	acc	627
Asn	Ile	Val	Gly	Gly	His	Gln	Ala	Ala	Met	Gln	Met	Leu	Lys	Asp	Thr	
		190				195					200					
atc	aat	gag	gag	gcc	gcc	gag	tgg	gac	aga	ctg	cac	ccc	gtg	cag	gcc	675
Ile	Asn	Glu	Glu	Ala	Ala	Glu	Trp	Asp	Arg	Leu	His	Pro	Val	Gln	Ala	
205					210					215					220	
gga	ccc	atc	ccc	cct	ggc	cag	atc	aga	gag	ccc	aga	ggc	agc	gac	atc	723
Gly	Pro	Ile	Pro	Pro	Gly	Gln	Ile	Arg	Glu	Pro	Arg	Gly	Ser	Asp	Ile	
				225					230					235		
gcc	ggc	acc	acc	tcc	acc	cct	caa	gaa	cag	ctg	cag	tgg	atg	acc	ggc	771
Ala	Gly	Thr	Thr	Ser	Thr	Pro	Gln	Glu	Gln	Leu	Gln	Trp	Met	Thr	Gly	
				240				245					250			
aac	cct	ccc	atc	cct	gtg	ggc	aac	atc	tac	aag	cgg	tgg	atc	atc	ctg	819
Asn	Pro	Pro	Ile	Pro	Val	Gly	Asn	Ile	Tyr	Lys	Arg	Trp	Ile	Ile	Leu	
		255					260					265				
ggc	ctg	aac	aag	att	gtg	cgg	atg	tac	agc	ccc	gtg	tcc	atc	ctg	gat	867
Gly	Leu	Asn	Lys	Ile	Val	Arg	Met	Tyr	Ser	Pro	Val	Ser	Ile	Leu	Asp	
		270				275					280					
atc	aag	cag	ggc	ccc	aag	gag	ccc	ttc	aga	gac	tac	gtg	gac	cgg	ttc	915
Ile	Lys	Gln	Gly	Pro	Lys	Glu	Pro	Phe	Arg	Asp	Tyr	Val	Asp	Arg	Phe	
285					290					295					300	
ttc	aag	gcc	ctg	aga	gcc	gag	cag	gcc	acc	cag	gac	gtg	aag	ggc	tgg	963
Phe	Lys	Ala	Leu	Arg	Ala	Glu	Gln	Ala	Thr	Gln	Asp	Val	Lys	Gly	Trp	
				305					310					315		
atg	acc	gag	acc	ctg	ctg	gtg	cag	aac	gcc	aac	ccc	gac	tgc	aag	agc	1011
Met	Thr	Glu	Thr	Leu	Leu	Val	Gln	Asn	Ala	Asn	Pro	Asp	Cys	Lys	Ser	
				320				325					330			
atc	ctg	aag	gcc	ctg	ggc	agc	ggc	gcc	aca	ctg	gag	gag	atg	atg	acc	1059
Ile	Leu	Lys	Ala	Leu	Gly	Ser	Gly	Ala	Thr	Leu	Glu	Glu	Met	Met	Thr	
		335					340					345				
gcc	tgc	cag	gga	gtg	ggc	gga	ccc	ggc	cac	aag	gcc	aga	gtg	ctg	gcc	1107
Ala	Cys	Gln	Gly	Val	Gly	Gly	Pro	Gly	His	Lys	Ala	Arg	Val	Leu	Ala	
		350				355					360					
gag	gcc	atg	agc	cag	gcc	cag	cag	acc	aac	atc	atg	atg	cag	cgg	ggc	1155
Glu	Ala	Met	Ser	Gln	Ala	Gln	Gln	Thr	Asn	Ile	Met	Met	Gln	Arg	Gly	
		365			370					375					380	
aac	ttc	aga	ggc	cag	aag	cgg	atc	aag	tgc	ttc	aac	tgc	ggc	aag	gag	1203
Asn	Phe	Arg	Gly	Gln	Lys	Arg	Ile	Lys	Cys	Phe	Asn	Cys	Gly	Lys	Glu	
				385					390					395		
ggc	cac	ctg	gcc	aga	aac	tgc	aga	gcc	ccc	agg	aag	aag	ggc	tgc	tgg	1251
Gly	His	Leu	Ala	Arg	Asn	Cys	Arg	Ala	Pro	Arg	Lys	Lys	Gly	Cys	Trp	
			400					405					410			
aag	tgt	ggc	aag	gaa	ggg	cac	cag	atg	aag	gac	tgc	acc	gag	agg	cag	1299
Lys	Cys	Gly	Lys	Glu	Gly	His	Gln	Met	Lys	Asp	Cys	Thr	Glu	Arg	Gln	
		415					420					425				
gcc	aat	ttc	ctg	ggc	aag	att	tgg	cct	agc	agc	aag	ggc	aga	ccc	ggc	1347
Ala	Asn	Phe	Leu	Gly	Lys	Ile	Trp	Pro	Ser	Ser	Lys	Gly	Arg	Pro	Gly	
		430				435						440				

2007256717 29 Apr 2013

aat ttc ccc cag agc aga ccc gag ccc acc gcc cct ccc gcc gag ctg Asn Phe Pro Gln Ser Arg Pro Glu Pro Thr Ala Pro Pro Ala Glu Leu 445 450	1395
ttc ggc atg ggc gag ggc atc gcc agc ctg ccc aag cag gag cag aag Phe Gly Met Gly Glu Gly Ile Ala Ser Leu Pro Lys Gln Glu Gln Lys 465 470 475	1443
gac aga gag cag gtg ccc ccc ctg gtg tcc ctg aag tcc ctg ttc ggc Asp Arg Glu Gln Val Pro Pro Leu Val Ser Leu Lys Ser Leu Phe Gly 480 485 490	1491
aac gat cct ctg agc cag gga tcc atg gcc ccc cag atc acc ctg tgg Asn Asp Pro Leu Ser Gln Gly Ser Met Ala Pro Gln Ile Thr Leu Trp 495 500 505	1539
cag aga ccc ctg gtg acc gtg aag atc ggc ggc cag ctg aag gaa gcc Gln Arg Pro Leu Val Thr Val Lys Ile Gly Gly Gln Leu Lys Glu Ala 510 515 520	1587
ctg ctg gat aca ggc gcc gat gat acc gtg ctg gag gac atc aac ctg Leu Leu Asp Thr Gly Ala Asp Asp Thr Val Leu Glu Asp Ile Asn Leu 525 530 535	1635
ccc ggc aag tgg aag cct aga atg atc ggc ggc atc ggc ggc ttc atc Pro Gly Lys Trp Lys Pro Arg Met Ile Gly Gly Ile Gly Gly Phe Ile 545 550 555	1683
aaa gtg aag cag tac gac cag atc ctg atc gag att tgc ggc aag aag Lys Val Lys Gln Tyr Asp Gln Ile Leu Ile Glu Ile Cys Gly Lys Lys 560 565 570	1731
gcc atc ggc acc gtg ctg gtg ggc ccc acc cct gtg aat atc atc ggc Ala Ile Gly Thr Val Leu Val Glu Pro Thr Pro Val Asn Ile Ile Gly 575 580 585	1779
cgg aac atg ctg acc cag atc ggc tgc acc ctg aac ttc ccc atc agc Arg Asn Met Leu Thr Gln Ile Gly Cys Thr Leu Asn Phe Pro Ile Ser 590 595 600	1827
ccc atc gag acc gtg ccc gtg acc ctg aag ccc ggc atg gat ggc ccc Pro Ile Glu Thr Val Pro Val Thr Leu Lys Pro Gly Met Asp Gly Pro 605 610 615 620	1875
aaa gtg aaa cag tgg ccc ctg acc gag gag aag att aag gcc ctg acc Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Thr 625 630 635	1923
gaa atc tgt acc gag atg gag aag gag ggc aag atc agc aag atc ggc Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly 640 645 650	1971
ccc gag aac ccc tac aac acc ccc atc ttc gcc atc aag aag aag gac Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala Ile Lys Lys Lys Asp 655 660 665	2019
agc acc aag tgg cgg aaa ctg gtg gac ttc cgg gag ctg aac aag agg Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg 670 675 680	2067
acc cag gac ttc tgg gaa gtg cag ctg ggc atc ccc cac cct gcc ggc Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly 685 690 695 700	2115
ctg aag aag aag aag tcc gtg aca gtg ctg gat gtg ggc gac gcc tac Leu Lys Lys Lys Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr 705 710 715	2163
ttc agc gtg ccc ctg gac gag aac ttc agg aag tac acc gcc ttc acc Phe Ser Val Pro Leu Asp Glu Asn Phe Arg Lys Tyr Thr Ala Phe Thr 720 725 730	2211
atc ccc agc acc aac aac gag acc ccc gga gtg aga tac cag tac aac Ile Pro Ser Thr Asn Asn Glu Thr Pro Gly Val Arg Tyr Gln Tyr Asn 735 740 745	2259
gtg ctg cct cag ggc tgg aag ggc agc ccc gcc atc ttc cag agc agc Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser 750 755 760	2307
atg acc aag atc ctg gag ccc ttc cgg agc aag aac ccc gag atc atc Met Thr Lys Ile Leu Glu Pro Phe Arg Ser Lys Asn Pro Glu Ile Ile 765 770 775 780	2355
atc tac cag tac atg gcc gcc ctg tat gtg ggc agc gat ctg gag atc Ile Tyr Gln Tyr Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile 785 790 795	2403
ggc cag cac agg acc aag atc gaa gag ctg agg gcc cac ctg ctg agc	2451

2007256717 29 Apr 2013

Gly Gln His Arg Thr Lys Ile Glu Glu Leu Arg Ala His Leu Leu Ser	
800 805 810	
tgg ggc ttc acc acc ccc gat aag aag cac cag aag gag ccc cct ttc	2499
Trp Gly Phe Thr Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe	
815 820 825	
ctg tgg atg ggc tac gag ctg cac ccc gat aag tgg acc gtg cag ccc	2547
Leu Trp Met Gly Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro	
830 835 840	
atc atg ctg ccc gat aag gag agc tgg acc gtg aac gac atc cag aaa	2595
Ile Met Leu Pro Asp Lys Glu Ser Trp Thr Val Asn Asp Ile Gln Lys	
845 850 855 860	
ctg gtg ggc aag ctg aat tgg gcc agc caa atc tac gcc ggc att aaa	2643
Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Ala Gly Ile Lys	
865 870 875	
gtg aag cag ctg tgc agg ctg ctg aga ggc gcc aaa gcc ctg aca gac	2691
Val Lys Gln Leu Cys Arg Leu Leu Arg Gly Ala Lys Ala Leu Thr Asp	
880 885 890	
atc gtg aca ctg aca gag gag gcc gag ctg gag ctg gcc gag aac agg	2739
Ile Val Thr Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg	
895 900 905	
gag atc ctg aag gac ccc gtg cac ggc gtg tac tac gac ccc agc aag	2787
Glu Ile Leu Lys Asp Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys	
910 915 920	
gac ctg gtg gcc gag att cag aag cag ggc cag gac cag tgg acc tac	2835
Asp Leu Val Ala Glu Ile Gln Lys Gln Gly Gln Asp Gln Trp Thr Tyr	
925 930 935 940	
caa atc tac cag gag cct ttc aag aac ctg aaa acc ggg aag tac gcc	2883
Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala	
945 950 955	
agg aag aga agc gcc cac acc aac gat gtg agg cag ctg gcc gaa gtg	2931
Arg Lys Arg Ser Ala His Thr Asn Asp Val Arg Gln Leu Ala Glu Val	
960 965 970	
gtg cag aaa gtg gct atg gag agc atc gtg atc tgg ggc aag acc ccc	2979
Val Gln Lys Val Ala Met Glu Ser Ile Val Ile Trp Gly Lys Thr Pro	
975 980 985	
aag ttc aag ctg ccc atc cag aag gag acc tgg gaa acc tgg tgg atg	3027
Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Met	
990 995 1000	
gac tac tgg cag gcc acc tgg att cct gag tgg gag ttc gtg aac	3072
Asp Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn	
1005 1010 1015	
acc ccc cct ctg gtg aag ctg tgg tat cag ctg gag aag gac ccc	3117
Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Lys Asp Pro	
1020 1025 1030	
atc ctg ggc gcc gag acc ttc tac gtg gac gga gcc gcc aat aga	3162
Ile Leu Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg	
1035 1040 1045	
gag acc aag ctg ggc aag gcc gcc tac gtg acc gag aga ggc aga	3207
Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asp Arg Gly Arg	
1050 1055 1060	
cag aaa gtg gtg tct ctg acc gag aca acc aac cag aaa acc gag	3252
Gln Lys Val Val Ser Leu Thr Glu Thr Thr Asn Gln Lys Thr Glu	
1065 1070 1075	
ctg cac gcc atc ctg ctg gcc ctg cag gac agc ggc agc gaa gtg	3297
Leu His Ala Ile Leu Leu Ala Leu Gln Asp Ser Gly Ser Glu Val	
1080 1085 1090	
aac atc gtg acc gac tcc cag tac gcc ctg ggc atc att cag gcc	3342
Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala	
1095 1100 1105	
cag ccc gat aga agc gag agc gag ctg gtg aac cag atc atc gag	3387
Gln Pro Asp Arg Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu	
1110 1115 1120	
aag ctg atc ggc aag gac aaa atc tac ctg agc tgg gtg ccc gcc	3432
Lys Leu Ile Gly Lys Asp Lys Ile Tyr Leu Ser Trp Val Pro Ala	
1125 1130 1135	
cac aag ggc atc ggc ggc aac gag cag gtg gac aag ctg gtg tcc	3477
His Lys Gly Ile Gly Gly Asn Glu Gln Val Asp Lys Leu Val Ser	
1140 1145 1150	

2007256717 29 Apr 2013

agc Ser 1155	ggc Gly Ile	atc Arg	cgg Lys	aaa Lys	gtg Val 1160	ctg Leu	ttt Phe	ctg Leu	gac Asp	ggc Gly 1165	atc Ile	gac Asp	aag Lys	gcc Ala	3522
cag Gln 1170	gag Glu	gac Asp	cac His	gag Glu	aga Arg 1175	tac Tyr	cac His	agc Ser	aac Asn	tgg Trp 1180	cgg Arg	aca Thr	atg Met	gcc Ala	3567
agc Ser 1185	gac Asp	ttc Phe	aac Asn	ctg Leu	cct Pro 1190	ccc Pro	atc Ile	gtg Val	gcc Ala	aag Lys 1195	gag Glu	atc Ile	gtg Val	gcc Ala	3612
agc Ser 1200	tgc Cys	gat Asp	aag Lys	tgt Cys	cag Gln 1205	ctg Leu	aag Lys	ggc Gly	gag Glu	gcc Ala 1210	atg Met	cac His	ggc Gly	cag Gln	3657
gtg Val 1215	gac Asp	tgc Cys	agc Ser	cct Pro	ggc Gly 1220	atc Ile	tgg Trp	cag Gln	ctg Leu	gcc Ala 1225	tgc Cys	acc Thr	cac His	ctg Leu	3702
gag Glu 1230	ggc Gly	aaa Lys	gtg Val	att Ile	ctg Leu 1235	gtg Val	gcc Ala	gtg Val	cac His	gtg Val 1240	gcc Ala	agc Ser	ggc Gly	tac Tyr	3747
atc Ile 1245	gag Glu	gcc Ala	gaa Glu	gtg Val	att Ile 1250	ccc Pro	gcc Ala	gag Glu	acc Thr	ggc Gly 1255	cag Gln	gag Glu	acc Thr	gcc Ala	3792
tac Tyr 1260	ttc Phe	ctg Leu	ctg Leu	aag Lys	ctg Leu 1265	gcc Ala	ggc Gly	aga Arg	tgg Trp	ccc Pro 1270	gtg Val	aaa Lys	gtg Val	gtg Val	3837
cac His 1275	acc Thr	gcc Ala	aac Asn	ggc Gly	agc Ser 1280	aac Asn	ttc Phe	acc Thr	tct Ser	gcc Ala 1285	gcc Ala	gtg Val	aag Lys	gcc Ala	3882
gcc Ala 1290	tgt Cys	tgg Trp	tgg Trp	gcc Ala	aat Asn 1295	atc Ile	cag Gln	cag Gln	gag Glu	ttc Phe 1300	ggc Gly	atc Ile	ccc Pro	tac Tyr	3927
aac Asn 1305	cct Pro	cag Gln	agc Ser	cag Gln	ggc Gly 1310	gtg Val	gtg Val	gcc Ala	agc Ser	atg Met 1315	aac Asn	aag Lys	gag Glu	ctg Leu	3972
aag Lys 1320	aag Lys	atc Ile	atc Ile	ggc Gly	cag Gln 1325	gtg Val	agg Arg	gac Asp	cag Gln	gcc Ala 1330	gag Glu	cac His	ctg Leu	aaa Lys	4017
aca Thr 1335	gcc Ala	gtg Val	cag Gln	atg Met	gcc Ala 1340	gtg Val	ttc Phe	atc Ile	cac His	aac Asn 1345	ttc Phe	aag Lys	cgg Arg	aag Lys	4062
ggc Gly 1350	ggc Gly	att Ile	ggc Gly	ggc Gly	tac Tyr 1355	agc Ser	gcc Ala	gga Gly	gag Glu	cgg Arg 1360	atc Ile	atc Ile	gac Asp	atc Ile	4107
atc Ile 1365	gcc Ala	acc Thr	gat Asp	atc Ile	cag Gln 1370	acc Thr	aag Lys	gaa Glu	ctg Leu	cag Gln 1375	aag Lys	cag Gln	atc Ile	acc Thr	4152
aag Lys 1380	att Ile	cag Gln	aac Asn	ttc Phe	aga Arg 1385	gtg Val	tac Tyr	tac Tyr	cgg Arg	gac Asp 1390	agc Ser	agg Arg	gac Asp	ccc Pro	4197
atc Ile 1395	tgg Trp	aag Lys	ggc Gly	cct Pro	gcc Ala 1400	aag Lys	ctg Leu	ctg Leu	tgg Trp	aag Lys 1405	ggc Gly	gaa Glu	ggc Gly	gcc Ala	4242
gtg Val 1410	gtg Val	atc Ile	cag Gln	gac Asp	aac Asn 1415	agc Ser	gac Asp	atc Ile	aaa Lys	gtg Val 1420	gtg Val	ccc Pro	cgg Arg	agg Arg	4287
aag Lys 1425	gcc Ala	aag Lys	att Ile	ctg Leu	cgg Arg 1430	gac Asp	tac Tyr	ggc Gly	aaa Lys	cag Gln 1435	atg Met	gcc Ala	ggc Gly	gat Asp	4332
gac Asp 1440	tgc Cys	gtg Val	gcc Ala	ggc Gly	agg Arg 1445	cag Gln	gat Asp	gag Glu	gac Asp	aga Arg 1450	tct Ser	atg Met	ggc Gly	ggc Gly	4377
aag Lys 1455	tgg Trp	tcc Ser	aag Lys	ggc Gly	agc Ser 1460	att Ile	gtg Val	ggc Gly	tgg Trp	ccc Pro 1465	gag Glu	atc Ile	cgg Arg	gag Glu	4422
aga Arg 1470	atg Met	aga Arg	aga Arg	gcc Ala	cct Pro 1475	gcc Ala	gcc Ala	gct Ala	cct Pro	gga Gly 1480	gtg Val	ggc Gly	gcc Ala	gtg Val	4467
tct	cag	gat	ctg	gat	aag	cac	ggc	gcc	atc	acc	agc	agc	aac	atc	4512

2007256717 29 Apr 2013

Ser 1485	Gln	Asp	Leu	Asp	Lys 1490	His	Gly	Ala	Ile	Thr 1495	Ser	Ser	Asn	Ile										
aac Asn 1500	aac Asn	ccc Pro	agc Ser	tgt Cys	gtg Val 1505	tgg Trp	ctg Leu	gag Glu	gcc Ala	cag Gln 1510	gaa Glu	gag Glu	gag Glu	gaa Glu	4557									
gtg Val 1515	ggc Gly	ttc Phe	cct Pro	gtg Val	aga Arg 1520	ccc Pro	cag Gln	gtg Val	ccc Pro	ctg Leu 1525	aga Arg	ccc Pro	atg Met	acc Thr	4602									
tac Tyr 1530	aag Lys	ggc Gly	gcc Ala	ttc Phe	gac Asp 1535	ctg Leu	agc Ser	cac His	ttc Phe	ctg Leu 1540	aag Lys	gag Glu	aag Lys	ggc Gly	4647									
ggc Gly 1545	ctg Leu	gac Asp	ggc Gly	ctg Leu	atc Ile 1550	tac Tyr	agc Ser	cgg Arg	aag Lys	cgg Arg 1555	cag Gln	gag Glu	atc Ile	ctg Leu	4692									
gat Asp 1560	ctg Leu	tgg Trp	gtg Val	tac Tyr	cac His 1565	acc Thr	cag Gln	ggc Gly	tac Tyr	ttc Phe 1570	ccc Pro	gac Asp	tgg Trp	cag Gln	4737									
aat Asn 1575	tac Tyr	acc Thr	cct Pro	ggc Gly	cct Pro 1580	gga Gly	gtg Val	cgg Arg	tat Tyr	ccc Pro 1585	ctg Leu	acc Thr	ttc Phe	ggc Gly	4782									
tgg Trp 1590	tgc Cys	ttc Phe	aag Lys	ctg Leu	gtg Val 1595	cct Pro	atg Met	gag Glu	ccc Pro	gac Asp 1600	gaa Glu	gtg Val	gag Glu	aag Lys	4827									
gcc Ala 1605	aca Thr	gag Glu	ggc Gly	gag Glu	aac Asn 1610	aac Asn	agc Ser	ctg Leu	ctg Leu	cac His 1615	cct Pro	atc Ile	tgc Cys	cag Gln	4872									
cac His 1620	ggc Gly	atg Met	gac Asp	gat Asp	gag Glu 1625	gag Glu	cgg Arg	gaa Glu	gtg Val	ctg Leu 1630	atc Ile	tgg Trp	aag Lys	ttc Phe	4917									
gac Asp 1635	agc Ser	agg Arg	ctg Leu	gcc Ala	ctg Leu 1640	aag Lys	cac His	aga Arg	gcc Ala	cag Gln 1645	gaa Glu	ctg Leu	cac His	cca Pro	4962									
gag Glu 1650	ttc Phe	tac Tyr	aag Lys	gac Asp	tgc Cys 1655	tgatgatcat aataatctag aa									5002									
<210> 13																								
<211> 1655																								
<212> PRT																								
<213> Artificial Sequence																								
<220>																								
<223> Synthetic Construct																								
<400> 13																								
Met	Ala	Ala	Arg	Ala	Ser	Ile	Leu	Ser	Gly	Gly	Lys	Leu	Asp	Ala	Trp									
1			5						10					15										
Glu	Lys	Ile	Arg	Leu	Arg	Pro	Gly	Gly	Lys	Lys	Lys	Tyr	Arg	Leu	Lys									
		20					25						30											
His	Leu	Val	Trp	Ala	Ser	Arg	Glu	Leu	Asp	Arg	Phe	Ala	Leu	Asn	Pro									
	35					40					45													
Ser	Leu	Leu	Glu	Thr	Thr	Glu	Gly	Cys	Gln	Gln	Ile	Met	Asn	Gln	Leu									
	50				55						60													
Gln	Pro	Ala	Val	Lys	Thr	Gly	Thr	Glu	Glu	Ile	Lys	Ser	Leu	Phe	Asn									
65				70				75						80										
Thr	Val	Ala	Thr	Leu	Tyr	Cys	Val	His	Gln	Arg	Ile	Asp	Val	Lys	Asp									
			85				90						95											
Thr	Lys	Glu	Ala	Leu	Asp	Lys	Ile	Glu	Glu	Ile	Gln	Asn	Lys	Ser	Lys									
		100					105						110											
Gln	Lys	Thr	Gln	Gln	Ala	Ala	Ala	Asp	Thr	Gly	Asp	Ser	Ser	Lys	Val									
	115				120						125													
Ser	Gln	Asn	Tyr	Pro	Ile	Ile	Gln	Asn	Ala	Gln	Gly	Gln	Met	Ile	His									
130					135						140													
Gln	Asn	Leu	Ser	Pro	Arg	Thr	Leu	Asn	Ala	Trp	Val	Lys	Val	Ile	Glu									
145				150						155				160										
Glu	Lys	Ala	Phe	Ser	Pro	Glu	Val	Ile	Pro	Met	Phe	Ser	Ala	Leu	Ser									
			165					170						175										

2007256717 29 Apr 2013

Glu Gly Ala Thr Pro Gln Asp Leu Asn Val Met Leu Asn Ile Val Gly
 180 185 190
 Gly His Gln Ala Ala Met Gln Met Leu Lys Asp Thr Ile Asn Glu Glu
 195 200 205
 Ala Ala Glu Trp Asp Arg Leu His Pro Val Gln Ala Gly Pro Ile Pro
 210 215 220
 Pro Gly Gln Ile Arg Glu Pro Arg Gly Ser Asp Ile Ala Gly Thr Thr
 225 230 235 240
 Ser Thr Pro Gln Glu Gln Leu Gln Trp Met Thr Gly Asn Pro Pro Ile
 245 250 255
 Pro Val Gly Asn Ile Tyr Lys Arg Trp Ile Ile Leu Gly Leu Asn Lys
 260 265 270
 Ile Val Arg Met Tyr Ser Pro Val Ser Ile Leu Asp Ile Lys Gln Gly
 275 280 285
 Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg Phe Phe Lys Ala Leu
 290 295 300
 Arg Ala Glu Gln Ala Thr Gln Asp Val Lys Gly Trp Met Thr Glu Thr
 305 310 315 320
 Leu Leu Val Gln Asn Ala Asn Pro Asp Cys Lys Ser Ile Leu Lys Ala
 325 330 335
 Leu Gly Ser Gly Ala Thr Leu Glu Glu Met Met Thr Ala Cys Gln Gly
 340 345 350
 Val Gly Gly Pro Gly His Lys Ala Arg Val Leu Ala Glu Ala Met Ser
 355 360 365
 Gln Ala Gln Gln Thr Asn Ile Met Met Gln Arg Gly Asn Phe Arg Gly
 370 375 380
 Gln Lys Arg Ile Lys Cys Phe Asn Cys Gly Lys Glu Gly His Leu Ala
 385 390 395 400
 Arg Asn Cys Arg Ala Pro Arg Lys Lys Gly Cys Trp Lys Cys Gly Lys
 405 410 415
 Glu Gly His Gln Met Lys Asp Cys Thr Glu Arg Gln Ala Asn Phe Leu
 420 425 430
 Gly Lys Ile Trp Pro Ser Ser Lys Gly Arg Pro Gly Asn Phe Pro Gln
 435 440 445
 Ser Arg Pro Glu Pro Thr Ala Pro Pro Ala Glu Leu Phe Gly Met Gly
 450 455 460
 Glu Gly Ile Ala Ser Leu Pro Lys Gln Glu Gln Lys Asp Arg Glu Gln
 465 470 475 480
 Val Pro Pro Leu Val Ser Leu Lys Ser Leu Phe Gly Asn Asp Pro Leu
 485 490 495
 Ser Gln Gly Ser Met Ala Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu
 500 505 510
 Val Thr Val Lys Ile Gly Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr
 515 520 525
 Gly Ala Asp Asp Thr Val Leu Glu Asp Ile Asn Leu Pro Gly Lys Trp
 530 535 540
 Lys Pro Arg Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val Lys Gln
 545 550 555 560
 Tyr Asp Gln Ile Leu Ile Glu Ile Cys Gly Lys Lys Ala Ile Gly Thr
 565 570 575
 Val Leu Val Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn Met Leu
 580 585 590
 Thr Gln Ile Gly Cys Thr Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr
 595 600 605
 Val Pro Val Thr Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln
 610 615 620
 Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Thr Glu Ile Cys Thr
 625 630 635 640
 Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro
 645 650 655

2007256717 29 Apr 2013

Tyr Asn Thr Pro Ile Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp
 660 665 670
 Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe
 675 680 685
 Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys
 690 695 700
 Lys Ser Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro
 705 710 715 720
 Leu Asp Glu Asn Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Thr
 725 730 735
 Asn Asn Glu Thr Pro Gly Val Arg Tyr Gln Tyr Asn Val Leu Pro Gln
 740 745 750
 Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Lys Ile
 755 760 765
 Leu Glu Pro Phe Arg Ser Lys Asn Pro Glu Ile Ile Ile Tyr Gln Tyr
 770 775 780
 Met Ala Ala Leu Tyr Val Gly Ser Asp Leu Glu Ile Gly Gln His Arg
 785 790 795 800
 Thr Lys Ile Glu Glu Leu Arg Ala His Leu Leu Ser Trp Gly Phe Thr
 805 810 815
 Thr Pro Asp Lys Lys His Gln Lys Glu Pro Pro Phe Leu Trp Met Gly
 820 825 830
 Tyr Glu Leu His Pro Asp Lys Trp Thr Val Gln Pro Ile Met Leu Pro
 835 840 845
 Asp Lys Glu Ser Trp Thr Val Asn Asp Ile Gln Lys Leu Val Gly Lys
 850 855 860
 Leu Asn Trp Ala Ser Gln Ile Tyr Ala Gly Ile Lys Val Lys Gln Leu
 865 870 875 880
 Cys Arg Leu Leu Arg Gly Ala Lys Ala Leu Thr Asp Ile Val Thr Leu
 885 890 895
 Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys
 900 905 910
 Asp Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Val Ala
 915 920 925
 Glu Ile Gln Lys Gln Gly Gln Asp Gln Trp Thr Tyr Gln Ile Tyr Gln
 930 935 940
 Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Lys Arg Ser
 945 950 955 960
 Ala His Thr Asn Asp Val Arg Gln Leu Ala Glu Val Val Gln Lys Val
 965 970 975
 Ala Met Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu
 980 985 990
 Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Met Asp Tyr Trp Gln
 995 1000 1005
 Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu
 1010 1015 1020
 Val Lys Leu Trp Tyr Gln Leu Glu Lys Asp Pro Ile Leu Gly Ala
 1025 1030 1035
 Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu
 1040 1045 1050
 Gly Lys Ala Gly Tyr Val Thr Asp Arg Gly Arg Gln Lys Val Val
 1055 1060 1065
 Ser Leu Thr Glu Thr Thr Asn Gln Lys Thr Glu Leu His Ala Ile
 1070 1075 1080
 Leu Leu Ala Leu Gln Asp Ser Gly Ser Glu Val Asn Ile Val Thr
 1085 1090 1095
 Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Arg
 1100 1105 1110
 Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Lys Leu Ile Gly
 1115 1120 1125

2007256717 29 Apr 2013

Lys Asp 1130	Lys Ile Tyr Leu Ser 1135	Trp Val Pro Ala His 1140	Lys Gly Ile
Gly Gly 1145	Asn Glu Gln Val Asp 1150	Lys Leu Val Ser Ser 1155	Gly Ile Arg
Lys Val 1160	Leu Phe Leu Asp Gly 1165	Ile Asp Lys Ala Gln 1170	Glu Asp His
Glu Arg 1175	Tyr His Ser Asn Trp 1180	Arg Thr Met Ala Ser 1185	Asp Phe Asn
Leu Pro 1190	Pro Ile Val Ala Lys 1195	Glu Ile Val Ala Ser 1200	Cys Asp Lys
Cys Gln 1205	Leu Lys Gly Glu Ala 1210	Met His Gly Gln Val 1215	Asp Cys Ser
Pro Gly 1220	Ile Trp Gln Leu Ala 1225	Cys Thr His Leu Glu 1230	Gly Lys Val
Ile Leu 1235	Val Ala Val His Val 1240	Ala Ser Gly Tyr Ile 1245	Glu Ala Glu
Val Ile 1250	Pro Ala Glu Thr Gly 1255	Gln Glu Thr Ala Tyr 1260	Phe Leu Leu
Lys Leu 1265	Ala Gly Arg Trp Pro 1270	Val Lys Val Val His 1275	Thr Ala Asn
Gly Ser 1280	Asn Phe Thr Ser Ala 1285	Ala Val Lys Ala Ala 1290	Cys Trp Trp
Ala Asn 1295	Ile Gln Gln Glu Phe 1300	Gly Ile Pro Tyr Asn 1305	Pro Gln Ser
Gln Gly 1310	Val Val Ala Ser Met 1315	Asn Lys Glu Leu Lys 1320	Lys Ile Ile
Gly Gln 1325	Val Arg Asp Gln Ala 1330	Glu His Leu Lys Thr 1335	Ala Val Gln
Met Ala 1340	Val Phe Ile His Asn 1345	Phe Lys Arg Lys Gly 1350	Gly Ile Gly
Gly Tyr 1355	Ser Ala Gly Glu Arg 1360	Ile Ile Asp Ile Ile 1365	Ala Thr Asp
Ile Gln 1370	Thr Lys Glu Leu Gln 1375	Lys Gln Ile Thr Lys 1380	Ile Gln Asn
Phe Arg 1385	Val Tyr Tyr Arg Asp 1390	Ser Arg Asp Pro Ile 1395	Trp Lys Gly
Pro Ala 1400	Lys Leu Leu Trp Lys 1405	Gly Glu Gly Ala Val 1410	Val Ile Gln
Asp Asn 1415	Ser Asp Ile Lys Val 1420	Val Pro Arg Arg Lys 1425	Ala Lys Ile
Leu Arg 1430	Asp Tyr Gly Lys Gln 1435	Met Ala Gly Asp Asp 1440	Cys Val Ala
Gly Arg 1445	Gln Asp Glu Asp Arg 1450	Ser Met Gly Gly Lys 1455	Trp Ser Lys
Gly Ser 1460	Ile Val Gly Trp Pro 1465	Glu Ile Arg Glu Arg 1470	Met Arg Arg
Ala Pro 1475	Ala Ala Ala Pro Gly 1480	Val Gly Ala Val Ser 1485	Gln Asp Leu
Asp Lys 1490	His Gly Ala Ile Thr 1495	Ser Ser Asn Ile Asn 1500	Asn Pro Ser
Cys Val 1505	Trp Leu Glu Ala Gln 1510	Glu Glu Glu Glu Val 1515	Gly Phe Pro
Val Arg 1520	Pro Gln Val Pro Leu 1525	Arg Pro Met Thr Tyr 1530	Lys Gly Ala
Phe Asp 1535	Leu Ser His Phe Leu 1540	Lys Glu Lys Gly Gly 1545	Leu Asp Gly
Leu Ile 1550	Tyr Ser Arg Lys Arg 1555	Gln Glu Ile Leu Asp 1560	Leu Trp Val
Tyr His 1565	Thr Gln Gly Tyr Phe 1570	Pro Asp Trp Gln Asn 1575	Tyr Thr Pro

2007256717 29 Apr 2013

Gly Pro Gly Val Arg Tyr Pro Leu Thr Phe Gly Trp Cys Phe Lys
 1580 1585 1590
 Leu Val Pro Met Glu Pro Asp Glu Val Glu Lys Ala Thr Glu Gly
 1595 1600 1605
 Glu Asn Asn Ser Leu Leu His Pro Ile Cys Gln His Gly Met Asp
 1610 1615 1620
 Asp Glu Glu Arg Glu Val Leu Ile Trp Lys Phe Asp Ser Arg Leu
 1625 1630 1635
 Ala Leu Lys His Arg Ala Gln Glu Leu His Pro Glu Phe Tyr Lys
 1640 1645 1650
 Asp Cys
 1655

<210> 14
 <211> 2050
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic construct

<220>
 <221> CDS
 <222> (16)..(2037)

<400> 14
 aagcttgccg ccacc atg agg gtg atg gag atc cag cgg aac tgc cag cac 51
 Met Arg Val Met Glu Ile Gln Arg Asn Cys Gln His
 1 5 10
 ctg ctg aga tgg ggc atc atg atc ctg ggc atg att atc atc tgc agc 99
 Leu Leu Arg Trp Gly Ile Met Ile Leu Gly Met Ile Ile Ile Cys Ser
 15 20 25
 acc gcc gac aac ctg tgg gtg acc gtg tac tac gcc gtg cct gtg tgg 147
 Thr Ala Asp Asn Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp
 30 35 40
 aga gat gcc gag acc acc ctg ttc tgc gcc agc gac gcc aag gcc tac 195
 Arg Asp Ala Glu Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr
 45 50 55 60
 agc acc gag aag cac aat gtg tgg gcc acc cac gcc tgc gtg cct acc 243
 Ser Thr Glu Lys His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr
 65 70 75
 gat ccc aac cct cag gag atc ccc ctg gac aac gtg acc gag gag ttc 291
 Asp Pro Asn Pro Gln Glu Ile Pro Leu Asp Asn Val Thr Glu Glu Phe
 80 85 90
 aac atg tgg aag aac aac atg gtg gac cag atg cac gag gac atc atc 339
 Asn Met Trp Lys Asn Asn Met Val Asp Gln Met His Glu Asp Ile Ile
 95 100 105
 agc ctg tgg gac cag agc ctg aag ccc tgc gtg cag ctg acc ccc ctg 387
 Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Gln Leu Thr Pro Leu
 110 115 120
 tgc gtg acc ctg aac tgc agc aac gcc aga gtg aac gcc acc ttc aac 435
 Cys Val Thr Leu Asn Cys Ser Asn Ala Arg Val Asn Ala Thr Phe Asn
 125 130 135 140
 tcc acc gag gac agg gag ggc atg aag aac tgc agc ttc aac atg acc 483
 Ser Thr Glu Asp Arg Glu Gly Met Lys Asn Cys Ser Phe Asn Met Thr
 145 150 155
 acc gag ctg cgg gat aag aag cag cag gtg tac agc ctg ttc tac cgg 531
 Thr Glu Leu Arg Asp Lys Lys Gln Gln Val Tyr Ser Leu Phe Tyr Arg
 160 165 170
 ctg gac atc gag aag atc aac agc agc aac aac aac agc gag tac cgg 579
 Leu Asp Ile Glu Lys Ile Asn Ser Ser Asn Asn Asn Ser Glu Tyr Arg
 175 180 185
 ctg gtg aac tgc aat acc agc gcc atc acc cag gcc tgc cct aag gtg 627
 Leu Val Asn Cys Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val
 190 195 200
 acc ttc gag ccc atc ccc atc cac tac tgc gcc cct gcc gcc ttc gcc 675
 Thr Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe Ala
 205 210 215 220
 atc ctg aag tgc aac gac acc gag ttc aat ggc acc ggc ccc tgc aag 723
 Ile Leu Lys Cys Asn Asp Thr Glu Phe Asn Gly Thr Gly Pro Cys Lys
 225 230 235

2007256717 29 Apr 2013

aat gtg agc acc gtg cag tgc acc cac ggc atc aag ccc gtg gtg tcc Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser 240 245 250	771
acc cag ctg ctg ctg aac ggc agc ctg gcc gag aga gaa gtg cgg atc Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Arg Glu Val Arg Ile 255 260 265	819
agg agc gag aac atc gcc aac aac gcc aag aac atc atc gtg cag ttc Arg Ser Glu Asn Ile Ala Asn Asn Ala Lys Asn Ile Ile Val Gln Phe 270 275 280	867
gcc agc ccc gtg aag atc aac tgc atc cgg ccc aac aac aat acc cgg Ala Ser Pro Val Lys Ile Asn Cys Ile Arg Pro Asn Asn Asn Thr Arg 285 290 300	915
aag agc tac aga atc ggc cct ggc cag acc ttc tac gcc acc gac att Lys Ser Tyr Arg Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Asp Ile 305 310 315	963
gtg ggc gac atc aga cag gcc cac tgc aac gtg tcc agg acc gac tgg Val Gly Asp Ile Arg Gln Ala His Cys Asn Val Ser Arg Thr Asp Trp 320 325 330	1011
aac aac acc ctg aga ctg gtg gcc aac cag ctg cgg aag tac ttc agc Asn Asn Thr Leu Arg Leu Val Ala Asn Gln Leu Arg Lys Tyr Phe Ser 335 340 345	1059
aac aag acc atc atc ttc acc aac agc agc ggc gga gac ctg gag atc Asn Lys Thr Ile Ile Phe Thr Asn Ser Ser Gly Gly Asp Leu Glu Ile 350 355 360	1107
acc acc cac agc ttc aat tgt ggc ggc gag ttc ttc tac tgc aac acc Thr Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Thr 365 370 375 380	1155
tcc ggc ctg ttc aat agc acc tgg acc acc aac aac atg cag gag tcc Ser Gly Leu Phe Asn Ser Thr Trp Thr Thr Asn Asn Met Gln Glu Ser 385 390 395	1203
aac gac acc agc aac ggc acc atc acc ctg ccc tgc cgg atc aag cag Asn Asp Thr Ser Asn Gly Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln 400 405 410	1251
atc atc cgg atg tgg cag cgc gtg ggc cag gcc atg tac gcc cct ccc Ile Ile Arg Met Trp Gln Arg Val Gly Gln Ala Met Tyr Ala Pro Pro 415 420 425	1299
atc gag ggc gtg att cgc tgc gag agc aac atc acc ggc ctg atc ctg Ile Glu Gly Val Ile Arg Cys Glu Ser Asn Ile Thr Gly Leu Ile Leu 430 435 440	1347
acc aga gat ggc ggc aac aac aat tcc gcc aac gag acc ttc aga cct Thr Arg Asp Gly Gly Asn Asn Asn Ser Ala Asn Glu Thr Phe Arg Pro 445 450 455 460	1395
ggc ggc gga gat atc cgg gac aac tgg cgg agc gag ctg tac aag tac Gly Gly Gly Asp Ile Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr 465 470 475	1443
aag gtg gtg aag atc gag ccc ctg ggc gtg gcc ccc acc aga gcc aag Lys Val Val Lys Ile Glu Pro Leu Val Ala Pro Thr Arg Ala Lys 480 485 490	1491
aga aga gtg gtg gag cgg gag aag aga gcc gtg ggc atc ggc gcc gtg Arg Arg Val Val Glu Arg Glu Lys Arg Ala Val Gly Ile Gly Ala Val 495 500 505	1539
ttt ctg ggc ttc ctg gga gcc gcc gga tct aca atg gga gcc gcc agc Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser 510 515 520	1587
atc acc ctg acc gtg cag gcc aga cag ctg ctg agc ggc atc gtg cag Ile Thr Leu Thr Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln 525 530 535 540	1635
cag cag agc aat ctg ctg aga gcc atc gag gcc cag cag cag ctg ctg Gln Gln Ser Asn Leu Leu Arg Ala Ile Glu Ala Gln Gln Gln Leu Leu 545 550 555	1683
aag ctg aca gtg tgg ggc atc aag cag ctg cag gcc agg gtg ctg gcc Lys Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu Ala 560 565 570	1731
gtg gag aga tac ctg agg gac cag cag ctc ctg ggc atc tgg ggc tgc Val Glu Arg Tyr Leu Arg Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys 575 580 585	1779
agc ggc aag ctg atc tgc acc acc aac gtg ccc tgg aat agc agc tgg	1827

2007256717 29 Apr 2013

Ser Gly 590	Lys Leu Ile Cys Thr 595	Thr Asn Val Pro Trp 600	Asn Ser Ser Trp	
agc aac aag agc tac gac gac atc tgg cag aac atg acc tgg ctg cag	Ser Asn Lys Ser Tyr Asp 610	Asp Ile Trp Gln Asn Met Thr Trp Leu Gln 620		1875
tgg gac aag gag atc agc aac tac acc gac atc atc tac agc ctg atc	Trp Asp Lys Glu Ile 625	Asp Ile Ile Tyr Ser Leu Ile 635		1923
gag gag agc cag aac cag cag gag aag aac gag cag gat ctg ctg gcc	Glu Glu Ser Gln Asn Gln Gln Glu Lys 645	Asn Glu Gln Asp Leu Leu Ala 650		1971
ctg gac aag tgg gcc aac ctg tgg aac tgg ttc gac atc agc aag tgg	Leu Asp Lys Trp Ala Asn Leu Trp Asn Trp Phe Asp 665	Ile Ser Lys Trp		2019
ctg tgg tac atc aga tct tgataatcta gaa	Leu Trp Tyr Ile Arg Ser 670			2050
<210> 15				
<211> 674				
<212> PRT				
<213> Artificial Sequence				
<220>				
<223> Synthetic Construct				
<400> 15				
Met Arg Val Met Glu Ile Gln Arg Asn Cys Gln His Leu Leu Arg Trp	1 5 10 15			
Gly Ile Met Ile Leu Gly Met Ile Ile Ile Cys Ser Thr Ala Asp Asn	20 25 30			
Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp Arg Asp Ala Glu	35 40 45			
Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Ser Thr Glu Lys	50 55 60			
His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp Pro Asn Pro	65 70 75 80			
Gln Glu Ile Pro Leu Asp Asn Val Thr Glu Glu Phe Asn Met Trp Lys	85 90 95			
Asn Asn Met Val Asp Gln Met His Glu Asp Ile Ile Ser Leu Trp Asp	100 105 110			
Gln Ser Leu Lys Pro Cys Val Gln Leu Thr Pro Leu Cys Val Thr Leu	115 120 125			
Asn Cys Ser Asn Ala Arg Val Asn Ala Thr Phe Asn Ser Thr Glu Asp	130 135 140			
Arg Glu Gly Met Lys Asn Cys Ser Phe Asn Met Thr Thr Glu Leu Arg	145 150 155 160			
Asp Lys Lys Gln Gln Val Tyr Ser Leu Phe Tyr Arg Leu Asp Ile Glu	165 170 175			
Lys Ile Asn Ser Ser Asn Asn Asn Ser Glu Tyr Arg Leu Val Asn Cys	180 185 190			
Asn Thr Ser Ala Ile Thr Gln Ala Cys Pro Lys Val Thr Phe Glu Pro	195 200 205			
Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys	210 215 220			
Asn Asp Thr Glu Phe Asn Gly Thr Gly Pro Cys Lys Asn Val Ser Thr	225 230 235 240			
Val Gln Cys Thr His Gly Ile Lys Pro Val Val Ser Thr Gln Leu Leu	245 250 255			
Leu Asn Gly Ser Leu Ala Glu Arg Glu Val Arg Ile Arg Ser Glu Asn	260 265 270			
Ile Ala Asn Asn Ala Lys Asn Ile Ile Val Gln Phe Ala Ser Pro Val	275 280 285			
Lys Ile Asn Cys Ile Arg Pro Asn Asn Asn Thr Arg Lys Ser Tyr Arg	290 295 300			

2007256717 29 Apr 2013

Ile Gly Pro Gly Gln Thr Phe Tyr Ala Thr Asp Ile Val Gly Asp Ile
 305 310 315 320
 Arg Gln Ala His Cys Asn Val Ser Arg Thr Asp Trp Asn Asn Thr Leu
 325 330 335
 Arg Leu Val Ala Asn Gln Leu Arg Lys Tyr Phe Ser Asn Lys Thr Ile
 340 345 350
 Ile Phe Thr Asn Ser Ser Gly Gly Asp Leu Glu Ile Thr Thr His Ser
 355 360 365
 Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Thr Ser Gly Leu Phe
 370 375 380
 Asn Ser Thr Trp Thr Thr Asn Asn Met Gln Glu Ser Asn Asp Thr Ser
 385 390 395 400
 Asn Gly Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Arg Met
 405 410 415
 Trp Gln Arg Val Gly Gln Ala Met Tyr Ala Pro Pro Ile Glu Gly Val
 420 425 430
 Ile Arg Cys Glu Ser Asn Ile Thr Gly Leu Ile Leu Thr Arg Asp Gly
 435 440 445
 Gly Asn Asn Asn Ser Ala Asn Glu Thr Phe Arg Pro Gly Gly Gly Asp
 450 455 460
 Ile Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys
 465 470 475 480
 Ile Glu Pro Leu Gly Val Ala Pro Thr Arg Ala Lys Arg Arg Val Val
 485 490 495
 Glu Arg Glu Lys Arg Ala Val Gly Ile Gly Ala Val Phe Leu Gly Phe
 500 505 510
 Leu Gly Ala Ala Gly Ser Thr Met Gly Ala Ala Ser Ile Thr Leu Thr
 515 520 525
 Val Gln Ala Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Ser Asn
 530 535 540
 Leu Leu Arg Ala Ile Glu Ala Gln Gln Gln Leu Leu Lys Leu Thr Val
 545 550 555 560
 Trp Gly Ile Lys Gln Leu Gln Ala Arg Val Leu Ala Val Glu Arg Tyr
 565 570 575
 Leu Arg Asp Gln Gln Leu Leu Gly Ile Trp Gly Cys Ser Gly Lys Leu
 580 585 590
 Ile Cys Thr Thr Asn Val Pro Trp Asn Ser Ser Trp Ser Asn Lys Ser
 595 600 605
 Tyr Asp Asp Ile Trp Gln Asn Met Thr Trp Leu Gln Trp Asp Lys Glu
 610 615 620
 Ile Ser Asn Tyr Thr Asp Ile Ile Tyr Ser Leu Ile Glu Glu Ser Gln
 625 630 635 640
 Asn Gln Gln Glu Lys Asn Glu Gln Asp Leu Leu Ala Leu Asp Lys Trp
 645 650 655
 Ala Asn Leu Trp Asn Trp Phe Asp Ile Ser Lys Trp Leu Trp Tyr Ile
 660 665 670
 Arg Ser